



**ULUSLARARASI İSTATİSTİKSEL SINIFLAMALARA YÖNELİK
KOD ATAMA SİSTEMİ (KASİS)**

Levent AHI

**DOKTORA TEZİ
YÖNETİM BİLİŞİM SİSTEMLERİ ANA BİLİM DALI**

**GAZİ ÜNİVERSİTESİ
BİLİŞİM ENSTİTÜSÜ**

OCAK 2021

Levent AHİ tarafından hazırlanan “ULUSLARARASI İSTATİSTİKSEL SINIFLAMALARA YÖNELİK KOD ATAMA SİSTEMİ (KASİS)” adlı tez çalışması aşağıdaki jüri tarafından OY BİRLİĞİ ile Gazi Üniversitesi Yönetim Bilişim Sistemleri Ana Bilim Dalında DOKTORA TEZİ olarak kabul edilmiştir.

Danışman: Prof. Dr. Ebru KILIÇ ÇAKMAK

Bilgisayar ve Öğretim Teknolojisi Eğitimi Ana Bilim Dalı, Gazi Üniversitesi

Bu tezin, kapsam ve kalite olarak Doktora Tezi olduğunu onaylıyorum.

.....

Başkan: Prof. Dr. M. Akif BAKIR

İstatistik Teorisi Ana Bilim Dalı, Gazi Üniversitesi

Bu tezin, kapsam ve kalite olarak Doktora Tezi olduğunu onaylıyorum.

.....

Üye: Prof. Dr. Cevriye TEMEL GENCER

Endüstri Mühendisliği Ana Bilim Dalı, Gazi Üniversitesi

Bu tezin, kapsam ve kalite olarak Doktora Tezi olduğunu onaylıyorum.

.....

Üye: Prof. Dr. Hakkı Okan YELOĞLU

Teknoloji ve Bilgi Yönetimi Ana Bilim Dalı, Başkent Üniversitesi

Bu tezin, kapsam ve kalite olarak Doktora Tezi olduğunu onaylıyorum.

.....

Üye: Doç. Dr. Furkan EMİRMAHMUTOĞLU

Ekonometri Ana Bilim Dalı, Ankara Hacı Bayram Veli Üniversitesi

Bu tezin, kapsam ve kalite olarak Doktora Tezi olduğunu onaylıyorum.

.....

Tez Savunma Tarihi: 12/1/2021

Jüri tarafından kabul edilen bu tezin Doktora Tezi olması için gerekli şartları yerine getirdiğini onaylıyorum.

.....

Prof. Dr. Aslıhan TÜFEKÇİ

Bilişim Enstitüsü Müdürü

ETİK BEYAN

Gazi Üniversitesi Bilişim Enstitüsü Tez Yazım Kurallarına uygun olarak hazırladığım bu tez çalışmada;

- Tez içinde sunduğum verileri, bilgileri ve dokümanları akademik ve etik kurallar çerçevesinde elde ettiğimi,
- Tüm bilgi, belge, değerlendirme ve sonuçları bilimsel etik ve ahlak kurallarına uygun olarak sunduğumu,
- Tez çalışmada yararlandığım eserlerin tümüne uygun atıfta bulunarak kaynak gösterdiğimi,
- Kullanılan verilerde herhangi bir değişiklik yapmadığımı,
- Bu tezde sunduğum çalışmanın özgün olduğunu,

bildirir, aksi bir durumda aleyhime doğabilecek tüm hak kayıplarını kabullendiğimi beyan ederim.

Levent AHİ

12/1/2021

ULUSLARARASI İSTATİSTİKSEL SINIFLAMALARA YÖNELİK
KOD ATAMA SİSTEMİ (KASİS)

(Doktora Tezi)

Levent AHI

GAZİ ÜNİVERSİTESİ

BİLİŞİM ENSTİTÜSÜ

Ocak 2021

ÖZET

İstatistiksel sınıflamaların, ülkelerin istatistik sistemlerinde çok büyük bir yeri ve önemi bulunmaktadır. İktisadi faaliyet, meslek, eğitim ve tüketim harcamaları sınıflaması bu sınıflamalara örnek olarak verilebilir. Bu sınıflamaları barındıran değişkenler ile yapılacak araştırmaların bizleri doğru sonuçlara ulaştırabilmesi, değişkenlerin sınıflamasının doğru olarak yapılmış olmasına bağlıdır. Veri hacmi büyüdükçe değişkenlerin sınıflamasının doğru yapılıp yapılmadığını etkin bir şekilde kontrol etmek klasik elle ve gözle yapılan yöntemlerle mümkün değildir. Bu yüzden bu işlemi yapabilecek otomatik bir sisteme ihtiyaç bulunmaktadır. Bu çalışmada, standart sınıflama sözlüğü olan tüm istatistiksel sınıflamalar için kullanılabilen bir sistem geliştirilmiş ve öncelikle geliştirilen sistem tanıtılmış daha sonra geliştirilen sistemin etkinliği Türkiye İstatistik Kurumu'nun (TÜİK) yapmış olduğu Hanehalkı Bütçe Araştırması (HBA) 2013-2018 yılları mikro veri setleri kullanılarak incelenmiştir. Bu veri seti, ülkemizdeki tüketim harcamaları istatistiklerinin ana veri kaynağıdır. Tüketim harcamalarının sınıflamasında uluslararası Bireysel Tüketimin Amaca Göre Sınıflaması (COICOP) kullanılmaktadır. Geliştirilen sistem ile daha önce anketör tarafından sınıflaması yapılmış tüketim harcamalarına tekrar kod ataması yapılarak sınıflanmış olduğu kategorinin doğru olup olmadığı belirlenerek sonuçlar yorumlanmıştır. Geliştirilen sistem, denetimli makine öğrenmesi yöntemlerini kullanan sistemlerden eğitim veri kümesine ihtiyaç duymaması ile ayrılmaktadır. Sistem çalışmaya başladıktan sonra her bir kayıta kendi öğrenmesini artırmakta ve öğrendiklerini sonraki kayıtlarda kullanmaktadır. Sistem denetimli makine öğrenmesi kullanan sistemlerde kullanılması zorunlu olan eğitim veri kümesindeki sınıflamanın doğru olarak yapılıp yapılmadığını kontrol ederek bu sistemlerin doğru şekilde öğrenmelerine de katkı sağlayabilmektedir. Aynı zamanda, sistem Tüketici Fiyat Endeksi'nde (TÜFE) kullanılabilen alternatif veri kaynaklarından en popüler olan barkod verilerine COICOP kodu atanması aşamasında kullanılabilir.

Bilim Kodu : 114607
Anahtar Kelimeler : İstatistiksel sınıflama, COICOP, HBA, otomatik kod atama, bulanık eşleştirme
Sayfa Adedi : 171
Danışman : Prof. Dr. Ebru KILIÇ ÇAKMAK

CODE ASSIGNMENT SYSTEM (KASIS)
FOR INTERNATIONAL STATISTICAL CLASSIFICATIONS
(Ph. D. Thesis)

Levent AHİ

GAZI UNIVERSITY
INSTITUTE OF INFORMATICS

January 2021

ABSTRACT

Statistical classifications have a great importance in the statistical systems of countries. Classification of economic activity, occupation, education and consumption expenditures can be given as examples for such classifications. Accuracy of studies using variables involving such classifications strongly depends on the implementation of coding correctly. As data size grows, it is usually not feasible to manually check whether not. As such, there is a clear need for an automated system to check the coding performed. In this study, a system that can be used for all statistical classifications with standard classification dictionary has been developed. The effectiveness of the proposed system has been examined using the 2013-2018 Household Budget Survey (HBS) micro data sets obtained from Turkish Statistical Institute (TURKSTAT). This data set can be regarded as the main source of the consumption expenditure statistics in Turkey. In this dataset, Classification of Individual Consumption by Purpose (COICOP) is used for classifying consumption expenditures. For each record, the codes assigned by the proposed system has been evaluated with the codes previously assigned by interviewers to observe accuracy. The proposed system distinguishes itself from supervised machine learning methods driven systems, as the former does not require training data set. The proposed system keeps learning in each iteration to exploit it in subsequent records. Also, the proposed system can be used as supporting tool for supervised machine learning methods driven systems by checking the codes assigned in the training data set. Furthermore, the proposed system can be used to assign COICOP codes for barcode data -- one of the most popular alternative data sources for Consumer Price Index (CPI).

Science Code : 114607
Key Words : Statistical classification, COICOP, HBS, automatic coding, fuzzy matching
Page Number : 171
Supervisor : Prof. Dr. Ebru KILIÇ ÇAKMAK

TEŞEKKÜR

Varlıkları ile bana güç veren ve çalışmamın her aşamasında beni destekleyen sevgili eşim Nazan AHI ve biricik oğlum Yiğit Efe AHI'ye teşekkürlerimi sunarım. İyi ki varsınız.

Bana her konuda şartsız ve koşulsuz destek veren babam Metin AHI'ye, annem Medine AHI'ye, ablalarım ve abime teşekkürlerimi sunarım. Bugüne kadar koşulsuz ve samimiyetle sağladığınız maddi ve manevi destekleriniz için sizlere minnettarım.

Çalışmamın her aşamasında beni yönlendiren, destekleyen ve cesaretlendiren değerli hocam Prof. Dr. Ebru KILIÇ ÇAKMAK'a teşekkürlerimi sunarım.

Tezimin bu aşamaya gelmesinde çok değerli katkıları bulunan Tez İzleme Komitesi Üyesi değerli hocalarım Prof. Dr. Cevriye TEMEL GENCER ve Prof. Dr. Hakkı Okan YELOĞLU'na teşekkürlerimi sunarım.

Bu çalışmanın oluşmasında katkı sağlayan çalışma arkadaşım Burcu ELLİDÖRT TUNÇ'a, çalışma adının KASİS olmasını öneren çalışma arkadaşım Duygu ÖZBAKIŞ'a ve burada adını tek tek sayamadığım aile büyüklerim ile Türkiye İstatistik Kurumu'ndaki tüm amirlerim, çalışma arkadaşlarım ve dostlarıma teşekkürlerimi sunarım.

İÇİNDEKİLER

	Sayfa
ÖZET	iv
ABSTRACT.....	v
TEŞEKKÜR.....	vi
İÇİNDEKİLER	vii
ÇİZELGELERİN LİSTESİ.....	xi
ŞEKİLLERİN LİSTESİ	xv
SİMGELER VE KISALTMALAR.....	xviii
1. GİRİŞ.....	1
1.1. Problem Durumu	1
1.2. Araştırmanın Amacı	6
1.3. Araştırmanın Önemi	8
1.4. Tanımlar	10
2. KAVRAMSAL ÇERÇEVE VE İLGİLİ ARAŞTIRMALAR	15
2.1. Anket Yöntemleri ve Kod Atama	15
2.2. Tüketici Fiyat Endeksi (TÜFE).....	18
2.3. Hanehalkı Bütçe Araştırması (HBA)	19
2.3.1. Tarihçe	19
2.3.2. Amaç	20
2.3.3. Kapsam.....	21
2.3.4. Alan uygulaması	21
2.4. Bireysel Tüketimin Amaca Göre Sınıflaması (COICOP)	21
2.4.1. Tarihçe	22
2.4.2. Sınıflamaya genel bakış ve kapsam	23
2.5. Kod Atama ve Kod Atama Sistemleri.....	25

	Sayfa
2.5.1. Elle yapılan kod atama	25
2.5.2. Bilgisayar destekli kod atama	26
2.5.3. Otomatik kod atama	27
2.6. Kayıt Bağlantısı.....	29
2.6.1. Deterministik bağlantı yöntemi.....	32
2.6.2. Bulanık eşleştirme yöntemi.....	33
3. YÖNTEM.....	43
3.1. Araştırmanın Modeli	43
3.1.1. Sorunları, fırsatları ve hedefleri belirleme	45
3.1.2. Bilgi gereksinimlerini belirleme	46
3.1.3. Sistem ihtiyaçlarını analiz etme	46
3.1.4. Önerilen sistemi tasarlama	46
3.1.5. Sistem geliştirme ve belgeleme.....	47
3.1.6. Sistemi test etme ve koruma	47
3.1.7. Sistemi uygulama ve değerlendirme	47
3.2. Evren ve Örneklem	47
3.3. Veri Toplama Araçları ve Uygulamanın Yapılışı	48
3.4. Verilerin Çözümlemesi ve Yorumlanması.....	50
4. KOD ATAMA SİSTEMİ (KASİS).....	51
4.1. Parametreler	51
4.1.1. Tablo	51
4.1.2. Ana grup.....	53
4.1.3. Bölge	54
4.1.4. Yıl.....	55
4.1.5. Ay.....	56

	Sayfa
4.2. Veri Setleri ve Anket Veri Giriş Programı.....	57
4.2.1. Tüketim veri seti	57
4.2.2. Sınıflama sözlüğü veri seti	57
4.2.3. Anket veri giriş programı	58
4.3. Katmanlar	59
4.3.1. Dil bilgisi katmanı.....	60
4.3.2. Eşleştirme katmanı	61
4.3.3. Sınıflama katmanı	62
4.3.4. Kod atama katmanı	71
4.4. Kullanıcı Ara Yüzü	72
4.4.1. Ana ekran	72
4.4.2. Arşivlenmiş veriyi sınıflama	73
4.4.3. Arşivlenmiş veriye daraltılmış listeden kod atama	74
4.4.4. Arşivlenmiş veriye geniş listeden kod atama.....	79
4.4.5. Madde tanımı girerek kod atama.....	82
4.4.6. Raporlar.....	88
5. BULGULAR	105
5.1. Dilbilgisi, Eşleştirme ve Sınıflama Katmanında Elde Edilen Bulgular	105
5.1.1. Anketör ile KASİS kodu arasında tablolara göre farklılık var mıdır?	106
5.1.2. Anketör ile KASİS kodu arasında ana gruplara göre farklılık var mıdır?	108
5.1.3. Anketör ile KASİS kodu arasında bölgelere göre farklılık var mıdır?	111
5.1.4. Anketör ile KASİS kodu arasında yıllara göre farklılık var mıdır?	114
5.1.5. Anketör ile KASİS kodu arasında aylara göre farklılık var mıdır?	116
5.1.6. KASİS sonuç verme süresi tablo ve kayıt sayısına göre farklı mıdır?.....	118
5.2. Kod Atama Katmanında Elde Edilen Bulgular	123
5.2.1 İki yöntem arasında tablolara göre farklılık var mıdır?	125

Sayfa

5.2.2. İki yöntem arasında ana gruplara göre farklılık var mıdır?.....	128
5.2.3. İki yöntem arasında bölgelere göre farklılık var mıdır?.....	131
5.2.4. İki yöntem arasında yıllara göre farklılık var mıdır?	136
5.2.5. İki yöntem arasında aylara göre farklılık var mıdır?.....	139
6. SONUÇ VE ÖNERİLER	143
KAYNAKLAR	149
EKLER.....	157
EK-1. İçerisinde “ekmek” ifadesi geçen COICOP kodlarının listesi	158
EK-2. Mikro verilere erişim izin yazısı	160
EK-3. Hanehalkı harcama kayıt defteri	161
EK-4. Ferdi harcama kayıt defteri	162
EK-5. İstatistiki Bölge Birimleri Sınıflaması (İBBS)	163
EK-6. KASİS program menüsü	164
ÖZGEÇMİŞ	170

ÇİZELGELERİN LİSTESİ

Çizelge	Sayfa
Çizelge 2.1. Kod atamasının olası yerleri ve kim tarafından / ne yapıldığı?	17
Çizelge 2.2. TÜFE’de farklı veri kaynaklarının özellikleri	18
Çizelge 2.2. (devam) TÜFE’de farklı veri kaynaklarının özellikleri	19
Çizelge 2.3. HBA’nın tarihçesi	20
Çizelge 2.4. COICOP sınıflamasının hiyerarşik yapısı	24
Çizelge 2.4. (devam) COICOP sınıflamasının hiyerarşik yapısı	25
Çizelge 2.5. Eşleşme kategorileri (Tip I ve Tip II bağlantı hataları)	35
Çizelge 2.6. Levenshtein mesafe ölçüsü hesaplama örneği	38
Çizelge 2.7. Genelleştirilmiş düzenleme mesafesi işlemleri ve maliyetleri	39
Çizelge 2.7. (devam) Genelleştirilmiş düzenleme mesafesi işlemleri ve maliyetleri	40
Çizelge 2.8. Genelleştirilmiş düzenleme mesafesi örneği	40
Çizelge 2.9. 2-gram değerleri örneği	41
Çizelge 2.10. 2-gram mesafe ölçüsü hesaplama örneği	41
Çizelge 4.1. KASİS tüketim veri seti ekran görüntüsü	57
Çizelge 4.2. KASİS 2013-2018 COICOP sözlüğü veri seti ekran görüntüsü	58
Çizelge 4.3. KASİS eşleştirme katmanı örneği	62
Çizelge 4.4. KASİS sözlük eşleşen kod örneği	63
Çizelge 4.5. KASİS yeni eşleşen kod örneği	63
Çizelge 4.6. KASİS bulanık eşleşen kod örneği 1	64
Çizelge 4.7. KASİS bulanık eşleşen kod örneği 2	65
Çizelge 4.8. İçerisinde “yumurta” kelimesi geçen sözlükteki kodlar	65
Çizelge 4.9. Yanlış yazılan kelime dönüşümleri	66
Çizelge 4.9. (devam) Yanlış yazılan kelime dönüşümleri	67
Çizelge 4.10. KASİS şüpheli kod örneği	68
Çizelge 4.11. KASİS yetersiz kod örneği	68

Çizelge	Sayfa
Çizelge 4.12. KASİS kabul kod örneği.....	69
Çizelge 4.13. KASİS bulanık kabul kod örneği.....	70
Çizelge 4.14. KASİS arşivlenmiş veriye kod atama sonuç ekran görüntüsü 1.....	77
Çizelge 4.15. KASİS arşivlenmiş veriye kod atama sonuç ekran görüntüsü 2.....	77
Çizelge 4.16. KASİS arşivlenmiş veriye kod atama sonuç ekran görüntüsü 3.....	77
Çizelge 4.17. KASİS arşivlenmiş veriye kod atama sonuç ekran görüntüsü 4.....	78
Çizelge 4.18. KASİS arşivlenmiş veriye kod atama sonuç ekran görüntüsü 5.....	78
Çizelge 4.19. KASİS arşivlenmiş veriye kod atama sonuç ekran görüntüsü 6.....	78
Çizelge 4.20. KASİS bulanık eşleştirme kod atama sonuç ekran görüntüsü 1	81
Çizelge 4.21. KASİS bulanık eşleştirme kod atama sonuç ekran görüntüsü 2.....	82
Çizelge 4.22. KASİS madde tanımı girerek kod atama sonuç ekran görüntüsü 1	84
Çizelge 4.23. KASİS madde tanımı girerek kod atama sonuç ekran görüntüsü 2.....	85
Çizelge 4.24. KASİS madde tanımı girerek kod atama sonuç ekran görüntüsü 3.....	86
Çizelge 4.25. KASİS madde tanımı girerek kod atama sonuç ekran görüntüsü 4.....	86
Çizelge 4.26. KASİS madde tanımı girerek kod atama sonuç ekran görüntüsü 5.....	87
Çizelge 4.27. KASİS raporlar sonuç ekran görüntüsü 1	90
Çizelge 4.28. KASİS raporlar sonuç ekran görüntüsü 2.....	92
Çizelge 4.29. KASİS raporlar sonuç ekran görüntüsü 3	94
Çizelge 4.30. KASİS raporlar sonuç ekran görüntüsü 4.....	95
Çizelge 4.31. KASİS raporlar sonuç ekran görüntüsü 5	97
Çizelge 4.32. KASİS raporlar sonuç ekran görüntüsü 6.....	99
Çizelge 4.33. KASİS raporlar sonuç ekran görüntüsü 7	100
Çizelge 4.34. KASİS raporlar sonuç ekran görüntüsü 8.....	101
Çizelge 4.35. KASİS raporlar sonuç ekran görüntüsü 9.....	102
Çizelge 4.36. KASİS raporlar sonuç ekran görüntüsü 10.....	103
Çizelge 4.37. KASİS raporlar sonuç ekran görüntüsü 11	104

Çizelge	Sayfa
Çizelge 4.38. KASİS raporlar sonuç ekran görüntüsü 12	104
Çizelge 5.1. Yıllara göre kayıt sayısı	105
Çizelge 5.2. Tablolara göre eşleşen kodlar	107
Çizelge 5.3. Tablolara göre kabul kodlar	108
Çizelge 5.4. Ana gruplara göre KASİS ile kontrol sonuçlarının dağılımı	108
Çizelge 5.5. Ana gruplara göre eşleşen kodlar	109
Çizelge 5.6. Ana gruplara göre kabul kodlar	110
Çizelge 5.7. Bölgelere göre KASİS ile kontrol sonuçlarının dağılımı	111
Çizelge 5.8. Bölgelere göre eşleşen kodlar	112
Çizelge 5.9. Bölgelere göre kabul kodlar	113
Çizelge 5.10. Yıllara göre eşleşen kodlar	115
Çizelge 5.11. Yıllara göre kabul kodlar	116
Çizelge 5.12. Aylara göre KASİS ile kontrol sonuçlarının dağılımı	116
Çizelge 5.13. Aylara göre eşleşen kodlar	117
Çizelge 5.14. Aylara göre kabul kodlar	118
Çizelge 5.15. İki farklı yöntemle göre yapılan KASİS kod atama sonuçları	124
Çizelge 5.16. Tablolara göre sistemin önerdiği kodların dağılımı	125
Çizelge 5.17. Tablolara göre sistemin önerdiği kod sayılarına göre dağılımı	126
Çizelge 5.18. Tablolara göre kodların ilk beş basamağına göre dağılımı	126
Çizelge 5.19. Tablolara göre kod atama sonuçları 1	127
Çizelge 5.20. Tablolara göre kod atama sonuçları 2	127
Çizelge 5.21. Ana gruplara göre sistemin önerdiği kodların dağılımı	128
Çizelge 5.22. Ana gruplara göre sistemin önerdiği kod sayılarına göre dağılımı	129
Çizelge 5.23. Ana gruplara göre kodların ilk beş basamağına göre dağılımı	129
Çizelge 5.24. Ana gruplara göre kod atama sonuçları 1	130
Çizelge 5.25. Ana gruplara göre kod atama sonuçları 2	131

Çizelge	Sayfa
Çizelge 5.26. Bölgelere göre sistemin önerdiği kodların dağılımı	132
Çizelge 5.27. Bölgelere göre sistemin önerdiği kod sayılarına göre dağılımı	133
Çizelge 5.28. Bölgelere göre kodların ilk beş basamağına göre dağılımı	134
Çizelge 5.29. Bölgelere göre kod atama sonuçları 1	135
Çizelge 5.30. Bölgelere göre kod atama sonuçları 2	136
Çizelge 5.31. Yıllara göre sistemin önerdiği kodların dağılımı.....	137
Çizelge 5.32. Yıllara göre sistemin önerdiği kod sayılarına göre dağılımı	137
Çizelge 5.33. Yıllara göre kodların ilk beş basamağına göre dağılımı.....	138
Çizelge 5.34. Yıllara göre kod atama sonuçları 1	138
Çizelge 5.35. Yıllara göre kod atama sonuçları 2.....	139
Çizelge 5.36. Aylara göre sistemin önerdiği kodların dağılımı	139
Çizelge 5.37. Aylara göre sistemin önerdiği kod sayılarına göre dağılımı.....	140
Çizelge 5.38. Aylara göre kodların ilk beş basamağına göre dağılımı	141
Çizelge 5.39. Aylara göre kod atama sonuçları 1	141
Çizelge 5.40. Aylara göre kod atama sonuçları 2	142

ŞEKİLLERİN LİSTESİ

Şekil	Sayfa
Şekil 2.1. COICOP sınıflamasının düzey 1 kategorileri	24
Şekil 2.2. Bulanık eşleştirme	33
Şekil 2.3. Bloklama örneği.....	34
Şekil 2.4. Jaro-Winkler mesafe ölçüsü hesaplama örneği	38
Şekil 3.1. Sistem geliştirme yaşam döngüsü.....	44
Şekil 4.1. KASİS'te kullanılan parametreler	51
Şekil 4.2. KASİS'te kullanılan tablo isimleri	52
Şekil 4.3. KASİS tablo parametresi ekran görüntüsü	52
Şekil 4.4. KASİS ana grup parametresi ekran görüntüsü	53
Şekil 4.5. Düzey 2 İstatistiki Bölge Birimleri.....	54
Şekil 4.6. KASİS bölge parametresi ekran görüntüsü	55
Şekil 4.7. KASİS yıl parametresi ekran görüntüsü	56
Şekil 4.8. KASİS ay parametresi ekran görüntüsü	56
Şekil 4.9. Anket veri giriş programı ekran görüntüsü örneği 1	58
Şekil 4.10. Anket veri giriş programı ekran görüntüsü örneği 2	59
Şekil 4.11. Anket veri giriş programı ekran görüntüsü örneği 3	59
Şekil 4.12. KASİS'in katmanları	60
Şekil 4.13. KASİS dil bilgisi katmanı algoritması.....	60
Şekil 4.14. KASİS dil bilgisi katmanında yapılan işlemler	61
Şekil 4.15. KASİS eşleştirme katmanı algoritması.....	61
Şekil 4.16. KASİS sınıflama katmanı algoritması	62
Şekil 4.17. Yetersiz kodlara önerilen kod sayılarının dağılımı.....	70
Şekil 4.18. KASİS kod atama katmanı algoritması	71
Şekil 4.19. KASİS ana ekran görüntüsü	72
Şekil 4.20. KASİS arşivlenmiş veriyi sınıflama giriş ekran görüntüsü	73

Şekil	Sayfa
Şekil 4.21. KASİS arşivlenmiş veriyi sınıflama sonuç ekran görüntüsü	74
Şekil 4.22. KASİS arşivlenmiş veriye kod atama giriş ekran görüntüsü 1	75
Şekil 4.23. KASİS arşivlenmiş veriye kod atama giriş ekran görüntüsü 2.....	76
Şekil 4.24. KASİS bulanık eşleştirme yöntemleriyle kod atama ekran görüntüsü 1	80
Şekil 4.25. KASİS bulanık eşleştirme yöntemleriyle kod atama ekran görüntüsü 2	81
Şekil 4.26. KASİS madde tanımı girerek kod atama giriş ekran görüntüsü 1	83
Şekil 4.27. KASİS madde tanımı girerek kod atama giriş ekran görüntüsü 2	83
Şekil 4.28. KASİS raporlar giriş ekran görüntüsü 1	88
Şekil 4.29. KASİS raporlar giriş ekran görüntüsü 2	89
Şekil 4.30. KASİS raporlar giriş ekran görüntüsü 3	90
Şekil 4.31. KASİS raporlar giriş ekran görüntüsü 4	91
Şekil 4.32. KASİS raporlar giriş ekran görüntüsü 5	92
Şekil 4.33. KASİS raporlar giriş ekran görüntüsü 6	93
Şekil 4.34. KASİS raporlar giriş ekran görüntüsü 7	93
Şekil 4.35. KASİS raporlar giriş ekran görüntüsü 8	94
Şekil 4.36. KASİS raporlar giriş ekran görüntüsü 9	95
Şekil 4.37. KASİS raporlar giriş ekran görüntüsü 10	96
Şekil 4.38. KASİS raporlar giriş ekran görüntüsü 11	97
Şekil 4.39. KASİS raporlar giriş ekran görüntüsü 12	98
Şekil 4.40. KASİS raporlar giriş ekran görüntüsü 13	98
Şekil 4.41. KASİS raporlar giriş ekran görüntüsü 14	99
Şekil 4.42. KASİS raporlar giriş ekran görüntüsü 15	100
Şekil 4.43. KASİS raporlar giriş ekran görüntüsü 16	101
Şekil 4.44. KASİS raporlar giriş ekran görüntüsü 17	102
Şekil 4.45. KASİS raporlar giriş ekran görüntüsü 18	103
Şekil 4.46. KASİS veri tabanı yapısı	104

Şekil	Sayfa
Şekil 5.1. Tablolara göre KASİS ile kontrol sonuçlarının dağılımı.....	106
Şekil 5.2. Yıllara göre KASİS ile kontrol sonuçlarının dağılımı	114
Şekil 5.3. Tablo ve kayıt sayılarına göre sonuç verme süresi	119
Şekil 5.4. Tablolara ve kayıt başına göre ortalama sonuç verme süresi	119
Şekil 5.5. Kayıt sayısı 1 - 100 arasında olan çalışmaların sonuç verme süresi	120
Şekil 5.6. Kayıt sayısı 101 - 500 arasında olan çalışmaların sonuç verme süresi	121
Şekil 5.7. Kayıt sayısı 501 – 1 000 arasında olan çalışmaların sonuç verme süresi ...	121
Şekil 5.8. Kayıt sayısı 1 001 – 5 000 arasında olan çalışmaların sonuç verme süresi .	122
Şekil 5.9. Kayıt sayısı 5 001 ve üzeri olan çalışmaların sonuç verme süresi	123

SİMGELER VE KISALTMALAR

Bu çalışmada kullanılmış simgeler ve kısaltmalar, açıklamaları ile birlikte aşağıda sunulmuştur.

Kısaltmalar	Açıklamalar
CAPI	Bilgisayar destekli yüz yüze görüşme (Computer-assisted personal interviewing)
CATI	Bilgisayar destekli telefon görüşmesi (Computer-assisted telephone interviewing)
CAQDAS	Bilgisayar Destekli Nitel Veri Analizi (Computer Aided Qualitative Data Analysis)
CAWI	Bilgisayar destekli web görüşmesi (Computer-assisted web interviewing)
COFOG	Devlet Fonksiyonlarının Sınıflaması (Classification of the Functions of Government)
COICOP	Bireysel Tüketimin Amaca Göre Sınıflaması (Classification of Individual Consumption According to Purpose)
COPNI	Hanehalkına Hizmet Veren Kar Amacı Gütmeyen Kuruluşların Sınıflaması (Classification of the Purposes of Non-profit Institutions)
COPP	Üreticilerin Giderlerinin Amaca Göre Sınıflaması (Classification of Outlays of Producers by Purpose)
DPT	Devlet Planlama Teşkilatı
ECOICOP	Avrupa Bireysel Tüketimin Amaca Göre Sınıflaması (European Classification of Individual Consumption According to Purpose)
Eurostat	Avrupa Birliği İstatistik Ofisi (European Statistical Office)
GSYİH	Gayri Safi Yurtiçi Hasıla
HBA	Hanehalkı Bütçe Araştırması

Kısaltmalar**Açıklamalar**

ICD	Uluslararası Hastalıklar Sınıflaması (International Classification of Diseases)
ICLS	Uluslararası Çalışma İstatistikçileri Konferansı (International Conference of Labour Statisticians)
ILO	Uluslararası Çalışma Örgütü (International Labour Organization)
IMF	Uluslararası Para Fonu (International Monetary Fund)
ISCO	Uluslararası Meslek Sınıflaması (International Standard Classification of Occupations)
İBBS	İstatistiki Bölge Birimleri Sınıflaması
KASİS	Kod Atama Sistemi
NAICS	Kuzey Amerika Endüstri Sınıflama Sistemi (North American Industry Classification System)
NUTS	İstatistiki Bölge Birimleri Sınıflandırması (The Nomenclature of Territorial Units for Statistics)
OECD	İktisadi İşbirliği ve Kalkınma Teşkilatı (Organisation for Economic Co-operation and Development)
PAPI	Kağıt ve kalem görüşmesi (Paper and pencil interviewing)
SAS	İstatistiksel Analiz Programı (Statistical Analysis Software)
SHARE	Avrupa'da Sağlık, Yaşlanma ve Emeklilik Anketi (The Survey of Health, Ageing and Retirement in Europe)
SNA	Ulusal Hesaplar Sistemi (System of National Accounts)
SPC	İstatistik Program Komitesi (Statistical Programme Committee)
SVM	Destek Vektör Makinesi (Support Vector Machine)
TGA	Tasarım ve geliştirme araştırması
TÜFE	Tüketici Fiyat Endeksi

Kısaltmalar**Açıklamalar**

TÜİK	Türkiye İstatistik Kurumu
UAVT	Ulusal Adres Veri Tabanı
UNESCO	Birleşmiş Milletler Eğitim, Bilim ve Kültür Kurumu (United Nations Educational, Scientific and Cultural Organization)
UNSC	Birleşmiş Milletler İstatistik Komisyonu (United Nations Statistical Commission)
UNSD	Birleşmiş Milletler İstatistik Bölümü (United Nations Statistical Division)
VCS	Kelimesi Kelimesine Kodlama Sistemi (Verbatim Coding System)

1. GİRİŞ

Bu bölüm problem durumu, araştırmanın amacı, araştırmanın önemi ve tanımlar kısmından oluşmaktadır.

1.1. Problem Durumu

Küreselleşme ile yaşanan kapsamlı değişim sürecinde ihtiyaçlar da çeşitlenmektedir. Bu değişim sürecinde, istatistik sistemi kendi payına düşeni almaktadır ve istatistik sistemine duyulan ihtiyaç da çeşitlenmektedir [1]. Bu durumun temel sebebi, her alanda belirlenen hedeflere ulaşıp ulaşılmadığının ispat edilmesi bir takım istatistiksel göstergelere dayandırılmasına bağlı olmasından kaynaklanmaktadır. Çevremizdeki tüm gelişmeleri ölçebildiğimiz ölçüde yönlendirilebiliriz. Bu nedenle, derlenmesinden yayınına kadar olan tüm süreçlerin belirli standartları sağlayan güvenilir bir istatistik sistemi bu ölçme görevini tam anlamıyla yerine getirebilmek için gereklidir [2].

İstenilen anda, istenilen formatta ve istenilen kapsamda istatistiklere erişim isteği günden güne artmaktadır. Verilerin kim tarafından nasıl derlendiği kadar verilerin güncel ve güvenilir olması da önemlidir. Bu nedenle, standartlaştırılmış veri toplama, analiz ve yayımlama konularında hem ulusal ve uluslararası standartların ve normların takip edilerek son teknolojinin bu süreçlere entegre edilmesi hassas bir konudur.

İstatistik sisteminin en önemli bileşenlerinden birisi göstergelerdir. İstatistiksel göstergelerden herkesin aynı şeyi anlayabilmesi ve aynı sonucu elde edebilmesi bu göstergelerin elde edilirken belirli koşulları ve standartları sağlamalarına bağlıdır. İstatistiksel gösterge üretirken sınıflamalardan sıklıkla faydalanılmaktadır. Bu anlamda, istatistiksel gösterge üretmek için kullanılan sınıflamalar farklı ülkelerde yaşayan insanlar arasında iletişimi sağlayan bir dil olmaktadır ve belirli standartları sağlamaları zorunlu olmaktadır [3]. Sınıflamalar bu alanda önemli görevler üstlenmektedir.

Etimolojik anlamda, sınıflama, önceden belirlenmiş standartlara göre karşılık gelen bir şeyi tanımlayan ve Yunanca “clasis” kelimesinin Latinceye uyarlanması sonucu ortaya çıkan bir terimdir [4]. Bir sınıflama sistemi ise, belgelerin veya verilerin yapılandırılmış bir

gösterimini yapmayı amaçlayan ve sistematik erişime izin vermek için karşılık gelen terimleri kullanan bir indeksleme dilidir [5]. Sınıflama kullanılmasının temel amacı, nesnelere kolayca ulaşmak, nesneleri belirli bir düzene koymak, olaylara daha bütüncül bakmak ve istenilen çıkarsamaları daha kolayca yapmaktır. Bu kapsamda oluşturulmuş çok sayıda istatistiksel sınıflama türü bulunmaktadır. Faaliyet sınıflamaları, ürün sınıflamaları, dış ticaret sınıflamaları, amaca göre sınıflamalar, coğrafi sınıflamalar, çevre sınıflamaları, eğitim sınıflamaları, sağlık sınıflamaları ve meslek sınıflamaları bu sınıflama türlerinden bazılarıdır [6]. Bu sınıflamalara göre yapılan gruplama sonucunda aynı gruptaki nesneler benzer özelliklere sahip olmaktadır.

Sınıflamalar, farklı ülkelerin istatistiksel göstergelerinin birbirleriyle karşılaştırılmasına imkan sağlamaktadır. Ortak bir sınıflamanın kullanılması, verilerin kararlaştırılan bir dizi düzenlemeye göre kategorize edilmesi anlamına gelir ve bu sayede kimin verileri derlediğinden bağımsız olarak veriler tüm kullanıcılar için tamamen karşılaştırılabilir olmaktadır. Ülkeler resmi istatistiklerinde, ekonomik faaliyetler, ürünler, meslekler veya coğrafi birimler gibi değişkenleri yapılandıran bir dizi sınıflama sistemini kullanmaktadırlar [7].

Sınıflamalar, ulusal veya uluslararası kaynaklı olabilmektedir. Ulusal sınıflama sistemlerine ek olarak, karşılaştırmalı uluslararası verilere talep artışı nedeniyle bir dizi uluslararası sınıflama sistemi de önem kazanmıştır. Bu yüzden, ulusal istatistik araştırmaları uluslararası sınıflama sistemlerini giderek daha fazla kullanmaktadırlar. Uluslararası sınıflamalar, Birleşmiş Milletler İstatistik Bölümü (UNSD), Avrupa Birliği İstatistik Ofisi (Eurostat), Uluslararası Çalışma Örgütü (ILO), Uluslararası Para Fonu (IMF), İktisadi İşbirliği ve Kalkınma Teşkilatı (OECD), Birleşmiş Milletler Eğitim, Bilim ve Kültür Örgütü (UNESCO) gibi özel konularda uzman kuruluşlar tarafından geliştirilmektedir. Uluslararası sınıflamaların varlığı, sıfırdan ulusal bir sınıflama yapmak yerine, ulusal istatistik ofislerine rehberlik ederek tanımların ve verilerin uluslararası karşılaştırılabilirliğini kolaylaştırmaktadır [6].

Ülkemizde ulusal sınıflamaların geliştirme sorumluluğu Türkiye İstatistik Kurumu'nda (TÜİK) bulunmaktadır. Geliştirilen sınıflamalar, TÜİK internet sitesindeki sınıflama sunucusu üzerinden paylaşılmaktadır. 5429 sayılı Türkiye İstatistik Kanunu'nun "Sınıflamalar" başlıklı 11. maddesinde yer alan "Kurum ve kuruluşlar, Başkanlık (TÜİK)

tarafından oluşturulan istatistik amaçlı tanım ve sınıflamaları kullanmakla; kendi ihtiyaçları doğrultusunda belirleyecekleri sınıflamaları kullanmaları halinde ise, Başkanlığın belirlediği sınıflamalara dönüştürülmesini sağlayacak tedbirleri almakla yükümlüdür” ifadesi ile sınıflamaların farklı kurumlar tarafından üretilmesinden doğacak sıkıntıların ortadan kaldırılması sağlanmıştır [8]. Kanunun bu maddesinden anlaşılmaktadır ki, sınıflamaların istatistik sisteminde üretilen istatistiklerin karşılaştırılabilirliği bakımından önemli bir işlevi bulunmaktadır.

Sınıflamaların kullanıldığı istatistikler idari kayıtlara bağlı olarak üretilebileceği gibi anket yaparak da üretilmektedir. Anketlerde, mümkün olduğunca cevabı metinsel olan sorular sorulmaktan kaçınılmasına rağmen, kapalı sorular sormak her zaman mümkün olmamaktadır [9]. Anketlerde, metinsel cevaplı değişkenlerin kullanılmasının ana nedeni, ankete cevap verenler bakımından daha uygun olmasıdır. Bu tür sorulara, cevap verenler kişisel olarak formüle edilmiş bir metinle cevap verebilme şansına sahip olmaktadır. Ayrıca, metinsel cevaplarda yapılacak hatanın, yanlış seçenek işaretlenmesinden dolayı anket sonuçları üzerinde çok daha az etkisi bulunmaktadır [10]. İstatistik üretebilmek için ihtiyaç duyulan formda ve sınıflamada bir cevap verilmesi istenmesi durumunda, cevap verenlerin iktisadi faaliyet sınıflaması, meslek sınıflaması ve eğitim sınıflaması vb. sınıflamalar hakkında bilgi sahibi olması gerekmektedir. Ancak, uzman olmayan birinden böyle bir şey beklemek çok zor, hatta imkansızdır. Bu yüzden her istatistiksel araştırmada metinsel cevapları doğru şekilde sınıflamak ve kod atamak kaçınılmaz bir iş olmaktadır. Bu bağlamda kod atama, bir veya birkaç kelimeden oluşan metinsel bir ifadeyi bir terminoloji ile ilgili olan sınıflamaya ait bir koda dönüştürme anlamına gelmektedir.

Sınıflama veya kod atama problemi, makine öğrenmesi ve veri madenciliği literatüründeki en temel problemlerden birisidir [11]. Özellikle, bankacılık, sigortacılık, tıp, biyoloji, kimya, sosyal medya ve pazarlama alanlarında sınıflama problemlerinin çözümü, makine öğrenmesinin önemli çalışma alanlarından olmuştur. Sınıflama probleminin çözümü sonucunda nesnelere veya gözlemlere sayısal veya karakter olarak tanımlanmış kodlar atanmış olmaktadır. Hacking ve Willenborg, Hollanda İstatistik Ofisi’nde kullanılan kod atama gerektiren sınıflamaları, eğitim, meslek, ürün, endüstri ve ölüm nedenleri olarak sıralamaktadır [10]. Groves ve diğerleri, kod atamasının gerekli olduğu sık kullanılan bazı sınıflamaları Uluslararası Standart Meslek Sınıflaması (ISCO), Kuzey Amerika Endüstri Sınıflama Sistemi (NAICS) ve Uluslararası Hastalık Sınıflaması (ICD) olarak vermektedir

[12]. Sınıflamalar, uzmanlık gerektiren alanlarda kullanılması için tanımlanmış olabileceği gibi tek seferlik araştırmaya özgü konularda da üretilmiş olabilmektedir. Esuli ve Sebastiani, bir pazar araştırmasında kod ataması yapmışlardır. Yaptıkları çalışma kapsamında, bir ürün veya marka adı ile cevap verilen “favori meşrubatınız nedir?” şeklindeki bir soruya verilen cevapları sınıflamışlardır [13]. Groves ve diğ. [12], kodlamanın yalnızca metinsel yanıtlar için yapılmadığına işaret etmektedir. Anketlerde toplanan sayısal olmayan veriler, örneğin görsel görüntüler, sesler, toprak, kan örnekleri veya coğrafi veriler sınıflamaya uygun bir koda atanabilmektedir.

Kod atamasının elle yapılması esnasında, bir takım hatalar yapılabilmektedir. Smith [14], hastane tıbbi kayıtlarındaki ICD kodlarının doğruluğu test etmek amacıyla, rastgele seçilen 174 hastanın kaydının geriye dönük olarak tekrar incelendiğini, elle atanan ICD kodlarında %29'luk hata oranı tespit edildiğini ve kod atamasını yapan personelin tıp eğitimi almış personelden ziyade tıbbi kayıt kodlayıcıları olduğu için hata oranının arttığını ortaya koymuştur.

Kodların elle atanması esnasında atamayı yapan personel yanlışlığının ortaya konulması açısından Rossi ve diğ. [15] bir araştırma yapmışlardır. Araştırmada, sağlık hizmetleri genelinde tıbbi kayıt istatistik kartlarında kodlama hatası oranının yaklaşık %30 olduğunu ve Padua Üniversitesi Hastanesi Klinik Cerrahi Enstitüsü tarafından hata oranını azaltmak amacıyla otomatik kodlama programı geliştirildiğini belirtmişlerdir. Hastanenin tüm bölümlerinden 4 776 tıbbi kayıt istatistik kartından rastgele seçilen 54'ü program tarafından otomatik olarak tekrar kodlanmıştır. Otomatik olarak kodlanan kartların kodlarında herhangi bir tutarsızlık gözlenmemiştir. Sonuç olarak, programın faydası açıkça görülmekte olup programın diğer birimlerde kullanılması önerilmektedir.

Kodların elle atanmasının negatif özelliklerinin yanına maliyetli olmasını da eklemek gerekmektedir. Appel ve Hellermann [16], iktisadi faaliyet ve meslek sınıflaması ile ilgili ofiste yapılan kod atama ve doğrulama maliyetlerinin tüm demografik araştırmalar için yılda yaklaşık 1,2 milyon dolar ve 1980 Nüfus Sayımı için ise maliyetin yaklaşık 7,7 milyon dolar olduğunu belirtmiştir.

İstatistiksel araştırmalarla genellikle, toplanmasını, yayılmasını ve yorumlanmasını kolaylaştırmak için uygun şekilde kategorize edilmesi gereken çok büyük miktarlarda veri

üretildiği için sürekli olarak bu yüksek kodlama maliyetlerine katlanmak durumunda kalınmaktadır. Yapılan araştırmalarda da vurgulandığı gibi verinin doğru sınıflanıp sınıflanmadığının yani atamayı yapan kişinin yanlışlığının olup olmadığının kontrolünün yapılması bu sınıflamalar sonucunda bir analiz ve çalışma yapılacaksa büyük önem arz etmektedir. Anketlere dayandırılarak yapılan ve ulaşılan sonuç bakımından benzer olması gereken çalışmalarda bile bazı sorunlar bulunmaktadır. Bu sorunların kaynağı genellikle veri ile dolayısıyla veri kalitesi ile ilgilidir. Veri kaynağındaki sorunları analiz tekniklerini kuvvetlendirerek aşabilmek mümkün değildir. Bu sorunların üstesinden gelebilmenin yolu en başından standartları iyi koymaktan ve süreci kontrol altında tutmaktan geçmektedir [17]. Herhangi bir ankette, ilgisiz veya dikkatsiz verilen cevaplar bir endişe kaynağıdır. Bu nedenle, araştırmacıların bu tür verileri düzeltmeleri veya elemeleri önemlidir [18]. Hatalardan arındırılmış veri kullanmak doğru sonuçlara ulaşabilmek açısından gereklidir [17]. Aynı durum, eğer ankette ulusal veya uluslararası bir sınıflama kullanılıyorsa, bilimsel bir çalışmaya başlamadan önce sınıflamaların doğru kullanıldığından ve kod atamasının doğru yapıldığından emin olunması gerekmektedir. Anketlerde kullanılan sınıflamalar sonucunda resmi istatistik kapsamında bir sonuç açıklanacaksa daha dikkatli olunması gerekmektedir.

Belloni ve diğ. [19], meslek kodlarındaki hataları incelemek amacıyla Hollanda'daki Avrupa'da Sağlık, Yaşlanma ve Emeklilik Anketi (SHARE) verilerindeki son ve şu anki mesleğe ilişkin açık uçlu soruları bir program kullanarak yeniden kodlamışlardır. Daha önce yapılan kod atamada yüksek miktarda hata tespit etmişlerdir. Anketlerde, kod atama kalitesinin önemli olduğunu ve genellikle ihmal edildiğini belirtmişlerdir. Yanlış kod atamalar nedeniyle oluşan hataların istatistiksel analizler yapılırken veya ekonometrik modellerde dikkate alınması gerektiğine vurgu yapmışlardır.

Kanada İstatistik Ofisi tarafından geliştirilmiş olan otomatik kod atama programının 1991 Kanada Nüfus Sayımı'nda dört aylık kullanımı sonucunda, kod atamadaki doğruluk oranı artmış, kodlayıcı sayısı 600'den 25'e inmiş ve 5,9 milyon ABD doları bütçeye sahip projede 3,5 milyon dolardan fazla tasarruf sağlanmıştır [20].

İrlanda örneğinde ise, elle yapılan meslek kod atamasının, çok sayıda kod atama personeline ihtiyaç duyması, yüksek maliyetli olması, kod atamasının kalitesi veya doğruluğunun yeterince tatmin edici düzeyde olmaması ve aynı tanıma farklı kodlayıcıların farklı kodlar

ataması dezavantajlarına sahip olduğu belirlenmiştir [21].

TÜİK’te, anketör tarafından kod ataması yapılmış kayıtların sınıflamasının doğru bir şekilde yapılıp yapılmadığını belirlemek amacıyla gözle tekrar kontrol edilmektedir. Çünkü kod atamasının elle yapılması esnasında hatalar yapılabilmektedir. Bu tür hataları Smith [14], Rossi ve diğ. [15] ve Belloni ve diğ. [19] araştırmaları ile ortaya koymuşlardır. An itibarıyla TÜİK’te kullanılan otomatik bir kontrol sistemi bulunmamaktadır. Veri hacmi büyüdükçe gözle yapılan kontrol işlemlerinin etkin ve verimli bir şekilde yapılabilmesi mümkün olamamaktadır. Appel ve Hellermann [16] yaptıkları araştırmada kodların elle atanmasında yapılan hataların yanında maliyetli olmasını ortaya koymuştur. Ulusal ve uluslararası alanda standart bir sınıflamanın kullanılmasının yanında günümüzde her istatistiğin otomasyon ile insandan bağımsız yöntemlerle yapılması maliyet ve doğruluk bakımından önemli hale gelmiştir. Bunun için gerek daha önce kod ataması yapılmamış kayıtlara otomatik şekilde kod ataması yapabilecek gerekse daha önce kod ataması yapılmış olan kayıtların doğru sınıfta sınıflanıp sınıflanmadığını kontrol edebilecek bir sisteme ihtiyaç duyulmaktadır. Böyle bir sistemin varlığı, anketör boyutunda yanlışlığın ortadan kalkmasını sağlayarak tasarruf sağlayacak ve doğru kod atama kabiliyetini artıracaktır.

1.2. Araştırmanın Amacı

Kod atamasının, istatistiksel süreçte kullanılan bir etkinlik olduğu düşünülebilir. Yapılması oldukça zor ve özel bir görevdir. Kod atamasının amacı, var olan bir metni uygun bir sınıfa veya gruba atamaktır. Bu süreçteki hedef, yazılan metne karşılık gelen çok sayıda sınıf arasından en uygununu seçerek uygun metni uygun sınıf içerisinde gruplamaktır [10]. Kod atama, bir doktorun kendisine çeşitli şikayetler ve semptomlar sunan hastaya tanı koymasına çok benzemektedir. Doktorun görevi, birkaç gözlem, hastadan gelen yanıtlar ve muhtemelen ek testlere (örneğin kan testleri) dayanarak teşhis koymak ve bu temelde bir tedavi seçmektir [22-24]. Benzer yaklaşımla, kod atama işlemi de eldeki bilgiler ışığında metinsel tanıma en uygun kodu seçerek gerekli atamayı yapmaktır.

Kod atama işlemi ilk başlarda kodlayıcı denilen ve sınıflama konusunda uzman kişiler tarafından yapılmaktaydı, ancak bu iş çok zaman alıcı ve maliyetli olduğu için verimli olarak yapılamamaktaydı. Ayrıca, sürecin standardizasyon ve atanmış kodun doğruluğunun sağlanması konusunda da bir garantisi bulunmamaktaydı. Aynı tanıma farklı kodlar

atanabilmektedir. Campanelli ve diğ. [25], orta düzey kodlayıcılar arasında aynı tanıma aynı meslek kodunu atama oranının %78 ve uzman kodlayıcılar arasında bu oranın %84 olduğunu ortaya koymuşlardır. Bu eksiklikleri giderebilmek amacıyla teknolojinin gelişmesi ile birlikte kod atamasında bilgisayarlar devreye girmiştir. Bilgisayarların bu alanda yerini almasıyla birlikte, sınıflama konusunda bilgi sahibi uzman kişilere ihtiyaç azalmış ve kullanılan bilgisayar programları sayesinde atanan kodun doğruluğu sağlanmıştır. Daha kısa zamanda daha çok miktarda kod atama işleminin gerçekleştirilmesi mümkün hale gelmiştir.

TÜİK'in yaptığı anketlerde kod atama görevi anketörler tarafından yerine getirilmektedir. Anketörlerin kod atama işlemini doğru bir şekilde yapıp yapmadıklarını kontrol etmek gerekmektedir. Hali hazırda kontrol işlemleri bir program veya sistem kullanılmadan gözle yapılmaktadır. Bu bağlamda anketörlerin kayıtları kodlarken yaptığı hatalı ve eksik işlemleri gözle kontrol ederek bulmak yerine bilgisayar programı kullanarak daha hızlı ve maliyetsiz olarak yapılabileceği düşünülmektedir. Bu çalışma ile anketörlerin atadığı kodların kontrolü maliyetsiz şekilde sağlanarak anketörlerin doğru kod verme performansının artırılması hedeflenmektedir. Ayrıca, yine çalışmada hiç kod ataması yapılmamış kayıtlara anketör veya kodlayıcı gibi davranarak otomatik bir şekilde kod ataması yapabileceğinin gösterilmesi amaçlanmıştır. Bu sayede anketör maliyetlerinin azaltılması ve kodların doğruluğunun artırılması planlanmaktadır.

Bu çalışmanın amacı ve temel motivasyon kaynağı, kod atama ve sınıflamada yaşanan sorunlara çözüm olacak bir bilgisayar programı için gerekli altyapının oluşturularak programın tasarımını, geliştirmesini ve değerlendirmesini yapmaktır. Geliştirilen sistem standart sınıflama sözlüğü olan tüm sınıflamalar için kullanılabilir olmakla birlikte çalışma sadece Bireysel Tüketimin Amaca Göre Sınıflaması (COICOP) üzerinedir. Bu kapsamda, çalışmada 2013-2018 yıllarında TÜİK tarafından uygulanan Hanehalkı Bütçe Araştırması (HBA) mikro veri setlerinde bulunan daha önce anketör tarafından ataması yapılan kodlar ile Kod Atama Sistemi'nin (KASİS) atadığı kodlar karşılaştırılarak sistemin performansı ve etkinliği test edilmiştir.

Araştırmanın amacına yönelik olarak cevaplanacak araştırma soruları şunlardır:

- a. Anketör ile KASİS'in atadığı kod arasında tablolara göre farklılık var mıdır?
- b. Anketör ile KASİS'in atadığı kod arasında ana gruplara göre farklılık var mıdır?

- c. Anketör ile KASİS'in atadığı kod arasında bölgelere göre farklılık var mıdır?
- d. Anketör ile KASİS'in atadığı kod arasında yıllara göre farklılık var mıdır?
- e. Anketör ile KASİS'in atadığı kod arasında aylara göre farklılık var mıdır?
- f. KASİS sonuç verme süresi tablo ve kayıt sayısına göre farklı mıdır?
- g. Daraltılmış liste ile geniş listeden kod ataması yapılması arasında tablolara göre farklılık var mıdır?
- h. Daraltılmış liste ile geniş listeden kod ataması yapılması arasında ana gruplara göre farklılık var mıdır?
- i. Daraltılmış liste ile geniş listeden kod ataması yapılması arasında bölgelere göre farklılık var mıdır?
- j. Daraltılmış liste ile geniş listeden kod ataması yapılması arasında yıllara göre farklılık var mıdır?
- k. Daraltılmış liste ile geniş listeden kod ataması yapılması arasında aylara göre farklılık var mıdır?

1.3. Araştırmanın Önemi

İstatistiksel sınıflamalar güvenilir, karşılaştırılabilir ve metodolojik olarak sağlam uluslararası standart ve normlara uygun istatistiklerin üretilmesi için kilit araçlar olduğu için bunların doğruluğunun sağlanması bakımından kod atama faaliyeti istatistiksel süreçte bir faaliyet olarak karşımıza çıkmaktadır. Kod atama işleminde, metinsel ifade sınıflamaya uygun bir çıktıya dönüştürülmektedir. Bu metinsel ifadelerin kodlara dönüştürülmesi esnasında insan faktörü devreye girdiği için metinsel ifade doğru olsa ve doğru kodu atamaya yetecek kadar ayrıntıya sahip olsa bile atanan sınıflama kodu hatalı olabilmektedir. Bir başka durumda, yeterli ayrıntı içermeyen kod atamalarının bulunmasıdır. Örneğin, içerisinde “ekmek” ifadesi geçen 62 farklı COICOP kodu bulunmaktadır (Ek-1). Bu şekilde yazılan bir tanımla bu 62 koddan birisinin anketör tarafından atanması durumunda kod atamasının doğru yapılmış olması olasılığı 62’de 1’dir. Diyelim ki, anketör tanım kısmına sadece “ekmek” yazarak doğru bir şekilde “0111301110 - Ekmek Francala” kodunu verdi. Bu kod atamasının doğru olup olmadığını kontrol eden bir başka kişinin gerçekten bu tanımın bu koda karşılık geldiğinden emin olması mümkün değildir. Doğal olarak, bu çeşit hata ve eksiklikleri barındıran yanlış sınıflanmış veri ile yapılan araştırmalar, değerlendirmeler ve sonuçlar hatalı olmaktadır. Geliştirilen sistem, her bir tanıma ve koda eşit mesafede

yaklaşmaktadır. Tıpkı bir kodlayıcı ve anketör gibi davranarak ve tanım ile kod arasındaki maksimum uyumu yakalayarak sistematik şekilde kod atamasını gerçekleştirmektedir.

Sistemin etkinliğinin test edileceği HBA verisi, kültür harcamaları, eğitim harcamaları, sağlık harcamaları ve Tüketici Fiyat Endeksi'nde (TÜFE) girdi verisi olarak kullanılmaktadır. Bu çalışmalarda HBA'da bulunan COICOP kodları temel alınarak hesaplamalar yapılmaktadır. Bu bakımdan düşünüldüğünde, geliştirilen program ile atanan her bir kodun doğruluğu denetlenebilir bir yapıya kavuşacak ve kodlarla ilgili yapılan hesaplamaların doğruluğu sağlanmış olacaktır. Anketör tarafından yapılan kod atamalarının doğruluğu geliştirilen sistem ile kontrol edilebilir ve sorgulanabilir bir yapıya kavuşacaktır. Anketörler hangi kodlarda hataya veya eksikliğe neden olduğunun geri bildirimini alarak gelecek çalışmalarında bu geri bildirimleri kullanır hale geleceklerdir.

HBA verisinin istatistik ofisleri dışında da aşağıda yer alan geniş bir kullanıcı yelpazesi bulunmaktadır [26].

- Ekonomik ve sosyal politika planlama amaçları için kullanan kamu kuruluşları,
- Hanelerin yaşam koşulları üzerine araştırma yapan üniversiteler ile araştırma kuruluşları,
- Özel firmalar ile danışmanlar ve
- Genel olarak kitle iletişim araçları aracılığıyla bilgi alan veya istatistik ofislerinin yayınlarını kullanan diğer kullanıcılar.

Bu kadar geniş bir kullanıcı ve kullanım alanı bulunan anket verilerindeki doğru sınıflanmış her bir değişkenin çok büyük önemi bulunmaktadır. Bu çalışma sonucunda kod atama tutarlılığı ile kod atama hassasiyeti artırılarak anket maliyetlerinde tasarruf sağlanacak ve anketörün kod atama esnasında oluşturduğu görüşme yükü azalacaktır. Yapılan çalışma ile iş süreçlerinde teknolojik ve yaratıcı bir çözüm geliştirilerek istatistiki sınıflamaların daha doğru, hassas ve hızlı olarak yapılması sağlanmış olacaktır. Ayrıca, çalışma ile literatüre aşağıdaki katkılar sağlanacaktır.

- İstatistiksel sınıflamalarda yapılan kod atama işleminin önemi dolayısıyla oluşturulan bu sistemin standart sınıflama sözlüğü olan veya gelecekte oluşacak diğer sınıflamalar

da uygulanabilir olması son derece önemlidir. Bu bakımdan geliştirilen bu otomatik kod atama sistemi hem şu an mevcut sınıflamalar için hem de gelecekte oluşacak yeni sınıflamalar için kullanılabilecektir. Bu kapsamda özgün bir çalışma gerçekleştirilmiş olacaktır,

- Metinsel ifadesi bulunan her türlü sınıflama için daha önce kod ataması yapılan kayıtların otomatik olarak kontrolünün yapılacağı ve geri bildirim sağlayacak bir sistem oluşturulmuş olacaktır,
- Anket yapılma yöntemlerinin hızla değiştiği günümüz koşullarında hangi anket yöntemi kullanılırsa kullanılsın anketörler veya kodlayıcılar tarafından kod ataması yapılmış kodların doğruluğu bu sistem aracılığıyla teyit edilebilecektir,
- Bu sistem, gerektiğinde yapılan kod atamasının düzeltilmesini sağlayarak kod atama kalitesini ve hassasiyetini artıracaktır,
- Bu sistem, makine öğrenmesinde kullanılacak eğitim veri setinin doğruluğunun teyit edilmesinde ve temizlenmesinde kullanılabilecektir. Ayrıca veri tabanında birikmiş doğruluğu bu sistem tarafından onaylanmış kayıtlar bir başka denetimli makine öğrenme uygulamasında eğitim veri seti olarak kullanılabilecektir ve
- TÜFE için alternatif veri toplama yöntemlerinin kullanılması durumunda fiyatları derlenen ürünlerin COICOP kodları bu sistem aracılığıyla hatasız olarak atanabilecektir.

1.4. Tanımlar

Açık uçlu soru: Cevap veren kişinin kendi cümleleri ile cevap vermesini sağlayan soru tipidir [27].

Kapalı uçlu soru: Açık uçlu soru aksine, cevap veren kişiyi belirlenmiş seçeneklerden birisini seçmeye zorlayan soru tipidir [27].

Makine öğrenmesi: Yapay zeka kavramı, bilgisayarların insan zekasını taklit ederek sorunları çözebilmesine karşılık gelmektedir. Makine öğrenmesi ile yapay zeka birbiriyle yakından ilişkilidir. Makine öğrenmesi sayesinde, önceki tecrübeler veya örnek veri setlere dayanarak bir işlemi optimize etmek için bilgisayarlar programlanmaktadır [28].

Bilgisayar destekli yüz yüze görüşme (CAPI): Bilgisayar ile yüz yüze görüşme yaparak anket yapma yöntemidir.

Bilgisayar destekli telefon görüşmesi (CATI): Bilgisayar ile telefon görüşmesi yaparak anket yapma yöntemidir.

Bilgisayar destekli web görüşmesi (CAWI): Bilgisayar ile web görüşmesi yaparak anket yapma yöntemidir.

Kalem ve kağıt görüşmesi (PAPI): Kalem ve kağıt kullanılarak anket yapma yöntemidir.

Barkod verileri (Scanner data): Tüketim mallarının satışına ilişkin ayrıntılı verilerdir [29].

Web kazıma (Web crawling veya web scraping): Web sayfalarından veri çekme işlemidir [30].

Hanehalkı: Tek başına yaşayan kişiler ile aralarında akrabalık bağı bulunsun ya da bulunmasın aynı konutta yaşayan, barınma, gıda vb. ihtiyaçlarını ortaklaşa karşılayan, hanehalkı hizmet veya yönetimine katılan hanehalkı fertlerinden oluşan topluluktur [31].

Hanehalkı ferdi: Hanehalkını meydana getiren topluluğun bir üyesi olup hanehalkının oturduğu adreste daimi olarak ikamet eden ya da ikamet etme niyetinde olan kişidir. Bununla birlikte askerde, hapiste olanlar ve huzurevlerinde kalan yaşlılar ve üniversite yurdunda kalan öğrenciler, hanede kalış süresi ne olursa olsun misafirler ve hanehalkından evlenme, askere gitme, çalışmaya gitme gibi sebeplerle kesin olarak ayrılanlar hanehalkı ferdi olarak kapsanmamaktadır [31].

Tüketim: Mal ve hizmetlerin insan ihtiyaçlarını ve arzularını karşılamak üzere kullanılmasıdır [31].

Harcama: Bir mal veya hizmet satın almak için yapılan parasal ödeme veya fedakarlıkların tümüdür [31].

Mal ve hizmet satın alışıları: Herhangi bir mal ya da hizmete belirli bir ücret ödeyerek

tüketmek ve özel kişi veya kuruluşlara hediye veya yardım olarak vermek için sahip olunmasıdır [31].

Hanenin üreterek tükettiği ürünler: Hanelerin kendi üretimleri sonucu elde ettikleri maddeleri, ya ürettikleri gibi ya da işleyerek tüketmeleridir. Örneğin; hanehalkının kendi ürettiği sütün ya süt olarak ya da peynir yapılarak tüketilmesi gibi [31].

İşyeri sahibi olanların işyerinden getirdiği ürünler ile çalışanlara işverenlerinin verdiği mal ve hizmetler: Hanede iktisaden faal olan fertlerin çalıştıkları işyerinde üretilen ya da satışa sunulan mal ve hizmetlerden hanede tüketilmesidir. Örneğin; ücretli olarak çalışan bir ferde işyerince verilen mal (yiyecek, giyecek vb.) ve hizmetler, ücretlinin elde ettiği aynı gelirden tüketimi; işveren ya da kendi hesabına çalışan bir ferdin de kendisine ait işyerinde ürettiği ya da sattığı mallardan hanesine tüketmek üzere getirdikleri aynı gelirden tüketimidir [31].

Hanehalkının özel kişi veya kuruluşlara vermek üzere hediye veya yardım şeklinde satın aldığı mal veya hizmetler: Hanehalkı fertleri tarafından, özel kişi veya kuruluşlara verilmek üzere hediye etmek veya yardımda bulunmak amacıyla satın alınan mal ve hizmetlerin değeridir [31].

İzafi kira: Konuta mülkiyet şekli ev sahibi, lojman ve diğer (babasına, akrabasına vb. kişilere ait yerlerde ikamet edip hiç bir şekilde ücret ödemeden ya da çok düşük değerinde kira ödeyenler) şekilde olan hanehalkının ikamet ettiği konutun kira değeridir [31].

Eşleşen kod: Anketörün atadığı kodun doğruluğu KASİS tarafından teyit edilmesi durumunda kod bu sınıfta yer almaktadır.

Sözlük eşleşen kod: Anketörün atadığı kodun doğruluğu KASİS tarafından teyit edilmesi ve anketörün yazdığı harcama tanımının sözlükte birebir yer alması durumunda kod bu sınıfta yer almaktadır.

Yeni eşleşen kod: Anketörün atadığı kodun doğruluğu KASİS tarafından teyit edilmesi ve anketörün yazdığı harcama tanımının sözlükte birebir yer almaması durumunda kod bu sınıfta yer almaktadır.

Bulanık eşleşen kod: Anketörün yazdığı harcama tanımındaki yazım yanlışı KASİS tarafından düzeltilerek anketörün atadığı kodun doğruluğu KASİS tarafından teyit edilmesi durumunda kod bu sınıfta yer almaktadır.

Şüpheli kod: Anketörün atadığı kodun hatalı olduğu KASİS tarafından teyit edilmesi durumunda kod bu sınıfta yer almaktadır.

Yetersiz kod: Anketörün yazdığı harcama tanımı ile atanabilecek birden fazla kodun olduğu KASİS tarafından teyit edilmesi durumunda kod bu sınıfta yer almaktadır.

Kabul kod: Anketörün atadığı kodun sistem tarafından önerilen kodların arasında olduğu, anketörün atadığı kod ile önerilen kodların ilk beş basamağının aynı olduğu ve önerilen kod sayısının beşten az olduğu KASİS tarafından teyit edilmesi durumunda kod bu sınıfta yer almaktadır.

Bulanık kabul kod: Anketörün yazdığı harcama tanımındaki yazım yanlışı KASİS tarafından düzeltilerek anketörün atadığı kodun kabul kod olduğu KASİS tarafından teyit edilmesi durumunda kod bu sınıfta yer almaktadır.

2. KAVRAMSAL ÇERÇEVE VE İLGİLİ ARAŞTIRMALAR

Bu bölümde araştırmanın kavramsal temelini oluşturan konular hakkında detaylı bilgi verilmiş ve ilgili çalışmalar aktarılmıştır.

2.1. Anket Yöntemleri ve Kod Atama

Günümüzde anketlerin yapılma şekilleri farklı boyutlara ulaşmıştır. Bilgisayar destekli yüz yüze görüşme yaparak anket yapma yöntemine CAPI denilmektedir. Bu yöntemde ankete cevap veren ile anketör bir arada bulunmaktadır. Bilgisayar destekli telefon görüşmesi yaparak anket yapma yöntemine CATI denilmektedir. Bu yöntemde anketör ile ankete cevap veren telefon aracılığıyla görüşmektedir. Bilgisayar destekli web görüşmesi yaparak anket yapma yöntemine CAWI denilmektedir. Bu yöntemde, ankete cevap veren bilgisayar aracılığıyla ankete cevap vermektedir. Diğer anket yöntemlerinden farklı olarak kağıt ve kalem kullanılarak yapılan PAPI yönteminin uygulanması için neredeyse hiçbir teknik uzmanlık gerektirmez ve anket tasarımının uygulanmasında büyük bir esneklik barındırmaktadır. Anket yöntemlerinin kendine özgü bazı üstün ve zayıf yönleri bulunmaktadır.

CAPI'nin üstünlükleri [32]:

- Hızlı olması,
- Kaliteli veri toplamaya olanak sağlaması,
- Veri girişinin anlık yapılabilmesi ve
- Soru akışında kolaylık sağlamasıdır.

Maliyetinin yüksek olması ise CAPI'nin zayıf yönüdür [32].

CATI'de süre kısıtı zayıf yön iken üstünlükleri [32]:

- Etkili iletişim,
- Kaliteli veri toplamaya olanak sağlaması,
- Hızlı olması,

- Cevaplayıcıların etkilenmeden cevaplarını iletebilmesi ve
- Veri giriş, düzeltme ve kod işlemlerinin daha hızlı gerçekleştirilmesidir.

CAWI'nin üstünlükleri [32]:

- Süreye cevaplayıcının karar vermesi,
- Sürenin etkin kullanılması,
- Azalan maliyetler,
- Anketin görselleştirilebilmesi,
- Soru akışında kolaylık sağlaması ve
- Lokasyondan bağımsız olmasıdır.

CAWI'nin zayıf yönleri ise [32]:

- Güven kaygıları,
- Cevaplayıcının kim olduğunun tespitinin imkansızlığı ve
- Cihaza bağımlı olunmasıdır.

PAPI, ileri programlama becerisine sahip olmayan biri tarafından bile tasarlanabilmektedir. Anketler, çok hızlı bir şekilde tekrarlanabildiği için anketlerin sürekli iyileştirilmesi veya yeniden düzenlenmesi beklenen pilot anketler için idealdir [33].

Anket uygulayarak veriye ulaşmak sıklıkla kullanılan bir yöntemdir. Ankette alınan metinsel bir cevaba karşılık gelen bir kodun atanması anket esnasında yapılabildiği gibi anketten sonra kodlayıcılar tarafından da yapılabilmektedir. Anket yöntemi ile kod atamasının nasıl yapıldığı ve yapılan kod atamasının avantaj – dezavantajları Çizelge 2.1'de yer almaktadır [10].

Çizelge 2.1. Kod atamasının olası yerleri ve kim tarafından / ne yapıldığı?

Kim / Ne?	Nerede?	Avantajları	Dezavantajları
Cevaplayıcı	CAWI	Doğrudan geribildirim	Sınıflama bilgisi yok
Anketör	CATI, CAPI, PAPI	Doğrudan geribildirim	Sınıflamanın yüzeysel bilgisi
Kodlayıcı	Ofis	Sınıflama uzmanı, Ek bilgileri de kullanabilir, Genel olarak, cevapları bir bilgisayar programından daha iyi yorumlayabilir.	Doğrudan geribildirim her zaman mümkün değildir, Geribildirim çok zaman alır, Kod ataması tutarsız olabilir.
Bilgisayar programı	Ofis	Hızlı ve tutarlı kod atama, Kod atama bilgisi bir sistemde tutulur ve bu nedenle transfer edilebilir, Gece ve gündüz çalışabilir.	Doğrudan geri bildirim yok, Çoğunlukla toplu halde kod ataması yapılır.

Anketin CAWI yöntemi ile yapılması durumunda, anket programı ile cevaplayıcı baş başa kalmaktadır. Cevaplayıcıya anket programı aracılığıyla anlık geri bildirimlerde bulunulmakta ancak cevaplayıcının sınıflama bilgisi olmadığı için bu geri bildirimleri yorumlayamamaktadır. Anketin CATI, CAPI ve PAPI yöntemlerinden biri ile yapılması durumunda, cevaplayıcı anketör ile baş başa olmaktadır. Veri girişi de eş zamanlı olacağı için, anketör cevaplayıcıya anket programının anlık geri bildirimlerini aktarabilmekte ve gerekli düzeltmeler yapılabilmektedir. Bu yöntemlerde sınıflama bilgisi bakımından uzman olmayan anketör bazen hatalı kayıtlara sebebiyet verebilmektedir.

Kod atamasının ofiste uzman kodlayıcılar tarafından yapılması durumunda, anketteki tüm bilgiler kullanılacağı için kod atama kalitesi yükselmektedir. Ancak, cevaplayıcı artık karşımızda olmadığı için ilave sorularımıza cevap verememekte ve cevaplayıcıya ilave sorular sormak için tekrar iletişime geçmek gerekmektedir. Bu durum maliyetlerimizi artırmaktadır. Ayrıca, aynı tanıma farklı kodlayıcılar farklı kod ataması yapabilmekte ve atanan kodlarda tam anlamıyla tutarlılık sağlanamamaktadır.

Kod atamasının ofiste otomatik program tarafından yapılması durumunda, hızlı ve tutarlı kod atama sayesinde kod atama kalitesi yükselmektedir. Program, mesai saatlerinden bağımsız olarak gece ve gündüz çalışabilmektedir. Ancak, bu yöntemde cevaplayıcı ile artık aynı ortamda olunmadığı için cevaplayıcı ilave sorulara cevap verememekte ve tekrar iletişime geçmek gerekmektedir. Bu durum maliyetleri artırıcı yönde etki etmektedir.

2.2. Tüketici Fiyat Endeksi (TÜFE)

TÜFE'nin ana veri kaynağı HBA verileridir. Ağırlıkların tespitinde ve hesaplamasında HBA'da bulunan COICOP kodları kullanılmaktadır. TÜFE, fiyatların değişimini ölçmektedir. TÜFE'nin veri kaynakları, HBA, kurumsal nüfus bireysel tüketim harcamaları anketi, turizm anketi ve idari kayıtlardan elde edilen harcama ve ciro bilgileridir. Sepet içerisinde bulunan maddeler ve ağırlıkları her yıl revize edilmektedir. 81 il merkezi ve 225 ilçede bulunan 28 019 işyerinden 553 064 fiyat derlenmekte ve 4 274 kiracı aylık olarak takip edilmektedir. İşyeri ve fiyat sayıları mevsimsel olarak değişkenlik göstermektedir. 26 bölgede yaklaşık 250 personel derlenen fiyatları anlık olarak veri tabanına göndermektedir [34].

TÜİK'te TÜFE kapsamında alternatif veri toplama yöntemleri konusunda çalışmalar yapılmaktadır. Tüketici fiyat endeksi hesabında web kazıma ve barkod verileri alternatif veri kaynağı olarak daha yaygın kullanılmaktadır. Son birkaç yılda, bu veri kaynaklarından fiyat endekslerinin derlenmesi için yeni yöntemler geliştirilmiştir. Çizelge 2.2'de gösterildiği üzere geleneksel fiyat verileri toplama ile bu alternatif veri kaynakları arasında önemli farklılıklar bulunmaktadır [30].

Çizelge 2.2. TÜFE'de farklı veri kaynaklarının özellikleri

Özellik	Geleneksel toplama	Web kazıma	Barkod verileri
Toplama	Çoğunlukla yerel yollarla toplanan, web siteleri, kataloglar ve telefon aracılığıyla merkezi toplama ile desteklenmektedir.	Perakendecilerin web sitesinden otomatik web kazıması yapılarak.	Satın alma noktasında toplanmaktadır. Doğrudan satıcı tarafından veya üçüncü bir tarafça tedarik edilmektedir.
Yer	Ülke çapında birden fazla perakendeciden ve birden fazla konumdan.	Çevrimiçi olan orta ve büyük ölçekli perakendecilerin fiyatları.	Genellikle verilerini toplayan orta ve büyük ölçekli perakendecilerin fiyatları.
Zamanlama	Ayın bir günü etrafında toplanan fiyatlar (endeks günü).	Günlük alınan fiyat.	Her işlem için toplanan fiyat ve miktar toplu biçimde temin edilebilir.
Örnekleme	Temsili ürünler seçilir.	Tüm ürünler için fiyatlar.	Tüm ürünler için toplanan tüm işlem verileri.
Kontroller	Gözle yapılan kontroller ve otomatik doğrulama.	Fiyatları ve sınıflamaları gözle kontrol etmek pratik değildir, makine öğrenimi gibi veri bilimi teknikleri kullanılmaktadır.	
Zaman serisi	Zamanla ürünleri eşleştir.	Yüksek ürün kayması.	

Çizelge 2.2. (devam) TÜFE’de farklı veri kaynaklarının özellikleri

Özellik	Geleneksel toplama	Web kazıma	Barkod verileri
Maliyet	Yüksek maliyet.	Fiyat teklifi başına potansiyel maliyet tasarrufu sağlamakta, ancak işleme, bakım ve depolama maliyetleri dikkate alınmalıdır. Ayrıca üçüncü bir şahıstan veri elde etmenin potansiyel maliyeti de vardır.	
Hacim	Küçük-orta çaplı toplama.	Geniş çaplı toplama.	
Zamanlılık	Yayın, veri toplandıktan sonraki yaklaşık bir ay sonra gerçekleşir.	Web kazıma işlemi bünyede yapılırsa, fiyat endeksleri, işlem hızına bağlı olarak veri toplandıktan sonra neredeyse anlık olarak hesaplanabilir.	Bilinmeyen, üçüncü taraf tedarikçi ile sözleşmeye bağlı olarak.

Hangi yöntem ile fiyat derlenirse derlensin fark etmez, tüketici fiyat istatistiklerine bir fiyatın konu olabilmesi ancak ve ancak bu fiyata ilişkin ürünlerin COICOP kodlarının doğru olarak verilmesine bağlıdır. Geleneksel veri toplama yönteminde kaynak olarak HBA veri seti kullanıldığı için COICOP kodları hazır olarak bulunmaktadır. Diğer yollarla veri toplama yöntemlerinde ise ürünler için COICOP kodlarının atanması gerekmektedir.

2.3. Hanehalkı Bütçe Araştırması (HBA)

HBA, hanelerin tüketim harcama seviye ve bileşimlerinin neler olduğu hakkında bilgi veren, en önemli kaynaklardan birisidir [31].

2.3.1. Tarihçe

TÜİK, tüketim harcama istatistiklerini hesaplayarak yayımlamak amacıyla değişik isimler altında, kapsam ve periyotlarda 1954 yılından itibaren anketler yapmıştır. Anket, 2002 yılından başlayarak belirli bir düzene oturarak her yıl yapılır hale gelmiştir. HBA’nın tarihçesine Çizelge 2.3’te yer verilmiştir [31].

Çizelge 2.3. HBA'nın tarihçesi

Yıl	Çalışma adı	Kapsam
1954	Aile Bütçesi Anketi	Ankara'daki memur aileleri
1964-70 1973-74 1978-79	Aile Bütçesi Anketi	İlk aşamada, Adana, Ankara, İzmir ve İstanbul illeri. İkinci aşamada, Samsun, Antalya, Diyarbakır, Bursa, Ordu, Erzurum, Eskişehir illeri.
1987	Hanehalkı Gelir ve Tüketim Harcamaları Anketi	Türkiye geneli
1994	Hanehalkı Gelir ve Tüketim Harcamaları Anketi	Türkiye geneli
2002-devam	Hanehalkı Bütçe Anketi	Türkiye geneli

2013-2014 HBA, Ocak–Aralık ayları arasında aylık 1 104, yıllık 13 248 hanehalkına uygulanmıştır. Cevaplı hanehalkı sayısı 2013 yılında 10 060 ve 2014 yılında ise 10 122 hanehalkı olmuştur [31].

2015 HBA, Ocak–Aralık ayları arasında yıllık 15 264 hanehalkı ile anket gerçekleştirilmiş ve cevaplı hane sayısı 11 491 olmuştur [31].

2016, 2017 ve 2018 HBA, yıllık 15 552 hanehalkı ile gerçekleştirilmiş ve cevaplı hane sayısı sırasıyla 12 096, 12 166 ve 11 828 olmuştur [31].

2.3.2. Amaç

HBA ile;

- Tüketici fiyat endekslerinde kullanılacak maddeler ile ağırlıkların belirlenmesi,
- Hanelerin tüketim harcamalarındaki değişikliklerin izlenmesi,
- Milli gelir hesaplamalarında kullanılacak verilerin elde edilmesi,
- Asgari ücret için verilerin elde edilmesi ve
- Mutlak yoksulluk sınırının belirlenmesi

amaçlanmaktadır [31].

2.3.3. Kapsam

Anketin kapsamı [31]:

- Hanehalkı odaklı,
- Tüketim odaklı,
- COICOP sınıflamasındaki 12 ana harcama grubu odaklı ve
- Belirli bir dönem odaklı

Tüketim harcaması bileşenleri [31]:

- Mal ve hizmet satın alıřları,
- Hanenin üreterek tükettiđi ürünler,
- İşyeri sahibi olanların işyerinden getirdiđi ürünler,
- Çalışanlara işverenlerinin verdiđi mal ve hizmetler,
- İzafi kira ve
- Son bir yıl içerisinde alınan dayanıklı tüketim malları.

2.3.4. Alan uygulaması

Hanehalklarından bilgiler; görüşme, kayıt ve gözlem metotları kullanılarak derlenmektedir. Her anketör, ortalama 6 hanehalkını ayda ortalama 6 defa ziyaret etmektedir [31]. Anket ayı öncesinde örnek haneye yapılan ilk ziyarette, hanehalkının fert bileřimi, yařadığı konutun özellikleri ve tüketim alışkanlıklarına ilişkin bilgiler alınmaktadır. Anket ayı içinde yapılan ziyaretlerde, hanenin anket ayında yaptıđı tüketim harcamaları kayıt ve görüşme yöntemleri ile elde edilmektedir. Anket ayı bitiminde yapılan son görüşmede ise, hanehalkı fertlerinin gelir bilgileri derlenmektedir [31].

2.4. Bireysel Tüketimin Amaca Göre Sınıflaması (COICOP)

COICOP, hanehalkı harcamalarının uluslararası referans sınıflamasıdır ve tüketim harcamaları kapsamında yer alan mal ve hizmetlerin bir çerçevesini sağlamaktadır. HBA'da, tüketim harcamalarının sınıflaması olarak COICOP sınıflaması kullanılmaktadır [35].

2.4.1. Tarihçe

COICOP, 1968'den bu yana Ulusal Hesaplar Sistemi'nin (SNA) işlevsel sınıflamalar olarak da bilinen amaca yönelik harcama sınıflamalarının bir parçasıdır. Hanehalkı tüketim harcamalarını uluslararası düzeyde sınıflama fikri, ilk kez Ekim 1923'te ILO tarafından düzenlenen Uluslararası Çalışma İstatistikçileri Konferansı'nda (ICLS) ortaya atılmıştır [35]. Bu konferansta, harcama ile ilgili bilgilerin gelirden daha ayrıntılı olması ve harcamaların az sayıda açık bir şekilde tanımlanmış gruplara göre derlenmesi gerektiği gündeme gelmiştir [36].

Ekim 1926'da yapılan 3. ICLS'te uluslararası harcama sınıflaması konusu tartışılmış ve tüketim harcamalarının gıda, giyim, konaklama, yakıt ve enerji, mobilya ve mefruşat ve çeşitli gruplarına ayrılmasına karar verilmiştir. 1951'de 7. ICLS'te öncekinden daha ayrıntılı olan harcama grupları kabul edilmiştir [35].

1953 SNA, mal ve hizmet türüne göre bir sınıflama sağlamaktadır. Sınıflama, hanehalkı nihai tüketimi için 12 kategorinin yanı sıra, ulusal konsepti uluslararası konseptte uyarlamak için üç ayrı kategoriyi de içermektedir. Sınıflama, ILO konferansı tarafından neredeyse 30 yıl önce önerilen liste gibi, ihtiyaçlar için gıda, giyim ve konaklama olmak üzere aynı temel kategorileri içermektedir. Ancak, konaklama daha ayrıntılı gruplara bölünmüştür. Mevcut kira, yakıt ve enerji, mobilya ve ev eşyası gruplarına ek olarak, ev faaliyetleri ayrı bir grup olarak oluşturulmuş ve çeşitli grubu kişisel bakım ve sağlık giderleri, ulaşım ve iletişim, rekreasyon ve eğlence ile çeşitli hizmetlere ayrılmıştır [35].

1968 SNA, hanehalkı eşyalarının ve hizmetlerin sınıflaması olarak adlandırılan sınıflamanın bir sonraki sürümünü sağlamaktadır. Bu sınıflama, tüketimin ana amacına göre belirlenen sekiz bölümden oluşmaktadır [35]:

1. Gıda, içecek ve tütün
2. Giyim ve ayakkabı
3. Kira, yakıt ve enerji
4. Mobilya, mobilya ve ev eşyası ve işletme
5. Tıbbi bakım ve sağlık giderleri
6. Ulaştırma ve iletişim

7. Rekreasyon, eğlence, eğitim ve kültürel hizmetler
8. Çeşitli mal ve hizmetler.

Sınıflama, mal ve hizmet ayrımı ve dayanıklı, yarı dayanıklı ve dayanıklı olmayan mallar arasında ayrım yapabilmek için tasarlanmıştır. 1968 SNA'da yer alan hanehalkı mal ve hizmetlerinin sınıflaması, Mart 1999'da Birleşmiş Milletler İstatistik Komisyonu (UNSC) tarafından diğer üç işlevsel sınıflama ile birlikte revize edilmiştir. Diğer üç sınıflama, Hanehalkına Hizmet Veren Kar Amacı Gütmeyen Kuruluşların Sınıflaması (COPNI), Devlet Fonksiyonlarının Sınıflaması (COFOG) ve Üreticilerin Giderlerinin Amaca Göre Sınıflaması'dır (COPP) [35].

2.4.2. Sınıflamaya genel bakış ve kapsam

SNA, ekonomik faaliyet ölçümlerinin nasıl derleneceğine ilişkin uluslararası kabul görmüş standart önerilerden oluşmaktadır. SNA, bir dizi uluslararası kabul görmüş kavram, tanım, sınıflama ve muhasebe kuralları bağlamında tutarlı ve bütünlüklü bir makroekonomik hesaplar dizisi olarak tanımlanmaktadır. COICOP, 1993 SNA'nın ayrılmaz bir parçasıdır, aynı zamanda HBA, TÜFE ve Gayri Safi Yurtiçi Hasıla (GSYİH) ve bileşenlerinin uluslararası karşılaştırmaları için de kullanılmaktadır [35].

Tüm bu çalışmalarda, ülkeler tarafından sınıflama daha alt kırılımlarda oluşturularak kullanılabilir. Ancak, bu farklı alanlardaki istatistiklerin ülkeler arasında karşılaştırılabilirliğinin sağlanabilmesi bakımından COICOP'un temel yapısının korunması özellikle istenmektedir. COICOP, ulusal istatistik ofislerinin kendi kullanımları için geliştirdikleri tüketim harcamalarının sınıflamasına dayanmaktadır. Her ne kadar COICOP belirli bir tüketici davranışı modeline sıkı sıkıya bağlı olmasa da, sınıflama gelir esnekliklerindeki farklılıkları geniş ölçüde yansıtacak şekilde tasarlanmıştır. Örneğin, düşük gelirli hanehalkları bütçelerinin nispeten yüksek oranlarını gıda, giyim ve konut harcamalarına ayırırken, daha zengin hanehalkları ulaşım, eğitim, sağlık ve eğlence konularında daha fazla harcama yapmaktadır. Ayrıca, COICOP sınıfları, hizmet (S), dayanıklı olmayan (ND), yarı dayanıklı (SD) ve dayanıklı (D) olarak ayrılmıştır [35].

Avrupa Bireysel Tüketimin Amaca Göre Sınıflaması (ECOICOP), COICOP sınıflamasının özel bir şeklidir. Orijinal ECOICOP sınıflaması beş seviyeden oluşan hiyerarşik bir yapıya

sahiptir. 12 adet düzey 1 kategorisi Şekil 2.1’de verilmiştir [6].



Şekil 2.1. COICOP sınıflamasının düzey 1 kategorileri

47 adet düzey 2, 117 adet düzey 3 ve 303 adet düzey 4 kategorisi bulunmaktadır. Örnek hiyerarşik yapı aşağıda Çizelge 2.4’te verilmiştir [6].

Çizelge 2.4. COICOP sınıflamasının hiyerarşik yapısı

Kod	Düzy	Tanım	
1	1	Gıda ve alkolsüz içecekler	
1.1	2	Gıda	
1.1.6	3	Meyveler	
1.1.6.1	4	Taze veya soğutulmuş meyveler	<i>Kapsananlar:</i> - karpuz ve kavun, üzüksü meyveler
1.1.6.2	4	Dondurulmuş meyveler	<i>Kapsananlar:</i> - dondurulmuş meyveler - dondurulmuş üzüksü meyveler
1.1.6.3	4	Kurutulmuş meyveler ve sert kabuklu meyveler	<i>Kapsananlar:</i> - kurutulmuş meyveler, meyve kabukları, meyve çekirdekleri - sert kabuklu meyveler ve yenilebilir tohumlar - kuru üzüksü meyveler

Çizelge 2.4. (devam) COICOP sınıflamasının hiyerarşik yapısı

Kod	Düzy	Tanım
1.1.6.4	4	Korunmuş meyveler ve meyve bazlı ürünler <i>Kapsananlar:</i> - korunmuş meyveler ve meyve bazlı ürünler - diyet müstahzarları ve yemeklere tat vermek için kullanılan yalnızca meyve bazlı malzemeler - konserve edilmiş veya kutulanmış meyveler

Avrupa Birliği ülkelerinde HBA anketleri centilmenlik sözleşmesi kapsamında yürütülmektedir. Bu anlaşmaya 1989 yılında yapılan İstatistik Program Komitesi (SPC) toplantısında ulaşılmıştır [37]. Eurostat, üye ülkeler ve aday ülkelerden HBA veri setleri beş senede bir, beş basamaklı düzey 4 bazında tüketim harcamalarını içerecek şekilde göndermelerini istemektedir. Anket, centilmenlik anlaşmasına dayanarak yapıldığından, her ülke anketin amacı, metodolojisi ve çalışma sıklığına kendisi karar vermektedir. Bu bakımdan her ne kadar anketin genel yapısı benzer olsa da ülkeler arasında farklılıklar bulunmaktadır.

TÜİK, HBA anketinde tüketim harcamaları sınıflaması olarak 10 basamaklı ECOICOP kodlarını kullanmaktadır ve her beş senede bir beş basamaklı düzey 4 bazında tüketim harcama veri setini Eurostat'a göndermektedir.

2.5. Kod Atama ve Kod Atama Sistemleri

Kod atama faaliyetini yerine getirmek için elle yapılan kod atama, bilgisayar destekli kod atama ve otomatik kod atama yöntemleri bulunmaktadır. Kod atama faaliyeti için hangi yöntemin en iyi olduğu sınıflamanın karmaşıklığına ve verinin büyüklüğüne bağlı olarak değişebilmektedir. Yöntemlerin avantajlarından faydalanabilmek amacıyla bu sistemlerin kombinasyonları uygulamada bir arada kullanılabilir [9].

2.5.1. Elle yapılan kod atama

Metinsel cevaplara kod atama işlemi birkaç yıl öncesine kadar elle olarak yapılmaktaydı. Sınıflamalar daha karmaşık hale geldikçe ve veri büyüdükçe kod atama işleminin klasik yöntemlerle sürdürülmesi bu işi çok maliyetli bir hale getirmiştir. Maliyete katlanılsa bile

yapılan kod atamasının kalitesi düşmüştür.

2.5.2. Bilgisayar destekli kod atama

Kod atamadaki etkinliği artırmak amacıyla zaman içerisinde daha yaygın olarak bilgisayarlar kullanılmaya başlanmıştır. Bilgisayar destekli kod atama, bilgisayarların kodlar ve tanımları içeren bir dosyada arama ve tarama yapmalarını tanımlamaktadır [10].

Anket sorularının çoğunda kod atama işlemi, cevap veren tarafından örtülü olarak yapılmaktadır. Yani, cevap verene kapalı bir soru sorulduktan sonra, verilen cevaba göre en uygun seçenek işaretlenmektedir. Bazen, sınıflama sözlüğü çok fazla sınıf içeriyorsa veya cevap veren için çok karmaşık ise, cevap verene açık uçlu bir soru sorulmakta ve cevap profesyonel bir kodlayıcı tarafından kategorize edilmektedir. Bu yöntemde, kodlayıcı işini ofiste yaparken cevap veren karşısında olmayacağı için bazı detaylar eksik kalabilmektedir. Bu duruma bir çözüm olarak, görüşme sırasında gerekli tüm bilgileri sorgulayabilen anketöre kod atama görevi verilmesi önerilmektedir [38].

Bilgisayar destekli kod atama, özel olarak tasarlanmış bilgisayar programlarıyla kod atama görevini kolaylaştırmak için kullanılmaktadır. Hangi kodun doğru olduğuna karar verme işi kodlayıcı da kalsa da, program genellikle az sayıda kod önererek kodlayıcıya yardım etmektedir. Meslek kodları için Bushnell, bir bilgisayar programının kod atama sürecini hızlandırabileceğini ve aynı zamanda kod atama kalitesini artırabileceğini göstermiştir [39]. Açık uçlu soruların anketlere entegrasyonu uzun süredir devam eden bir metodolojik kaygıdır. Bu sorunu giderebilmek amacıyla çeşitli program çözümleri önerilmiştir. Fielding ve diğerleri, açık uçlu sorularına verilen cevapların yeterince analiz edilmemesi ve hatta cevapların doğru bir şekilde kaydedilmemesine Bilgisayar Destekli Nitel Veri Analizi (CAQDAS) programının çözüm olacağından bahsetmiştir [40]. Program, açık uçlu sorulara verilen cevapların analizine ilgi duyan araştırmacılara yöneliktir. Esuli ve Sebastiani, kullanıcı tarafından belirlenen herhangi bir kod çerçevesinde metinsel cevapları otomatik olarak kodlayan ve kullanımı son derece pratik ve kullanışlı olan Kelimesi Kelimesine Kodlama Sistemi'ni (VCS) tanıtmıştır [13].

TÜİK, yaptığı tüm anketlerde alan uygulaması sırasında bilgisayar destekli kod atama yöntemini kullanmaktadır. Bilgisayar destekli kod atama, kodlayıcılar tarafından gözlem

bazında gerçekleştirilen ve kod atama görevini kolaylaştıran tüm teknikleri içermektedir. Bu yaklaşım, bazen otomatik bir kod atama sisteminden daha uygun bir kod bulma şansı sağlamakla birlikte kodlayıcı sınıflama hakkında üst düzey bilgisini gerektirmektedir. Üst düzey bilgisi olan kodlayıcılar tarafından kod atama işlemi bir şekilde yerine getirilse bile verinin büyük miktarlarda olması durumunda yapılan kod atamasının doğruluğunu ve kalitesini ölçmek için klasik kontrol yöntemlerin etkisiz kalacağı ve yine otomatik bir kontrol sistemine ihtiyaç duyulacağı kolayca anlaşılabilmektedir.

2.5.3. Otomatik kod atama

Otomatik kod atama, herhangi bir müdahale gerektirmeyen bir kod atama algoritmasının uygulanması anlamına gelmektedir. Otomatik kod atamada genellikle bir listedeki kayıtlara kod ataması yapılmaktadır. Bu türden bir algoritma başarılı bir şekilde çalıştığında, atanan kod daima tek bir kod olmaktadır. Ancak otomatik kod atamada ele alınması gereken iki büyük sorun bulunmaktadır [10]:

1. Yazılı metnin yorumlanması ve
2. Tanımların uyması gereken karmaşık sınıflama.

Clarke ve Brooker, metinsel bir tanıma kod atama işleminin doğrudan insan katılımı olmadan bilgisayar tarafından yapılmasını otomatik kod atama olarak tanımlamaktadır [41]. Bilgisayar programı hangi kodun metne en uygun olduğunu seçebilmelidir. Bu aşamada, yazılı metinle ilgili aşağıdaki sorunlar ortaya çıkabilmektedir [10]:

1. Yazım sorunları
2. Dil bilgisi problemleri (kelimeler arasındaki ilişkiler, söz dizilimleri)
3. Anlamsal problemler (kelimelerin anlamı, kavramlar, cümle parçaları, tek bir cümle, birkaç cümle)
4. Yorumlama problemleri (sınıflamadaki hangi kod metne en iyi şekilde uyar).

Metinle ilgili diğer bir sorun da, sınıflama açısından bakıldığında, bir metnin eksik ifadeler içermesi veya cevabın iki veya daha fazla kodla ilgili olmasıdır. Bu sorun, metnin beklenenden daha ayrıntılı ifade içermesi nedeniyle de ortaya çıkabilmektedir [10]. Metinsel bir cevaba otomatik olarak kod atamak için, çoğu yöntem bir sözlükten veya

benzer anlamdaki diğer cevaplardan yararlanmaktadır [9]. Bu işlemlerin tümü doğrudan insan katılımı olmayan bir bilgisayar tarafından yapıldığından, otomatik kod atama olarak adlandırılmaktadır.

Otomatik kod atama, basit olarak mantıksal kurallar kullanılarak da yapılabilir. Bu durumda, tam olarak belirtilen koşul gerçekleşince öngörülen kod atanmaktadır. Meslek kod ataması için, çeşitli yazarlar bu tekniği uygulamışlardır. Geis [42], Drasch ve diğerleri [43], Jung ve diğerleri [44] ve Conrad [38] bu tekniği kod atamanın ilk adımı olarak kullanmışlardır. Bu sistemlerde 1 000 tane kural olsa dahi, meslek kodlarına doğru bir şekilde kod ataması yapılmasının bir garantisi bulunmamaktadır. Kuralda yer almayan duruma uygun yeni bir kayıt geldiğinde sistem bir sonuç üretememektedir. Hartmann ve Schütz daha yüksek başarı oranları için daha fazla kural oluşturmuşlardır [45].

Otomatik kod atama uygulamalarında son yıllarda denetimli makine öğrenmesi teknikleri kullanılmaktadır. İstatistiksel bağlamda, makine öğrenmesi, istatistiki verilerin işlenmesi veya işlenmesine yardımcı olmak için algoritmalar yoluyla mevcut bilgilerin kullanıldığı yapay zekanın bir uygulaması olarak tanımlanabilir. Makine öğrenmesi otomasyon kavramlarını içerirken, insan rehberliği de gerektirmektedir [46]. Sınıflama problemlerinin çözümünde makine öğrenimi algoritmasının seçiminde dikkat edilmesi gereken kriterlerden biri, algoritmanın genele uygulanabilme performansıdır. Bu performans ne kadar iyi olursa algoritmanın başarısı o denli iyi olmaktadır.

Bethmann ve diğ. [47], Alman panel araştırmalarında otomatik meslek kod ataması için iki tür olasılıksal denetimli makine öğrenme algoritması uygulamış ve eğitim verisi olarak yaklaşık 300 000 adet elle kod ataması yapılmış meslek kodu ve tanımlarını kullanmışlardır. Yazarlar, algoritmanın girdi bilgisi olarak kullanılan eğitim verisinin kaliteli olması durumunda, meslek kodlarının otomatik olarak yüksek başarı ile kodlanabileceği sonucuna varmışlardır.

İrlanda Merkez İstatistik Ofisi, daha önce kod ataması yapılmış kayıtları eğitim verisi olarak kullanarak, COICOP için otomatik bir kod atama sistemi üzerinde çalışmaktadır [46]. Lüksemburg İstatistik Ofisi, barkod verileri projesinde, perakende satış işlemlerini TÜFE'nin derlenmesinde kullanılan COICOP'a göre sınıflamak için kullanmaktadır. Proje, maliyetleri arttırmadan daha büyük miktarda veri kullanımına izin vermektedir [48].

Kanada İstatistik Ofisi, ödeme işlemlerini COICOP'a göre sınıflamak için makine öğrenmesi yöntemlerini kullanmak için çalışma yapmaktadır. Projenin amacı, ödeme işlemleri verisinin perakende satış, hanehalkı tüketimi, dijital işlemler ve turizm istatistikleri ile ilgili bilgiler üretmek için kullanılıp kullanılamayacağını belirlemektir [48]. Yeni Zelanda İstatistik Ofisi, sayımlarındaki kod atama kalitesini iyileştirmek için Destek Vektör Makinesi (SVM) algoritmasının kullanılması potansiyelini araştırmaktadır. Census 2013 verilerinden eğitim ve test için her biri 10 000 boyutunda iki ayrı gözlem seti kullanarak meslek ve okul sonrası yeterlilik değişkenlerine kod atamak için SVM algoritmasını kullanmışlardır. Her iki değişken için de test verilerinde %50 doğruluk oranına ulaşmışlardır ve SVM algoritmasını otomatik kod atama için kullanabilmek amacıyla daha fazla araştırma yapılması gerektiğine karar vermişlerdir [46]. De Matteis ve diğ. [49], bireysel iş geçmişlerinin hızlı ve makul bir şekilde toplanmasını ve kod atamasını sağlamak için bilgisayar tabanlı otomatik programların verimliliğini standart mesleki kod atamanın geçerliliği ile birleştiren yenilikçi bir web tabanlı araç geliştirmişlerdir.

2.6. Kayıt Bağlantısı

Kayıt bağlantısı, bir veya daha fazla veri tabanından gelen kayıtların aynı kayıt olduğuna karar verme işlemidir. Yani, “1234 West Fifth Street” te yaşayan “Glenn Thomas Wright” kaydı ile “1234 Fifth Street” de yaşayan “Glen T Wright” kaydının aynı kişiyi ifade ettiğini ve “678 Ninth Street” de yaşayan “Denis Hulett” ile aynı kişi olmadığını belirleme işlemidir [50]. Mikro veri bağlantısı, aynı varlık veya nesne hakkında bir bileşik kayıt oluşturmak için iki veya daha fazla kaydın birleştirilmesi olarak tanımlanır. Bir mikro veri bağlantısının çıktısı, girdi olarak kullanılan birden fazla veri tabanındaki bilgileri içermelidir [51].

Mikro veri bağlantısı olarak kabul edilen durumlar:

Aynı varlık veya nesneye ilişkin bilgilerin iki veya daha fazla veri tabanından eşleştirilmesi mikro veri bağlantısı olarak kabul edilmektedir. Girdi olarak anket verileri, nüfus sayımı verileri veya idari kayıtlar veya bunların bir kombinasyonu olabildiği gibi girdi dosyalarının aynı zaman referans dönemi olma zorunluluğu bulunmamaktadır [51].

Mikro veri bağlantısı olarak kabul edilmeyen durumlar:

- Ana veri tabanının bir alt kümesini içeren veri dosyası bir mikro veri bağlantısı değildir, çünkü çıktı dosyası birden fazla girdi dosyasından bilgi içermemektedir [51].
- İmpütasyon için kullanılacak benzer bir varlığı veya nesneyi tanımlayan ancak aynı varlığı veya nesneyi eşleştirmek için hiçbir girişimde bulunulmamış yöntemler [51].
- Coğrafi düzeyde toplu bilgi içeren veri dosyaları ile fert düzeyinde anket veya idari kayıt bağlantıları. Bu duruma, bir kişi hakkındaki bilgilerin, o kişinin yaşadığı mahallenin sayım özet bilgileri ile eşleştirilmesi örnek olarak verilebilir [51].

Kayıt bağlantısı, bu çalışma kapsamında anketörün yazdığı tanım ile standart sınıflama sözlüğündeki tanımın aynı olup olmadığına veya eşleşip eşleşmediğine karar verilmesi aşamasında kullanılmaktadır. Kayıt bağlantısı terimi, ilk olarak sağlık alanında bireysel hasta dosyalarındaki hasta adı, doğum tarihi ve diğer bilgiler kullanılarak bir araya getirilmesi aşamasında ortaya çıkmıştır. Son yıllarda, bilgisayar bilimleri, istatistik ve yöneylem araştırması alanındaki çığır açan fikirler bu alandaki bilgisayar sistemleri ve programlarının gelişmesine ışık tutmuştur [52].

Modern kayıt bağlantısı fikri, frekans oranları ile eşleşen ve eşleşmeyenleri tanımlayan karar kuralları ile bilinen genetikçi Howard Newcombe'ya dayanmaktadır. Newcombe'nun fikirleri birçok epidemiyolojik uygulamada kullanılan programlarda ulusal sağlık dosyaları kullanılarak hesaplanan önsel frekansların oranlarına dayanmaktadır [52]. Fellegi ve Sunter ise, kayıt bağlantısının biçimsel matematiksel temellerini sağlamıştır. Onların teorisi, Newcombe tarafından kullanılan karar kurallarının optimal olduğunu göstermiştir ve doğrudan eşleştirilen dosyalardan eşleşme olasılıklarını tahmin etmenin çeşitli yollarını ortaya koymuşlardır [52].

Kayıt bağlantısı, nesne tanımlama [53], veri temizleme [54], yaklaşık bağlantı veya yaklaşık birleştirme [55], bulanık eşleştirme [56] ve varlık çözümlemesi [57] kavramları ile de tanımlanmaktadır. Temel olarak deterministik ve bulanık olmak üzere iki tür kayıt bağlantı yöntemi bulunmaktadır [58]. Bu yöntemlerin yanı sıra, veri tabanı ve veri madenciliği araştırmalarında kurala dayalı kayıt karşılaştırma yöntemleri [59], makine öğrenmesi [60] ve Bayesian karar modeli gibi yeni teknikler geliştirilmiştir [61].

Kayıt bağlantısında yöntem seçimi, bağlantı değişkenlerinin mevcudiyeti ve kalitesine bağlıdır. Değişkenler yüksek kalitede olduğunda kayıt bağlantısında deterministik bir yaklaşım sıklıkla tercih edilmektedir [62]. Daha düşük bir veri kalitesi olduğunda ise bulanık yaklaşım sıklıkla seçilmektedir [63-65]. Uygulamada, özellikle veri hacmi arttıkça iki yöntemin kombinasyonu birlikte kullanılmaktadır [66].

Farklı veri kaynaklarından gelen kayıtları eşleştirmek için bu veri kaynaklarında ortak olan değişkenler kullanılmaktadır. Bu değişkenler çeşitli türlerde olabilmektedir. Her türün farklı bir zorluğu bulunmaktadır. Örneğin, iki ismin benzerliğini değerlendirmek, iki farklı zaman diliminin benzerliğini değerlendirmekten farklı bir yaklaşım gerektirmektedir. İlki, anlamsal ve morfoloji hakkında biraz bilgi gerektirirken, ikincisi doğrudan gözlemlenebilir. Değişkenler arasındaki benzerliği değerlendirme görevi, değişkenlerdeki hatalar veya tutarsızlıklar göz önüne alındığında çok önemli bir görev olmaktadır [66].

Genel olarak, bağlantı değişkenleri aşağıdaki gibi kategorize edilebilir [66]:

- Metinsel (isim, adres, posta kodu, bir tarihin metinsel gösterimi vb.),
- Sayısal (yaş, tansiyon, kolesterol seviyesi vb. ölçüm değerleri),
- Kategorik (cinsiyet, etnik grup, eğitim düzeyi, medeni durum vb.).

Çoğu durumda, kayıtlardaki hata veya tutarsızlıklar değişkenlerin türüne özgü olmaktadır. Metinsel değişkenler yazım hatalarına daha yatkındır ve metinsel değişkenler yazılırken çeşitli hatalar oluşabilmektedir [66]:

- Yanlışlıkla fazladan harf eklenebilir (ekleme),
- Harfler yanlış bir sıraya yerleştirilebilir (aktarma),
- Bazı harfler yanlışlıkla silinebilir (silme) ve
- Bazı harfler rastgele değiştirilebilir (değiştirme).

Sayısal değişkenlerde, kişisel tercihlere bağlı olarak yapılan yuvarlama nedeniyle tutarsızlıklar olması daha olasıdır. Kategorik değişkenler kısa bir koddan oluştuğundan yazılırken daha tutarlı olmakta ve bu durum olası hataları en aza indirmektedir. Ancak, yazımda hatalar veya tutarsızlıklar meydana geldiğinde, bunların yanlış sınıflama nedeniyle etkisi ciddi olmaktadır. Hataya neden olabilecek tüm faktörlere bağlı olarak, bağlantı

değişkenlerinde rastgele ve sistematik hata olmak üzere iki tür potansiyel hata olabilmektedir [66].

Rasgele hatalar, tanımlayıcı değerine bağlı olmayan ve dolayısıyla herhangi bir kayıta ortaya çıkabilecek hatalar olarak tanımlanmaktadır. Ancak bu hataların nasıl gerçekleştiği rastgele değildir [66]. Örneğin, metinsel değişkenlerde çoğu hatanın metnin orta konumundan kaynaklandığı varsayılmaktadır, çünkü insanlar genellikle ilk karakterleri daha dikkatli bir şekilde yazmaya meyillidirler [67].

Sistematik hatalar, belirli değerlere sahip kayıtlarda ortaya çıkması daha muhtemel olan hatalar olarak tanımlanmaktadır. Bu yüzden bunların oluşumu belirli bir tanımlayıcıya bağlıdır. Örneğin, bir yabancıнын adını yazmak, tanıdık bir isim yazmaktan daha fazla hatayla sonuçlanmaktadır [68]. Aynı şekilde, evli kadınlar kızlık soyadı altında bir kayıt sisteminde ve eşinin soyadı altında diğer kayıt sisteminde tutulabilmektedir. Bu gibi tutarsızlıkları tespit etmek gerçekten zor olmaktadır [66].

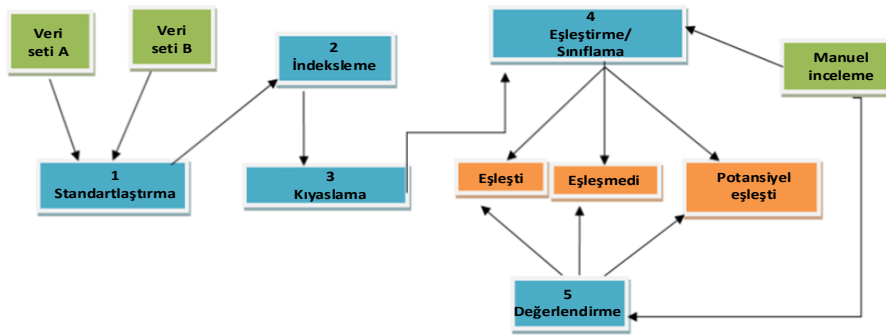
2.6.1. Deterministik bağlantı yöntemi

Deterministik bağlantı yöntemi, genellikle bir kimlik numarası gibi yüksek kaliteli bir tanımlayıcı mevcut olduğunda kullanılmaktadır. Kimlik numarası yoksa, yaş, cinsiyet, ikamet yeri vb. değişkenler benzersiz bir kimlik numarası oluşturacak şekilde birleştirilebilir, ancak birleştirilmiş bu değişkenlerin hataya sebep olabileceği gözden kaçırılmamalıdır. Bu yöntem, verilerde herhangi bir hata olmadığını varsaymakla birlikte yazım yanlışları için de bir yöntem önerilmiştir. Örneğin, ad ve soyad fonetik bir koda dönüştürülebilir veya sadece adın ilk beş harfini kullanarak isimler kısaltılabilir. Eşleştirme değişkenlerinin çoğu aynıysa, eşleşmenin sağlanacağı karar kuralları da belirlenebilir. Deterministik eşleşmede, tüm eşleşme değişkenleri eşleşme belirlenirken eşit derecede katkıya sahip olmaktadır. Bir diğer önemli özellik, deterministik eşleşmenin sonucunda eşleşme var veya eşleşme yok şeklinde iki olası sonuçtan biri ortaya çıkmaktadır. Genel olarak hesaplama yükünü azaltmak amacıyla bulanık eşleştirmeye geçilmeden önce deterministik bir eşleştirme işleminin gerçekleştirilmesi faydalı olmaktadır [69].

2.6.2. Bulanık eşleştirme yöntemi

Bulanık eşleştirme yöntemi, farklı dosyalardaki aynı varlığa veya nesneye ait olma olasılığı en yüksek olan kayıtları birbirine bağlamaya çalışan bir süreç olarak tanımlanmaktadır. Deterministik bağlantı yönteminde veri kümelerini birbirine bağlamak için benzersiz tanımlayıcılar kullanılmakta, bulanık eşleştirme yönteminde ise bağlantıların tanımlanması ve değerlendirilmesi için bir araya getirilen birtakım tanımlayıcılar kullanılmaktadır. Bulanık eşleştirme yöntemi genellikle benzersiz bir tanımlayıcı mevcut olmadığında veya değişkenler yetersiz kalitede olduğunda kullanılmaktadır. Yöntem, adını Fellegi ve Sunter tarafından geliştirilen olasılık çerçevesinden almakta ve bu yöntemde hesaplamaları gerçekleştirmek için gelişmiş bir program gerekmektedir [58].

Şekil 2.2’de, çoğu istatistik ofisi tarafından kullanılan bulanık eşleştirme sürecinin şematik gösterimi yer almaktadır. Uygulamada, iki veri tabanının bağlanması zorunlu değildir, aynı anda birden fazla veri tabanı da bağlanabilmektedir [58].

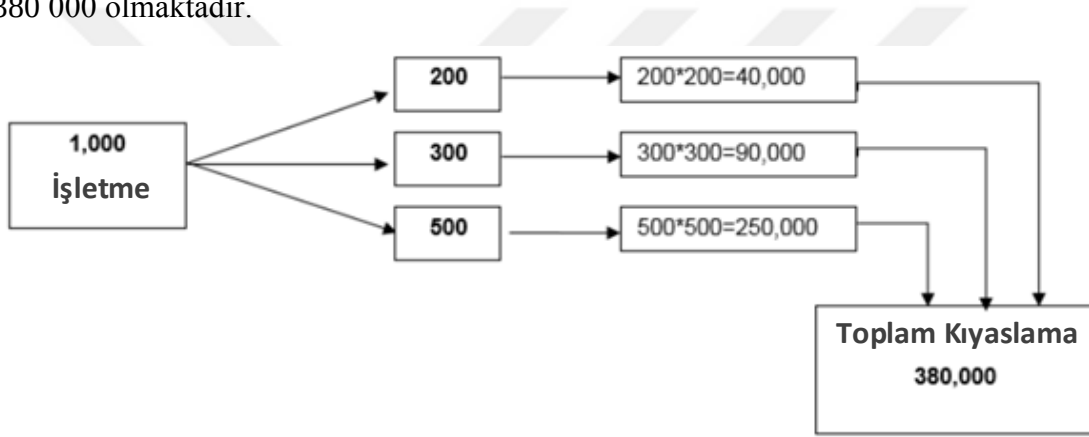


Şekil 2.2. Bulanık eşleştirme

Adım 1 - Standartlaştırma: Bu süreç, söz diziliminin standartlaştırılması, karakterlerin tutarlı bir biçimde tanımlanabilir alanlara dağıtılması, verilerin işlenmesi, ayrıştırılması, dönüştürülmesi ve düzenlenmesini kolaylaştırmak için yapılan biçimsel işlemlerden oluşmaktadır. Dönüşüm, küçük harflerin büyük harfe dönüştürülmesini veya tam tersini, sayısal bir karakterin alfa nümerik bir karaktere dönüştürülmesini ve yanlışlıkla girilen bir gözlemin hariç tutulmasını içerebilir. Bu nedenle, bu adım veri tabanının hazırlanmasında esastır [58].

Adım 2 - İndeksleme: İndeksleme süreci, eşleştirme yapmak için aday çiftleri üretmekten ibarettir. Amaç, karşılaştırılacak gözlem çiftlerinin sayısını en aza indirebilen bir teknik kullanmaktır [58].

Bloklama yönteminde, bir veri tabanı diğer veri tabanındaki 1 000 gözlemin her biriyle karşılaştırılacak 1 000 gözlem içeriyorsa, bir milyon gözlemin sadece bir çift gözlem seçebilmek için karşılaştırılması gerekmektedir. Aynı sayıda çalışanı olan işletmeleri ortak bloklara ayırmak ve ardından aynı bloktaki işletmeleri karşılaştırmak daha basit bir yol olmaktadır. Şekil 2.3'te, 1 000 işletmenin karşılaştırılması için ihtiyaç duyulan örnek gösterilmektedir [58]. Bu durumda bloklama yöntemi sonrasında kıyaslanacak kayıt sayısı 380 000 olmaktadır.



Şekil 2.3. Bloklama örneği

Adım 3 - Kıyaslama: Bu işlem, Adım 2'de seçilen gözlem çiftlerinin karşılaştırılmasından oluşmaktadır. Cebirsel fonksiyonlar kullanılarak hesaplanan bir puan atayarak gözlem çiftleri karşılaştırılmaktadır. Bu işlemler, alfa nümerik karakter dizilerden oluşan gözlem çiftleri arasındaki benzerlikleri karşılaştırmayı mümkün kılmaktadır [58].

Adım 4 - Eşleştirme ve sınıflama: Bu işlem, kayıt sınıflama ve bağlantı işlemidir. Bazen, iki işlem iki farklı adıma ayrılabilir. Sınıflamanın amacı, gözlem çiftlerini eşleştirdi, eşleşmedi veya potansiyel eşleştirdi olarak ayrımlarının yapılabilmesini sağlamaktır. Her ne kadar sınıflama aşaması istatistiksel bir yöntemle dayansa da, “potansiyel eşleştirdi” kategorisinde sınıflanan gözlem çiftlerini “eşleştirdi” veya “eşleşmedi” kategorisine yerleştirebilmek için insan müdahalesi gerekmektedir [58].

Adım 5 - Değerlendirme: Bir önceki adımda sınıflaması yapılmış gözlem çiftlerinin eşleşme kalitesi değerlendirilmelidir. Çizelge 2.5'te, Tip I ve Tip II hatalarına ilişkin bilgiler yer almaktadır [58].

Çizelge 2.5. Eşleşme kategorileri (Tip I ve Tip II bağlantı hataları)

	Gözlem aynı birim ile ilişkilidir.	Gözlem aynı birim ile ilişkili değildir.
Gözlem eşleşti	Gözlem doğru bir şekilde eşleştirilmiştir (gerçek eşleşme)	Gözlem yanlışlıkla eşleştirildi Tip I hatası veya yanlış negatif eşleşme
Gözlem eşleşmedi	Gözlem hatayla eşleşmedi Tip II hatası veya yanlış pozitif eşleşme	Gözlem doğru bir şekilde eşleşmedi (gerçek eşleşme yok)

Genel olarak, deterministik veya bulanık eşleştirme yöntemi işlemlerinin sonucunda yapılması gereken gözle doğrulama, bağlantı kalitesini sağlamada önemli bir adım olarak görülmektedir. Metinsel veriler tüm şekillerde, boyutlarda ve karmaşıklıkta olabilmektedir. Veri kaynakları tutarlı veriyi içerdiğinde ve eksik veri olmadığında, eşleştirme sürecinde nadiren sorun ortaya çıkmaktadır. Ancak, birden fazla kaynaktan gelen veriler, yinelenen gözlemler, eksik değerler, geçersiz değerler, büyük harf kullanımı ve noktalama işaretleri eşleştirme işleminde genellikle güvenilir sonuçlar doğurmaktadır. Bu ve benzeri sorunlar olduğunda, veri kayıtlarının eşleştirilmesi işlemine başlamadan önce, öncelikle tüm veri düzensizlikleri temizlenmeli ve standart hale getirilmelidir [70].

Farklı karmaşık verilerin sürekli olarak çoğalması ve depolanmasıyla birlikte, bu bilgilerin doğru ve temiz olması hiç bu kadar önemli olmamıştı. Bununla birlikte, bu verilerde basit bir birleştirme işlemi yapılacak aynı değişkenlere sahip olmaması çoğu zaman söz konusudur. Bu amaçla, bulanık eşleştirme algoritmaları kullanmak zorunlu olmaktadır. Basitçe söylemek gerekirse, bulanık eşleştirme, işyeri adları, kişi adları veya adres bilgileri gibi veri öğeleri arasındaki benzerliği belirlemek için bulanık eşleştirme için oluşturulan algoritmik süreçlerin uygulanmasıdır. Bu algoritmalar, veri temizleme, tekilleştirme veya farklı veri kümelerinin eşleştirilmesinde yardımcı olmak için kesin olmayan eşleşme olasılığını tahmin etmek için kullanılmaktadır [71].

Eşleştirme yöntemleri

Metin eşleşmesi, iki metnin aynı olup olmadığının belirlenmesi işlemi olarak tanımlanabilir [72]. Metinde yapılabilecek varyasyonlar ve hatalar oldukça yaygın olduğundan [73], karşılaştırma iyi bir eşleşme kalitesiyle sonuçlanmaz. Genel olarak, 1 (iki metin aynıdır) ve 0 (iki metin tamamen farklıdır) arasında normalleştirilmiş bir benzerlik ölçüsü kullanılmaktadır [74].

Metinleri eşleştirmek için kullanılan iki ana yaklaşım bulunmaktadır [74].

- Fonetik kod atama ve
- Desen eşleştirmesi.

Fonetik kod atama

Tüm fonetik kod atama tekniklerinde yaygın olarak, bir metnin nasıl telaffuz edildiğine yani bir metnin söylendiği şekline göre bir koda dönüştürülmeye çalışılmaktadır. Doğal olarak, bu süreç dile bağlı olmaktadır. Bu anlamda geliştirilen çoğu teknik sadece İngilizce dili düşünülerek geliştirilmiştir. Bu yüzden bu çalışmada fonetik kod atama teknikleri kullanılmamıştır.

Desen eşleştirmesi

Desen eşleştirme teknikleri yaygın olarak yaklaşık kelime eşleştirmesinde kullanılmaktadır [75]. Yaklaşık kelime eşleştirmesinde ise sıklıkla mesafe algoritmaları kullanılmaktadır. Mesafe algoritmaları, bilgisayar ve matematik dünyasında farklı iki karakter dizisi arasındaki benzerliği hesaplayan algoritmalarlardır. Bu algoritmaların bir kısmı benzerliği hesaplarken diğer bir kısmı farklılığı hesaplar da genelde mesafe algoritmaları olarak bilinirler. Günümüzde bu algoritmalar, arama motorları veya kelime işlemci programlarının dışında; kopya içerik tespiti (intihal), parmak izi tanıma, virüs tarama, DNA/RNA analizi gibi alanlarda yoğun bir şekilde kullanılmaktadır [76].

Jaro-Winkler mesafe algoritması, iki metin arasındaki benzerliği hesaplamak için kayıt bağlantısında yaygın olarak kullanılmaktadır. Diğer popüler yöntemler; Jaro, Levenshtein ve n-gram mesafe algoritmalarıdır [66].

Jaro ve Jaro-Winkler mesafe algoritması

Jaro-Winkler mesafe ölçüsü, Amerika'nın nüfus sayımında isim karşılaştırması yapmak için geliştirilmiştir. Jaro mesafesi, eşleşme sayısı ve yer değiştirmelerin sayısını hesaplamak için iki adım içermektedir. Winkler ayarı ise, her iki metnin ilk karakterleri için eşleşme puanına dayanan yeniden skorlamayı içermektedir [77].

Diyelim ki, cs1 ve cs2 karakter dizilerini karşılaştırmak istiyoruz. Jaro-Winkler mesafesi aşağıdaki adımlarla tanımlanmaktadır [77].

Adım 1: Eşleşmeler: Bir kelimedeki karakterlerin başka bir kelimedeki karakterlere göre hizalama aşamasıdır. Karakterlerin eşleştirilebileceği maksimum mesafe (dizi indeksi ile ölçülmektedir) [77]:

Eşleşmeler (cs1, cs2) = maksimum (cs1 uzunluğu (), cs2 uzunluğu ()) / 2-1

Adım 2: Yer değiştirmeler: Eşleştirmeden sonra, her iki kelimedeki eşleşen karakterlerin alt dizileri çıkarılmaktadır. Jaro mesafesi, eşleşen karakter ve yer değiştirmelerin sayısı cinsinden tanımlanmaktadır [77]:

Jaro mesafesi (cs1, cs2)

$$= 1/3 * ((\text{eşleşmeler (cs1, cs2)} / \text{cs1 uzunluğu ()}) + \text{eşleşmeler (cs1, cs2)} / \text{cs2 uzunluğu ()} + (\text{eşleşmeler (cs1, cs2)} - \text{yer değiştirmeler (cs1, cs2)}) / \text{eşleşmeler (cs1, cs2)})$$

Yani, ölçü aşağıdaki üç değerlerin ortalamasından oluşmaktadır [77]:

- (a) eşleşenlerin ilk kelimedeki yüzdesi,
- (b) eşleşenlerin ikinci kelimedeki yüzdesi ve
- (c) yer değiştirmemiş eşleşmelerin yüzdesi.

Adım 3: Winkler modifikasyonu: Jaro mesafesi ölçüsüne yapılan Winkler modifikasyonu, başlangıçta eşleşen karakter için Jaro skorları artırıcı etkisi bulunmaktadır [77].

Jaro-Winkler Mesafesi (cs1, cs2)

$$= \text{Jaro mesafesi (cs1, cs2)} + 0,1 * \text{Kelimelerin başında eşleşen karakter sayısı (cs1, cs2)} * (1,0 - \text{Jaro mesafesi (cs1, cs2)})$$

Şekil 2.4'te yer alan MARTHA ve MARHTA kelimeleri için Jaro-Winkler mesafe ölçüsü aşağıdaki şekilde hesaplanmaktadır [77].

cs1	M	A	R	T	H	A
cs2	M	A	R	H	T	A

Şekil 2.4. Jaro-Winkler mesafe ölçüsü hesaplama örneği

$$\text{Jaro Mesafesi (MARTHA, MARHTA)} = 1/3 * (6/6 + 6/6 + (6 - 1)/6) = 0,944$$

$$\text{Jaro-Winkler Mesafesi (MARTHA, MARHTA)} = 0,944 + 0,1 * 3 * (1,0 - 0,944) = 0,961$$

Levenshtein mesafe algoritması

Levenshtein mesafesi, bir metni diğer metne dönüştürmek amacıyla yapılması gereken en az düzenleme (ilave, silme ve değiştirme) sayısı olarak tanımlanmaktadır. Çizelge 2.6'da yer alan "Cuma" ve "Pazar" kelimeleri Levenshtein mesafesi ölçüsü aşağıdaki şekilde hesaplanmaktadır [78].

Çizelge 2.6. Levenshtein mesafe ölçüsü hesaplama örneği

		C	U	M	A
	0	1	2	3	4
P	1	1	2	3	4
A	2	2	2	3	3
Z	3	3	3	3	4
A	4	4	4	4	3
R	5	5	5	5	4

Tablodaki değerler aşağıdaki şekilde okunabilmektedir [78]:

P-CUM: 3, P harfini C ile değiştirip UM eklendiği için toplam 3 işlem gerekmektedir.

PA-CUM: 3, P harfi C ile A harfi U ile değiştikten sonra M harfi eklenmiştir.

PAZ-CUM: 3, P harfi C ile A harfi U ile ve Z harfi M ile değişmiştir, toplam maliyet 3'tür.

PAZA-CUM: 4, P harfi C ile A harfi U ile ve Z harfi M ile değişmiştir, ilave olarak A harfi eklenmiştir.

PAZAR – CUM: 5, P harfi C ile A harfi U ile ve Z harfi M ile değişmiştir ilave olarak A ve R harfi eklenmiştir.

PAZAR – CUMA: 4, P harfi C ile A harfi U ile ve Z harfi M ile değişmiştir ilave olarak R harfi eklenmiştir. Levenshtein mesafe ölçüsü 4'tür.

Genelleştirilmiş düzenleme mesafesi

Genelleştirilmiş düzenleme mesafesi, iki kelime arasındaki farklılığın bir ölçütü olan Levenshtein mesafesi ölçütünün bir genellemesidir. Levenshtein düzenleme mesafesi, kelime-1'i kelime-2'ye dönüştürmek için gereken karakterleri silme, ekleme veya değiştirme sayısıdır. Genelleştirilmiş düzenleme mesafesi işlemleri ve maliyetleri Çizelge 2.7'de açıklanmıştır [79].

Çizelge 2.7. Genelleştirilmiş düzenleme mesafesi işlemleri ve maliyetleri

İşlem	Maliyet	İşlem Tanımı
1	50	Çıktı kelimesi giriş kelimesinden daha uzun olduğunda, işaretçiyi hareket ettirmeden çıktı kelimesinin sonuna herhangi bir karakter eklenmesi.
2	10	İşaretçiyi hareket ettirmeden çıktı kelimesinin sonuna bir boşluk karakteri eklenmesi, işaretçideki karakter boşluk karakteri ise, çıktı kelimesini değiştirmeden işaretçinin bir konum ilerletilmesi, işaretçideki karakter boşluk karakteri ise, çıktı kelimesinin sonuna bir boşluk karakteri eklenmesi ve işaretçinin bir konum ilerletilmesi.
3	100	İşaretçiyi, çıktı kelimesini değiştirmeden bir konum kadar ilerletilmesi.
4	20	İşaretçiyi hareket ettirmeden çıktı kelimesinin sonuna işaretçiye karakter eklenmesi.
5	200	Çıktı kelimesi boş olduğunda, çıktı kelimesini değiştirmeden işaretçinin bir konum ilerletilmesi.

Çizelge 2.7. (devam) Genelleştirilmiş düzenleme mesafesi işlemleri ve maliyetleri

6	200	İşaretçi bir konumdayken, işaretçiyi hareket ettirmeden çıktı kelimesinin sonuna herhangi bir karakter eklenmesi.
7	200	İşaretçi bir konumdayken ve çıktı kelimesi boş olduğunda, çıktı kelimesinin sonuna herhangi bir karakter eklenmesi ve işaretçinin bir konum ilerletilmesi.
8	100	İşaretçiyi hareket ettirmeden çıktı kelimesinin sonuna herhangi bir karakter eklenmesi.
9	0	İşaretçideki karakteri giriş kelimesinden çıktı kelimesinin sonuna kopyalanması ve işaretçinin bir konum ilerletilmesi.
10	30	İşaretçiyi hareket ettirmeden çıktı kelimesinin sonuna bir noktalama karakterinin eklenmesi, işaretçideki karakter noktalama işareti ise, çıktı kelimesini değiştirmeden işaretçinin bir konum ilerletilmesi, işaretçideki karakter bir noktalama işareti ise, çıktı kelimesinin sonuna bir noktalama işaretinin eklenmesi ve işaretçinin bir konum ilerletilmesi.
11	100	Çıktı kelimesinin sonuna herhangi bir karakter eklenmesi ve işaretçinin bir konum ilerletilmesi.
12	20	İşaretçideki karakter, giriş esinde izleyen karakterle aynı olduğunda, çıktı kelimesini değiştirmeden işaretçinin bir konum ilerletilmesi.
13	20	İşaretçiyi izleyen karakteri giriş kelimesinden çıktı kelimesine kopyalanması, sonra işaretçideki karakteri giriş kelimesinden çıktı kelimesine kopyalanması ve işaretçinin iki pozisyon ilerletilmesi.
14	10	Çıktı kelimesi giriş kelimesinden daha kısaysa, çıktı kelimesini değiştirmeden işaretçinin bir konum ilerletilmesi.

Genelleştirilmiş düzenleme mesafesi ile ilgili örnek Çizelge 2.8’de verilmiştir [79].

Çizelge 2.8. Genelleştirilmiş düzenleme mesafesi örneği

Kelime 1	Kelime 2	Mesafe	İşlem
baboon	baboon	0	9
baXboon	baboon	100	8
baoon	baboon	100	3
baXoon	baboon	100	11
baboonX	baboon	50	1
baboo	baboon	10	14
babboon	baboon	20	4
babon	baboon	20	12
baobon	baboon	20	13
baboon	baboon	10	2
bab,oon	baboon	30	10
bXaoon	baboon	200	8,3
bXaYoon	baboon	200	8,11
bXoon	baboon	200	3,11
Xbaboon	baboon	200	6
aboon	baboon	200	13,3
Xaboon	baboon	200	7
axoon	baboon	300	5,11
axoo	baboon	310	5,11,14
axon	baboon	320	5,11,12
baby	baboon	120	11,14,14
balloon	baboon	200	11,8

n-gram algoritması

n-gram algoritması, tekrar oranı ile ilgili bir yöntemdir. “n” sayısı, tekrarın kontrol edildiği değeri göstermektedir [80].

“wwwbilgisayarkavramlaricomssadievrenseker” kelimesinin 2-gram değerleri Çizelge 2.9’da verilmiştir:

Çizelge 2.9. 2-gram değerleri örneği

2-gram	Frekans	Olasılık	2-gram	Frekans	Olasılık	2-gram	Frekans	Olasılık
ww	2	0,051	ka	1	0,026	ad	1	0,026
wb	1	0,026	av	1	0,026	di	1	0,026
bi	1	0,026	vr	2	0,051	ie	1	0,026
il	1	0,026	ra	1	0,026	ev	1	0,026
lg	1	0,026	am	1	0,026	re	1	0,026
gi	1	0,026	ml	1	0,026	en	1	0,026
is	1	0,026	la	1	0,026	ns	1	0,026
sa	2	0,051	ri	1	0,026	se	1	0,026
ay	1	0,026	ic	1	0,026	ek	1	0,026
ya	1	0,026	co	1	0,026	ke	1	0,026
ar	2	0,051	om	1	0,026	er	1	0,026
rk	1	0,026	ms	1	0,026			
Toplam	15	0,385	Toplam	13	0,333	Toplam	11	0,282
						Genel Toplam	39	1

Çizelge 2.10’da verilen “soğan” ve “YESİL SOAN” 2-gram mesafe ölçüsü, “SO” ve “AN” ifadeleri her iki kelimede ortaktır. Bu yüzden bu değerın maksimum uzunluğa sahip ifadenin bir eksiğine bölümünden elde edilen değeri $2/9=0,22222$ ’dir.

Çizelge 2.10. 2-gram mesafe ölçüsü hesaplama örneği

Uzunluk=5	<table><tr><td>sogan</td></tr><tr><td>SO</td></tr><tr><td>OG</td></tr><tr><td>GA</td></tr><tr><td>AN</td></tr></table>	sogan	SO	OG	GA	AN	<table><tr><td>YESIL SOAN</td></tr><tr><td>YE</td></tr><tr><td>ES</td></tr><tr><td>SI</td></tr><tr><td>IL</td></tr><tr><td>L</td></tr><tr><td>S</td></tr><tr><td>LS</td></tr><tr><td>SO</td></tr><tr><td>OA</td></tr><tr><td>AN</td></tr></table>	YESIL SOAN	YE	ES	SI	IL	L	S	LS	SO	OA	AN	Uzunluk=10
sogan																			
SO																			
OG																			
GA																			
AN																			
YESIL SOAN																			
YE																			
ES																			
SI																			
IL																			
L																			
S																			
LS																			
SO																			
OA																			
AN																			



3. YÖNTEM

Bu bölümde, araştırmanın modeli, evren ve örnekleme, veri elde etmek için kullanılan veri toplama araçları, uygulamanın yapılışı ve elde edilen verilerin çözümlenmesi ve yorumlanması anlatılmıştır.

3.1. Araştırmanın Modeli

Bu çalışma, nicel araştırma yöntemlerinden yeni bilgi ve var olan uygulamanın geçerliliğini sağlamak için tasarlanan bir araştırma türü olan tasarım ve geliştirme araştırması kapsamında yürütülmüştür. Bu geliştirmeye yönelik araştırma yaklaşımı, yenilikçi süreçlerin, tasarımların ve ürünlerin araştırma sürecinde geliştirilmesi, uygulanması ve değerlendirilmesini kapsamaktadır. Yeni ürünün, aracın, modelin ve süreçlerin ortaya konulduğu ve bu ürünlerin etkisinin ve verimliliğinin uygulamalı olarak incelendiği araştırmalara tasarım ve geliştirme araştırması (TGA) denilmektedir [81,82].

En basit haliyle, TGA, belirli tasarım ve geliştirme çabalarının ya da bir bütün olarak tasarım ve geliştirme sürecinin ya da belirli süreç bileşenlerinin çalışmasının süreci ve etkisinin incelenmesidir. Bu tür araştırmalar, başkalarının tasarım ve geliştirme çalışmalarını inceleyen bir durumu içerebilir. Bununla birlikte, aynı zamanda bir kişinin tasarım ve geliştirme faaliyetlerini üstlendiği ve süreci aynı zamanda incelediği bir durumu da içerebilir. Her iki durumda da, tasarım ve geliştirme yapmakla süreci incelemek arasında bir ayrım vardır.

TGA, Tip 1 ve Tip 2 olmak üzere iki kategoriye ayrılmaktadır. Ürünler ve araçların geliştirilmesiyle ilgili araştırmalar Tip 1 ve tasarım ve geliştirme modellerinin geliştirilmesiyle ilgili araştırmalar Tip 2 başlığı altında yer almaktadır [81,82].

TGA, gerçek hayatta karşılaşılan problemlere yönelik, somut, uygulanabilir çözümler, ürünler araştırmaya yönelik bir yöntemdir. Bu nedenle TGA özellikle birçok aşamadan oluşan ve ürün geliştirme hedefi olan projeler için bir bilimsel araştırma yöntemidir [83].

TGA sürecinde sistem geliştirme yaşam döngüsü temel alınmıştır. Sistem geliştirme yaşam döngüsü, bir sistemin üretilmesi ile ilgili tüm faaliyetleri kapsamaktadır. Tüm yaratıcı süreçler gibi sistem geliştirme süreci de karmaşıktır ve çok sayıda insanın kararına ve yargılarına bağlı olmaktadır. İdeal bir sistem geliştirme süreci yoktur ve çoğu kurum veya kuruluş kendi sistem geliştirme süreçlerini kendileri geliştirmişlerdir. Süreçler, bir kurumdaki insanların yeteneklerinden ve geliştirilmekte olan sistemlerin özelliklerinden yararlanmak için sürekli evrim geçirmektedir [84].

Sistemin üretimde ve kullanımda geçirdiği süreç sistem geliştirme yaşam döngüsü olarak adlandırılmaktadır. Süreç döngü şeklinde ele alındığı için değişiklikler ve genişlemeler sürece anında entegre edilebilmektedir. Döngü içerisinde her hangi bir aşamada geriye dönmek ve tekrar ilerlemek söz konusudur. Sistem yaşam döngüsü çift yönlüdür ve doğrusal değildir [85].

Sistem geliştirme yaşam döngüsü, zaman ve bütçe dahilinde yüksek kaliteli ürünler sunmak için planlı ve sistematik bir düzen sağlamaktadır. İzlenecek gerçek yaşam döngüsü, organizasyonun sistem geliştirme yaşam döngüsü standardına bağlı olmaktadır. Sistem geliştirme yaşam döngüsü aşamalarının, literatürdeki hemen her çalışmada birbirinden farklı olduğu dikkat çekmektedir. Ancak yine de her bir yaşam döngüsü, farklı isim ya farklı yapılarda sunulsa da, aynı temel özellikleri içermektedir [86].

Bu çalışmada, Şekil 3.1’de verilen yedi aşamadan oluşan sistem geliştirme yaşam döngüsü adımlarının kullanılması benimsenmiştir [87].



Şekil 3.1. Sistem geliştirme yaşam döngüsü

Bu çalışma kapsamında aşağıdaki yaşam döngüsü adımları kapsamında belirtilen işlemler yerine getirilmiştir.

3.1.1. Sorunları, fırsatları ve hedefleri belirleme

Sorunlar

Karşılaştırılabilir ve kapsayıcı bir veri ortaya koymak istendiğinde ulusal veya uluslararası bir sınıflama kullanılması günümüzde artık bir zorunluluktur. Anketör tarafından yapılan kod atamasında doğru kodun verilip verilmediğinin kontrolünün elle veya gözle yapılan klasik yöntemlerle yapılması maliyetli bir iştir. Ayrıca bu işlem sonucunda tam olarak güvenilir bir sonuca erişilmesi beklenemez. Bunun için her koda aynı hassasiyet ve standart ile yaklaşan otomatik bir sistemle bu kontrollerin yapılması gerekmektedir.

Anket yapma ve veri derleme yöntemleri giderek farklılaşmakta ve gelişmektedir. Anketör ve kodlayıcı istihdam etmek maliyetli bir kaynaktır. Bu kaynak maliyetinden kurtulma yönünde anket yapma yöntemlerinde yenilikler sürekli denenmektedir. Anket yapma veya veri derleme yöntemlerinde anketörün veya kodlayıcının bulunmadığı yeni yöntemlerin ortaya çıkması bu yeniliklerdendir. Anketörün ve kodlayıcının bulunmadığı bu yöntemlerde kod atama görevinin otomatik bir sistem tarafından yerine getirilmesi gerekmektedir.

Fırsat

Teknolojinin gelişmesi ile birlikte bilgisayarların verimliliği ve kapasitesi artmıştır. İleri düzey programlama dilleri kullanılabilir hale gelmiştir. Bu sayede bir kod atama sisteminin oluşturulması ve kullanılması mümkün hale gelmiştir.

Hedef

Sistemin kurulmasındaki hedef, doğru kod ataması yapabilecek veya yapılmış kod atamasını kontrol ederek hatayı ortadan kaldıracak bir sistem geliştirebilmektir. Bu sayede, insandan uzaklaşan bir sistemin kurularak anketörün neden olduğu hataların önüne geçilmesi ve anketörsüz de kod atama işleminin gerçekleştirilmesi hedeflenmektedir.

3.1.2. Bilgi gereksinimlerini belirleme

Mevcut durumda kod atama işlemleri anketörler tarafından ofiste bilgisayar destekli olarak yapılmaktadır. Anketörler tarafından atanan kodların doğruluğu gözle kontrol edildikten sonra hatalı kayıtlar düzeltilmektedir. Sistem tarafından yapılan kontrol işlemi sonucunda anketörün hangi kayıt ile ilgili nasıl bir kontrol ve nasıl bir düzeltme yapacağı liste halinde gösterilmektedir. Bu bakımdan anketörlerin mevcut durumdan farklı bir bilgi ihtiyacı bulunmamaktadır.

3.1.3. Sistem ihtiyaçlarını analiz etme

Oluşturulacak sistemin ilk olarak anketörlerin kod atamasına yardımcı olacak sonuçlar ortaya koyması ve nihai olarak da anketörsüz doğrudan kod ataması gerçekleştirmesi beklenmektedir. Bu sistemin etkinliğinin testi öncelikle resmi olarak yayınlanmış veriler kullanılmaktadır. Daha sonra sistemin anketörlerin veri girişi yaptıkları aşamada canlı veri üzerinden geri bildirim yapması planlanmaktadır.

3.1.4. Önerilen sistemi tasarlama

Ankette, “Harcamanın adı, cinsi ve ayrıntılı tanımı nedir?” sorusuna cevap olarak yazılan tanım ile COICOP kodları anketör tarafından verilmektedir. Verilen bu COICOP kodlarının doğruluğunu analiz ederek bunun sonucunda anketör tarafından yapılan kod atamasının “eşleşen”, “şüpheli”, “kabul” ve “yetersiz” olduğu sonucunu çıktı olarak veren İstatistiksel Analiz Programı (SAS) 9.3 W32_7PRO platformunda istatistiksel analiz için kullanılan bir bilgisayar programlama dili olan SAS dilinde bir kod atama sistemi oluşturulması planlanmaktadır.

SAS programlama dilinin izin verdiği ölçüde kullanıcı dostu bir ara yüz tasarlanmaya çalışılacak olup ara yüzde temel olarak aşağıdaki üç ana işlem gerçekleştirilmesi planlanmaktadır.

- Sınıflama,
- Kod atama,
- Raporlar

Sistem parametreleri olarak tablo, ana grup, bölge, yıl ve ay belirlenmiştir. Sistemdeki tüm işlemler ve raporlamalar bu parametrelere değer atanarak gerçekleştirilecektir. Sistemde harcama tanımlarının kontrol edileceği dört katman oluşturulması planlanmaktadır. Birinci katmanda, tanımlar dilbilgisi yönünden kontrol edilerek gereken dönüşümler yapılacaktır. İkinci katmanda, harcama tanımı ve sözlük tanımı eşleştirilmeye çalışılacaktır. Üçüncü katmanda, atanan kod “eşleşen”, “şüpheli”, “kabul” ve “yetersiz” olarak sınıflanacaktır. Dördüncü katmanda ise kod atama işlemleri gerçekleştirilecektir. Anketörün yazdığı harcama tanımındaki yazım yanlışlarını düzeltmek ve kod ataması yapmak için bulanık eşleştirme algoritmaları kullanılacaktır.

3.1.5. Sistem geliştirme ve belgeleme

Bu aşamada planlanan sistemin oluşturulması esnasında yürütülen faaliyetleri açıklayan dokümanların oluşturulması işlemleri yapılmıştır.

3.1.6. Sistemi test etme ve koruma

Kurulan sistem anketörlere açılmadan önce hali hazırda veri tabanında tutulan kayıtlar üzerinden kod ataması yapılarak elde edilen bulguların tutarlılığı test edilmiştir. Karşılaşılan sorunlar ve çözüm önerileri konusunda anketörlerin bilgilendirilmesi sağlanarak sürekli gelişen ve tutarlılığı artan bir yapı oluşturulması planlanmaktadır.

3.1.7. Sistemi uygulama ve değerlendirme

Anketörlerin sistemi kullanmaya başlaması ile birlikte yapılan her işlemin ve özet bilgilerin veri tabanında tutulması sağlanmıştır. Bu şekilde sistemde sonradan ortaya çıkacak problemlerin izlenmesi ve çözüm konusunda somut veriler elde edilmesi planlanmaktadır. Öncelikle eski yapı ile yeni yapı birlikte kullanılarak sistemin verimliliğinin test edilmesi ve bu yolla yeni sisteme geçmek için gerekli önlemlerin alınması planlanmaktadır.

3.2. Evren ve Örneklem

Bu çalışma kapsamında kullanılacak veriler ayrı bir anket çalışması ile derlenmeyecek olup, daha önce TÜİK tarafından yapılmış olan 2013-2018 yılları arası HBA mikro veri setleri

TÜİK'ten izin alınarak kullanılmıştır (Ek-2).

Anket çalışmasında, Türkiye Cumhuriyeti sınırları içinde bulunan tüm yerleşim yerleri ve bu yerleşim yerlerinde yaşayan tüm hanehalkı fertleri kapsama alınmıştır. Ancak kurumsal nüfus olarak tanımlanan yaşlılar evi, huzur evleri, hapishane, askeri kışla, özel nitelikli hastane, otel, çocuk yuvalarında bulunan nüfus ile pratik nedenlerden dolayı göçer nüfus kapsam dışı tutulmuştur [88-93].

Kentsel yerlerden ve belediye teşkilatı olan kırsal yerler ve köylerden yerleşim yerinin büyüklüğüne orantılı olarak 2007 yılında oluşturulan ve anket yılının Ağustos ve Şubat aylarında güncellenen Ulusal Adres Veri Tabanı (UAVT) kullanılarak bloklar belirlenmiş ve her bloktan hanehalkları tabakalı iki aşamalı küme örnekleme yöntemiyle seçilmiştir. Örneğe çıkan bloklardan 36 hane seçilerek A ve B setlerine ayrılmıştır. Araştırma, yıl içinde dört dönem ve her dönemde iki alt örnek kullanılacak şekilde tasarlanmıştır. Birinci aşama örnekleme birimi, ortalama 100 haneden oluşan kümeler ve ikinci aşama örnekleme birimi, her bloktan her set için seçilen 18 hanehalkından oluşmaktadır. Çalışmanın her iki aşamasında da sistematik yolla örnek birimler seçilmektedir [88-93].

3.3. Veri Toplama Araçları ve Uygulamanın Yapılışı

Hanehalklarından bilgiler; görüşme, kayıt ve gözlem metotları kullanılarak derlenmektedir. Anket çalışmasında üç ayrı soru formu kullanılmaktadır.

- HBA soru formu,
- Hanehalkı harcama kayıt defteri,
- Ferdi harcama kayıt defteri.

HBA soru formu: Anket çalışmasına ilişkin alandan derlenen tüm bilgilerin kayıt edildiği soru formudur. Anketör, hanehalkı fertleriyle yapılan yüz yüze görüşmelerde soruları sorup tablet bilgisayarlara ilgili cevapları kaydetmektedirler.

Hanehalkı harcama kayıt defteri (Ek-3): Anket çalışması boyunca hanehalklarının yapacakları harcamaları (ekmek, süt, domates, çorap, ayakkabı, pijama, kazak, kira, su

faturası, gazete, ulaşım, lokanta vb. harcamalar) kayıt etmeleri için hazırlanmış formlardır. Her hafta bitiminde hanehalkının doldurduğu harcama kayıt defteri alınarak yeni haftanın defteri hanehalkına bırakılmaktadır. Anketör, hanehalkından kontrol ederek teslim aldığı harcama kayıt defterinin veri girişini ofiste yapmaktadır.

Ferdi harcama kayıt defteri (Ek-4): Bu form, harcama kayıt defteri ile aynı şekilde dizayn edilmiş olup, çalışan, öğrenci olan veya hane dışında kişisel harcamalarının yoğun olduğu düşünülen hanedeki 15 ve daha yukarı yaştaki fertlere bu harcamalarını (sigara, çay, lokanta vb. yerlerde yenen öğle yemekleri, dolmuş-otobüs ücreti, öğrencilerin harçlıkları ile kantinden satın aldıkları tost vb.) unutmadan kaydetmeleri için yanlarında taşıyabilecekleri şekilde küçük bir defter şeklinde hazırlanmış formdur.

HBA uygulamasında ilk olarak ankete seçilen hanelerle bir ön görüşme yapıp anket çalışması hakkında bilgi verilmesi gerekmektedir. Her ne kadar anket öncesinde hanelere mektup ve broşür gönderilerek çalışma hakkında bilgi verilse de hanehalkına çalışmayı tekrar yüz yüze anlatmak faydalı olmaktadır. Bu görüşmeler genellikle kontrolör ve anketörün haneyi adresinde ziyaret etmesiyle yapılmaktadır. Bu görüşmelerde, hanehalklarından alınan konut bilgileri, tüketim alışkanlıkları ve hanede yaşan fert bilgileri tablet bilgisayarlar aracılığıyla bu anket için oluşturulmuş veri giriş programına aktarılmaktadır. Bu görüşmede hanehalkının yaptığı harcamaların hanehalkı harcama kayıt defterlerine nasıl kayıt edileceği ve hanede 15 ve daha yukarı yaşta olup bireysel harcama yapan fertler varsa bu fertlerin ferdi harcama kayıt defterini nasıl doldurulacağı konusunda açıklamalarda bulunmaktadır. Hanehalkından daha sonraki görüşme için randevu alınmaktadır.

Anketör, anket ayı öncesinde yaptığı ilk görüşmeden sonra aylık harcamaları en doğru şekilde derleyebilmek için örnek haneye anket ayı içinde en az dört ziyaret yapmaktadır. Bu ziyaretlerde, önceki haftaya ait hanehalkı harcama kayıt defteri kontrol edilerek teslim alınmakta ve yeni defterin hanehalkına teslimi gerçekleştirilmektedir.

Anketör, bir sonraki ayın ilk haftasında haneleri ziyaret ederek, hanehalkının anket ayı içinde yaptığı son hanehalkı harcama kayıt defterini ve ilk görüşmede verdiği ferdi harcama kayıt defterini kontrol ederek geri almaktadır. Bu görüşmede ayrıca tüketim dışı harcamalar, istihdam ve gelir durumu bilgileri sorgulanarak tablet bilgisayar aracılığıyla veri giriş

programına kaydedilmektedir.

Bir anketörün, bir anket ayında hanelere yapması gereken ziyaret ile iş planı aşağıdaki şekilde olmaktadır:

- Tanışma/ilk görüşme, hanehalkı harcama kayıt defterinin bırakılması (Anket ayı öncesi),
- İlk defterin alınıp, 2. defterin bırakılması (5-6. günler),
- İlk defterin ofiste veri girişinin yapılması,
- İkinci defterin alınıp, 3. defterin bırakılması (12-13. günler),
- İkinci defterin ofiste veri girişinin yapılması,
- Üçüncü defterin alınıp, 4. defterin bırakılması (18-19. günler),
- Üçüncü defterin ofiste veri girişinin yapılması,
- Dördüncü defterin alınıp, 5. defterin bırakılması (25-26. günler),
- Dördüncü defterin ofiste veri girişinin yapılması,
- Son defterin alınması ve gelir görüşmesi (Sonraki ayın 1-10. günleri),
- Son defterin ofiste veri girişinin yapılması ve anketin son halinin analiz için veri tabanına aktarma.

Anketör, anket süresince örnek hanehalkının kaydetmiş olduğu tüm mal ve hizmet harcamalarını, hanehalkı fertlerinin elde ettiği parasal ve mala dayalı gelirlerini, oturduğu konuta ilişkin bilgileri ve tüketim alışkanlıklarına ilişkin gözlemlerini sohbet şeklindeki sorularıyla denetlemektedir. Hanehalkından aldığı ve soru kağıdına kaydettiği gelir ve harcamaları ile hanehalkının oturduğu konutun özellikleri, konutun bulunduğu semt, hanehalkının tüketim alışkanlıkları arasında devamlı ilgi kurmaktadır.

3.4. Verilerin Çözümlemesi ve Yorumlanması

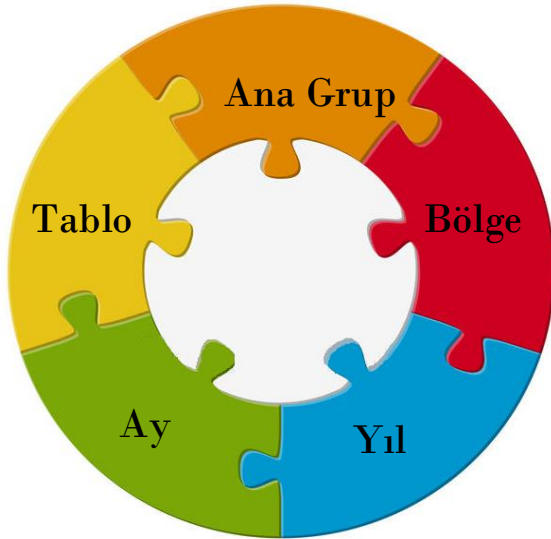
Araştırma sonucunda elde edilen veriler SAS programı kullanılarak analiz edilmiştir. Araştırma soruları doğrultusunda çözümlemeler gerçekleştirilerek yorumlar yapılmıştır.

4. KOD ATAMA SİSTEMİ (KASİS)

Bu bölümde, KASİS'in parametreleri, kullanılan veri setleri ve anket veri giriş programı, katmanları ve veri tabanı ile bilgiler verilmiştir.

4.1. Parametreler

KASİS'te, tablo, ana grup, bölge, yıl ve ay olmak üzere kullanılan beş adet parametre Şekil 4.1'de verilmiştir. Bu parametreler, kontrolü yapılan kodlarla ilgili sonuçların araştırma soruları kapsamında belirli bir sistematik içerisinde sunulabilmesi için önemlidir.



Şekil 4.1. KASİS'te kullanılan parametreler

4.1.1. Tablo

Tablo parametresinde, dayanıklı, stok, satın alış ve aynı olmak üzere kullanılan dört adet tablo ismi Şekil 4.2'de gösterilmiştir.



Şekil 4.2. KASİS'te kullanılan tablo isimleri

1. COICOP DAYANIKLI tablosu: Hanehalklarının son 11 ay içinde satın aldığı sıklıkla yapılmayan harcama kayıtlarını kapsamaktadır.

2. COICOP STOK tablosu: Hanehalklarının anket ayından önce tüketmek amacıyla aldığı stok mal kayıtlarını kapsamaktadır.

3. COICOP SATINALIS tablosu: Hanehalklarının satın aldığı mal ve hizmet kayıtlarını kapsamaktadır.

4. COICOP AYNI tablosu: Hanehalklarının aynı olarak elde ettiği kayıtları kapsamaktadır.

KASİS'teki tablo parametresine ilişkin ekran görüntüsüne Şekil 4.3'te yer verilmiştir.

ULUSLARARASI İSTATİSTİKSEL SINIFLAMALARA YÖNELİK KOD ATAMA SİSTEMİ
Copyright © 2019 by Levent AHI. All Rights reserved.

Saturday 21NOV2020

(İSTEDİĞİNİZ SEÇENEĞİN YANINDA HERHANGİ BİR TUŞA BASARAK SEÇİMİNİZİ YAPINIZ.)

TABLO

- _ COICOP_DAYANIKLI
- _ COICOP_STOK
- _ COICOP_SATINALIS
- _ COICOP_AYNI

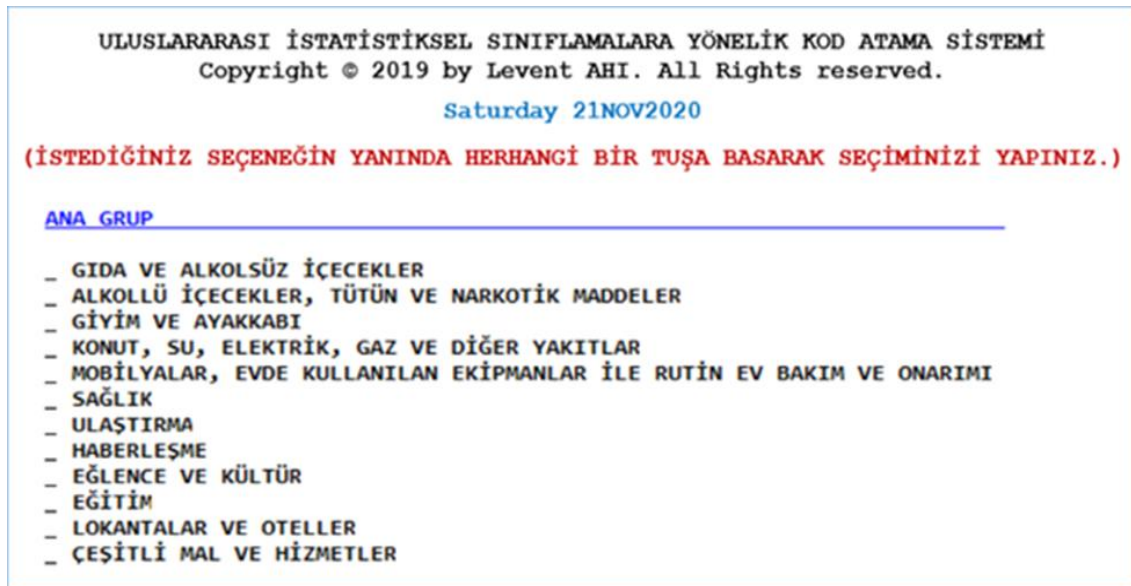
Şekil 4.3. KASİS tablo parametresi ekran görüntüsü

4.1.2. Ana grup

Ana grup parametresinde, COICOP sınıflamasına uygun olarak aşağıdaki ana harcama grupları kullanılmıştır.

1. Gıda ve Alkolsüz İçecekler
2. Alkollü İçecekler, Tütün ve Narkotik Maddeler
3. Giyim ve Ayakkabı
4. Konut, Su, Elektrik, Gaz ve Diğer Yakıtlar
5. Mobilyalar, Evde Kullanılan Ekipmanlar İle Rutin Ev Bakım ve Onarımı
6. Sağlık
7. Ulaştırma
8. Haberleşme
9. Eğlence ve Kültür
10. Eğitim
11. Lokantalar ve Oteller
12. Çeşitli Mal ve Hizmetler

KASİS'teki ana grup parametresine ilişkin ekran görüntüsüne Şekil 4.4'te yer verilmiştir.



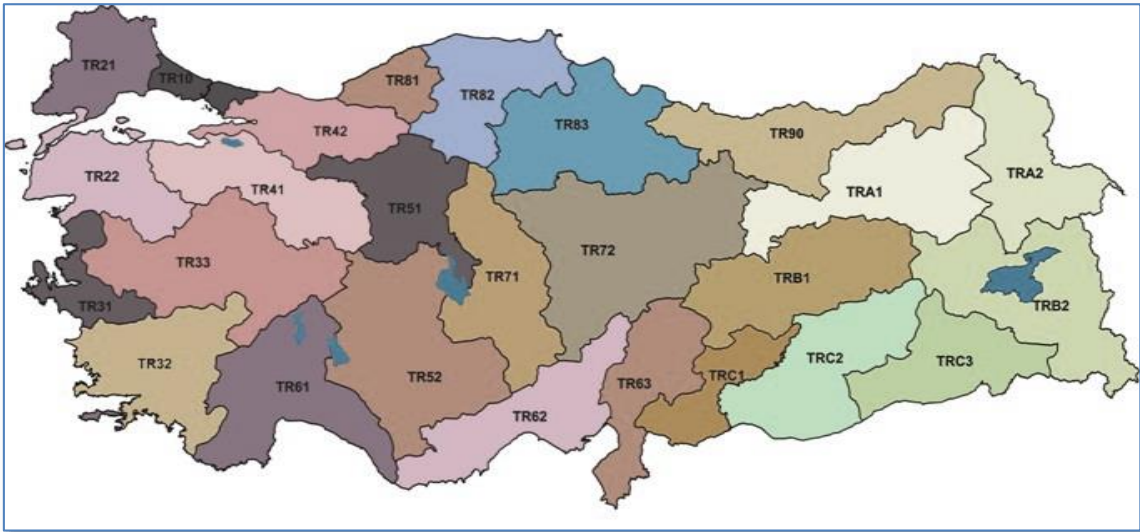
Şekil 4.4. KASİS ana grup parametresi ekran görüntüsü

4.1.3. Bölge

Bölge parametresinde İstatistiki Bölge Birimleri Sınıflaması (İBBS) Düzey 2 sınıflaması kullanılmıştır. TÜİK ve Devlet Planlama Teşkilatı (DPT) tarafından hazırlanan İBBS raporu 2002/4720 sayılı Bakanlar Kurulu kararı ile resmi gazetede yayımlanmıştır [94].

Bu karara göre, “bölgesel istatistiklerin toplanması, geliştirilmesi, bölgelerin sosyo-ekonomik analizlerinin yapılması, bölgesel politikaların çerçevesinin belirlenmesi ve Avrupa’da kullanılan İstatistiki Bölge Birimleri Sınıflandırması’na (NUTS) uygun karşılaştırılabilir istatistiki veri tabanı oluşturulması amacıyla ülke genelinde İBBS tanımlanmıştır”. İBBS’de iller Düzey 3 olarak tanımlanmış; ekonomik, sosyal ve coğrafi yönden benzerlik gösteren komşu iller ise bölgesel kalkınma planları ve nüfus büyüklükleri de dikkate alınarak Düzey 1 ve Düzey 2 olarak gruplandırılmak suretiyle hiyerarşik İBBS yapılmıştır [94].

Düzey 3 kapsamındaki İstatistiki Bölge Birimleri 81 adet olup il düzeyindedir. Her il bir İstatistiki Bölge Birimini tanımlamaktadır. Düzey 2 İstatistiki Bölge Birimleri, Düzey 3 kapsamındaki komşu illerin gruplandırılması sonucu tanımlanmış olup, 26 adettir. Düzey 1 İstatistiki Bölge Birimleri ise Düzey 2 İstatistiki Bölge Birimlerinin gruplandırılması sonucu tanımlanmış olup, 12 adettir [94]. Düzey 2 İstatistiki Bölge Birimleri harita üzerinde Şekil 4.5’te gösterilmiştir. Düzey 1, Düzey 2 ve Düzey 3 bazında İBBS Ek-5’te verilmiştir.



Şekil 4.5. Düzey 2 İstatistiki Bölge Birimleri

KASİS'teki bölge parametresine ilişkin ekran görüntüsüne Şekil 4.6'da yer verilmiştir.

ULUSLARARASI İSTATİSTİKSEL SINIFLAMALARA YÖNELİK KOD ATAMA SİSTEMİ
Copyright © 2019 by Levent AHI. All Rights reserved.

Saturday 21NOV2020

(İSTEDİĞİNİZ SEÇENEĞİN YANINDA HERHANGİ BİR TUŞA BASARAK SEÇİMİNİZİ YAPINIZ.)

BÖLGE

- _ ADANA
- _ ANKARA
- _ ANTALYA
- _ BALIKESİR
- _ BURSA
- _ DENİZLİ
- _ DİYARBAKIR
- _ EDİRNE
- _ ERZURUM
- _ GAZİANTEP
- _ HATAY
- _ İSTANBUL
- _ İZMİR
- _ KARS
- _ KASTAMONU
- _ KAYSERİ
- _ KOCAELİ
- _ KONYA
- _ MALATYA
- _ MANİSA
- _ NEVŞEHİR
- _ SAMSUN
- _ SİİRT
- _ TRABZON
- _ VAN
- _ ZONGULDAK

Şekil 4.6. KASİS bölge parametresi ekran görüntüsü

4.1.4. Yıl

Anket yapılan yılı ifade etmektedir. Yıl parametresi olarak, 2013-2018 yılları arasındaki yıllar kullanılmıştır. KASİS'teki yıl parametresine ilişkin ekran görüntüsüne Şekil 4.7'de yer verilmiştir.

ULUSLARARASI İSTATİSTİKSEL SINIFLAMALARA YÖNELİK KOD ATAMA SİSTEMİ
Copyright © 2019 by Levent AHI. All Rights reserved.

Saturday 21NOV2020

(İSTEDİĞİNİZ SEÇENEĞİN YANINDA HERHANGİ BİR TUŞA BASARAK SEÇİMİNİZİ YAPINIZ.)

YIL

- 2013
- 2014
- 2015
- 2016
- 2017
- 2018

Şekil 4.7. KASİS yıl parametresi ekran görüntüsü

4.1.5. Ay

Anket yapılan ayı ifade etmektedir. Ay parametresi olarak, 12 takvim ayı kullanılmıştır ve her bir ay tekli veya çoklu veya tüm aylar aynı anda seçilebilmektedir. KASİS'teki ay parametresine ilişkin ekran görüntüsüne Şekil 4.8'de yer verilmiştir.

ULUSLARARASI İSTATİSTİKSEL SINIFLAMALARA YÖNELİK KOD ATAMA SİSTEMİ
Copyright © 2019 by Levent AHI. All Rights reserved.

Saturday 21NOV2020

(İSTEDİĞİNİZ SEÇENEĞİN YANINDA HERHANGİ BİR TUŞA BASARAK SEÇİMİNİZİ YAPINIZ.)

AY

- 1
- 2
- 3
- 4
- 5
- 6
- 7
- 8
- 9
- 10
- 11
- 12
- HEPSİ

Şekil 4.8. KASİS ay parametresi ekran görüntüsü

KASİS'in ayrıntılı program menüsüne Ek-6'da yer verilmiştir.

4.2. Veri Setleri ve Anket Veri Giriş Programı

Çalışmada altlık olarak kullanılan tüketim ve sınıflama sözlüğü veri seti olmak üzere iki adet veri seti bulunmaktadır.

4.2.1. Tüketim veri seti

Çizelge 4.1’de ekran görüntüsü verilen tüketim veri setinde; 2013-2018 yılları arasında 12 676 591 kayıt bulunmaktadır. Tüketim veri setinde sekiz değişken bulunmaktadır. BOLGE_KOD, REFERANS_YIL, REFERANS_AY, COICOP_KOD ve OLCU_BIRIM nümerik değişkenler ve TABLO, COICOP_TANIM ve ANA_GRUP karakter değişkenlerdir.

Çizelge 4.1. KASİS tüketim veri seti ekran görüntüsü

BOLGE_KOD	REFERANS_YIL	REFERANS_AY	TABLO	COICOP_TANIM	ANA_GRUP	COICOP_KOD	OLCU_BIRIM
1	2013	1	COICOP_AYNI	pirinç berzani	1	111101020	1
1	2013	1	COICOP_AYNI	ekmek normal ekmek	1	111201010	1
1	2013	1	COICOP_AYNI	ekmek normal ekmek	1	111201010	1
1	2013	1	COICOP_AYNI	ekmek normal ekmek	1	111201010	1
1	2013	1	COICOP_AYNI	ekmek normal ekmek	1	111201010	1
1	2013	1	COICOP_AYNI	ekmek normal ekmek	1	111201010	1
56	2015	6	COICOP_STOK	çay ithal	1	121201030	1500
56	2015	6	COICOP_STOK	buğday	1	111210010	1500
56	2015	6	COICOP_STOK	şeker toz	1	118101010	1500
56	2015	6	COICOP_STOK	salça domates	1	117309090	1500
56	2015	6	COICOP_STOK	pirinç osmancık	1	111101140	1500
56	2015	6	COICOP_STOK	ayçiçek yağı	1	115401010	2000
56	2015	6	COICOP_STOK	un buğday unu	1	111201010	1500
56	2015	6	COICOP_STOK	otlu peynir	1	114505160	1500
56	2015	6	COICOP_STOK	kuru soğan	1	117170010	1500
56	2015	6	COICOP_STOK	yeşil sabun	12	1213204020	1500

4.2.2. Sınıflama sözlüğü veri seti

Çizelge 4.2’de ekran görüntüsü bulunan 2013-2018 COICOP sözlüğü veri setinde 61 187 kayıt bulunmaktadır. Veri setinde YIL, KOD, AD, SOZLUK_TANIM, SOZLUK_TANIM_DUZ, SOZLUK_OLCU_BIRIM, SOZLUK_OLCU_BIRIM_TANIM ve ANA_GRUP olmak üzere sekiz karakter değişken bulunmaktadır.

Çizelge 4.2. KASİS 2013-2018 COICOP sözlüğü veri seti ekran görüntüsü

YIL	KOD	AD	SOZLUK_TANIM	SOZLUK_OLCU_BIRIM	SOZLUK_OLCU_BIRIM_TANIM	ANA_GRUP
2013	1011101010	Anaokulu Ücreti	anaokulu ücreti	36	AYLIK	10
2013	1011102010	Özel İlköğretim Ücreti	özel ilköğretim ücreti	36	AYLIK	10
2013	1011102010	Özel İlköğretim Ücreti	özel ilköğretim ücreti	60	YILLIK	10
2013	1011103010	Okul Öncesi Eğitim ve İlköğretim Kayıt Ücreti	okul öncesi eğitim ve ilköğretim kayıt ücreti	41	DEFA	10
2013	1011104010	Karne Parası (Okul Öncesi Eğitim ve İlköğretim)	karne parası (okul öncesi eğitim ve ilköğretim)	41	DEFA	10
2013	1011105010	Diploma Ücreti (Okul Öncesi Eğitim ve İlköğretim)	diploma ücreti (okul öncesi eğitim ve ilköğretim)	41	DEFA	10
2013	1011106010	Etüd ücreti (Okul Öncesi Eğitim ve İlköğretim)	etüd ücreti (okul öncesi eğitim ve ilköğretim)	36	AYLIK	10
2013	1011106010	Etüd ücreti (Okul Öncesi Eğitim ve İlköğretim)	etüd ücreti (okul öncesi eğitim ve ilköğretim)	41	DEFA	10
2013	1011107010	İlköğretim İçin Kurs Ücreti (Okulda)	ilköğretim için kurs ücreti (okulda)	36	AYLIK	10

4.2.3. Anket veri giriş programı

HBA’da TÜİK tarafından geliştirilmiş anket veri giriş programı kullanılmaktadır. Anketör bu veri giriş programı aracılığıyla hanelerin yaptığı harcama kayıtlarını ofiste kaydetmektedirler. Anketör tarafından veri giriş programına 2. sırada “makarna” harcamasını kaydetmek için tanım alanına “makarna” yazıldıktan sonra bilgisayar destekli kod atama mantığına uygun olarak kod alanında, içerisinde makarna geçen harcama tanımları ve kodları ekrana gelmektedir. Ekran görüntüsüne Şekil 4.9’da yer verilmiştir.

SIRA_NO2	DEFTER	NO	SATINALIS_AD_CINS	SATINALIS_MARKA
1	1	1	EKMEK	
2	1	2	makarna	

Şekil 4.9. Anket veri giriş programı ekran görüntüsü örneği 1

Şekil 4.10’da gösterildiği üzere, içerisinde makarna ifadesi geçen alternatif kodlar bulunmaktadır.

KOD	AD
111601010	Makarna (Çubuk, Fiyonk, Erişte vb.) (Sade) (Kilogram-1500)
111601020	Makarna (Çubuk, Fiyonk, Erişte vb.) (Kepekli, Sebzeli, Sütü, Yumurtalı vb.) (Kilogram-1500)
111601030	Makarna Spagetti (Sade) (Kilogram-1500)
111601040	Makarna Spagetti (Kepekli, Sebzeli, Sütü, Yumurtalı vb.) (Kilogram-1500)
111601050	Makarna Kuskus (Kilogram-1500)
111601060	Makarna Lazanya (Kilogram-1500)
111601070	Makarna Kesme (Ev Makarnası-Erişte) (Kilogram-1500)

Şekil 4.10. Anket veri giriş programı ekran görüntüsü örneği 2

Şekil 4.10’da gösterildiği üzere alternatif kodlardan hanenin yaptığı harcamaya uygun olan “111601030- Makarna Spagetti (Sade)” kodu seçilmiştir. Şekil 4.11’de, harcamaya ilişkin miktarı, ölçü birimi ve toplam değer girilerek bu harcama için veri giriş işlemi bitirilmiş olmaktadır.

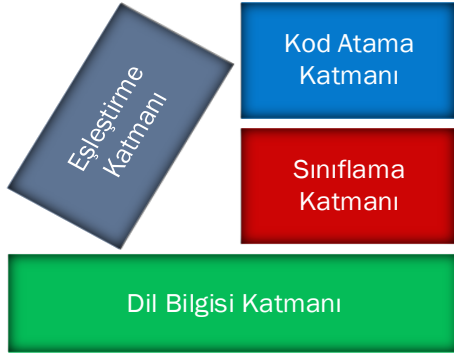
MİKTAR	SATINALIS_OLCU_BIRIM		SATINALIS_TOPLAM_DEGER
1	1500	Kilogram	3

Şekil 4.11. Anket veri giriş programı ekran görüntüsü örneği 3

Anketörün bu şekilde veri giriş programına yaptığı kayıt veri tabanına gönderimi sağlandıktan sonra Çizelge 4.1’de ekran görüntüsü örneği verilen tüketim veri setine ilave edilmektedir. KASİS, bu veri setinde COICOP_TANIM değişkeninde yer alan anketörün yazdığı tanıma Çizelge 4.2’de ekran görüntüsü örneği verilen COICOP sözlüğü veri setinde yer alan SOZLUK_TANIM değişkeninde yer alan tanımlar ile eşleştirmeye çalışmaktadır.

4.3. Katmanlar

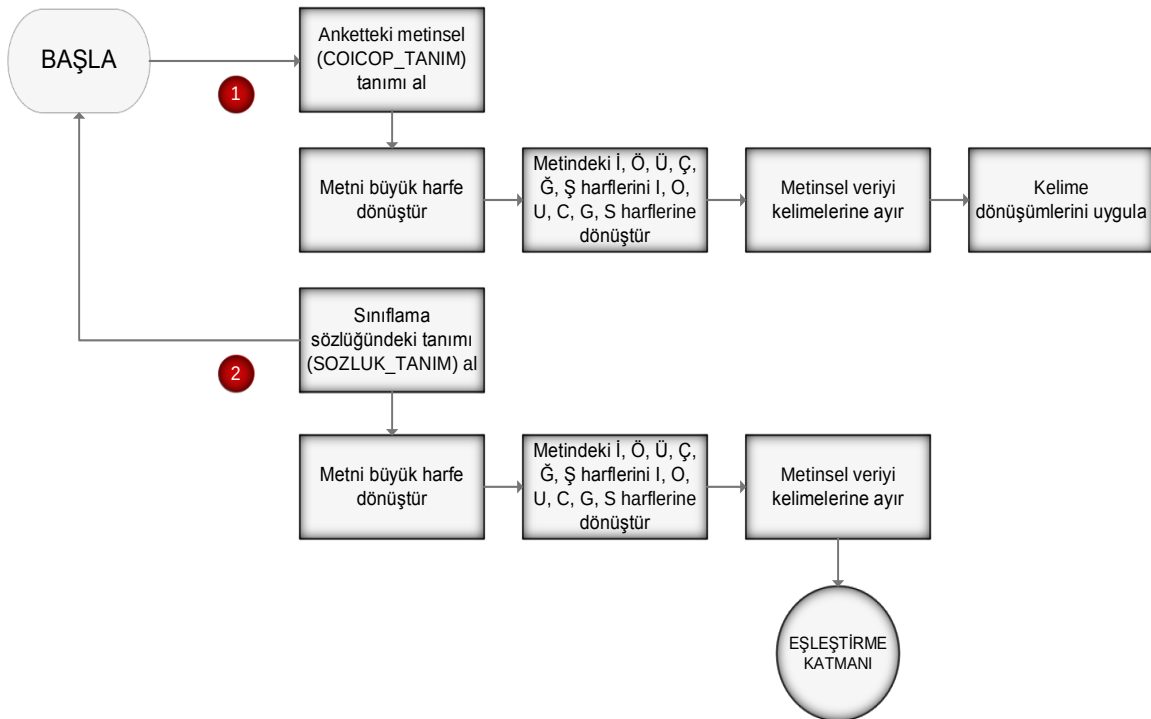
KASİS’in Şekil 4.12’de görüleceği üzere dört farklı katmanı bulunmaktadır.



Şekil 4.12. KASİS'in katmanları

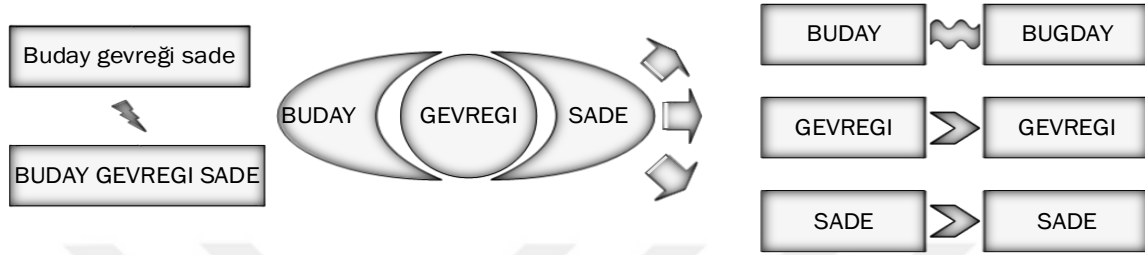
4.3.1. Dil bilgisi katmanı

Katmanların birincisi, dil bilgisi katmanıdır. Şekil 4.13'te yer alan katmanda anketör tarafından yazılmış harcama tanımını içeren tüketim veri setindeki COICOP_TANIM değişkeni büyük harfe dönüştürülür ve Türkçe karakterlerden arındırılır. Aynı işlemler sınıflama sözlüğündeki tanımlara da uygulanarak sözlükte yer alan ve anketör tarafından yazılmış harcama tanımları kelimelerine ayrılır. Kelimelerine ayrılmış olan COICOP_TANIM kelimelerine, yazım yanlışlarını gidermek amacıyla veri tabanında kayıtlı olan kelime dönüşümleri ile Bölüm 4.3.3'te açıklanan işlemler uygulanır.



Şekil 4.13. KASİS dil bilgisi katmanı algoritması

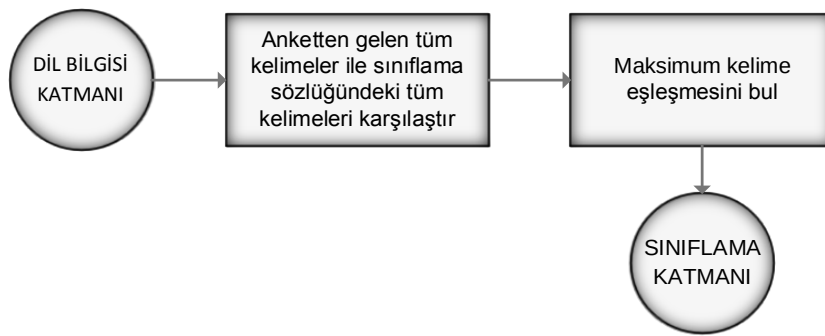
Dilbilgisi katmanında yapılan işlemlere örnek olarak verilen Şekil 4.14'e göre, anketör "Buday gevreği sade" tanımını yazarak COICOP kod ataması yapmıştır. KASİS, dil bilgisi katmanında önce bu tanım büyük harfe dönüştürülerek Türkçe harflerinden arındırılmaktadır. Daha sonra tanım kelimelerine ayrılarak "buday" kelimesindeki yazım yanlışları düzeltilerek "buğday"a dönüştürülmektedir.



Şekil 4.14. KASİS dil bilgisi katmanında yapılan işlemler

4.3.2. Eşleştirme katmanı

Katmanlardan ikincisi, Şekil 4.15'te verilen eşleştirme katmanıdır. Bu katmanda kelimelerine ayrılmış harcama tanımı ile kelimelerine ayrılmış sınıflama sözlüğündeki tanımlar karşılaştırılmaktadır. Anketörün yazdığı harcama tanımının kelimelerinin sözlükteki hangi kod veya kodların tanımının kelimelerini maksimum olarak içerdiği bulunmaktadır.



Şekil 4.15. KASİS eşleştirme katmanı algoritması

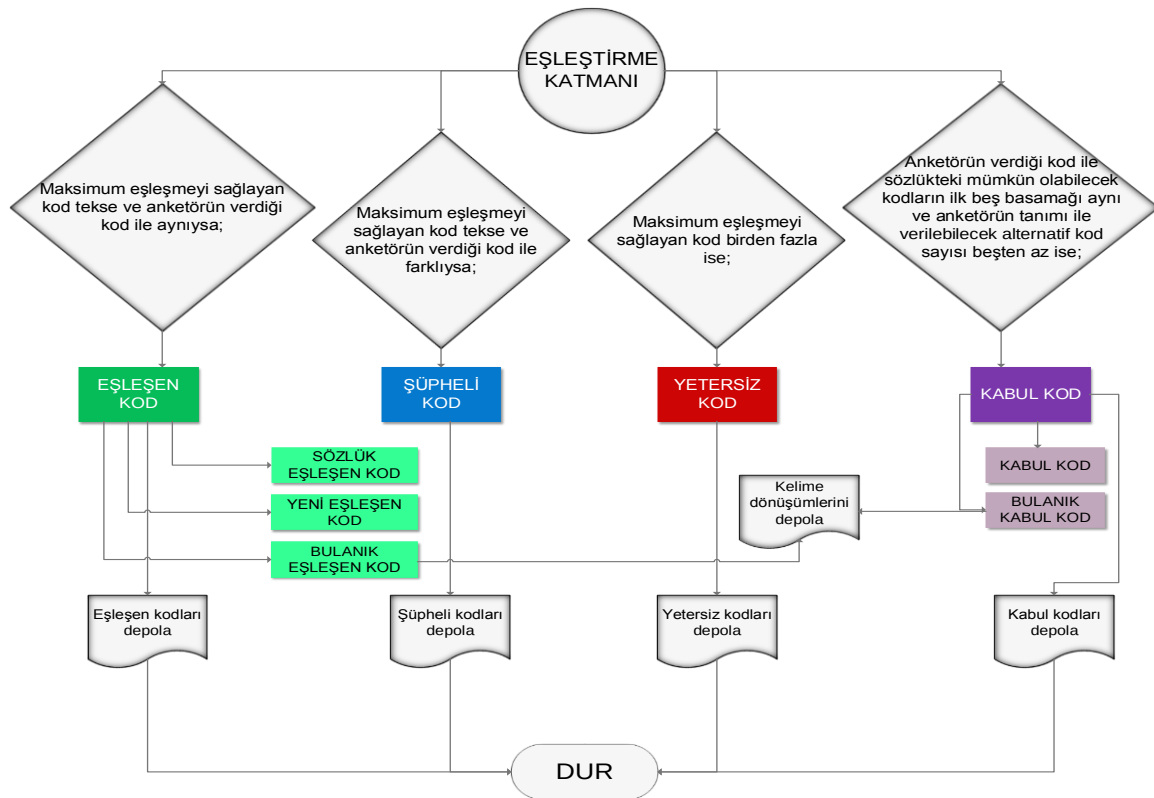
Çizelge 4.3'te dil bilgisi katmanından geçerek gelen "bulgur" harcama tanımı sözlükteki tanımlar ile karşılaştırıldığında bu tanımın yer aldığı 16 tane alternatif kod belirlenmiştir.

Çizelge 4.3. KASİS eşleştirme katmanı örneği

ANKETOR_TANIM	SOZLUK_KOD	SOZLUK_TANIM
bulgur	1111102090	Bulgur Pilavı (Restoranda) (Porsiyon - 4370)
bulgur	1111131090	Bulgur Pilavı (Cafe, Bar, Pastane vb. Yerlerde) (Porsiyon - 4370)
bulgur	1112002090	Bulgur Pilavı (Kantinlerde) (Porsiyon - 4370)
bulgur	111209010	Bulgur İnce (Kilogram - 1500)
bulgur	111209020	Bulgur İri (Kilogram - 1500)
bulgur	111209030	Bulgur Domatesli (Kilogram - 1500)
bulgur	111209040	Bulgur Sebzeli (Kilogram - 1500)
bulgur	111209050	Bulgur Simit (Yöresel) (Kilogram - 1500)
bulgur	111209060	Bulgur Esmer İri Pilavlık (Kilogram - 1500)
bulgur	111209070	Bulgur Esmer İnce Çiğköftelik (Kilogram - 1500)
bulgur	111209080	Bulgur Kepekli İri Pilavlık (Kilogram - 1500)
bulgur	111209090	Bulgur Kepekli İnce Çiğköftelik (Kilogram - 1500)
bulgur	111209100	Bulgur Başbaşı (Kilogram - 1500)
bulgur	111209110	Bulgur Firik (Kilogram - 1500)
bulgur	111209980	Bulgur Diğer (Kilogram - 1500)
bulgur	119109010	Kolay Çiğ Köfte Seti (Baharat ve Bulgurlu Karışım) (Paket - 4360)

4.3.3. Sınıflama katmanı

Katmanlardan üçüncüsü, sınıflama katmanıdır. Şekil 4.16'da yer alan bu katmanda maksimum kelime eşleşmeleri bulunan harcama tanımları ve sözlük tanımları sınıflanmaktadır. Burada dört olası durum ortaya çıkmaktadır:



Şekil 4.16. KASİS sınıflama katmanı algoritması

1. Eşleşen kod: Anketörün yazdığı harcama tanımındaki kelimeleri maksimum olarak içeren COICOP kodu tekse ve bu tanıma karşılık gelen kod anketörün verdiği harcama kodu ile aynıysa kayıt “eşleşen” olarak sınıflanmaktadır. Eşleşen koddaki; sözlük eşleşen kod, yeni eşleşen kod ve bulanık eşleşen kod olmak üzere üç farklı tür bulunmaktadır.

Sözlük eşleşen kod: Anketörün yazdığı harcama tanımı, COICOP sözlüğündeki tanımla sadece tek bir kayıtlı birebir eşleşiyorsa bu kayıt KASİS’in eşleştirme katmanına girmeden direkt olarak bu sınıfta yer almaktadır. Sözlük eşleşen kod ile ilgili örnekler Çizelge 4.4’te yer almaktadır.

Çizelge 4.4. KASİS sözlük eşleşen kod örneği

ANKETOR_KOD	ANKETOR_TANIM	SOZLUK_KOD	SOZLUK_TANIM	DURUM
111101010	pirinç baldo	111101010	pirinç baldo	SÖZLÜK EŞLEŞEN
111101020	pirinç berzani	111101020	pirinç berzani	SÖZLÜK EŞLEŞEN
111201010	ekmek normal ekmek	111201010	ekmek normal ekmek	SÖZLÜK EŞLEŞEN
111201160	ekmek köy ekmeği	111201160	ekmek köy ekmeği	SÖZLÜK EŞLEŞEN
111405990	kurabiye diğer	111405990	kurabiye diğer	SÖZLÜK EŞLEŞEN
111501010	un buğday unu	111501010	un buğday unu	SÖZLÜK EŞLEŞEN
111509010	bulgur ince	111509010	bulgur ince	SÖZLÜK EŞLEŞEN
111509020	bulgur iri	111509020	bulgur iri	SÖZLÜK EŞLEŞEN

Yeni eşleşen kod: İlk etapta anketörün yazdığı harcama tanımı COICOP sözlüğünde tanımla birebir eşleşmiyorsa, KASİS’in eşleştirme ve sınıflama katmanlarından geçerek eşleşen olarak sınıflanıyorsa bu kaydın durumu “yeni eşleşen” olmaktadır. Yeni eşleşen kod ile ilgili örnekler Çizelge 4.5’te yer almaktadır.

Çizelge 4.5. KASİS yeni eşleşen kod örneği

ANKETOR_KOD	ANKETOR_TANIM	SOZLUK_KOD	SOZLUK_TANIM	DURUM
111406050	simit	111406050	simit (eve alınan)	YENİ EŞLEŞEN
114505050	peynir çökelek	114505050	peynir çökelek (minci vb.)	YENİ EŞLEŞEN
115301040	zeytin yağı sızma	115301040	zeytinyağı (sızma)	YENİ EŞLEŞEN
116101020	ortakal sofralık taze	116101020	portakal sofralık (taze)	YENİ EŞLEŞEN
116101020	portakal sofralık	116101020	portakal sofralık (taze)	YENİ EŞLEŞEN
116102010	greyfurt	116102010	greyfurt (taze)	YENİ EŞLEŞEN
117115010	ispanak taze	117115010	ispanak (taze sebze)	YENİ EŞLEŞEN
118101020	toz şeker	118101020	toz şeker (açık)	YENİ EŞLEŞEN
117336040	zeytin siyah	117336040	zeytin siyah (sofralık)	YENİ EŞLEŞEN
117408010	pancar taze	117408010	pancar (taze sebze)	YENİ EŞLEŞEN

Bulanık eşleşen kod: Anketörün yazdığı harcama tanımı ne COICOP sözlüğünde tanımla

birebir eşleşmiş ne de KASİS'in eşleştirme ve sınıflama katmanlarından geçerek eşleşmiştir.

Sınıflama katmanında yapılan işlemlere örnek olarak verilen Çizelge 4.6'ya göre, kelimeler sınıflama sözlüğündeki kelimeler ile karşılaştırıldığında anketörün yazarken yazım yanlışı yaptığı “sakaız” kelimesinin sözlükteki “sakız” kelimesi ile bulanık eşleştirme yöntemi ile eşleştirilmiştir. Anketörün yazdığı tanım “sakaız tekli” yetersiz olarak sınıflanması gerekirken bulanık eşleştirme yöntemi ile “sakaız” kelimesinin “sakız” ifadesine dönüşümü gerçekleştirilerek kaydın eşleşmesi sağlanmıştır. Bulanık eşleştirme yöntemi kullanılarak eşleştirildiği için kaydın durumu “bulanık eşleşen” olmaktadır. Tekrar aynı yazım yanlışı tekrar olabilir diye kelime dönüşümleri veri tabanına “sakaız” ile “sakız” dönüşümü ilave edilmektedir.

Çizelge 4.6. KASİS bulanık eşleşen kod örneği 1

ANKETOR_KOD	ANKETOR_TANIM	SOZLUK_KOD	SOZLUK_TANIM	DURUM
111509010	bulgur inci	111509010	bulgur ince	BULANIK EŞLEŞEN
121101030	kahvesi çekilmiş	121101030	kahve çekilmiş	BULANIK EŞLEŞEN
1011106010	okulda etüd ücreti	1011106010	etüd ücreti (okul öncesi eğitim ve ilköğretim)	BULANIK EŞLEŞEN
913102020	hesap makinası	913102020	hesap makinesi	BULANIK EŞLEŞEN
111202020	bebek bisküvisi	111202020	bisküvi bebe	BULANIK EŞLEŞEN
1213211030	bebek şampuanı	1213211030	şampuan (bebe ve çocuk)	BULANIK EŞLEŞEN
1213217070	bebek kolonyası	1213217070	bebe kolonyası (johnson, dalin vb.)	BULANIK EŞLEŞEN
312231070	şalvar emprime kumaş	312231070	şalvar empirme kumaş (kadın için)	BULANIK EŞLEŞEN
512102020	yolluk makina	512102020	yolluk makine halısı	BULANIK EŞLEŞEN
561103150	jel deterjanı	561103150	jel deterjan (cif vb.) (hijyenik zeminler için)	BULANIK EŞLEŞEN
623305010	çıkıkçı ücreti	623305010	kırkçı ve çıkıkçıya ödenen ücret	BULANIK EŞLEŞEN
118407010	sakaız tekli	118407010	sakız (tek)	BULANIK EŞLEŞEN

Sınıflama katmanında yapılan işlemlere örnek olarak verilen Çizelge 4.7'ye göre, anketör tarafından “yumurta çift5lik” tanımı yazılarak “114701020 - yumurta çiftlik” kodu verilmiştir. Dil bilgisi katmanında, anketörün yazdığı tanım önce büyük harfe dönüştürülerek Türkçe harflerden arındırılmış ve “YUMURTA CIFT5LIK” halini almıştır. Sözlükteki tanımlara da aynı işlemler yapılmış ve anketör tanımı ile sözlük tanımındaki kelimeler birer birer karşılaştırılmıştır. Anketör tanımında yazım yanlışı bulunduğu için “CIFT5LIK” kelimesi sözlükte bulunamamıştır ve geçerli tanım olarak elimizde bulunan “YUMURTA” kelimesinin geçtiği kayıtlara Çizelge 4.8'de yer verilmiştir.

Çizelge 4.7. KASİS bulanık eşleşen kod örneği 2

ANKETOR_KOD	ANKETOR_TANIM	SOZLUK_KOD	SOZLUK_TANIM	DURUM
520905070	makrome ipliği	520905070	makrame ipliği	BULANIK EŞLEŞEN
532102080	tost makinası	532102080	elektrikli tost makinesi	BULANIK EŞLEŞEN
114505110	paynir salamura	114505110	peynir köy (salamura - taze)	BULANIK EŞLEŞEN
116131010	trabzon hurması	116131010	hurma (trabzon) (taze)	BULANIK EŞLEŞEN
552102040	su hortumu	552102040	hortum (sulama)	BULANIK EŞLEŞEN
118407010	yumurta çiftlik	114701020	yumurta çiftlik	BULANIK EŞLEŞEN

Çizelge 4.8. İçerisinde “yumurta” kelimesi geçen sözlükteki kodlar

MADDEKOD	MADDE TANIM
111301020	Makarna (Çubuk, Fiyonk, Erişte vb.) (Kepekli, Sebzeli, Sütü, Yumurtalı vb.)
111301040	Makarna Spagetti (Kepekli, Sebzeli, Sütü, Yumurtalı vb.)
112501310	Sakatat Koç Yumurtası
114701010	Yumurta Köy
114701020	Yumurta Çiftlik
114701030	Yumurta Kırık
114701040	Yumurta Çift Sarılı
114701050	Yumurta Sıvı Yumurta
114702010	Yumurta Bildircin
114799990	Yumurta Diğer
118302070	Çikolata Sürprizli Yumurta Çikolata (Kinder Sürpriz, Toto vb.)
454101090	Kömür Ücreti Yumurta
532102100	Elektrikli Yumurta Pişirme Makinesi
541103310	Yumurtalık (Seramik)
541312340	Pasta Fırçası (Yumurta, Yağ vb. Sürmek İçin)
541312480	Yumurta Dilimleyicisi
541312490	Yumurtalık
541312500	Yumurtalık
1111102560	Yumurta (Restoranda)
1111202560	Yumurta (Cafe, Bar, Pastane vb. Yerlerde)
1111304170	Gobit Ekmek Arası Yumurta, Soğan (Seyyar Satıcı vb. Yerlerde)
1111310100	Haşlanmış Tek Yumurta (Seyyar Satıcı vb. Yerlerde)
1112102560	Yumurta (Kantinlerde)

Verilen kodda bir hata olmadığını gözümüzle teyit edebilmemize rağmen, tanımdaki yazım yanlışından dolayı kayıt, “yetersiz” olarak sınıflanmıştır. KASİS bu yazım yanlışını belirlemiş ve “çiftlik” kelimesinin “çiftlik” olduğuna karar vererek dönüşümü gerçekleştirmiştir. Kaydın durumu “bulanık eşleşen” olmuştur ve tekrar aynı yazım yanlış olabilir diye kelime dönüşümleri veri tabanına “çiftlik” ile “çiftlik” dönüşümü ilave edilmiştir. Aynı yazım yanlış tekrar geldiğinde KASİS direkt olarak “çiftlik” kelimesini “çiftlik” olarak dönüşümünü sağlamakta ve bundan sonraki katmanlarda işlemlerine devam etmektedir.

Burada anlatılan bulanık eşleştirme algoritması olarak Bölüm 2.6.2’de yer alan genelleştirilmiş düzenleme mesafesini hesaplayan COMPGED fonksiyonu kullanılmıştır.

Bir ifadenin diğer bir ifade ile aynı olduğunu belirlerken belirli bir maliyeti göz önünde bulundurmak ve bunun için bir eşik değeri belirlemek gerekmektedir. Uygun eşik değerin ne olması gerektiği öznel bir süreçtir. Eşik değerini belirlerken ne tür sonuçların tolere edilebileceğine ve ne tür sonuçların eşleştirilmeyeceğine karar vermek gerekmektedir [95].

Bu yöntemde, COMPGED fonksiyon değeri 100 ve 100'den küçük olan eşleşmeler dikkate alınmıştır. Bu değerlerin uygulanması sonucunda, örnek olması açısından KASİS tarafından belirlenmiş dört kelime ile ilgili yapılan yanlış yazımlar liste halinde Çizelge 4.9'da verilmiştir.

Çizelge 4.9. Yanlış yazılan kelime dönüşümleri

RESTORANDA 59	CIFTLIK 57	DETERJANI 56	AYCICEK 50
RASTORANDA	C,FTLIK	DATERJANI	AAYCICEK
RESORANDA	C,IFTLIK	DDETERJANI	ACICEK
RESOTORANDA	C,IFTLIK	DEDERJANI	ACYICEK
RESOTRANDA	CCIFTLIK	DEDERJANI	AGCICEK
RESRORANDA	CFILTIK	DEERJANI	AKCICEK
RESRTORANDA	CFITLIK	DEETERJANI	AQYCICEK
RESSTORANDA	CFTLIK	DEETRANI	ASYCICEK
REST6ORANDA	CI,FTLIK	DERTERJANI	ATYCICEK
RESTARANDA	CICFTLIK	DETARJANI	AY,ICEK
RESTIORANDA	CIFFLIK	DETEJANI	AYACICEK
RESTOANDA	CIFFTLIK	DETERAJANI	AYC,CEK
RESTOARANDA	CIFITLIK	DETERAJNI	AYC,ICEK
RESTOARNADA	CIFLIK	DETERANI	AYCACEK
RESTOARNDA	CIFLTIK	DETEREJANI	AYCCEK
RESTOERANDA	CIFLTLIK	DETERFJANI	AYCCICEK
RESTOIRANDA	CIFRLIK	DETERHANI	AYCECEK
RESTOORANDA	CIFRTLK	DETERJAANI	AYCEICEK
RESTORAANDA	CIFT5LIK	DETERJABI	AYCI,CEK
RESTORADA	CIFTCIK	DETERJABNI	AYCIACEK
RESTORAMDA	CIFTIK	DETERJAJI	AYCICAK
RESTORAN6A	CIFTIL	DETERJANII	AYCICCEK
RESTORANA	CIFTILIK	DETERJANIN	AYCICECK
RESTORANAD	CIFTILK	DETERJANIU	AYCICEEK
RESTORANADA	CIFTKLIK	DETERJANIZ	AYCICEKJ
RESTORANDA,	CIFTL,IK	DETERJANKI	AYCICEKK
RESTORANDAA	CIFTL,K	DETERJANNI	AYCICEL
RESTORANDAN	CIFTLI,K	DETERJANO	AYCICEY
RESTORANDAQ	CIFTLIHK	DETERJANOI	AYCICIEK

Çizelge 4.9. (devam) Yanlış yazılan kelime dönüşümleri

RESTORANDA 59	CİFTLİK 57	DETERJANI 56	AYCİCEK 50
RESTORANDAU	CİFTLİJK	DETERJANR	AYCİCİK
RESTORANDDA	CİFTLİK,	DETERJENİ	AYCİCK
RESTORANDE	CİFTLİK4	DETERJHANI	AYCİCKE
RESTORANDO	CİFTLİKJ	DETERJANI	AYCİCOEK
RESTORANDS	CİFTLİKK	DETERJİNİ	AYCİCRK
RESTORANTA	CİFTLİKL	DETERJNA	AYCİCEK
RESTORANTDA	CİFTLİL	DETERJNAİ	AYCİECK
RESTORBND	CİFTLİLİK	DETERJNANI	AYCİEK
RESTOREANDA	CİFTLİSK	DETERJNİ	AYCİCEK
RESTORNA	CİFTLİT	DETERJQANI	AYCİOCEK
RESTORNAD	CİFTLİTK	DETERKANI	AYCİOEK
RESTORNADA	CİFTLK	DETERKJANI	AYCİSCEK
RESTORND	CİFTLKİK	DETERRJANI	AYCİCEK
RESTORONDA	CİFTLLİK	DETERSANİ	AYCİEK
RESTORTANDA	CİFTLSK	DETERTANI	AYCİCEK
RESTOTANDA	CİFTLUK	DETERTJANI	AYSİCEK
RESTOTRANDA	CİFTTLİK	DETERYANI	AYTCİCEK
RESTPRANDA	CİFYLİK	DETERJAN	AYYCİCEK
RESTRANDA	CİGFTLİK	DETERJANI	AYYCİEK
RESTROANDA	CİİFTLİK	DETERRJANI	AZYCİCEK
RESTRORANDA	CİİFTLLİK	DETRJANI	YACİCEK
RESTURANDA	CİRTLİK	DETERJANI	
RETORANDA	CİSFTLİK	DEYERJANI	
RETSORAN	CİTFİLİK	DEYTERJANI	
RETSORANDA	CİFTLİK	DFETERJANI	
ROSTORANDA	CİFTTLİK	DTERJANI	
RRESTORANDA	CİTLİK	DTERJANI	
RSTORANDA	CİTTLİK		
RTESTORANDA			
RWSTORANDA			

Çizelge 4.9’a göre, anketörler “restoranda” kelimesini yazarken 59 farklı yazım yanlış yapmışlardır. Benzer şekilde “çiftlik” kelimesi için 57, “deterjanı” kelimesi için 56 ve “ayçiçek” kelimesi için 50 farklı yazım yanlış yapmıştır. Bu kelimeler burada yer alan şekilde yazılmaları halinde kelimelere ilgili dönüşüm uygulanarak doğru şekilde KASİS’te işlemlere devam edilmesi sağlanmaktadır.

2. Şüpheli kod: Anketörün yazdığı harcama tanımındaki kelimeleri en çok içeren COICOP kodu tekse ve bu kod anketörün verdiği harcama kodu ile farklıysa kayıt “şüpheli” olarak

sınıflanmaktadır. Şüpheli kod ile ilgili örnek Çizelge 4.10’da yer almaktadır. “davul fırın” tanımı yazılarak “fırın elektrikli” kodu verilmiştir. Bu yüzden, bu kodun veya tanımın düzeltilmesi önerilmektedir.

Çizelge 4.10. KASİS şüpheli kod örneği

ANKETOR_KOD	ANKETOR_TANIM	ANKETOR_KOD_TANIM	ONERILEN_KOD	ONERILEN_TANIM
531301010	davul fırın	fırın elektrikli	531301060	fırın davul
531501040	şarjlı süpürge	elektrikli süpürge upright	532107020	elektrikli el süpürgesi (şarjlı)
531501990	şarjlı süpürge	elektrikli süpürge diğer	532107020	elektrikli el süpürgesi (şarjlı)
511118020	kanepe ve çekyat tek kişilik	kanepe ve çekyat çift kişilik	511118010	kanepe ve çekyat tek kişilik

3. Yetersiz kod: Anketörün yazdığı harcama tanımındaki kelimeleri içeren COICOP kodu birden fazlaysa kayıt “yetersiz” olarak sınıflanmaktadır. Yetersiz kod ile ilgili örnek Çizelge 4.11’de yer almaktadır. “makarna erişte paket” tanımına karşılık gelen altı farklı kod tanımı bulunmaktadır. Bu yüzden harcama tanımına “sade” gibi bir tanımın da ilave edilmesi kaydın “eşleşen” olarak sınıflanmasına yeterli olacaktır.

Çizelge 4.11. KASİS yetersiz kod örneği

ANKETOR_KOD	ANKETOR_TANIM	SOZLUK_KOD	SOZLUK_TANIM
111601010	makarna erişte paket	111601010	makarna (cubuk, fiyonk, eriste vb.) (sade)
111601010	makarna erişte paket	111601020	makarna (cubuk, fiyonk, eriste vb.) (kepekli, sebzeli, sutlu, yumurtalı vb.)
111601010	makarna erişte paket	111601070	makarna kesme (ev makarnasi-eriste)
111601010	makarna erişte paket	111601100	makarna cabuk eriste (peynirli, domatesli)
111601010	makarna erişte paket	111601110	makarna noodle (cin eristesli)
111601010	makarna erişte paket	119107010	makarna, spagetti sosu (bolonez sos vb.) (hazir pakette toz)

4. Kabul kod: Yetersiz olarak sınıflanan kayıtların içerisinde belirlenen kriterleri sağlayan kayıtlar varsa bu kayıtlar “kabul” olarak sınıflanmaktadır. Anketörün verdiği kod sistem tarafından önerilen alternatif kodların arasında varsa, anketörün verdiği kod ile alternatif olarak önerilen kodların ilk beş basamağı aynı ve anketörün tanımı ile verilebilecek alternatif kod sayısı beşten azsa kayıt “kabul” olarak sınıflanmaktadır. Kabul kodda; kabul kod ve bulanık kabul kod olmak üzere iki farklı tür bulunmaktadır.

Kabul kod ile ilgili örnek Çizelge 4.12’de yer almaktadır. “kraker peynirli” tanımına üç farklı

kod ve kod tanımı karşılık gelmektedir. Alternatif üç kodun da ilk beş basamağı “1112” ile başladığı için kod tanımı ile verilen kod, kabul olma kriterlerini karşıladığı için kayıt “kabul” olarak sınıflanmıştır. Eğer bu kriterler uygulanmamış olsaydı bu kayıt “yetersiz” olarak sınıflanmalıydı.

Çizelge 4.12. KASİS kabul kod örneği

ANKETOR_KOD	ANKETOR_TANIM	ACIKLAMA
111301010	makarna sade	####SOZLUK_KOD=111301010, SOZLUK_TANIM=makarna (çubuk, fiyonk, erişte vb.) (sade) ####SOZLUK_KOD=111301030, SOZLUK_TANIM=makarna spaghetti (sade)
111203040	kraker peynirli	####SOZLUK_KOD=111203020, SOZLUK_TANIM=kraker çubuk kraker (sade, susamlı, peynirli, baharatlı vb.) ####SOZLUK_KOD=111203040, SOZLUK_TANIM=kraker peynirli kraker (çizi kraker vb.) ####SOZLUK_KOD=111203050, SOZLUK_TANIM=kraker peynirli sandviç kraker (eti kraki vb.)
111205020	gofret çikolata kaplı	####SOZLUK_KOD=111205020, SOZLUK_TANIM=gofret çikolata kaplı (ülker çikolatalı gofret, çökonat vb) ####SOZLUK_KOD=111205040, SOZLUK_TANIM=gofret çikolata kaplı çtır top (kingtop, çokotop vb.)

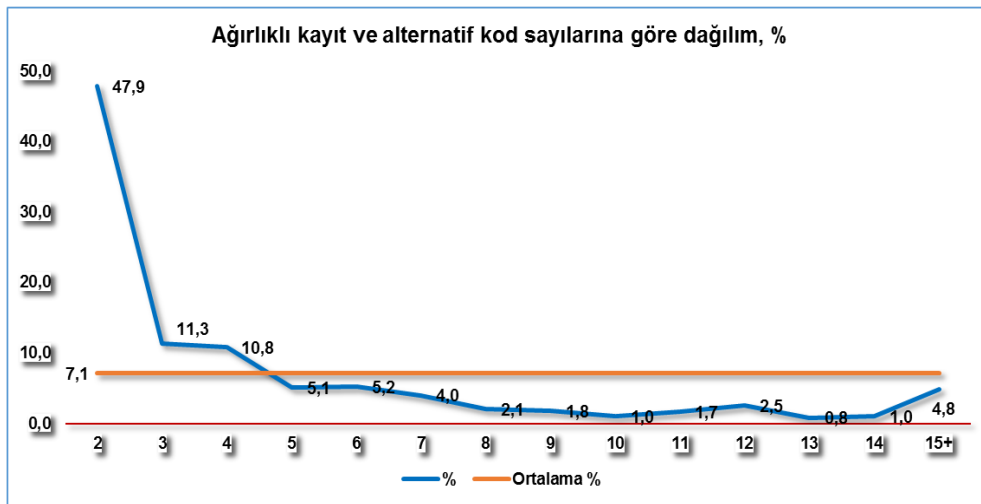
Bulanık kabul kod: Bu kodda, anketörün yazdığı harcama tanımı ne COICOP sözlüğünde tanımla birebir eşleşmiş ne de KASİS’in eşleştirme ve sınıflama katmanlarından geçerek eşleşmiştir.

Bulanık kabul kod ile ilgili örnek Çizelge 4.13’te yer almaktadır. Kelimeler sınıflama sözlüğündeki kelimeler ile karşılaştırıldığında anketörün yazarken yazım yanlışı yaptığı “tleefonu” kelimesinin sözlükteki “telefon” kelimesi ile bulanık eşleştirme yöntemi ile eşleştirilmiştir. Anketörün yazdığı tanım “cep tleefonu kontörü” yetersiz olarak sınıflanması gerekirken bulanık eşleştirme yöntemi ile “tleefonu” kelimesinin “telefonu” ifadesine dönüşümü gerçekleştirilmiştir. Ancak bu dönüşüm bile tanımın sadece bir kodla eşleşmesine yetmemiştir. “cep telefonu kontörü” tanımı sözlükte iki farklı kod tanımına karşılık gelmektedir. Kabul olma kriterlerini karşıladığı ve bulanık eşleştirme yöntemi kullanılarak eşleştirildiği için kaydın durumu “bulanık kabul” olmaktadır. Tekrar aynı yazım yanlışı olabilir diye kelime dönüşümleri veri tabanına “tleefonu” ile “telefonu” dönüşümü ilave edilmektedir.

Çizelge 4.13. KASİS bulanık kabul kod örneği

ANKETOR_KOD	ANKETOR_TANIM	ACIKLAMA	DONUSEN	DONUSUM
830201020	cep tleefonu kontoru	###SOZLUK_KOD=830201020, SOZLUK_TANIM=cep telefonu görüşme ücreti (kontürlü kart) ###SOZLUK_KOD=830203020, SOZLUK_TANIM=cep telefonu tahsis ücreti (kontürlü hat)	TLEEFONU KONTORLU	TELEFONU KONTURLU
830201020	cep telefonu kontoru	###SOZLUK_KOD=830201020, SOZLUK_TANIM=cep telefonu görüşme ücreti (kontürlü kart) ###SOZLUK_KOD=830203020, SOZLUK_TANIM=cep telefonu tahsis ücreti (kontürlü hat)	TELEFRONU KONTORLU	TELEFONU KONTURLU

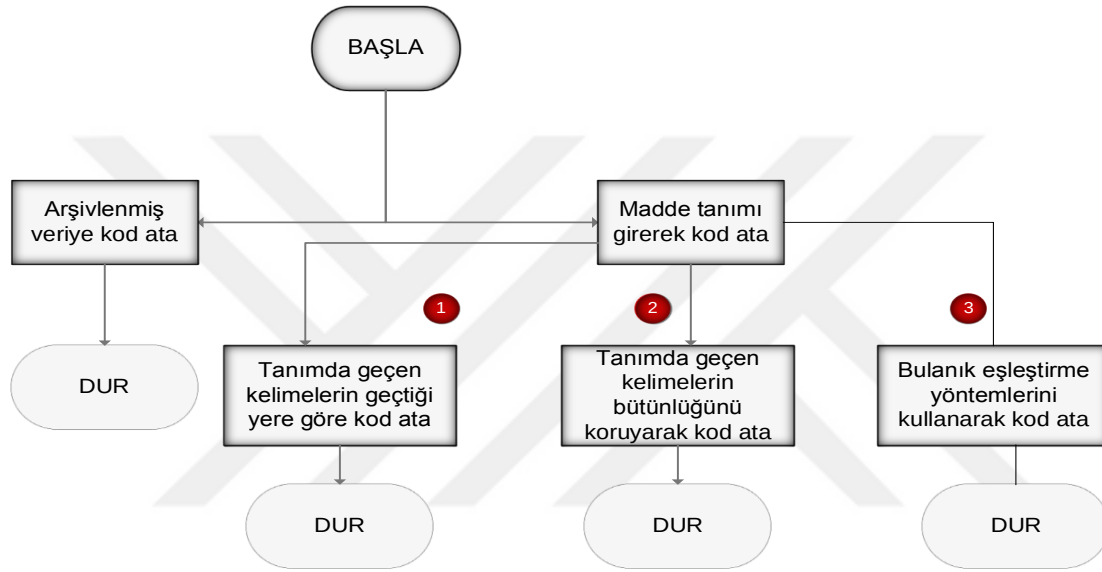
Eurostat, üye ve aday ülkelerden HBA veri setlerini beş senede bir beş basamaklı tüketim harcamalarını içerecek şekilde göndermelerini istemektedir. Ayrıca, HBA'nın mevcut örnek hacmi tüketim harcama tahminlerinin en fazla beş basamaklı olarak kullanılmasına imkan sağlamaktadır. Bu yüzden, kabul kod sınıflamasının ilk şartı önerilen kodların ilk beş basamağının aynı olması kriterine karar verilmiştir. Kabul kod olarak sınıflaması yapılacak kayıtlar için belirlenmiş ikinci şart olan alternatif kod sayısının beşten az olma kriterine, yetersiz kod olarak sınıflanmış harcama kayıtlarına sistemin önerdiği alternatif kod sayılarının Şekil 4.17'de verilen dağılıma göre karar verilmiştir. Alternatif olarak önerilen kod sayısı beşten itibaren, yetersiz kayıtların oranı, %7,1 ortalama oranın altına düşmektedir ve tekrar bu oranın üzerine çıkmamaktadır. Program tarafından önerilen alternatif kod sayısının beşten az olma ve ilk beş basamağı aynı olma kriterlerini yetersiz kayıtların %70'i sağlamaktadır.



Şekil 4.17. Yetersiz kodlara önerilen kod sayılarının dağılımı

4.3.4. Kod atama katmanı

Katmanlardan sonuncusu, kod atama katmanıdır. Şekil 4.18’de verilen bu katmanda daha önceki katmanlarda sınıflaması yetersiz ve şüpheli olarak yapılmış kayıtlara bulanık eşleştirme algoritmaları kullanılarak kod ataması yapılmaktadır. Ayrıca madde tanımı elle girilebilen beş kayda harcama kod ataması da yapılabilmektedir. Burada üç seçenek bulunmaktadır.



Şekil 4.18. KASİS kod atama katmanı algoritması

1. Tanımda geçen kelimelerin bütünlüğünü korumadan kod ata: Bu atama yönteminde aranan kelime hangi sözlük tanımında bulunursa eşleştirme sağlanmaktadır. Örneğin “elma” harcama tanımının arandığı varsayarsak, “elmas” sözlük tanımının eşleşmesi gibi. Bu seçenekte, KASİS yazılan tanımı dil bilgisi ve eşleştirme katmanlarında tanımda geçen kelimelerin bütünlüğünü korumadan işlemlere tabi tutmaktadır. Eşleştirme katmanında oluşan alternatif kodların arasından bulanık eşleştirme yöntemleriyle uygun kod ve kodlar seçilmektedir.

2. Tanımda geçen kelimelerin bütünlüğünü koruyarak kod ata: Bu atama yönteminde kelime bütünlüğü korunarak sözlük tanımında bulunursa eşleştirme sağlanmaktadır. Örneğin top harcama tanımının arandığı varsayarsak, “laptop” sözlük tanımının değil sadece “top” sözlük tanımının eşleşmesi gibi. Bu seçenekte, KASİS yazılan tanımı dil bilgisi ve eşleştirme katmanlarında tanımda geçen kelimelerin bütünlüğünü koruyarak işlemlere tabi tutmaktadır.

Eşleştirme katmanında oluşan alternatif kodların arasından bulanık eşleştirme yöntemleriyle uygun kod ve kodlar seçilmektedir.

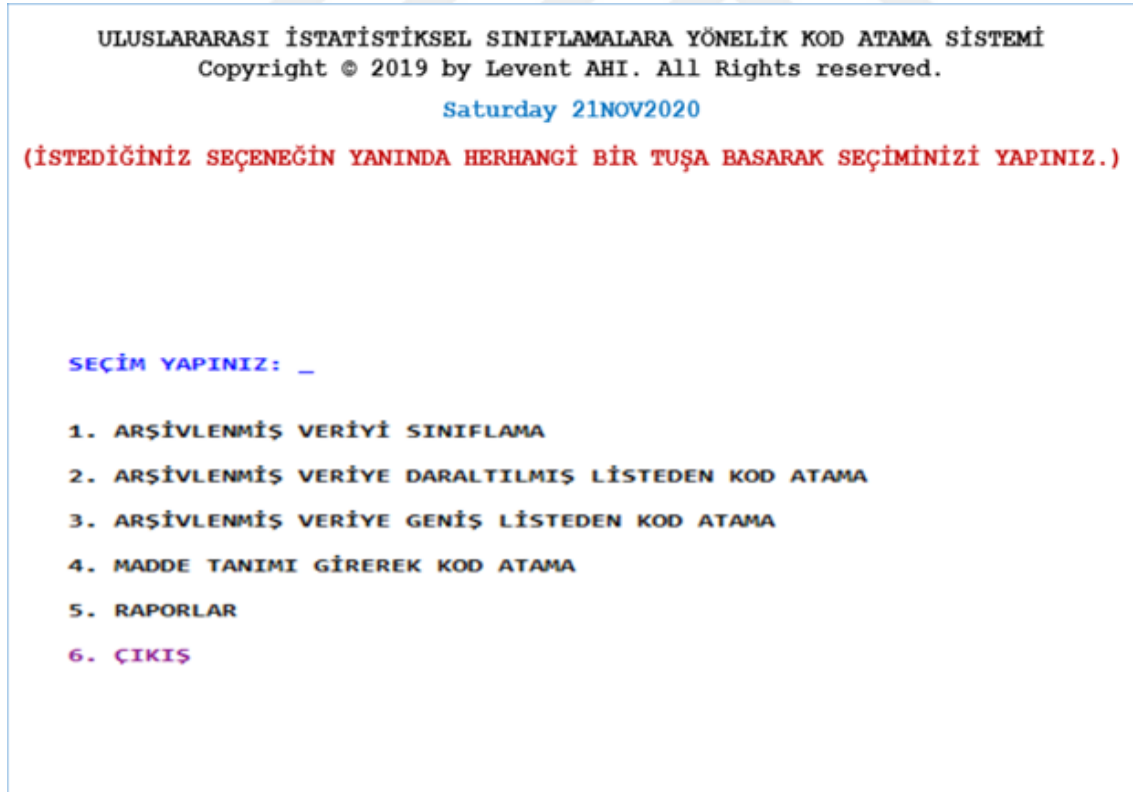
3. Bulanık eşleştirme yöntemlerini kullanarak kod ata: Bu atama yönteminde, tanım sadece dil bilgisi katmanından geçirildikten sonra bulanık eşleştirme yöntemleri kullanılarak en olası kod ataması yapılmaktadır.

4.4. Kullanıcı Ara Yüzü

Bu bölümde, KASİS'in kullanıcı ara yüzü hakkında bilgiler verilmektedir.

4.4.1. Ana ekran

KASİS'in ana ekran görüntüsü Şekil 4.19'da verilmiştir.



Şekil 4.19. KASİS ana ekran görüntüsü

Bu ekran aracılığıyla aşağıdaki işlemler yapılabilmektedir.

1. Arşivlenmiş veriyi sınıflama
2. Arşivlenmiş veriye daraltılmış listeden kod atama
3. Arşivlenmiş veriye geniş listeden kod atama
4. Madde tanımını girerek kod atama
5. Raporlar
6. Çıkış

4.4.2. Arşivlenmiş veriyi sınıflama

Bu seçenekte, 2013-2018 yılları HBA verisi seçilen parametreler kapsamında dil bilgisi, eşleştirme ve sınıflama katmanlarından geçirilerek sonuç ekrana yansıtılmaktadır. KASİS'in arşivlenmiş veriyi sınıflama giriş ekran görüntüsü Şekil 4.20'de verilmiştir.

ULUSLARARASI İSTATİSTİKSEL SINIFLAMALARA YÖNELİK KOD ATAMA SİSTEMİ

Copyright © 2019 by Levent AHI. All Rights reserved.

Saturday 21NOV2020

(İSTEDİĞİNİZ SEÇENEĞİN YANINDA HERHANGİ BİR TUŞA BASARAK SEÇİMİNİZİ YAPINIZ.)

TABLO	YIL	AY	BÖLGE
- COICOP_DAYANIKLI	- 2013	- 1	- ADANA
- COICOP_STOK	- 2014	- 2	- ANKARA
- COICOP_SATINALIS	- 2015	- 3	- ANTALYA
- COICOP_AYNI	- 2016	- 4	- BALIKESİR
	- 2017	- 5	- BURSA
	- 2018	- 6	- DENİZLİ
		- 7	- DİYARBAKIR
		- 8	- EDİRNE
		- 9	- ERZURUM
		- 10	- GAZİANTEP
		- 11	- HATAY
		- 12	- İSTANBUL
		- HEPSİ	- İZMİR
			- KARS
			- KASTAMONU
			- KAYSERİ
			- KOCAELİ
			- KONYA
			- MALATYA
			- MANİSA
			- NEVŞEHİR
			- SAMSUN
			- SİİRT
			- TRABZON
			- VAN
			- ZONGULDAK

Şekil 4.20. KASİS arşivlenmiş veriyi sınıflama giriş ekran görüntüsü

Açılan bu ekranda; tablo, ana grup, bölge, yıl ve ay parametreleri için seçim yapılmaktadır. Parametrelere sırasıyla; COICOP_SATINALIS > Çeşitli mal ve hizmetler > Trabzon > 2013 > 8 girilerek “çalıştır” seçildiği zaman oluşan sonuç ekranı Şekil 4.21’de verilmiştir.

ULUSLARARASI İSTATİSTİKSEL SINIFLAMALARA YÖNELİK KOD ATAMA SİSTEMİ

Copyright © 2019 by Levent AHI. All Rights reserved.

Saturday 21NOV2020

(İSTEDİĞİNİZ SEÇENEĞİN YANINDA HERHANGİ BİR TUŞA BASARAK SEÇİMİNİZİ YAPINIZ.)

TOPLAM_KAYIT_SAYISI	CREATED_DATE	SURE	ESLESEN_YUZDE	SUPHELI_YETERSIZ_YUZDE	TOPLAM_YUZDE
162	2020-11-21 17:45:40	0:01:06	98.8	1.2	100.0

SOZLUK_ESLESEN_KAYIT_SAYISI	YENI_ESLESEN_KAYIT_SAYISI	BULANIK_ESLESEN_KAYIT_SAYISI	ESLESEN_TOPLAM
36	123	.	159

KABUL_KAYIT_SAYISI	BULANIK_KABUL_KAYIT_SAYISI	KABUL_TOPLAM
1	.	1

YETERSIZ_KAYIT_SAYISI	SUPHELI_KAYIT_SAYISI	SUPHELI_YETERSIZ_TOPLAM
.	2	2

DOGRULAMA_SONUCU
BAŞARILI

Şekil 4.21. KASİS arşivlenmiş veriyi sınıflama sonuç ekran görüntüsü

Sistem, bu isteğe 1 dakika 6 saniyede cevap vermiştir. 162 kaydın %98,8’i eşleşen kayıt olarak, %1,2’si şüpheli ve yetersiz kayıt olarak sınıflanmıştır. Eşleşen kayıtların dağılımına bakıldığında; 36 kaydın sözlük eşleşen ve 123 kaydın yeni eşleşen olarak sınıflandığı görülmektedir. Kabul kayıt sayısının 1 ve şüpheli kayıt sayısının 2 olduğu belirlenmiştir. Veri tabanlarına çalıştırma sonuçlarının doğru bir şekilde aktarıldığı, doğrulama sonucunun “başarılı” olmasından anlaşılmaktadır.

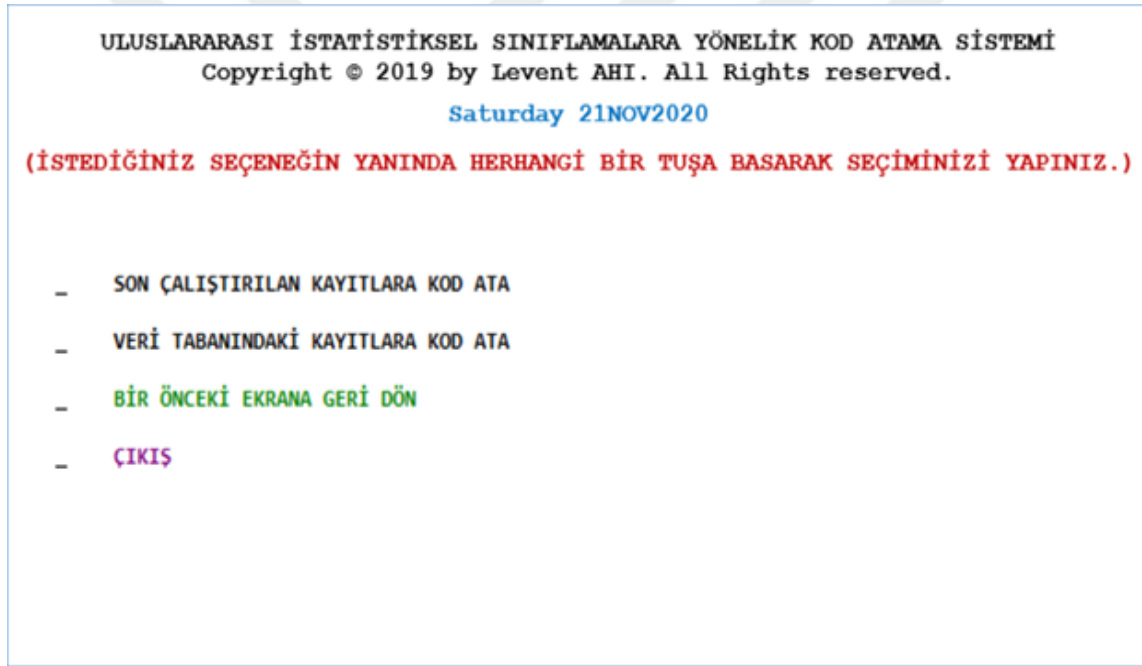
4.4.3. Arşivlenmiş veriye daraltılmış listeden kod atama

Sınıflama katmanında yetersiz veya şüpheli olarak sınıflanmış kayıtlar için eşleştirme katmanında kaç tane alternatif kodun ve tanımın karşılık geldiği ve bunların neler olduğu bilinmektedir. Bu yöntemde bulanık eşleştirme algoritmaları eşleştirme katmanında

belirlenen daraltılmış olan bu kod ve tanımlar ile sınırlı kalmak koşuluyla uygulanmıştır. KASİS'in arşivlenmiş veriye daraltılmış listeden kod atama ekran görüntüsü Şekil 4.22'de verilmiştir.

Açılan ekran aracılığıyla aşağıdaki işlemler yapılabilmektedir.

- Son çalıştırılan kayıtlara kod ata
- Veri tabanındaki kayıtlara kod ata
- Bir önceki ekrana geri dön
- Çıkış



Şekil 4.22. KASİS arşivlenmiş veriye kod atama giriş ekran görüntüsü 1

Şekil 4.22'de gösterilen ekran görüntüsünden “SON ÇALIŞTIRILAN KAYITLARA KOD ATA” veya “VERİ TABANINDAKİ KAYITLARA KOD ATA” seçimi yapıldığındaki ekran görüntüsü Şekil 4.23'te verilmiştir.

ULUSLARARASI İSTATİSTİKSEL SINIFLAMALARA YÖNELİK KOD ATAMA SİSTEMİ
Copyright © 2019 by Levent AHI. All Rights reserved.

Saturday 21NOV2020

(İSTEDİĞİNİZ SEÇENEĞİN YANINDA HERHANGİ BİR TUŞA BASARAK SEÇİMİNİZİ YAPINIZ.)

TABLO	YIL	AY	BÖLGE
- COICOP_DAYANIKLI	- 2013	- 1	- ADANA
- COICOP_STOK	- 2014	- 2	- ANKARA
- COICOP_SATINALIS	- 2015	- 3	- ANTALYA
- COICOP_AYNI	- 2016	- 4	- BALIKESİR
	- 2017	- 5	- BURSA
	- 2018	- 6	- DENİZLİ
		- 7	- DİYARBAKIR
		- 8	- EDİRNE
		- 9	- ERZURUM
		- 10	- GAZİANTEP
		- 11	- HATAY
		- 12	- İSTANBUL
		- HEPSİ	- İZMİR
			- KARS
			- KASTAMONU
			- KAYSERİ
			- KOCAELİ
			- KONYA
			- MALATYA
			- MANİSA
			- NEVŞEHİR
			- SAMSUN
			- SİİRT
			- TRABZON
			- VAN
			- ZONGULDAK

Şekil 4.23. KASİS arşivlenmiş veriye kod atama giriş ekran görüntüsü 2

Açılan bu ekranda; tablo, ana grup, bölge, yıl ve ay parametreleri için seçim yapılmaktadır. Parametrelere sırasıyla; COICOP_DAYANIKLI > Konut, su elektrik, gaz ve diğer yakıtlar > Antalya > 2013 > 5 girilerek “kayıtları göster” seçildiği zaman oluşan sonuç ekranına aşağıdaki çizelgelerde yer verilmiştir. Çizelge 4.14’te, genelleştirilmiş düzenleme mesafesi, Çizelge 4.15’te, Levenshtein düzenleme mesafesi, Çizelge 4.16’da, Jaro düzenleme mesafesi, Çizelge 4.17’de, Jaro-Winkler düzenleme mesafesi, Çizelge 4.18’de, 2-gram düzenleme mesafesi sonuçları ile Çizelge 4.19’da, önerilerin özet şekilde yer aldığı sonuçlar verilmiştir.

Çizelge 4.14. KASİS arşivlenmiş veriye kod atama sonuç ekran görüntüsü 1

ANKETOR_KOD	ANKETOR_TANIM	_1_GED_KOD	_1_GED_TANIM	_1_GED
431106030	pimapen pencere	431109060	KAPI, PENCERE IZOLASYON BANDI	0
432101990	badana ücreti malzemeve işçilik	432101990	BADANACI VE BOYACI ÜCRETİ DİĞER (MALZEME VE İSCİLİK ÜCRETİ)	1
432102990	badana boya işçilik ücreti	432102990	BADANACI VE BOYACI ÜCRETİ DİĞER (İSCİLİK ÜCRETİ)	1
432102990	boyacı ücreti	432102990	BADANACI VE BOYACI ÜCRETİ DİĞER (İSCİLİK ÜCRETİ)	1
432107990	konut tamiratı (panjur taktırma)malzeme ve işçilik	432107080	CAM TAKTIRMA ÜCRETİ-KONUT İCİN (MALZEME VE İSCİLİK ÜCRETİ)	0
454102010	tahta parçası	454103070	TAHTA PARÇALARI (YAKACAK İCİN)	0

Çizelge 4.15. KASİS arşivlenmiş veriye kod atama sonuç ekran görüntüsü 2

ANKETOR_KOD	ANKETOR_TANIM	_2_LEV_KOD	_2_LEV_TANIM	_2_LEV
431106030	pimapen pencere	431109080	KORKULUK (PENCERE İCİN)	0
432101990	badana ücreti malzemeve işçilik	432101990	BADANACI VE BOYACI ÜCRETİ DİĞER (MALZEME VE İSCİLİK ÜCRETİ)	1
432102990	badana boya işçilik ücreti	432102990	BADANACI VE BOYACI ÜCRETİ DİĞER (İSCİLİK ÜCRETİ)	1
432102990	boyacı ücreti	432102990	BADANACI VE BOYACI ÜCRETİ DİĞER (İSCİLİK ÜCRETİ)	1
432107990	konut tamiratı (panjur taktırma)malzeme ve işçilik	432107080	CAM TAKTIRMA ÜCRETİ-KONUT İCİN (MALZEME VE İSCİLİK ÜCRETİ)	0
454102010	tahta parçası	454103070	TAHTA PARÇALARI (YAKACAK İCİN)	0

Çizelge 4.16. KASİS arşivlenmiş veriye kod atama sonuç ekran görüntüsü 3

ANKETOR_KOD	ANKETOR_TANIM	_3_JARO_KOD	_3_JARO_TANIM	_3_JARO
431106030	pimapen pencere	431109050	KAPI PENCERE BANDI İCİN FITİL	0
432101990	badana ücreti malzemeve işçilik	432101990	BADANACI VE BOYACI ÜCRETİ DİĞER (MALZEME VE İSCİLİK ÜCRETİ)	1
432102990	badana boya işçilik ücreti	432102990	BADANACI VE BOYACI ÜCRETİ DİĞER (İSCİLİK ÜCRETİ)	1
432102990	boyacı ücreti	432102990	BADANACI VE BOYACI ÜCRETİ DİĞER (İSCİLİK ÜCRETİ)	1
432107990	konut tamiratı (panjur taktırma)malzeme ve işçilik	432107080	CAM TAKTIRMA ÜCRETİ-KONUT İCİN (MALZEME VE İSCİLİK ÜCRETİ)	0
454102010	tahta parçası	454103070	TAHTA PARÇALARI (YAKACAK İCİN)	0

Çizelge 4.17. KASİS arşivlenmiş veriye kod atama sonuç ekran görüntüsü 4

ANKETOR_KOD	ANKETOR_TANIM	_4_JAROWINK_KOD	_4_JAROWINK_TANIM	_4_JAROWINK
431106030	pimapen pencere	431109050	KAPI PENCERE BANDI ICIN FITIL	0
432101990	badana ücreti malzemeve işçilik	432101990	BADANACI VE BOYACI UCURETI DIGER (MALZEME VE ISCILIK UCURETI)	1
432102990	badana boya işçilik ücreti	432102990	BADANACI VE BOYACI UCURETI DIGER (ISCILIK UCURETI)	1
432102990	boyacı ücreti	432102990	BADANACI VE BOYACI UCURETI DIGER (ISCILIK UCURETI)	1
432107990	konut tamirati (panjur taktirma)malzeme ve işçilik	432107080	CAM TAKTIRMA UCURETI-KONUT ICIN (MALZEME VE ISCILIK UCURETI)	0
454102010	tahta parçası	454103070	TAHTA PARCALARI (YAKACAK ICIN)	0

Çizelge 4.18. KASİS arşivlenmiş veriye kod atama sonuç ekran görüntüsü 5

ANKETOR_KOD	ANKETOR_TANIM	_5_NGRAM_KOD	_5_NGRAM_TANIM	_5_NGRAM
431106030	pimapen pencere	431106030	PENCERE SATIN ALMA PVC	1
432101990	badana ücreti malzemeve işçilik	432101990	BADANACI VE BOYACI UCURETI DIGER (MALZEME VE ISCILIK UCURETI)	1
432102990	badana boya işçilik ücreti	432101160	YAGLI BOYA (BADANACI VE BOYACI UCURETI - MALZEME VE ISCILIK UCURETI))	0
432102990	boyacı ücreti	432102990	BADANACI VE BOYACI UCURETI DIGER (ISCILIK UCURETI)	1
432107990	konut tamirati (panjur taktirma)malzeme ve işçilik	432107080	CAM TAKTIRMA UCURETI-KONUT ICIN (MALZEME VE ISCILIK UCURETI)	0
454102010	tahta parçası	454103070	TAHTA PARCALARI (YAKACAK ICIN)	0

Çizelge 4.19. KASİS arşivlenmiş veriye kod atama sonuç ekran görüntüsü 6

ANKETOR_KOD	ANKETOR_TANIM	SOZLUK_KOD	SOZLUK_TANIM	KOD_ATAMA_DURUM	ORAN_YUZDE
431106030	pimapen pencere			TEK BİR KOD ÖNERİLEMEDİ. ÖNERİLEN KODLARIN İLK BEŞ BASAMAĞINA BAKARAK KARAR VERİLMELİ !!!	
432101990	badana ücreti malzemeve işçilik	432101990	badanacı ve boyacı ücreti diğer (malzeme ve işçilik ücreti)	ÖNERİLEN KOD	100
432102990	badana boya işçilik ücreti	432102990	badanacı ve boyacı ücreti diğer (işçilik ücreti)	ÖNERİLEN KOD	80
432102990	boyacı ücreti	432102990	badanacı ve boyacı ücreti diğer (işçilik ücreti)	ÖNERİLEN KOD	100
432107990	konut tamirati (panjur taktirma)malzeme ve işçilik	432107080	cam taktirma ücreti-konut için (malzeme ve işçilik ücreti)	ÖNERİLEN KOD	100
454102010	tahta parçası	454103070	tahta parçaları (yakacak için)	ÖNERİLEN KOD	100

Eşleştirme katmanında belirlenmiş alternatif kodların arasından bulanık eşleştirme yöntemleriyle uygun kod veya kodlar seçilmektedir. Beş bulanık eşleştirme algoritmasından üç veya daha fazlasının önerdiği kod nihai olarak önerilmektedir. Bu kriterlere uymayan durumlar için ekrana “TEK BİR KOD ÖNERİLEMEDİ. ÖNERİLEN KODLARIN İLK BEŞ BASAMAĞINA BAKARAK KARAR VERİLMELİ !!!” uyarısı gelmektedir.

Anketör tarafından oluşturulmuş “432102990 - badana boya işçilik ücreti” kaydı için KASİS tarafından önerilen kodlar verilmiştir.

- Genelleştirilmiş düzenleme mesafesi, 432102990 - badanacı ve boyacı ücreti diğer (işçilik ücreti),
- Levenshtein düzenleme mesafesi, 432102990 - badanacı ve boyacı ücreti diğer (işçilik ücreti),
- Jaro düzenleme mesafesi, 432102990 - badanacı ve boyacı ücreti diğer (işçilik ücreti),
- Jaro-Winkler düzenleme mesafesi, 432102990 - badanacı ve boyacı ücreti diğer (işçilik ücreti),
- 2-gram düzenleme mesafesi, 432101160 - yağlı boya (badanacı ve boyacı ücreti - malzeme ve işçilik ücreti).

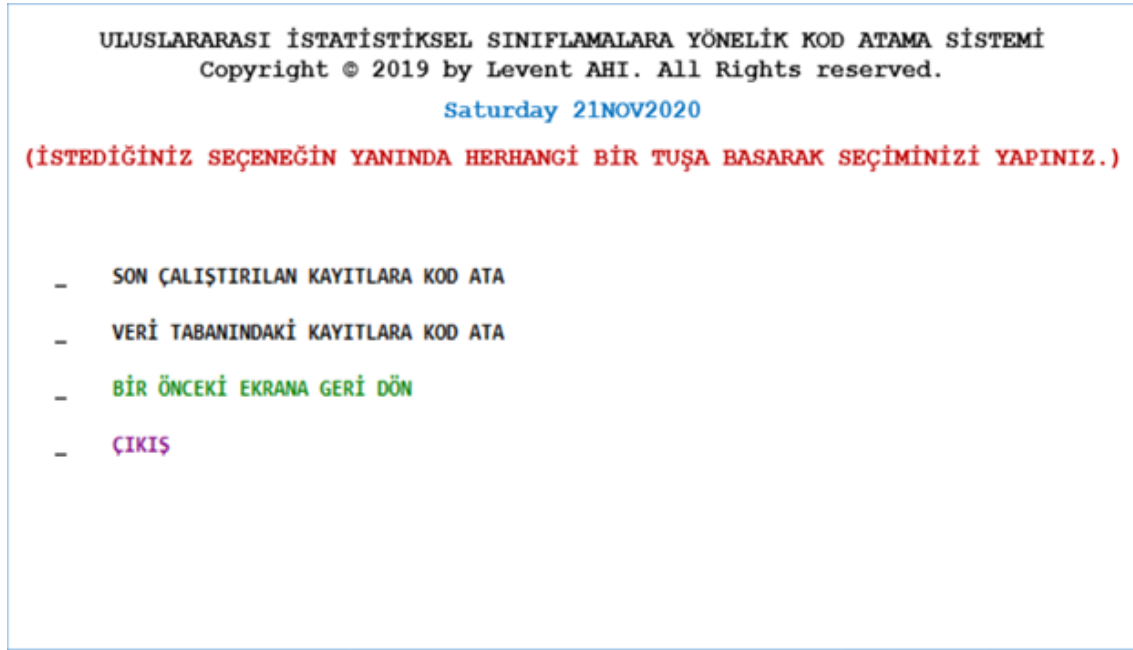
Beş bulanık eşleştirme algoritmasından dört tanesi “432102990 - badanacı ve boyacı ücreti diğer (işçilik ücreti)” kodu üzerinde mutabık kaldığı için sistem tarafından bu kod önerilmiştir. Bu kod anketörün verdiği kod ile aynı olduğu için anketörün oluşturduğu kaydın doğru olduğu sonucuna varılmaktadır.

4.4.4. Arşivlenmiş veriye geniş listeden kod atama

Bu yöntemde, listede daraltma uygulanmadan sözlükte bulunan tüm kod ve tanımlar için bulanık eşleştirme algoritmaları uygulanmıştır. KASİS’in arşivlenmiş veriye bulanık eşleştirme yöntemleriyle kod atama ekran görüntüsü Şekil 4.24’te verilmiştir.

Bu ekran aracılığıyla aşağıdaki işlemler yapılabilmektedir.

- Son alıřtırılan kayıtlara kod ata
- Veri tabanındaki kayıtlara kod ata
- Bir nceki ekrana geri dn
- ıkıř



Őekil 4.24. KASİS bulanık eőleőtirme yöntemleriyle kod atama ekran grnts 1

Őekil 4.24’te gsterilen ekran grntsnden “SON ALIŐTIRILAN KAYITLARA KOD ATA” veya “VERİ TABANINDAKİ KAYITLARA KOD ATA” seimi yapıldıėındaki ekran grnts Őekil 4.25’te verilmiőtir.

ULUSLARARASI İSTATİSTİKSEL SINIFLAMALARA YÖNELİK KOD ATAMA SİSTEMİ
Copyright © 2019 by Levent AHI. All Rights reserved.

Saturday 21NOV2020

(İSTEDİĞİNİZ SEÇENEĞİN YANINDA HERHANGİ BİR TUŞA BASARAK SEÇİMİNİZİ YAPINIZ.)

TABLO	YIL	AY	BÖLGE
- COICOP_DAYANIKLI	- 2013	- 1	- ADANA
- COICOP_STOK	- 2014	- 2	- ANKARA
- COICOP_SATINALIS	- 2015	- 3	- ANTALYA
- COICOP_AYNI	- 2016	- 4	- BALIKESİR
	- 2017	- 5	- BURSA
	- 2018	- 6	- DENİZLİ
		- 7	- DİYARBAKIR
		- 8	- EDİRNE
		- 9	- ERZURUM
		- 10	- GAZİANTEP
		- 11	- HATAY
		- 12	- İSTANBUL
		- HEPSİ	- İZMİR
			- KARS
			- KASTAMONU
			- KAYSERİ
			- KOCAELİ
			- KONYA
			- MALATYA
			- MANİSA
			- NEVŞEHİR
			- SAMSUN
			- SİİRT
			- TRABZON
			- VAN
			- ZONGULDAK

ANA GRUP

- GIDA VE ALKOLSÜZ İÇECEKLER
- ALKOLLÜ İÇECEKLER, TÜTÜN VE NARKOTİK MADDELER
- GİYİM VE AYAKKABI
- KONUT, SU, ELEKTRİK, GAZ VE DİĞER YAKITLAR
- MOBİLYALAR, EVDE KULLANILAN EKİPMANLAR İLE RUTİN EV BAKIM VE ONARIMI
- SAĞLIK
- ULAŞTIRMA
- HABERLEŞME
- EĞLENCE VE KÜLTÜR
- EĞİTİM
- LOKANTALAR VE OTELLER
- ÇEŞİTLİ MAL VE HİZMETLER

Şekil 4.25. KASİS bulanık eşleştirme yöntemleriyle kod atama ekran görüntüsü 2

Açılan bu ekranda; tablo, ana grup, bölge, yıl ve ay parametreleri için seçim yapılmaktadır. Parametrelere sırasıyla; COICOP_STOK > Gıda ve alkolsüz içecekler > Denizli > 2017 > 9 girilerek “kayıtları göster” seçildiği zaman oluşan sonuç ekranı Çizelge 4.20 ve Çizelge 4.21’de verilmiştir.

Çizelge 4.20. KASİS bulanık eşleştirme kod atama sonuç ekran görüntüsü 1

ANKETOR_KOD	ANKETOR_TANIM	_1_GED_KOD	_1_GED_TANIM	_1_GED
111201010	un buğday unu (eriş un)	111201010	UN BUGDAY UNU	1
111601010	makarna erişte beşler	111601090	MAKARNA DIYET	0
ANKETOR_KOD	ANKETOR_TANIM	_2_LEV_KOD	_2_LEV_TANIM	_2_LEV
111201010	un buğday unu (eriş un)	111201020	UN BUGDAY UNU KEPEKLI UN	0
111601010	makarna erişte beşler	111698980	MAKARNA CESITLERI DIGER	0
ANKETOR_KOD	ANKETOR_TANIM	_3_JARO_KOD	_3_JARO_TANIM	_3_JARO
111201010	un buğday unu (eriş un)	111201010	UN BUGDAY UNU	1
111601010	makarna erişte beşler	111424010	KATMER	0
ANKETOR_KOD	ANKETOR_TANIM	_4_JAROWINK_KOD	_4_JAROWINK_TANIM	_4_JAROWINK
111201010	un buğday unu (eriş un)	111204010	UN YULAF UNU	0
111601010	makarna erişte beşler	111424010	KATMER	0
ANKETOR_KOD	ANKETOR_TANIM	_5_NGRAM_KOD	_5_NGRAM_TANIM	_5_NGRAM
111201010	un buğday unu (eriş un)	111201980	UN BUGDAY UNU DIGER	0
111601010	makarna erişte beşler	111601110	MAKARNA NOODLE (CIN ERISTESI)	0

Çizelge 4.21. KASİS bulanık eşleştirme kod atama sonuç ekran görüntüsü 2

ANKETOR_KOD	ANKETOR_TANIM	KOD_ATAMA_DURUM
111201010	un buğday unu (eriş un)	TEK BİR KOD ÖNERİLEMEDİ. ÖNERİLEN KODLARIN İLK BEŞ BASAMAĞINA BAKARAK KARAR VERİLMELİ !!!
111601010	makarna erişte beşler	TEK BİR KOD ÖNERİLEMEDİ. ÖNERİLEN KODLARIN İLK BEŞ BASAMAĞINA BAKARAK KARAR VERİLMELİ !!!

COICOP sözlüğünün tümü içerisinde bulanık eşleştirme yöntemleriyle uygun kod ve kodlar seçilmektedir. Beş bulanık eşleştirme algoritmasından üç veya daha fazlasının önerdiği kod nihai olarak önerilmektedir. Bu kriterlere uymayan durumlar için ekrana “TEK BİR KOD ÖNERİLEMEDİ. ÖNERİLEN KODLARIN İLK BEŞ BASAMAĞINA BAKARAK KARAR VERİLMELİ !!!” uyarısı gelmektedir.

Anketör tarafından oluşturulmuş “111201010 - un buğday unu (eriş un)” kaydı için KASİS tarafından önerilen kodlar verilmiştir.

- Genelleştirilmiş düzenleme mesafesi, 111201010 - un buğday unu,
- Levenshtein düzenleme mesafesi, 111201020 - un buğday unu kepekli un,
- Jaro düzenleme mesafesi, 111201010 - un buğday unu,
- Jaro-Winkler düzenleme mesafesi, 111204010 - un yulaf unu,
- 2-gram düzenleme mesafesi, 111201980 - un buğday unu diğer.

Beş bulanık eşleştirme algoritmasından üç veya daha fazlasının önerdiği ortak bir kod olmadığı için sistemin üzerinde mutabık kaldığı bir kod bulunmamaktadır. Ancak, uzman görüşüyle sistemin önerdiği kodların içerisinde bir tanesi seçilebilir veya sistemin önerdiği kodların ilk beş basamağı ile anketörün atadığı kodun ilk beş basamağı aynı olduğu için anketörün verdiği kodun doğru olduğu sonucuna varılabilir.

4.4.5. Madde tanımı girerek kod atama

Madde tanımı girerek kod atama giriş ekran görüntüsü Şekil 4.26’da verilmiştir. Bu ekranda beş maddeye kadar harcama tanımı girilerek sorgulama yapılabilmektedir.

ULUSLARARASI İSTATİSTİKSEL SINIFLAMALARA YÖNELİK KOD ATAMA SİSTEMİ
Copyright © 2019 by Levent AHI. All Rights reserved.
Saturday 21NOV2020

(İSTEDİĞİNİZ SEÇENEĞİN YANINDA HERHANGİ BİR TUŞA BASARAK SEÇİMİNİZİ YAPINIZ.)

KOD ATAMASI YAPMAK İSTEDİĞİNİZ MADDE TANIMINI/TANIMLARINI YAZINIZ.

BİRİNCİ MADDE TANIMI : _____

İKİNCİ MADDE TANIMI : _____

ÜÇÜNCÜ MADDE TANIMI : _____

DÖRDÜNCÜ MADDE TANIMI : _____

BEŞİNCİ MADDE TANIMI : _____

Şekil 4.26. KASİS madde tanımı girerek kod atama giriş ekran görüntüsü 1

Bu ekrana “elma” tanımı yazıldıktan sonra oluşan ekran görüntüsüne Şekil 4.27’de yer verilmektedir.

ULUSLARARASI İSTATİSTİKSEL SINIFLAMALARA YÖNELİK KOD ATAMA SİSTEMİ
Copyright © 2019 by Levent AHI. All Rights reserved.
Saturday 21NOV2020

(İSTEDİĞİNİZ SEÇENEĞİN YANINDA HERHANGİ BİR TUŞA BASARAK SEÇİMİNİZİ YAPINIZ.)

KOD ATAMASI YAPMAK İSTEDİĞİNİZ MADDE TANIMINI/TANIMLARINI YAZINIZ.

BİRİNCİ MADDE TANIMI : elma _____

İKİNCİ MADDE TANIMI : _____

ÜÇÜNCÜ MADDE TANIMI : _____

DÖRDÜNCÜ MADDE TANIMI : _____

BEŞİNCİ MADDE TANIMI : _____

SEÇİM YAPINIZ: _

1. KELİMELERİN GEÇTİĞİ YERE GÖRE DAR LİSTEDEN KOD ATA
ÖRNEK: ARANAN KELİME TOP - LAPTOP, TOPRAK, ORTOPEDİK...

2. KELİMELERİN BÜTÜNLÜĞÜNÜ KORUYARAK DAR LİSTEDEN KOD ATA
ÖRNEK: ARANAN KELİME TOP - OYUNCAK TOP...

3. GENİŞ LİSTEDEN KOD ATA

4. BİR ÖNCEKİ EKRANA GERİ DÖN

5. ÇIKIŞ

Şekil 4.27. KASİS madde tanımı girerek kod atama giriş ekran görüntüsü 2

Bu ekran aracılığıyla aşağıdaki işlemler yapılabilmektedir.

- Kelimelerin geçtiği yere göre dar listeden kod ata
- Kelimelerin bütünlüğünü koruyarak dar listeden kod ata
- Geniş listeden kod ata
- Bir önceki ekrana geri dön
- Çıkış

Kelimelerin geçtiği yere göre dar listeden kod ata

Bu seçenek ile devam edildiğinde; elma tanımına karşılık önerilen 19 farklı kod Çizelge 4.22’de verilmiştir.

Çizelge 4.22. KASİS madde tanımı girerek kod atama sonuç ekran görüntüsü 1

TANIM	SOZLUK_KOD	SOZLUK_TANIM
ELMA	1111120110	MEYVE VE BITKİ CAYI (IHLAMUR, ADACAYI, SEFTALI, ELMA VB.) (RESTORANDA)
ELMA	1111152110	MEYVE VE BITKİ CAYI (IHLAMUR, ADACAYI, SEFTALI, ELMA VB.) (CAFE, KAHVEHANE, BAR, PASTANE VB. YERLERDE)
ELMA	1111210050	HOROZ SEKERI, ELMA SEKERI VB. (SEYYAR SATICI VB. YERLERDE)
ELMA	1111211060	MEYVE VE BITKİ CAYI (IHLAMUR, ADACAYI, SEFTALI, ELMA VB.) (SEYYAR SATICI VB. YERLERDE)
ELMA	1112020110	MEYVE VE BITKİ CAYI (IHLAMUR, ADACAYI, SEFTALI, ELMA VB.) (KANTINLERDE, CAY OCAGINDA)
ELMA	111414020	KURABIYE MARMELATLI (ELMA, KAYISI, VISNE VB)
ELMA	116121010	ELMA (TAZE)
ELMA	116315010	ELMA KURUSU
ELMA	117120010	YER ELMASI (TAZE SEBZE)
ELMA	118201060	RECEL ELMA
ELMA	118208060	PEKMEZ ELMA PEKMEZI
ELMA	118405060	ELMA SEKERI
ELMA	119101020	SIRKE ELMA SIRKESI
ELMA	121202020	MEYVE CAYI (ELMA, PORTAKAL, BOGURTLEN, KUSBURNU VB.) (POSET)
ELMA	122302050	TAZE SIKILMIS ELMA SUYU
ELMA	1231103010	MUCEVHER ELMAS
ELMA	212201010	SARAP ELMA SARABI
ELMA	532901030	ELEKTRIKLI KATI MEYVE SIKACAGI (ELMA, HAVUC VB.)
ELMA	611003020	ADACAYI YAGI (ELMA YAGI) (SAGLIKAMACLI)

Çizelge 4.23'te ise, elma tanımı için eşleştirme katmanında oluşan alternatif kodların arasından bulanık eşleştirme yöntemleriyle belirlenerek önerilen kodlar verilmiştir.

Çizelge 4.23. KASİS madde tanımı girerek kod atama sonuç ekran görüntüsü 2

TANIM	_1_GED_KOD	_1_GED_TANIM
ELMA	116121010	ELMA (TAZE)
TANIM	_2_LEV_KOD	_2_LEV_TANIM
ELMA	118201060	RECEL ELMA
TANIM	_3_JARO_KOD	_3_JARO_TANIM
ELMA	116121010	ELMA (TAZE)
TANIM	_4_JAROWINK_KOD	_4_JAROWINK_TANIM
ELMA	116121010	ELMA (TAZE)
TANIM	_5_NGRAM_KOD	_5_NGRAM_TANIM
ELMA	116121010	ELMA (TAZE)

Çizelgeye göre;

- Genelleştirilmiş düzenleme mesafesi, 116121010 - elma (taze),
- Levenshtein düzenleme mesafesi, 118201060 - reçel elma,
- Jaro düzenleme mesafesi, 116121010 - elma (taze),
- Jaro-Winkler düzenleme mesafesi, 116121010 - elma (taze),
- 2-gram düzenleme mesafesi, 116121010 - elma (taze)

kodunu önermiştir.

Beş bulanık eşleştirme algoritmasından dört tanesi elma tanımına 116121010 – elma (taze) kodunu önerdiği için nihai olarak sistem tarafından bu kod önerilmiştir.

Kelimelerin bütünlüğünü koruyarak dar listeden kod ata

Bu seçenek ile devam edildiğinde; elma tanımına karşılık önerilen 14 farklı kod Çizelge 4.24'te verilmiştir.

Çizelge 4.24. KASİS madde tanımı girerek kod atama sonuç ekran görüntüsü 3

TANIM	SOZLUK_KOD	SOZLUK_TANIM_DUZ
ELMA	1111120110	MEYVE VE BITKI CAYI (IHLAMUR, ADACAYI, SEFTALI, ELMA VB.) (RESTORANDA)
ELMA	1111152110	MEYVE VE BITKI CAYI (IHLAMUR, ADACAYI, SEFTALI, ELMA VB.) (CAFE, KAHVEHANE, BAR, PASTANE VB. YERLERDE)
ELMA	1111210050	HOROZ SEKERI, ELMA SEKERI VB. (SEYYAR SATICI VB. YERLERDE)
ELMA	1111211060	MEYVE VE BITKI CAYI (IHLAMUR, ADACAYI, SEFTALI, ELMA VB.) (SEYYAR SATICI VB. YERLERDE)
ELMA	1112020110	MEYVE VE BITKI CAYI (IHLAMUR, ADACAYI, SEFTALI, ELMA VB.) (KANTINLERDE, CAY OCAGINDA)
ELMA	116121010	ELMA (TAZE)
ELMA	116315010	ELMA KURUSU
ELMA	118201060	RECEL ELMA
ELMA	118208060	PEKMEZ ELMA PEKMEZI
ELMA	118405060	ELMA SEKERI
ELMA	119101020	SIRKE ELMA SIRKESI
ELMA	122302050	TAZE SIKILMIS ELMA SUYU
ELMA	212201010	SARAP ELMA SARABI
ELMA	611003020	ADACAYI YAGI (ELMA YAGI) (SAGLIKAMACLI)

Çizelge 4.25'te ise elma tanımı için eşleştirme katmanında oluşan alternatif kodların arasından bulanık eşleştirme yöntemleriyle belirlenerek önerilen kodlar verilmiştir.

Çizelge 4.25. KASİS madde tanımı girerek kod atama sonuç ekran görüntüsü 4

TANIM	_1_GED_KOD	_1_GED_TANIM
ELMA	116121010	ELMA (TAZE)
TANIM	_2_LEV_KOD	_2_LEV_TANIM
ELMA	118201060	RECEL ELMA
TANIM	_3_JARO_KOD	_3_JARO_TANIM
ELMA	116121010	ELMA (TAZE)
TANIM	_4_JAROWINK_KOD	_4_JAROWINK_TANIM
ELMA	116121010	ELMA (TAZE)
TANIM	_5_NGRAM_KOD	_5_NGRAM_TANIM
ELMA	116121010	ELMA (TAZE)

Çizelgeye göre;

- Genelleştirilmiş düzenleme mesafesi, 116121010 - elma (taze),
- Levenshtein düzenleme mesafesi, 118201060 - reçel elma,

- Jaro düzenleme mesafesi, 116121010 - elma (taze),
- Jaro-Winkler düzenleme mesafesi, 116121010 - elma (taze),
- 2-gram düzenleme mesafesi, 116121010 - elma (taze)

kodunu önermiştir.

Beş bulanık eşleştirme algoritmasından dört tanesi elma tanımına 116121010 - elma (taze) kodunu önerdiği için nihai olarak sistem tarafından bu kod önerilmiştir.

Geniş listeden kod ata

Bu seçenek ile devam edildiğinde; elma tanımına karşılık yöntemlerin önerdiği kodlar Çizelge 4.26'da verilmiştir.

Çizelge 4.26. KASİS madde tanımı girerek kod atama sonuç ekran görüntüsü 5

TANIM	_1_GED_KOD	_1_GED_TANIM
ELMA	552101080	EL MALASI
TANIM	_2_LEV_KOD	_2_LEV_TANIM
ELMA	111406010	ACMA
TANIM	_3_JARO_KOD	_3_JARO_TANIM
ELMA	552101080	EL MALASI
TANIM	_4_JAROWINK_KOD	_4_JAROWINK_TANIM
ELMA	552101080	EL MALASI
TANIM	_5_NGRAM_KOD	_5_NGRAM_TANIM
ELMA	116121010	ELMA (TAZE)

Çizelgeye göre;

- Genelleştirilmiş düzenleme mesafesi, 552101080 - el malası,
- Levenshtein düzenleme mesafesi, 111406010 - açma,
- Jaro düzenleme mesafesi, 552101080 - el malası,
- Jaro-Winkler düzenleme mesafesi, 552101080 - el malası,
- 2-gram düzenleme mesafesi, 116121010 - elma (taze)

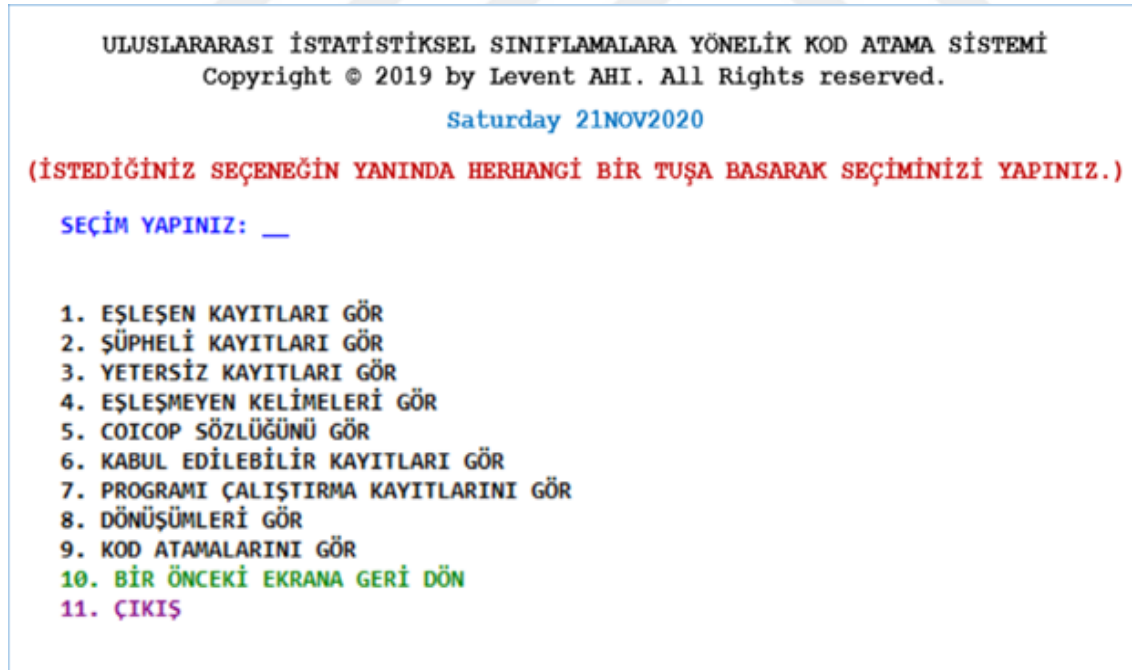
kodunu önermiştir.

Beş bulanık eşleştirme algoritmasından üç tanesi elma tanımına 552101080 - el malası kodunu önerdiği için nihai olarak sistem tarafından bu kod önerilmiştir.

Aynı tanıma ilişkin üç farklı yöntem ile yapılan kod atama işlemi, dil bilgisi ve eşleştirme katmanlarından geçirildikten sonra oluşan daraltılmış listeden kod atama yapılmasının ne kadar önemli olduğunu bize göstermektedir. Geniş listeden kod atama seçeneğini kullanarak yapılan kod atamasında ise dil bilgisi ve eşleştirme katmanları devre dışı bırakıldığı için bulanık eşleştirme algoritmaları tüm tanımları tarayarak en uygun kodu bulmaya çalışmaktadır.

4.4.6. Raporlar

Bu seçenekte, veri tabanlarına aktarılmış kayıtlara ilişkin raporlar ekrana yansıtılmaktadır. Geçmişte yapılmış her türlü işlemin kayıtlarına bu ekran aracılığıyla erişebilmek mümkün olmaktadır. KASİS'in raporlar giriş ekran görüntüsü Şekil 4.28'de verilmiştir.



Şekil 4.28. KASİS raporlar giriş ekran görüntüsü 1

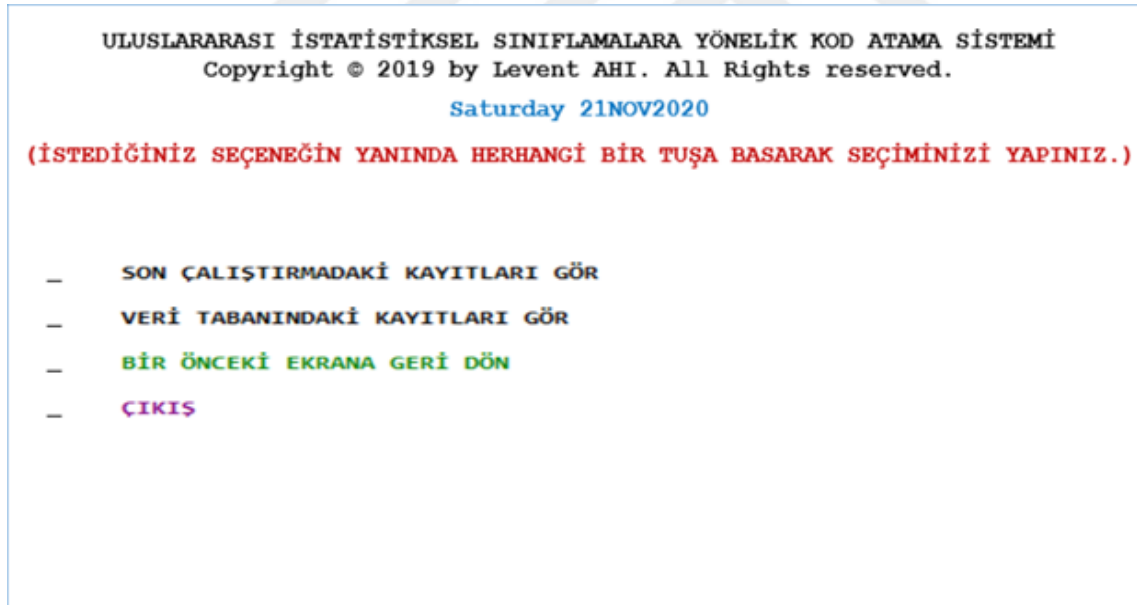
Bu ekran aracılığıyla aşağıdaki işlemler yapılabilmektedir.

- Eşleşen kayıtları gör

- Şüpheli kayıtları gör
- Yetersiz kayıtları gör
- Eşleşmeyen kelimeleri gör
- COICOP sözlüğünü gör
- Kabul edilebilir kayıtları gör
- Programı çalıştırma kayıtlarını gör
- Dönüşümleri gör
- Kod atamalarını gör
- Bir önceki ekrana geri dön
- Çıkış

Eşleşen kayıtları gör

KASİS'in eşleşen kayıtları gör ekran görüntüsü Şekil 4.29'da verilmiştir.



Şekil 4.29. KASİS raporlar giriş ekran görüntüsü 2

Şekil 4.29'da gösterilen ekran görüntüsünden “SON ÇALIŞTIRMADAKİ KAYITLARI GÖR” veya “VERİ TABANINDAKİ KAYITLARI GÖR” seçimi yapıldığındaki ekran görüntüsü Şekil 4.30'da verilmiştir.

ULUSLARARASI İSTATİSTİKSEL SINIFLAMALARA YÖNELİK KOD ATAMA SİSTEMİ
Copyright © 2019 by Levent AHI. All Rights reserved.
Saturday 21NOV2020

(İSTEDİĞİNİZ SEÇENEĞİN YANINDA HERHANGİ BİR TUŞA BASARAK SEÇİMİNİZİ YAPINIZ.)

TABLO	YIL	AY	BÖLGE
- COICOP_DAYANIKLI	- 2013	- 1	- ADANA
- COICOP_STOK	- 2014	- 2	- ANKARA
- COICOP_SATINALIS	- 2015	- 3	- ANTALYA
- COICOP_AYNI	- 2016	- 4	- BALIKESİR
	- 2017	- 5	- BURSA
	- 2018	- 6	- DENİZLİ
		- 7	- DİYARBAKIR
		- 8	- EDİRNE
		- 9	- ERZURUM
		- 10	- GAZİANTEP
		- 11	- HATAY
		- 12	- İSTANBUL
		- HEPSİ	- İZMİR
			- KARS
			- KASTAMONU
			- KAYSERİ
			- KOCAELİ
			- KONYA
			- MALATYA
			- MANİSA
			- NEVŞEHİR
			- SAMSUN
			- SİİRT
			- TRABZON
			- VAN
			- ZONGULDAK

ANA GRUP

- GIDA VE ALKOLSÜZ İÇECEKLER
- ALKOLLÜ İÇECEKLER, TÜTÜN VE NARKOTİK MADDELER
- GİYİM VE AYAKKABI
- KONUT, SU, ELEKTRİK, GAZ VE DİĞER YAKITLAR
- MOBİLYALAR, EVDE KULLANILAN EKİPMANLAR İLE RUTİN EV BAKIM VE ONARIMI
- SAĞLIK
- ULAŞTIRMA
- HABERLEŞME
- EĞLENCE VE KÜLTÜR
- EĞİTİM
- LOKANTALAR VE OTELLER
- ÇEŞİTLİ MAL VE HİZMETLER

Şekil 4.30. KASİS raporlar giriş ekran görüntüsü 3

Açılan bu ekranda; tablo, ana grup, bölge, yıl ve ay parametreleri için seçim yapılmaktadır. Parametrelere sırasıyla; COICOP_DAYANIKLI > Konut, su, elektrik, gaz ve diğer yakıtlar > Balıkesir > 2016 > 2 girilerek “kayıtları göster” seçildiği zaman oluşan sonuç ekranı Çizelge 4.27’de verilmiştir.

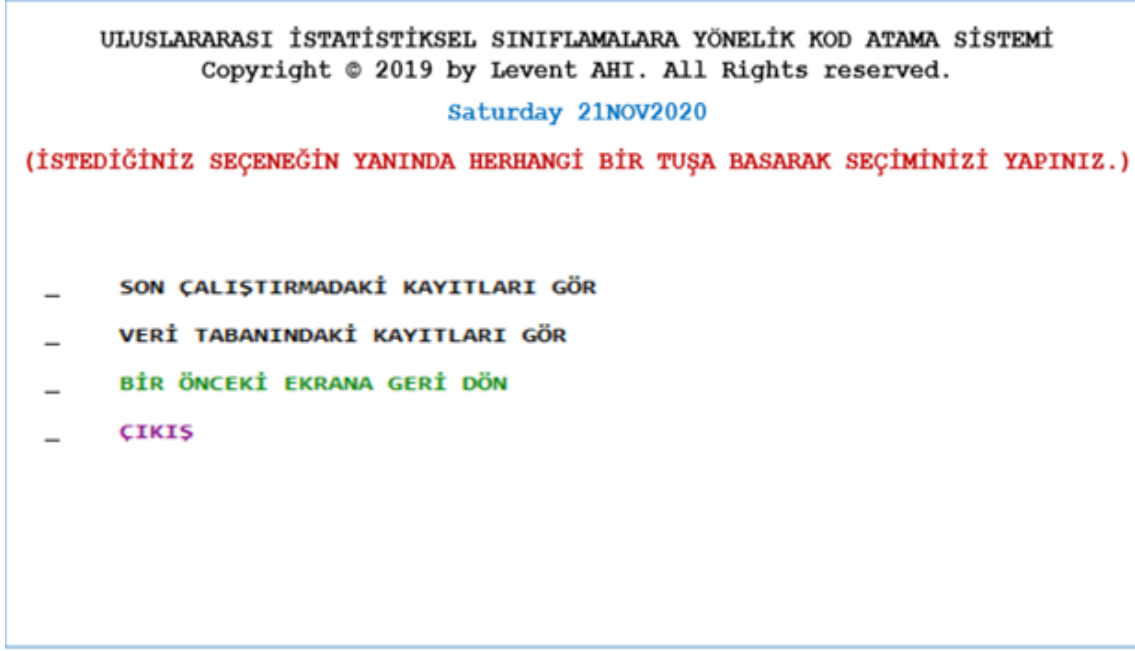
Çizelge 4.27. KASİS raporlar sonuç ekran görüntüsü 1

TABLO	ANA_GRUP	BOLGE_KOD	REFERANS_YIL	REFERANS_AY
COICOP_DAYANIKLI	4	10	2016	2
ANKETOR_KOD	ANKETOR_TANIM	SOZLUK_KOD	SOZLUK_TANIM	DURUM
431001050	kireç	431001050	kireç (badana ve boya malzemesi)	YENİ EŞLEŞEN
431001050	kireç malzeme	431001050	kireç (badana ve boya malzemesi)	YENİ EŞLEŞEN
431001050	kireç malzeme parası	431001050	kireç (badana ve boya malzemesi)	YENİ EŞLEŞEN
431001050	kireç malzemesi	431001050	kireç (badana ve boya malzemesi)	YENİ EŞLEŞEN
431001070	boya plastik	431001070	plastik boya (badana ve boya malzemesi)	YENİ EŞLEŞEN
431001070	plastik boya	431001070	plastik boya (badana ve boya malzemesi)	YENİ EŞLEŞEN
431001100	boya sentetik	431001100	sentetik boya (badana ve boya malzemesi)	YENİ EŞLEŞEN
431006030	pencere satın alma	431006030	pencere satın alma pvc	YENİ EŞLEŞEN
431006030	pvc pencere	431006030	pencere satın alma pvc	YENİ EŞLEŞEN
454101040	kömür linyit	454101040	kömür ücreti linyit	YENİ EŞLEŞEN
454101040	kömür ücreti linyit	454101040	kömür ücreti linyit	SÖZLÜK EŞLEŞEN
454902010	odun ücreti	454902010	odun ücreti	SÖZLÜK EŞLEŞEN

Raporda, sözlük eşleşen, yeni eşleşen ve bulanık eşleşen ayırımında kayıtların detayları ve bu kayıtların hangi tarih ve saatte eşleştirildiği bilgisi de yer almaktadır.

Şüpheli kayıtları gör

KASİS'in şüpheli kayıtları gör ekran görüntüsü Şekil 4.31'de verilmiştir.



Şekil 4.31. KASİS raporlar giriş ekran görüntüsü 4

Şekil 4.31'de gösterilen ekran görüntüsünden “SON ÇALIŞTIRMADAKİ KAYITLARI GÖR” veya “VERİ TABANINDAKİ KAYITLARI GÖR” seçimi yapıldığındaki ekran görüntüsü Şekil 4.32’de verilmiştir.

ULUSLARARASI İSTATİSTİKSEL SINIFLAMALARA YÖNELİK KOD ATAMA SİSTEMİ
Copyright © 2019 by Levent AHI. All Rights reserved.

Saturday 21NOV2020

(İSTEDİĞİNİZ SEÇENEĞİN YANINDA HERHANGİ BİR TUŞA BASARAK SEÇİMİNİZİ YAPINIZ.)

TABLO	YIL	AY	BÖLGE
- COICOP_DAYANIKLI	- 2013	- 1	- ADANA
- COICOP_STOK	- 2014	- 2	- ANKARA
- COICOP_SATINALIS	- 2015	- 3	- ANTALYA
- COICOP_AYNI	- 2016	- 4	- BALIKESİR
	- 2017	- 5	- BURSA
	- 2018	- 6	- DENİZLİ
		- 7	- DİYARBAKIR
		- 8	- EDİRNE
		- 9	- ERZURUM
		- 10	- GAZİANTEP
		- 11	- HATAY
		- 12	- İSTANBUL
		- HEPSİ	- İZMİR
			- KARS
			- KASTAMONU
			- KAYSERİ
			- KOCAELİ
			- KONYA
			- MALATYA
			- MANİSA
			- NEVŞEHİR
			- SAMSUN
			- SİİRT
			- TRABZON
			- VAN
			- ZONGULDAK

ANA GRUP

- GIDA VE ALKOLSÜZ İÇECEKLER
- ALKOLLÜ İÇECEKLER, TÜTÜN VE NARKOTİK MADDELER
- GİYİM VE AYAKKABI
- KONUT, SU, ELEKTRİK, GAZ VE DİĞER YAKITLAR
- MOBİLYALAR, EVDE KULLANILAN EKİPMANLAR İLE RUTİN EV BAKIM VE ONARIMI
- SAĞLIK
- ULAŞTIRMA
- HABERLEŞME
- EĞLENCE VE KÜLTÜR
- EĞİTİM
- LOKANTALAR VE OTELLER
- ÇEŞİTLİ MAL VE HİZMETLER

Şekil 4.32. KASİS raporlar giriş ekran görüntüsü 5

Açılan bu ekranda; tablo, ana grup, bölge, yıl ve ay parametreleri için seçim yapılmaktadır. Parametrelere sırasıyla; COICOP_AYNI > Giyim ve ayakkabı > Nevşehir > 2014 > 1 girilerek “kayıtları göster” seçildiği zaman oluşan sonuç ekranı Çizelge 4.28’de verilmiştir.

Çizelge 4.28. KASİS raporlar sonuç ekran görüntüsü 2

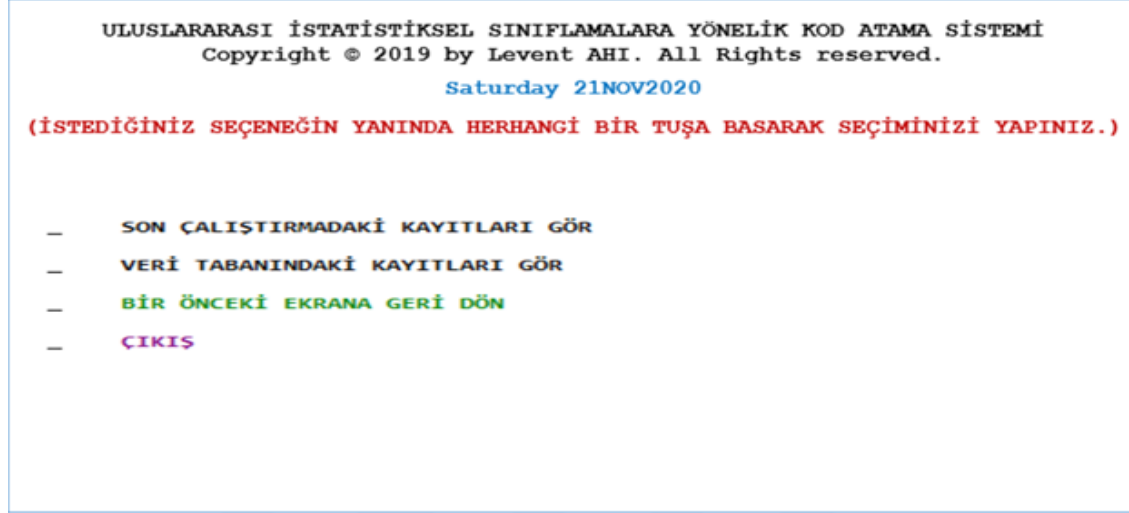
TABLO	ANA_GRUP	BOLGE_KOD	REFERANS_YIL	REFERANS_AY
COICOP_AYNI	3	50	2014	1

ANKETOR_KOD	ANKETOR_TANIM	ANKETOR_KOD_TANIM	ONERILEN_KOD	ONERILEN_TANIM
312114050	süveter moher erkek için	süveter orlon (erkek için)	312114040	süveter moher (erkek için)

Raporda, şüpheli koda karşılık KASİS’in hangi kodu önerdiği ve bu kodun hangi tarih ve saatte şüpheli olarak sınıflandığı bilgisi de yer almaktadır.

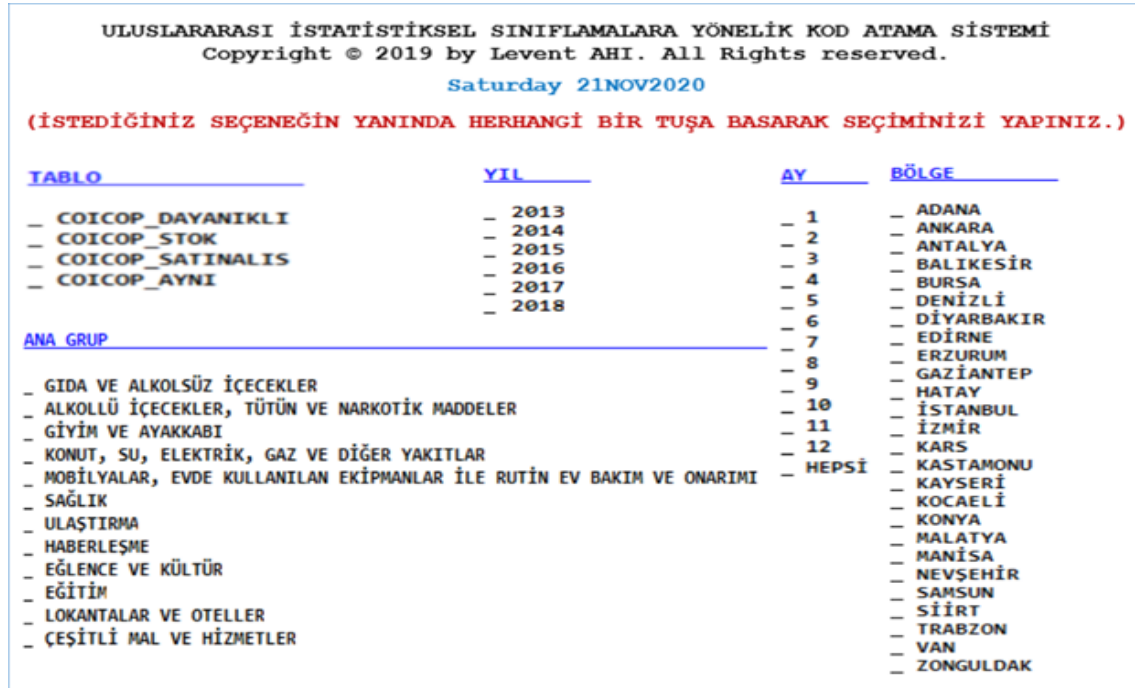
Yetersiz kayıtları gör

KASİS'in yetersiz kayıtları gör ekran görüntüsü Şekil 4.33'te verilmiştir.



Şekil 4.33. KASİS raporlar giriş ekran görüntüsü 6

Şekil 4.33'te gösterilen ekran görüntüsünden “SON ÇALIŞTIRMADAKİ KAYITLARI GÖR” veya “VERİ TABANINDAKİ KAYITLARI GÖR” seçimi yapıldığındaki ekran görüntüsü Şekil 4.34'te verilmiştir.



Şekil 4.34. KASİS raporlar giriş ekran görüntüsü 7

Açılan bu ekranda; tablo, ana grup, bölge, yıl ve ay parametreleri için seçim yapılmaktadır. Parametrelere sırasıyla; COICOP_STOK > Konut, su, elektrik, gaz ve diğer yakıtlar > Siirt > 2016 > 6 girilerek “kayıtları göster” seçildiği zaman oluşan sonuç ekranı Çizelge 4.29’da verilmiştir.

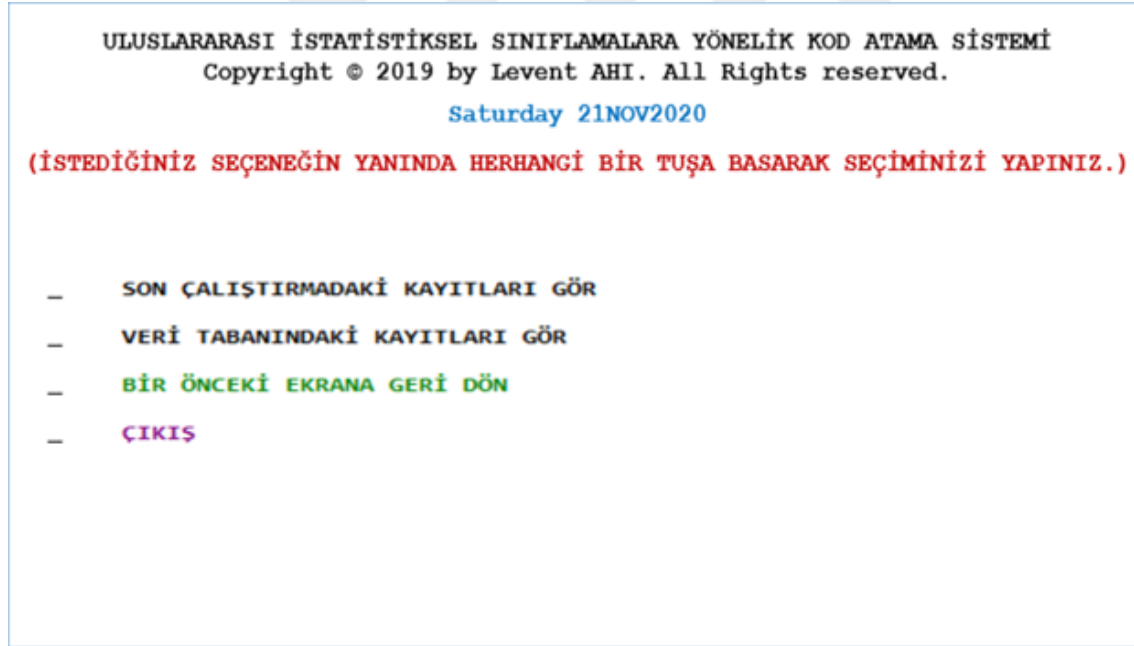
Çizelge 4.29. KASİS raporlar sonuç ekran görüntüsü 3

TABLO	ANA_GRUP	BOLGE_KOD	REFERANS_YIL	REFERANS_AY
COICOP_STOK	4	56	2016	6

ANKETOR_KOD	ANKETOR_TANIM
454101080	KÖMÜR ÜCRETİ

Eşleşmeyen kelimeleri gör

KASİS’in eşleşmeyen kelimeleri gör ekran görüntüsü Şekil 4.35’te verilmiştir.



Şekil 4.35. KASİS raporlar giriş ekran görüntüsü 8

Şekil 4.35’te gösterilen ekran görüntüsünden “SON ÇALIŞTIRMADAKİ KAYITLARI GÖR” veya “VERİ TABANINDAKİ KAYITLARI GÖR” seçimi yapıldığındaki ekran görüntüsü Şekil 4.36’da verilmiştir.

ULUSLARARASI İSTATİSTİKSEL SINIFLAMALARA YÖNELİK KOD ATAMA SİSTEMİ
Copyright © 2019 by Levent AHI. All Rights reserved.
Saturday 21NOV2020

(İSTEDİĞİNİZ SEÇENEĞİN YANINDA HERHANGİ BİR TUŞA BASARAK SEÇİMİNİZİ YAPINIZ.)

TABLO	YIL	AY	BÖLGE
- COICOP_DAYANIKLI	- 2013	- 1	- ADANA
- COICOP_STOK	- 2014	- 2	- ANKARA
- COICOP_SATINALIS	- 2015	- 3	- ANTALYA
- COICOP_AYNI	- 2016	- 4	- BALIKESİR
	- 2017	- 5	- BURSA
	- 2018	- 6	- DENİZLİ
		- 7	- DİYARBAKIR
		- 8	- EDİRNE
		- 9	- ERZURUM
		- 10	- GAZİANTEP
		- 11	- HATAY
		- 12	- İSTANBUL
		- HEPSİ	- İZMİR
			- KARS
			- KASTAMONU
			- KAYSERİ
			- KOCAELİ
			- KONYA
			- MALATYA
			- MANİSA
			- NEVŞEHİR
			- SAMSUN
			- SİİRT
			- TRABZON
			- VAN
			- ZONGULDAK

ANA GRUP

- GIDA VE ALKOLSÜZ İÇECEKLER
- ALKOLLÜ İÇECEKLER, TÜTÜN VE NARKOTİK MADDELER
- GİYİM VE AYAKKABI
- KONUT, SU, ELEKTRİK, GAZ VE DİĞER YAKITLAR
- MOBİLYALAR, EVDE KULLANILAN EKİPMANLAR İLE RUTİN EV BAKIM VE ONARIMI
- SAĞLIK
- ULAŞTIRMA
- HABERLEŞME
- EĞLENCE VE KÜLTÜR
- EĞİTİM
- LOKANTALAR VE OTELLER
- ÇEŞİTLİ MAL VE HİZMETLER

Şekil 4.36. KASİS raporlar giriş ekran görüntüsü 9

Açılan bu ekranda; tablo, ana grup, bölge, yıl ve ay parametreleri için seçim yapılmaktadır. Parametrelere sırasıyla; COICOP_DAYANIKLI > Eğlence ve kültür > Adana > 2013 > 4 girilerek “kayıtları göster” seçildiği zaman oluşan sonuç ekranı Çizelge 4.30’da verilmiştir.

Çizelge 4.30. KASİS raporlar sonuç ekran görüntüsü 4

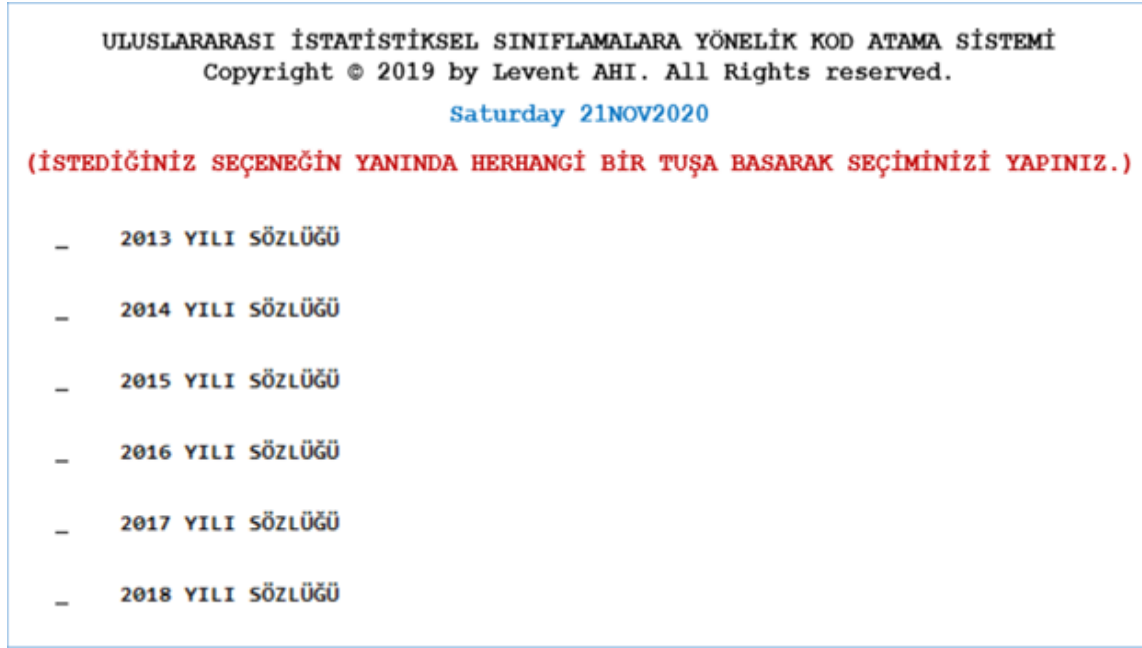
TABLO	ANA_GRUP	BOLGE_KOD	REFERANS_YIL	REFERANS_AY
COICOP_DAYANIKLI	9	1	2013	4

ANKETOR_KOD	ANKETOR_TANIM	SOZLUK_KOD	SOZLUK_TANIM
961106020	yurt içi turlar	961101010	kültürel amaçlı turlar (yurt içi günübirlik)
961106020	yurt içi turlar	961101020	ziyaret amaçlı turlar (yurt içi günübirlik)
961106020	yurt içi turlar	961101030	sportif amaçlı turlar (yurt içi günübirlik)
961106020	yurt içi turlar	961102010	tatil amaçlı turlar (yurt içi haftasonu)
961106020	yurt içi turlar	961102020	kültürel amaçlı turlar (yurt içi haftasonu)
961106020	yurt içi turlar	961102030	ziyaret amaçlı turlar (yurt içi haftasonu)
961106020	yurt içi turlar	961102040	sportif amaçlı turlar (yurt içi haftasonu)
961106020	yurt içi turlar	961103010	tatil amaçlı turlar (yurt içi bir hafta ve daha fazla süreli)
961106020	yurt içi turlar	961103020	kültürel amaçlı turlar (yurt içi bir hafta ve daha fazla süreli)
961106020	yurt içi turlar	961103030	ziyaret amaçlı turlar (yurt içi bir hafta ve daha fazla süreli)
961106020	yurt içi turlar	961103040	sportif amaçlı turlar (yurt içi bir hafta ve daha fazla süreli)
961106020	yurt içi turlar	961104010	tatil amaçlı turlar (yurt içi bayramlar, resmi tatiller vb.)
961106020	yurt içi turlar	961104020	kültürel amaçlı turlar (yurt içi bayramlar, resmi tatiller vb.)
961106020	yurt içi turlar	961104030	ziyaret amaçlı turlar (yurt içi bayramlar, resmi tatiller vb.)
961106020	yurt içi turlar	961104040	sportif amaçlı turlar (yurt içi bayramlar, resmi tatiller vb.)

Bu raporda, yetersiz olan kayıtlara karşılık kaç tane alternatif kod karşılık gelmekte olduğu görüntülenmektedir.

COICOP sözlüğünü gör

KASİS'in COICOP sözlüğünü gör ekran görüntüsü Şekil 4.37'de verilmiştir.



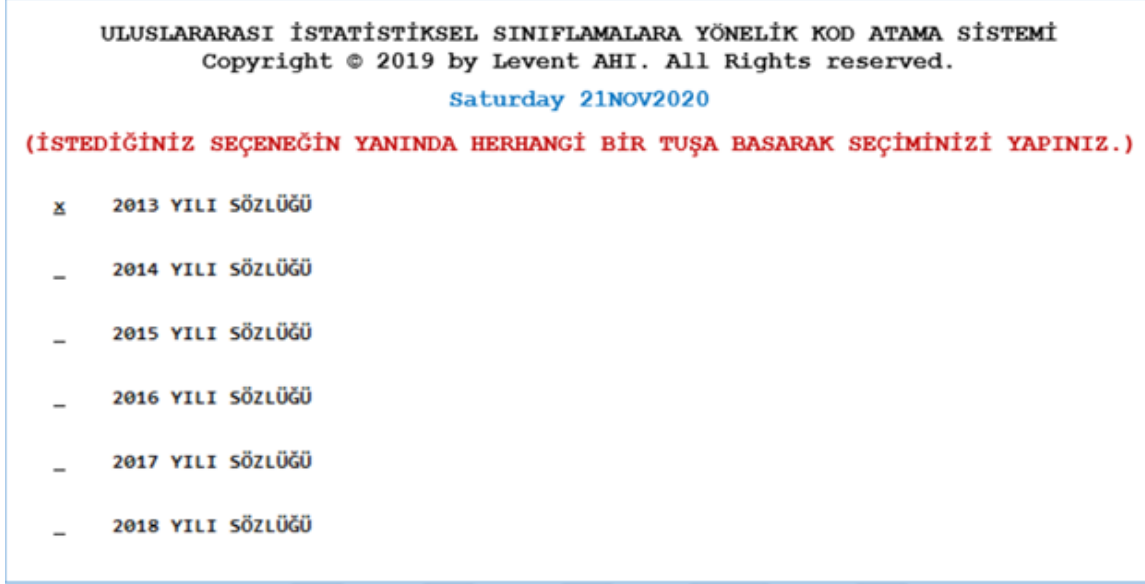
Şekil 4.37. KASİS raporlar giriş ekran görüntüsü 10

Bu ekranda;

- 2013 yılı sözlüğü
- 2014 yılı sözlüğü
- 2015 yılı sözlüğü
- 2016 yılı sözlüğü
- 2017 yılı sözlüğü
- 2018 yılı sözlüğü
- Bir önceki ekrana geri dön
- Çıkış

seçimi yapılabilir.

“2013 YILI SÖZLÜĞÜ”, “2014 YILI SÖZLÜĞÜ”, “2015 YILI SÖZLÜĞÜ”, “2016 YILI SÖZLÜĞÜ”, “2017 YILI SÖZLÜĞÜ” veya “2018 YILI SÖZLÜĞÜ” seçimi yapıldığındaki ekran görüntüsü Şekil 4.38’de verilmiştir.



Şekil 4.38. KASİS raporlar giriş ekran görüntüsü 11

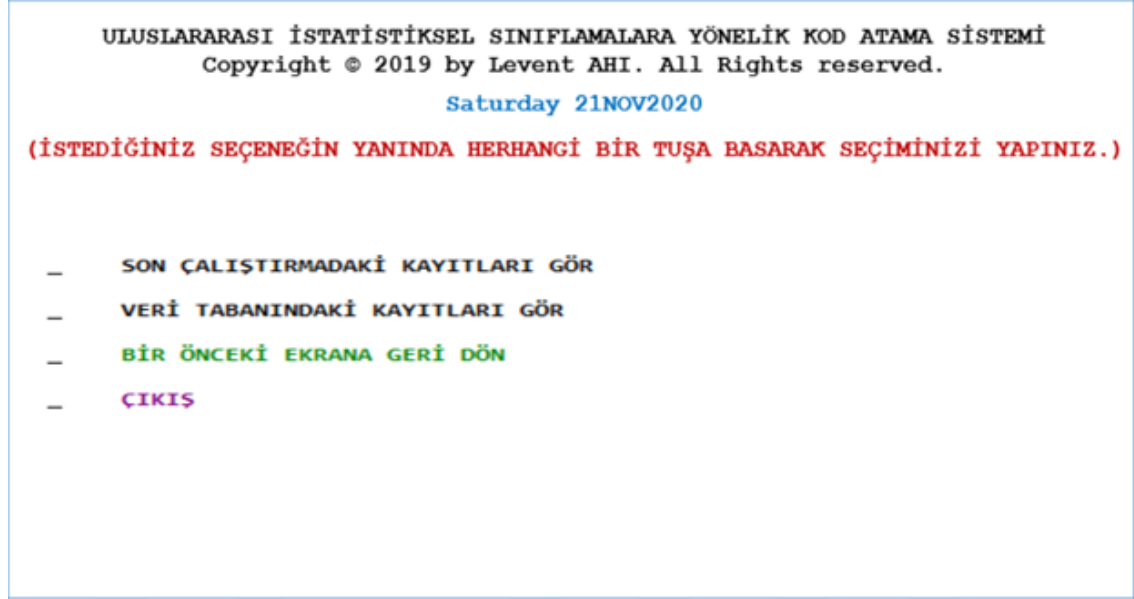
“2013 YILI SÖZLÜĞÜ” seçildiği zaman oluşan sonuç ekranının bir bölümü Çizelge 4.31’de verilmiştir.

Çizelge 4.31. KASİS raporlar sonuç ekran görüntüsü 5

YIL	SOZLUK_KOD	TANIM	SOZLUK_OLCU_BIRIM	SOZLUK_OLCU_BIRIM_TANIM	ANA_GRUP	hbs2	hbs3	hbs4	hbs5
2013	111101010	Pirinç Baldo	1 KG		1	1	11	111	1111
2013	111101020	Pirinç Berzani	1 KG		1	1	11	111	1111
2013	111101030	Pirinç Boyabat	1 KG		1	1	11	111	1111
2013	111101040	Pirinç Çorum	1 KG		1	1	11	111	1111
2013	111101050	Pirinç Karacadağ	1 KG		1	1	11	111	1111
2013	111101060	Pirinç Kepekli	1 KG		1	1	11	111	1111
2013	111101070	Pirinç Kirik	1 KG		1	1	11	111	1111
2013	111101080	Pirinç Maratelli	1 KG		1	1	11	111	1111
2013	111101090	Pirinç Terme	1 KG		1	1	11	111	1111
2013	111101100	Pirinç Tosya	1 KG		1	1	11	111	1111
2013	111101110	Pirinç Viyolin	1 KG		1	1	11	111	1111

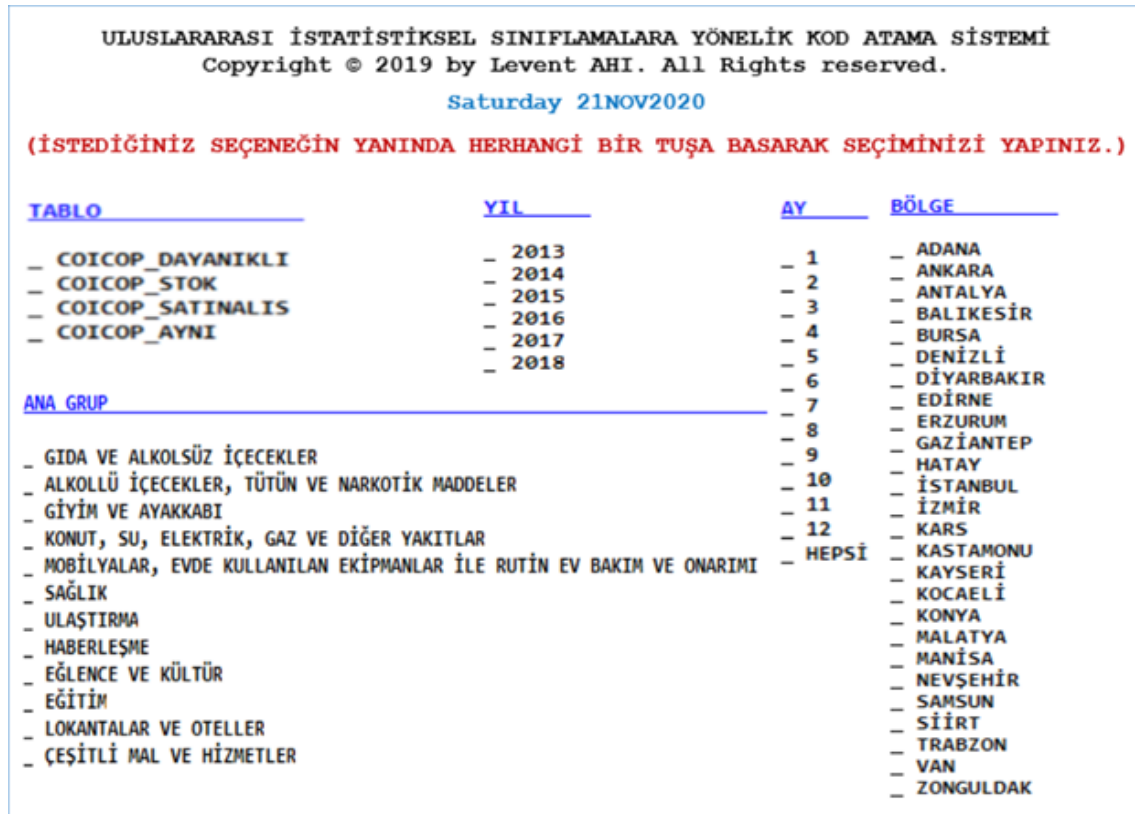
Kabul edilebilir kayıtları gör

KASİS’in kabul edilebilir kayıtları gör ekran görüntüsü Şekil 4.39’da verilmiştir.



Şekil 4.39. KASİS raporlar giriş ekran görüntüsü 12

Şekil 4.39’da gösterilen ekran görüntüsünden “SON ÇALIŞTIRMADAKİ KAYITLARI GÖR” veya “VERİ TABANINDAKİ KAYITLARI GÖR” seçimi yapıldığındaki ekran görüntüsü Şekil 4.40’ta verilmiştir.



Şekil 4.40. KASİS raporlar giriş ekran görüntüsü 13

Açılan bu ekranda; tablo, ana grup, bölge, yıl ve ay parametreleri için seçim yapılmaktadır. Parametrelere sırasıyla; COICOP_SATINALIS > Mobilyalar, evde kullanılan ekipmanlar ile rutin ev bakım ve onarımı > Manisa > 2016 > 11 girilerek “kayıtları göster” seçildiği zaman oluşan sonuç ekranı Çizelge 4.32’de verilmiştir.

Çizelge 4.32. KASİS raporlar sonuç ekran görüntüsü 6

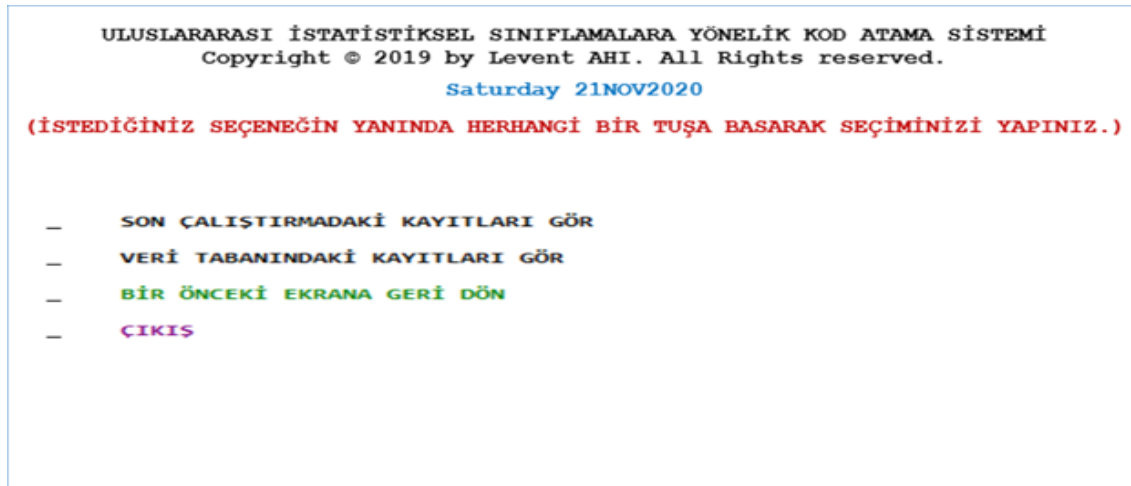
TABLO	ANA_GRUP	BOLGE_KOD	REFERANS_YIL	REFERANS_AY
COICOP_SATINALIS	5	45	2016	11

ANKETOR_KOD	ANKETOR_TANIM	ACIKLAMA
540304250	plastik kova	###SOZLUK_KOD=540304230, SOZLUK_TANIM=paspas kovası, ev temizlik aletleri (plastik) ###SOZLUK_KOD=540304250, SOZLUK_TANIM=su kovası (plastik)
561102010	bulaşık makinesi deterjanı toz	###SOZLUK_KOD=561102010, SOZLUK_TANIM=bulaşık deterjanı toz (makine için) ###SOZLUK_KOD=561102110, SOZLUK_TANIM=bulaşık deterjanı tuz ve parlatici katkı (toz) (makine için)
561102040	bulaşık deterjanı tablet,	###SOZLUK_KOD=561102040, SOZLUK_TANIM=bulaşık deterjanı tablet (makine için) ###SOZLUK_KOD=561102210, SOZLUK_TANIM=bulaşık deterjanı, tuz ve parlatici set (makine için)
561201030	faraş	###SOZLUK_KOD=561201030, SOZLUK_TANIM=faraş (kürek) ###SOZLUK_KOD=561201040, SOZLUK_TANIM=faraşlı süpürge

Bu raporda, kabul edilebilir kriterler kapsamında kabul edilen kayıtlar, “kabul” ve “bulanık kabul” ayrımında gösterilmektedir. Ayrıca harcama tanımının hangi kod tanımlarının kapsamına girdiklerini de raporda görebilmek mümkündür.

Programı çalıştırma kayıtlarını gör

KASİS’in programı çalıştırma kayıtlarını gör ekran görüntüsü Şekil 4.41’de verilmiştir.



Şekil 4.41. KASİS raporlar giriş ekran görüntüsü 14

Şekil 4.41’de gösterilen ekran görüntüsünden “SON ÇALIŞTIRMADAKİ KAYITLARI GÖR” veya “VERİ TABANINDAKİ KAYITLARI GÖR” seçimi yapıldığındaki ekran görüntüsü Şekil 4.42’de verilmiştir.

ULUSLARARASI İSTATİSTİKSEL SINIFLAMALARA YÖNELİK KOD ATAMA SİSTEMİ
Copyright © 2019 by Levent AHI. All Rights reserved.

Saturday 21NOV2020

(İSTEDİĞİNİZ SEÇENEĞİN YANINDA HERHANGİ BİR TUŞA BASARAK SEÇİMİNİZİ YAPINIZ.)

TABLO	YIL	AY	BÖLGE
- COICOP_DAYANIKLI	- 2013	- 1	- ADANA
- COICOP_STOK	- 2014	- 2	- ANKARA
- COICOP_SATINALIS	- 2015	- 3	- ANTALYA
- COICOP_AYNI	- 2016	- 4	- BALIKESİR
	- 2017	- 5	- BURSA
	- 2018	- 6	- DENİZLİ
		- 7	- DİYARBAKIR
		- 8	- EDİRNE
		- 9	- ERZURUM
		- 10	- GAZİANTEP
		- 11	- HATAY
		- 12	- İSTANBUL
		- HEPSİ	- İZMİR
			- KARS
			- KASTAMONU
			- KAYSERİ
			- KOCAELİ
			- KONYA
			- MALATYA
			- MANİSA
			- NEVŞEHİR
			- SAMSUN
			- SİİRT
			- TRABZON
			- VAN
			- ZONGULDAK

ANA GRUP

- GIDA VE ALKOLSÜZ İÇECEKLER
- ALKOLLÜ İÇECEKLER, TÜTÜN VE NARKOTİK MADDELER
- GİYİM VE AYAKKABI
- KONUT, SU, ELEKTRİK, GAZ VE DİĞER YAKITLAR
- MOBİLYALAR, EVDE KULLANILAN EKİPMANLAR İLE RUTİN EV BAKIM VE ONARIMI
- SAĞLIK
- ULAŞTIRMA
- HABERLEŞME
- EĞLENCE VE KÜLTÜR
- EĞİTİM
- LOKANTALAR VE OTELLER
- ÇEŞİTLİ MAL VE HİZMETLER

Şekil 4.42. KASİS raporlar giriş ekran görüntüsü 15

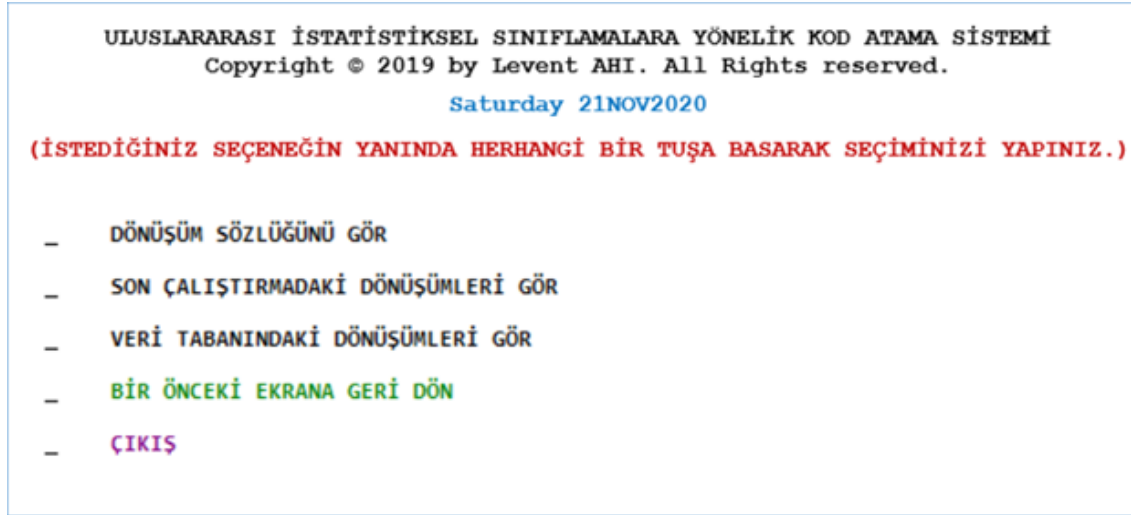
Açılan bu ekranda; tablo, ana grup, bölge, yıl ve ay parametreleri için seçim yapılmaktadır. Parametrelere sırasıyla; COICOP_SATINALIS > Gıda ve alkolsüz içecekler > İstanbul > 2016 > 4 girilerek “kayıtları göster” seçildiği zaman oluşan sonuç ekranı Çizelge 4.33’te verilmiştir.

Çizelge 4.33. KASİS raporlar sonuç ekran görüntüsü 7

TABLO	ANA_GRUP	BOLGE_KOD	REFERANS_YIL	REFERANS_AY
COICOP_SATINALIS	1	34	2016	4
TOPLAM_KAYIT_SAYISI	SUPHELI_KAYIT_SAYISI	YETERSİZ_KAYIT_SAYISI		
18123	52	193		
YENİ_ESLESEN_KAYIT_SAYISI	SOZLUK_ESLESEN_KAYIT_SAYISI	BULANIK_ESLESEN_KAYIT_SAYISI		
14013	3362	19		
KABUL_KAYIT_SAYISI	BULANIK_KABUL_KAYIT_SAYISI	SURE		
463	21	00:12:21		
ESLESEN_YUZDE	SUPHELI_YETERSİZ_YUZDE	TOPLAM_YUZDE		
98.6	1.4	100		

Dönüşümleri gör

KASİS'in dönüşümleri gör ekran görüntüsü Şekil 4.43'te verilmiştir.



Şekil 4.43. KASİS raporlar giriş ekran görüntüsü 16

“DÖNÜŞÜM SÖZLÜĞÜNÜ GÖR” seçimi yapıldığında oluşan kayıtların bir kısmının ekran görüntüsü Çizelge 4.34’te verilmiştir.

Çizelge 4.34. KASİS raporlar sonuç ekran görüntüsü 8

INPUT	OUTPUT1	OUTPUT2
A.CIK	ACIK	
AACIK	ACIK	
AADANA	ADANA	
AAKKABI	AYAKKABI	
AARBA	ARABA	
AARMUT	ARMUT	
AATURKA	ALATURKA	
AAYAKKABI	AYAKKABI	
AAYCICEK	AYCICEK	
ABLDO	BALDO	
ABNYO	BANYO	
ABOJUR	ABAJUR	
ABONBAN	ABONMAN	
ABONMEN	ABONMAN	
ACCIK	ACIK	
ACIBADEMLI	ACIBADEM	

Şekil 4.43’te gösterilen ekran görüntüsünden “SON ÇALIŞTIRMADAKİ KAYITLARI GÖR” veya “VERİ TABANINDAKİ KAYITLARI GÖR” seçimi yapıldığındaki ekran görüntüsü Şekil 4.44’te verilmiştir.

ULUSLARARASI İSTATİSTİKSEL SINIFLAMALARA YÖNELİK KOD ATAMA SİSTEMİ
Copyright © 2019 by Levent AHI. All Rights reserved.
Saturday 21NOV2020

(İSTEDİĞİNİZ SEÇENEĞİN YANINDA HERHANGİ BİR TUŞA BASARAK SEÇİMİNİZİ YAPINIZ.)

TABLO	YIL	AY	BÖLGE
- COICOP_DAYANIKLI	- 2013	- 1	- ADANA
- COICOP_STOK	- 2014	- 2	- ANKARA
- COICOP_SATINALIS	- 2015	- 3	- ANTALYA
- COICOP_AYNI	- 2016	- 4	- BALIKESİR
	- 2017	- 5	- BURSA
	- 2018	- 6	- DENİZLİ
		- 7	- DİYARBAKIR
		- 8	- EDİRNE
		- 9	- ERZURUM
		- 10	- GAZİANTEP
		- 11	- HATAY
		- 12	- İSTANBUL
		- HEPSİ	- İZMİR
			- KARS
			- KASTAMONU
			- KAYSERİ
			- KOCAELİ
			- KONYA
			- MALATYA
			- MANİSA
			- NEVŞEHİR
			- SAMSUN
			- SİİRT
			- TRABZON
			- VAN
			- ZONGULDAK

Şekil 4.44. KASİS raporlar giriş ekran görüntüsü 17

Açılan bu ekranda; tablo, ana grup, bölge, yıl ve ay parametreleri için seçim yapılmaktadır. Parametrelere sırasıyla; COICOP_SATINALIS > Mobilyalar, evde kullanılan ekipmanlar ile rutin ev bakım ve onarımı > Ankara > 2013 > 5 girilerek “kayıtları göster” seçildiği zaman oluşan sonuç ekranı Çizelge 4.35’te verilmiştir.

Çizelge 4.35. KASİS raporlar sonuç ekran görüntüsü 9

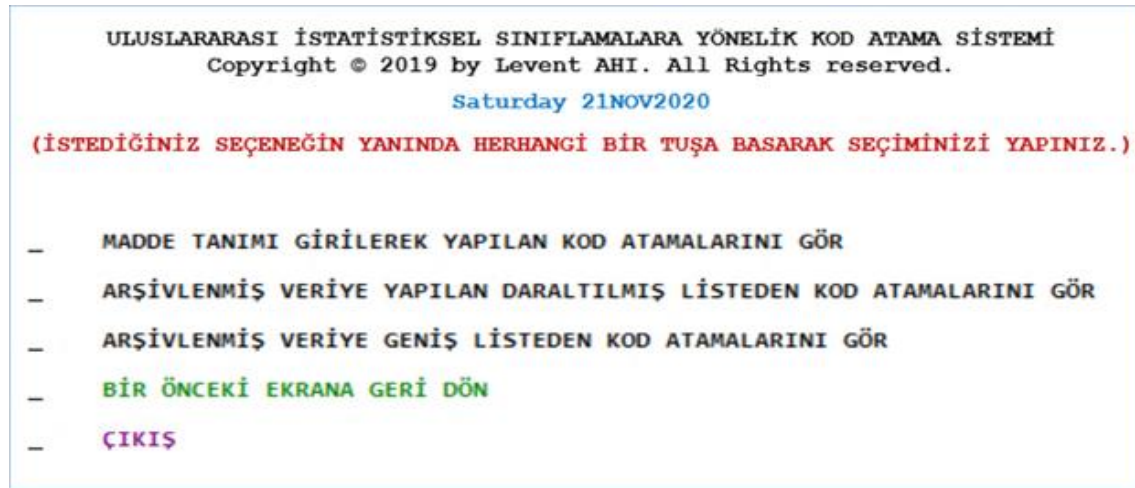
TABLO	ANA_GRUP	BOLGE_KOD	REFERANS_YIL	REFERANS_AY
COICOP_SATINALIS	5	6	2013	5

ANKETOR_KOD	ANKETOR_TANIM	DONUSEN	DONUSUM
561203020	buzdolabu poşeti	BUZDOLABU	BUZDOLABI

Çizelge 4.35’te görüleceği üzere; “buzdolabu poşeti” tanımında yer alan “buzdolabu” kelimesi “buzdolabi” kelimesine dönüştürülerek kod eşleştirilmesi sağlanmıştır.

Kod atamalarını gör

KASİS'in madde tanımı girilerek, arşivlenmiş veriye yapılan daraltılmış listeden veya arşivlenmiş veriye geniş listeden yapılan kod atamalarının görüntülediği ekran görüntüsü Şekil 4.45'te verilmiştir.



Şekil 4.45. KASİS raporlar giriş ekran görüntüsü 18

“MADDE TANIMI GİRİLEREK YAPILAN ATAMALARI GÖR” seçimi yapıldığındaki ekran görüntüsü Çizelge 4.36’da verilmiştir.

Çizelge 4.36. KASİS raporlar sonuç ekran görüntüsü 10

TANIM	SOZLUK_KOD	SOZLUK_TANIM	KELIME_GECEN_YERE_GORE	KELIME_BUTUNLUGU	BULANIK_ESLESTIRME
ELMA	116121010	elma (taze)	ÖNERİLEN KOD		
ELMA	552101080	el malası			ÖNERİLEN KOD
PEKMEZ	118208980	pekmez diğer	ÖNERİLEN KOD		
civi	561207010	çivi			ÖNERİLEN KOD
elma	116121010	elma (taze)	ÖNERİLEN KOD		
elma	552101080	el malası			ÖNERİLEN KOD
karpuz	116136010	karpuz			ÖNERİLEN KOD
raptiye	561207060	raptiye			ÖNERİLEN KOD
zıma	722401030	şanzıman yağ katkısı	ÖNERİLEN KOD		
zıma	954908430	zımba			ÖNERİLEN KOD

“ARŞİVLENMİŞ VERİYE YAPILAN DARALTIILMIŞ LİSTEDEN KOD ATAMA” seçimi yapıldığındaki ekran görüntüsünün bir kısmı Çizelge 4.37’de verilmiştir.

Çizelge 4.37. KASİS raporlar sonuç ekran görüntüsü 11

ANKETOR_KOD	ANKETOR_TANIM	SOZLUK_KOD	SOZLUK_TANIM	KOD_ATAMA_DURUM	YUZDE_ORAN
511115030	ranza(iki katlı yavru yataklı)	511114040	yatak bazası yavrulu	ÖNERİLEN KOD	100
511122030	sehpa servis sehpa	511122980	servis sehpa diğer	ÖNERİLEN KOD	80
531101030	buzdolabı nofrost	531101030	buzdolabı no-frost	ÖNERİLEN KOD	60

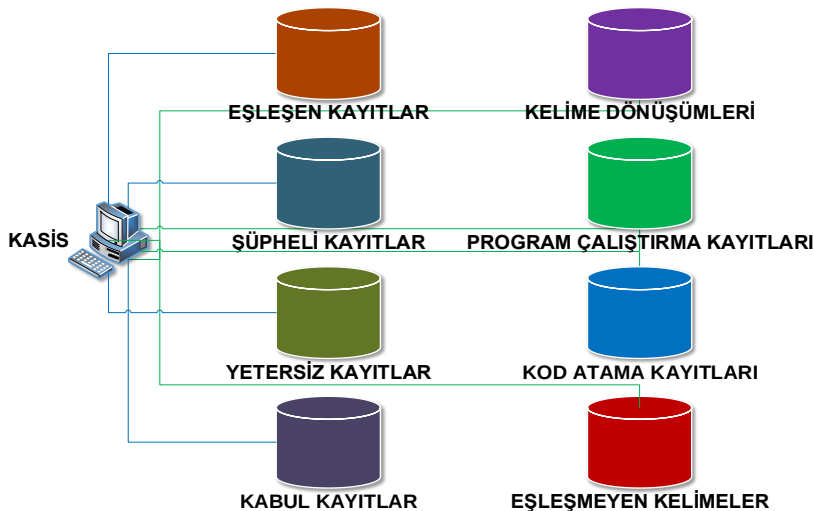
“ARŞİVLENMİŞ VERİYE GENİŞ LİSTEDEN KOD ATAMA” seçimi yapıldığındaki ekran görüntüsünün bir kısmı Çizelge 4.38’de verilmiştir.

Çizelge 4.38. KASİS raporlar sonuç ekran görüntüsü 12

ANKETOR_KOD	ANKETOR_TANIM	SOZLUK_KOD	SOZLUK_TANIM	KOD_ATAMA_DURUM	YUZDE_ORAN
511115030	ranza(iki katlı yavru yataklı)	511115030	ranza	ÖNERİLEN KOD	60
511122030	sehpa servis sehpa	520304060	sehpa örtüsü	ÖNERİLEN KOD	60
531101030	buzdolabı nofrost	531101030	buzdolabı no-frost	ÖNERİLEN KOD	60

4.5. Veri Tabanları

KASİS tarafından kayıtlar Şekil 4.46’da verilen veri tabanlarına ilave edilmektedir. Bu veri tabanlarındaki kayıtlara istenildiği zaman raporlar aracılığıyla ulaşılabilir. Ayrıca, veri tabanında mevcut olan kayıt KASİS’e tekrar geldiği zaman eşleştirme ve sınıflama katmanlarından geçmeden direkt olarak kaydın sınıflaması yapılabilir. Bu durum, sistemin sonuç verme süresini olumlu yönde etkileyerek sistemin çalışma performansını artırmaktadır.



Şekil 4.46. KASİS veri tabanı yapısı

5. BULGULAR

Bu bölüm, araştırma soruları göz önünde bulundurularak iki alt bölüme ayrılmıştır. Birinci alt bölümde anketörün atadığı kodların doğruluğu kontrol edilerek elde edilen bulgulara yer verilmiştir. İkinci alt bölümde ise KASİS tarafından yetersiz ve şüpheli olarak sınıflanmış olan kayıtlara yeniden kod ataması yapılarak anketörün atamış olduğu kodlar ile karşılaştırılmış ve elde edilen bulgulara yer verilmiştir.

Bu çalışma kapsamında yaklaşık 13 milyon COICOP kodunun doğru verilip verilmediğinin kontrolü KASİS'in dilbilgisi, eşleştirme ve sınıflama katmanlarında gerçekleştirilmiştir. Yıllara göre kontrol edilen kayıt sayısı Çizelge 5.1'de verilmiştir.

Çizelge 5.1. Yıllara göre kayıt sayısı

Yıl	Kayıt sayısı
2013	1 935 743
2014	1 952 803
2015	2 064 452
2016	2 193 165
2017	2 225 172
2018	2 305 256
Toplam	12 676 591

5.1. Dilbilgisi, Eşleştirme ve Sınıflama Katmanında Elde Edilen Bulgular

Bu alt bölümde, KASİS'in dört katmanından üçü olan dilbilgisi, eşleştirme ve sınıflama katmanlarında anketörün yazdığı tanım ve karşılığında verdiği kod kontrol edilerek araştırma soruları kapsamında bulgulara yer verilmiştir. Anketörün atadığı kodların kontrol edilmesi sonusunda kodlar temel olarak dört grupta sınıflanmaktadır.

1. Eşleşen kod: Anketörün atadığı kodun doğru olduğu KASİS tarafından teyit edilmesi durumunda kod bu sınıfta yer almaktadır. Anketörün yazdığı harcama tanımının sözlükte birebir yer alması durumunda “sözlük eşleşen kod”, anketörün yazdığı harcama tanımının sözlükte birebir yer almaması durumunda “yeni eşleşen kod” ve anketörün yazdığı harcama tanımındaki yazım yanlışının KASİS tarafından düzeltilerek eşleşen kod olması durumunda “bulanık eşleşen kod” olarak sınıflanmaktadır.

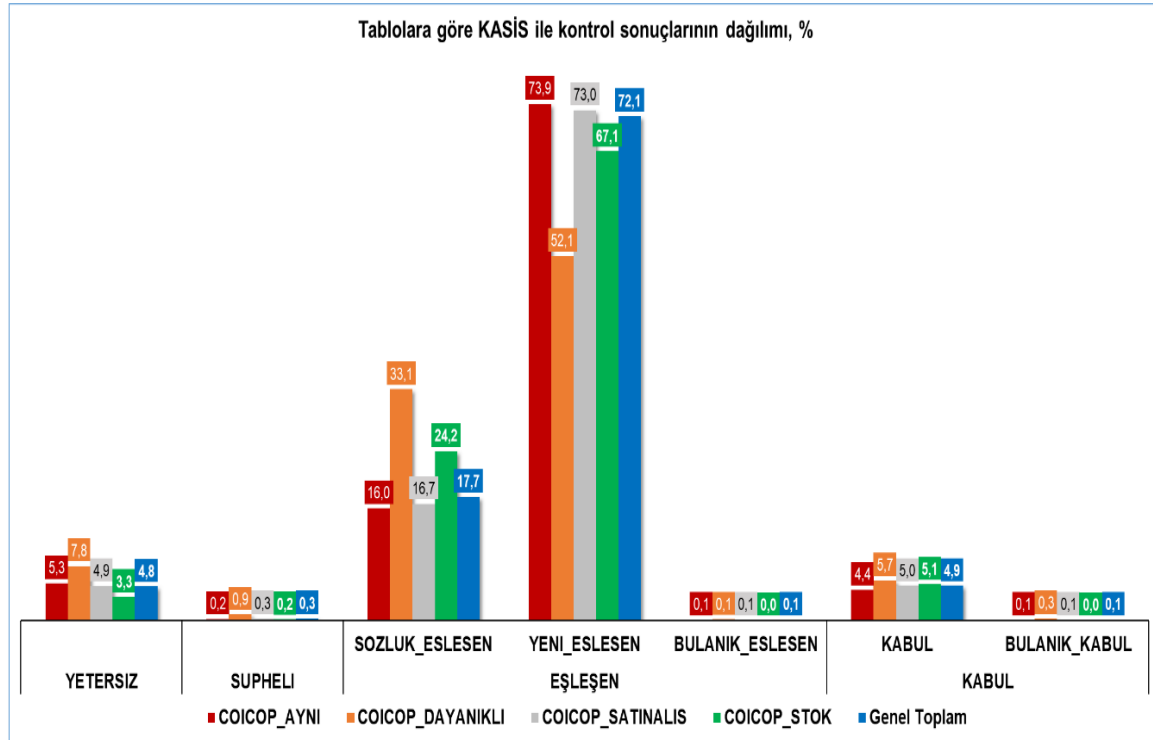
2. *Şüpheli kod*: Anketörün atadığı kodun hatalı olduğu KASİS tarafından teyit edilmesi durumunda kod bu sınıfta yer almaktadır.

3. *Yetersiz kod*: Anketörün yazdığı harcama tanımı ile atanabilecek birden fazla kodun olduğu KASİS tarafından teyit edilmesi durumunda kod bu sınıfta yer almaktadır.

4. *Kabul kod*: Anketörün atadığı kodun sistem tarafından önerilen kodların arasında olduğu, anketörün atadığı kod ile önerilen kodların ilk beş basamağının aynı olduğu ve önerilen kod sayısının beşten az olduğu KASİS tarafından teyit edilmesi durumunda kod bu sınıfta yer almaktadır. Anketörün yazdığı harcama tanımındaki yazım yanlışının KASİS tarafından düzeltilerek kabul kod olması durumunda “bulanık kabul kod” diğer durumlarda “kabul kod” olarak sınıflanmaktadır.

5.1.1. Anketör ile KASİS kodu arasında tablolara göre farklılık var mıdır?

KASİS programı aracılığıyla anketörün verdiği kodlar kontrol edilmiş ve tablolara göre Şekil 5.1’de verilen dağılım elde edilmiştir.



Şekil 5.1. Tablolara göre KASİS ile kontrol sonuçlarının dağılımı

Şekil 5.1'e göre; yetersiz kod oranı, dayanıklı tablosunda %7,8, aynı tablosunda %5,3, satın alış tablosunda %4,9 ve stok tablosunda %3,3'tür. Dayanıklı tablosunda diğer tablolara nispeten daha yüksek oranda yetersiz kod belirlenmiştir.

Şüpheli kod olarak sınıflanmış kayıt oranı dayanıklı tablosunda %0,9, satın alış tablosunda %0,3 ve aynı ve stok tablolarında ise %0,2'dir. Dayanıklı tablosunda diğer tablolara nispeten daha yüksek oranda şüpheli kod belirlenmiştir.

Yetersiz ve şüpheli kodların toplamına bakıldığında; en başarılı tablo %3,5 ile stok tablosu olurken en başarısız tablo %8,7 ile dayanıklı tablosu olmuştur.

Tablolara göre eşleşen kodların dağılımı Çizelge 5.2'de verilmiştir.

Çizelge 5.2. Tablolara göre eşleşen kodlar

Tablo	Eşleşen kodlar içerisindeki pay (%)	Tüm kodlar içerisindeki pay (%)
COICOP_AYNI	%6,0	%90,0
COICOP_DAYANIKLI	%1,8	%85,3
COICOP_SATINALIS	%82,7	%89,8
COICOP_STOK	%9,6	%91,3
Toplam	%100,0	%89,9
Ortalama	%25,0	%89,1

Çizelge 5.2'de verilen eşleşen kodlar içerisindeki pay dağılımında, eşleşen kodların toplamı 100 kabul edildiğinde tabloların bu dağılıma yaptığı katkı gösterilmiştir. Tüm kodlar içerisindeki pay dağılımında ise, her bir tablodaki kayıt sayısı kendi içerisinde 100 kabul edildiğinde tabloda yer alan eşleşen kodların payları verilmiştir. Çizelgeye göre; stok kodlarının %91,3'ü, aynı kodlarının %90'ı, satın alış kodlarının %89,8'i ve dayanıklı kodlarının %85,3'ü eşleşen olarak sınıflanmıştır. Tüm kodlar içerisindeki payların ortalaması %89,1'dir. Dayanıklı kodları bu oranın altında, diğer üç tablonun kodları ise bu oranın üzerinde yer almıştır. Eşleşen kodlar içerisinde, satın alış kodları %82,7, stok kodları %9,6, aynı kodları %6 ve dayanıklı kodları %1,8'lik kısmı oluşturmaktadır. Eşleşen kodlar içerisindeki payların ortalaması %25'tir. Satın alış dışındaki tablolar bu oranın altında yer almıştır.

Tablolara göre kabul kodların dağılımı Çizelge 5.3'te verilmiştir.

Çizelge 5.3. Tablolara göre kabul kodlar

Tablo	Kabul kodlar içerisindeki pay (%)	Tüm kodlar içerisindeki pay (%)
COICOP_AYNI	%5,3	%4,4
COICOP_DAYANIKLI	%2,2	%6,0
COICOP_SATINALIS	%82,7	%5,0
COICOP_STOK	%9,8	%5,2
Toplam	%100,0	%5,0
Ortalama	%25,0	%5,2

Çizelge 5.3'te verilen kabul kodlar içerisindeki pay dağılımında, kabul kodların toplamı 100 kabul edildiğinde tabloların bu dağılıma yaptığı katkı gösterilmiştir. Tüm kodlar içerisindeki pay dağılımında ise, her bir tablodaki kayıt sayısı kendi içerisinde 100 kabul edildiğinde tabloda yer alan kabul kodların payları verilmiştir. Çizelgeye göre; dayanıklı kodlarının %6'sı, stok kodlarının %5,2'si, satın alış kodlarının %5'i ve ayni kodlarının %4,4'ü kabul olarak sınıflanmıştır. Tüm kodlar içerisindeki payların ortalaması %5,2'dir. Ayni ve satın alış kodları bu oranın altında, dayanıklı ve stok kodları ise bu oranın üzerinde yer almıştır. Kabul kodlar içerisinde, satın alış kodları %82,7, stok kodları %9,8, ayni kodları %5,3 ve dayanıklı kodları %2,2'lik kısmı oluşturmaktadır. Kabul kodlar içerisindeki payların ortalaması %25'tir. Satın alış dışındaki tablolar bu oranın altında yer almıştır.

5.1.2. Anketör ile KASİS kodu arasında ana gruplara göre farklılık var mıdır?

KASİS programı aracılığıyla anketörün verdiği kodlar kontrol edilmiş ve ana gruplara göre elde edilen sonuçlara Çizelge 5.4'te yer verilmiştir.

Çizelge 5.4. Ana gruplara göre KASİS ile kontrol sonuçlarının dağılımı

Ana grup	YETERSİZ	SUPHELİ	EŞLEŞEN			KABUL	
			SOZLUK_ESLESEN	YENI_ESLESEN	BULANIK_ESLESEN	KABUL	BULANIK_KABUL
Gıda ve alkolsüz içecekler	3,1	0,2	21,6	71,3	0,1	3,6	0,1
Alkollü içecek, sigara ve tütün	1,6	0,1	3,6	93,5	0,0	1,3	0,0
Giyim ve ayakkabı	9,7	0,3	3,0	82,3	0,1	4,5	0,1
Konut ve kira	3,7	0,4	20,3	73,1	0,0	2,4	0,0
Mobilya ve ev eşyaları	4,4	0,3	27,7	59,3	0,2	8,0	0,1
Sağlık	4,1	0,5	19,6	73,1	0,4	2,2	0,0
Ulaştırma	6,2	0,4	1,6	84,2	0,0	7,6	0,0
Haberleşme	10,9	0,4	1,7	84,3	0,1	2,2	0,5
Eğlence ve kültür	5,6	1,0	20,4	70,3	0,1	2,6	0,0
Eğitim hizmetleri	13,5	1,2	21,9	58,1	0,1	5,2	0,0
Lokanta ve oteller	22,1	0,1	0,2	59,8	0,0	17,7	0,0
Çeşitli mal ve hizmetler	2,1	0,4	17,9	71,0	0,1	8,5	0,0
Genel Toplam	4,8	0,3	17,7	72,1	0,1	4,9	0,1

Çizelge 5.4'e göre; yetersiz kod oranı, lokanta ve oteller ana grubunda %22,1, eğitim hizmetleri ana grubunda %13,5, haberleşme ana grubunda %10,9 ve giyim ve ayakkabı ana grubunda %9,7'dir. Yetersiz kodların yüzdesel olarak en az olduğu ana gruplar sırasıyla, %1,6 ile alkollü içecek, sigara ve tütün ana grubu, %2,1 ile çeşitli mal ve hizmetler ana grubu ve %3,1 ile gıda ve alkolsüz içecekler ana grubudur.

Şüpheli kod olarak sınıflanmış kayıt oranı eğitim hizmetleri ana grubunda %1,2 ve eğlence ve kültür ana grubunda %1'dir. Şüpheli kodların yüzdesel olarak en az olduğu ana gruplar sırasıyla, %0,1 ile alkollü içecek, sigara ve tütün ana grubu ile lokanta ve oteller ana grubu ve %0,2 ile gıda ve alkolsüz içecekler ana grubudur. Eğitim hizmetleri ve eğlence ve kültür ana grubunda diğer ana gruplara oranla daha yüksek şüpheli kod belirlenmiştir.

Yetersiz ve şüpheli kodların toplamına bakıldığında; en başarılı ana grup %1,7 ile alkollü içecek, sigara ve tütün ana grubu olurken en başarısız ana grup %22,2 ile lokanta ve oteller ana grubu olmuştur.

Ana gruplara göre eşleşen kodların dağılımı Çizelge 5.5'te verilmiştir.

Çizelge 5.5. Ana gruplara göre eşleşen kodlar

Ana grup	Eşleşen kodlar içerisindeki pay (%)	Tüm kodlar içerisindeki pay (%)
Gıda ve alkolsüz içecekler	%65,2	%93,0
Alkollü içecek, sigara ve tütün	%6,0	%97,1
Giyim ve ayakkabı	%2,3	%85,4
Konut ve kira	%3,2	%93,5
Mobilya ve ev eşyaları	%5,0	%87,2
Sağlık	%1,7	%93,1
Ulaştırma	%3,9	%85,8
Haberleşme	%1,2	%86,1
Eğlence ve kültür	%2,2	%90,8
Eğitim hizmetleri	%0,2	%80,0
Lokanta ve oteller	%4,5	%60,1
Çeşitli mal ve hizmetler	%4,4	%89,0
Toplam	%100,0	%89,9
Ortalama	%8,3	%86,7

Çizelge 5.5'te verilen eşleşen kodlar içerisindeki pay dağılımında, eşleşen kodların toplamı 100 kabul edildiğinde tabloların bu dağılıma yaptığı katkı gösterilmiştir. Tüm kodlar içerisindeki pay dağılımında ise, her bir tablodaki kayıt sayısı kendi içerisinde 100 kabul

edildiğinde tabloda yer alan eşleşen kodların payları verilmiştir. Çizelgeye göre; alkollü içecek, sigara ve tütün ana grubu kodlarının %97,1'i, konut ve kira ana grubu kodlarının %93,5'i, sağlık ana grubu kodlarının %93,1'i ve gıda ve alkolsüz içecekler ana grubu %93'ü eşleşen olarak sınıflanmıştır. Tüm kodlar içerisindeki payların ortalaması %86,7'dir. Haberleşme ana grubu %86,1, ulaştırma ana grubu %85,8, giyim ve ayakkabı ana grubu %85,4, eğitim hizmetleri ana grubu %80 ve lokanta ve oteller ana grubu %60,1 ile bu oranın altında yer almıştır. Eşleşen kodlar içerisinde, gıda ve alkolsüz içecekler ana grubu kodları %65,2, alkollü içecek, sigara ana grubu kodları %6 ve mobilya ve ev eşyaları ana grubu kodları %5'lik kısmı oluşturmaktadır. Eşleşen kodlar içerisindeki payların ortalaması %8,3'tür. Gıda ve alkolsüz içecekler ana grubu dışındaki ana gruplar bu oranın altında yer almıştır.

Ana gruplara göre kabul kodların dağılımı Çizelge 5.6'da verilmiştir.

Çizelge 5.6. Ana gruplara göre kabul kodlar

Ana grup	Kabul kodlar içerisindeki pay (%)	Tüm kodlar içerisindeki pay (%)
Gıda ve alkolsüz içecekler	%46,0	%3,7
Alkollü içecek, sigara ve tütün	%1,4	%1,3
Giyim ve ayakkabı	%2,2	%4,6
Konut ve kira	%1,5	%2,5
Mobilya ve ev eşyaları	%8,4	%8,2
Sağlık	%0,7	%2,3
Ulaştırma	%6,3	%7,6
Haberleşme	%0,7	%2,6
Eğlence ve kültür	%1,1	%2,6
Eğitim hizmetleri	%0,3	%5,2
Lokanta ve oteller	%23,8	%17,7
Çeşitli mal ve hizmetler	%7,6	%8,5
Toplam	%100,0	%5,0
Ortalama	%8,3	%5,6

Çizelge 5.6'da verilen kabul kodlar içerisindeki pay dağılımında, kabul kodların toplamı 100 kabul edildiğinde tabloların bu dağılıma yaptığı katkı gösterilmiştir. Tüm kodlar içerisindeki pay dağılımında ise, her bir tablodaki kayıt sayısı kendi içerisinde 100 kabul edildiğinde tabloda yer alan kabul kodların payları verilmiştir. Çizelgeye göre; lokanta ve oteller ana grubu kodlarının %17,7'si, çeşitli mal ve hizmetler ana grubu kodlarının %8,5'i, mobilya ve ev eşyaları ana grubu kodlarının %8,2'si ve ulaştırma kodlarının %7,6'sı kabul olarak sınıflanmıştır. Tüm kodlar içerisindeki payların ortalaması %5,6'dır. Bu ana gruplar

dışındaki ana gruplar bu oranın altında yer almıştır. Kabul kodlar içerisinde, gıda ve alkolsüz içecekler ana grubu kodları %46, lokanta ve oteller ana grubu kodları %23,8 ve mobilya ve ev eşyaları ana grubu kodları %8,4'lük kısmı oluşturmaktadır. Kabul kodlar içerisindeki payların ortalaması %8,3'tür. Bu ana gruplar dışındaki ana gruplar bu oranın altında yer almıştır.

5.1.3. Anketör ile KASİS kodu arasında bölgelere göre farklılık var mıdır?

KASİS programı aracılığıyla kontrol edilen kodların bölgelere göre elde edilen sonuçlara Çizelge 5.7'de yer verilmiştir.

Çizelge 5.7. Bölgelere göre KASİS ile kontrol sonuçlarının dağılımı

Bölge	YETERSİZ	SUPHELI	EŞLEŞEN			KABUL	
			SOZLUK ESLESEN	YENİ ESLESEN	BULANIK ESLESEN	KABUL	BULANIK KABUL
Adana	4,1	0,1	21,4	70,1	0,0	4,2	0,1
Ankara	0,5	0,1	23,1	74,5	0,0	1,8	0,0
Antalya	5,2	0,3	14,7	73,4	0,2	6,1	0,1
Balıkesir	8,8	0,2	11,9	69,1	0,0	9,9	0,0
Bursa	6,3	0,4	14,4	72,9	0,1	5,9	0,1
Denizli	6,9	0,5	14,4	72,4	0,1	5,6	0,1
Diyarbakır	7,5	0,2	15,5	66,6	0,1	10,1	0,1
Edirne	2,2	0,3	24,8	68,1	0,1	4,6	0,0
Erzurum	10,7	0,3	18,0	64,5	0,1	6,4	0,1
Gaziantep	5,1	0,1	20,5	66,7	0,0	7,4	0,1
Hatay	0,8	0,1	24,3	71,6	0,0	3,1	0,0
İstanbul	3,4	0,3	15,2	76,9	0,1	4,1	0,1
İzmir	8,0	0,3	11,9	73,5	0,2	5,9	0,2
Kars	4,5	0,1	21,4	67,5	0,0	5,9	0,0
Kastamonu	2,8	0,3	21,6	72,6	0,1	2,7	0,0
Kayseri	2,5	0,3	18,2	74,2	0,1	4,6	0,1
Kocaeli	4,7	0,3	13,3	75,7	0,1	5,8	0,1
Konya	10,3	0,4	16,0	66,1	0,2	6,8	0,2
Malatya	4,9	0,4	12,3	70,2	0,2	11,8	0,1
Manisa	4,3	0,2	21,4	69,5	0,1	4,3	0,1
Nevşehir	3,9	0,4	12,1	79,1	0,1	4,2	0,2
Samsun	1,0	0,1	24,3	72,4	0,0	2,1	0,0
Siirt	16,8	0,5	14,8	61,0	0,1	6,8	0,1
Trabzon	1,1	0,1	22,6	72,9	0,0	3,2	0,0
Van	21,5	0,3	14,7	57,8	0,1	5,5	0,0
Zonguldak	2,9	0,3	21,8	70,9	0,1	4,1	0,0
Genel Toplam	4,8	0,3	17,7	72,1	0,1	4,9	0,1

Çizelge 5.7'ye göre; yetersiz kod oranı, Van bölgesinde %21,5, Siirt bölgesinde %16,8 ve Erzurum bölgesinde %10,7'dir. Yetersiz kodların yüzdesel olarak en az olduğu bölgeler sırasıyla, %0,5 ile Ankara bölgesi, %0,8 ile Hatay bölgesi ve %1 ile Samsun bölgesidir. Siirt ve Van bölgesi kayıtlarında diğer bölgelere oranla daha yüksek oranda yetersiz kod belirlenmiştir.

Şüpheli kod olarak sınıflanmış kayıt oranı Kars bölgesinde %0,7 ve Denizli ve Siirt bölgelerinde %0,5'tir. Şüpheli kodların yüzdesel olarak en az olduğu bölgeler sırasıyla, %0,1 ile Adana, Ankara, Gaziantep, Hatay, Samsun ve Trabzon bölgeleridir.

Yetersiz ve şüpheli kodların toplamına bakıldığında; en başarılı bölge %0,6 ile Ankara bölgesi olurken en başarısız bölge %21,8 ile Van bölgesi olmuştur.

Bölgelere göre eşleşen kodların dağılımı Çizelge 5.8'de verilmiştir.

Çizelge 5.8. Bölgelere göre eşleşen kodlar

Bölge	Eşleşen kodlar içerisindeki pay (%)	Tüm kodlar içerisindeki pay (%)
Adana	%6,2	%91,5
Ankara	%10,0	%97,7
Antalya	%3,1	%88,2
Balıkesir	%2,3	%81,0
Bursa	%6,0	%87,4
Denizli	%5,1	%87,0
Diyarbakır	%1,1	%82,1
Edirne	%3,2	%92,9
Erzurum	%1,6	%82,6
Gaziantep	%2,5	%87,3
Hatay	%3,9	%96,0
İstanbul	%15,4	%92,2
İzmir	%6,9	%85,6
Kars	%1,4	%88,9
Kastamonu	%1,6	%94,2
Kayseri	%3,3	%92,5
Kocaeli	%4,4	%89,1
Konya	%3,1	%82,3
Malatya	%2,0	%82,7
Manisa	%3,8	%91,1
Nevşehir	%1,7	%91,3
Samsun	%3,2	%96,8
Siirt	%1,3	%75,9
Trabzon	%3,4	%95,6
Van	%1,3	%72,6
Zonguldak	%2,3	%92,8
Toplam	%100,0	%89,9
Ortalama	%3,8	%88,4

Çizelge 5.8'de verilen eşleşen kodlar içerisindeki pay dağılımında, eşleşen kodların toplamı 100 kabul edildiğinde tabloların bu dağılıma yaptığı katkı gösterilmiştir. Tüm kodlar

içerisindeki pay dağılımında ise, her bir tablodaki kayıt sayısı kendi içerisinde 100 kabul edildiğinde tabloda yer alan eşleşen kodların payları verilmiştir. Çizelgeye göre; Ankara bölgesi kodlarının %97,7'si, Samsun bölgesi kodlarının %96,8'i ve Hatay bölgesi kodlarının %96'sı eşleşen olarak sınıflanmıştır. Tüm kodlar içerisindeki payların ortalaması %88,4'tür. Van bölgesi %72,6, Siirt bölgesi %75,9 ve Balıkesir bölgesi %81 ile bu oranın altında kalan 12 bölgeden birisi olmuştur. Eşleşen kodlar içerisinde, İstanbul bölgesi kodları %15,4, Ankara bölgesi kodları %10 ve İzmir bölgesi kodları %6,9'luk kısmı oluşturmaktadır. Eşleşen kodlar içerisindeki payların ortalaması %3,8'dir. Diyarbakır bölgesi %1,1, Siirt ve Van bölgeleri %1,3 ile bu oranın altında kalan 17 bölge içerisinde yer almıştır. Siirt ve Van bölgeleri eşleşme oranları bakımından diğer bölgelerden geride kalmışlardır.

Bölgelere göre kabul kodların dağılımı Çizelge 5.9'da verilmiştir.

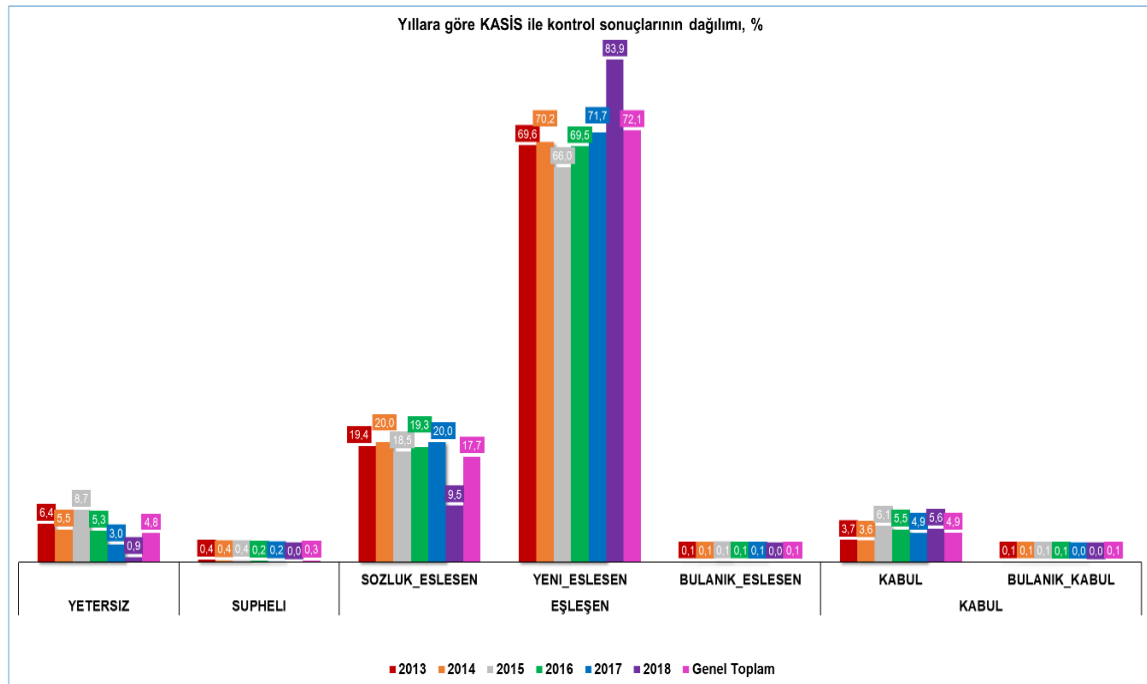
Çizelge 5.9. Bölgelere göre kabul kodlar

Bölge	Kabul kodlar içerisindeki pay (%)	Tüm kodlar içerisindeki pay (%)
Adana	%5,2	%4,3
Ankara	%3,2	%1,8
Antalya	%3,9	%6,2
Balıkesir	%5,2	%10,0
Bursa	%7,3	%6,0
Denizli	%6,0	%5,7
Diyarbakır	%2,3	%10,1
Edirne	%2,8	%4,6
Erzurum	%2,3	%6,4
Gaziantep	%3,8	%7,5
Hatay	%2,3	%3,1
İstanbul	%12,3	%4,1
İzmir	%8,8	%6,1
Kars	%1,7	%5,9
Kastamonu	%0,8	%2,8
Kayseri	%3,0	%4,7
Kocaeli	%5,3	%5,9
Konya	%4,7	%7,0
Malatya	%5,3	%12,0
Manisa	%3,3	%4,4
Nevşehir	%1,5	%4,3
Samsun	%1,3	%2,2
Siirt	%2,1	%6,9
Trabzon	%2,0	%3,2
Van	%1,8	%5,5
Zonguldak	%1,8	%4,1
Toplam	%100,0	%5,0
Ortalama	%3,8	%5,6

Çizelge 5.9’da verilen kabul kodlar içerisindeki pay dağılımında, kabul kodların toplamı 100 kabul edildiğinde tabloların bu dağılıma yaptığı katkı gösterilmiştir. Tüm kodlar içerisindeki pay dağılımında ise, her bir tablodaki kayıt sayısı kendi içerisinde 100 kabul edildiğinde tabloda yer alan kabul kodların payları verilmiştir. Çizelgeye göre; Malatya bölgesi kodlarının %12’si, Diyarbakır bölgesi kodlarının %10,1’i ve Balıkesir bölgesi kodlarının %10’u kabul olarak sınıflanmıştır. Ortalama kabul oranı %5,6’dır. Ankara bölgesi %1,8, Samsun bölgesi %2,2 ve Kastamonu bölgesi %2,8 ile bu oranın altında kalan 13 bölgeden birisi olmuştur. Kabul kodlar içerisinde, İstanbul bölgesi kodları %12,3, İzmir bölgesi kodları %8,8 ve Bursa bölgesi kodları %7,3’lük kısmı oluşturmaktadır. Kabul kodlar içerisindeki payların ortalaması %3,8’dir. Kastamonu bölgesi %0,8, Samsun %1,3 ve Nevşehir bölgeleri %1,5 ile bu oranın altında kalan 15 bölge içerisinde yer almıştır. Kabul koda düşen kayıtlar istenilen kaliteyi sağlasada kabul kodun bir adım ötesi tanımın yetersiz olarak sınıflanmasıdır. Malatya, Diyarbakır ve Balıkesir bölgeler kabul oranı bakımından diğer bölgelerin önünde yer almaktadır.

5.1.4. Anketör ile KASİS kodu arasında yıllara göre farklılık var mıdır?

KASİS programı aracılığıyla anketörün verdiği kodlar kontrol edilmiş ve yıllara göre elde edilen sonuçlara Şekil 5.2’de yer verilmiştir.



Şekil 5.2. Yıllara göre KASİS ile kontrol sonuçlarının dağılımı

Şekil 5.2'ye göre; yetersiz kod oranı, 2015 yılında %8,7, 2013 yılında %6,4 ve 2014 yılında %5,5'tir. Yetersiz kodların yüzdesel olarak en az olduğu yıllar sırasıyla, %0,9 ile 2018 yılı, %3 ile 2017 yılı ve %5,3 ile 2016 yılıdır.

Şüpheli kod olarak sınıflanmış kayıt oranı 2013, 2014 ve 2015 yıllarında %0,4 ve 2016 ve 2017 yıllarında %0,2 ve 2018 yılında %0'dır. Yıllara göre şüpheli kod dağılımına bakıldığında son üç yılın kayıtlarının daha iyi kodlandığı söylenebilir.

Yetersiz ve şüpheli kodların toplamına bakıldığında; en başarılı yıl %0,9 ile 2018 yılı olurken en başarısız yıl %9,1 ile 2015 yılı olmuştur.

Yıllara göre eşleşen kodların dağılımı Çizelge 5.10'da verilmiştir.

Çizelge 5.10. Yıllara göre eşleşen kodlar

Yıl	Eşleşen kodlar içerisindeki pay (%)	Tüm kodlar içerisindeki pay (%)
2013	%15,2	%89,3
2014	%15,5	%90,4
2015	%15,3	%84,7
2016	%17,1	%88,9
2017	%17,9	%91,9
2018	%18,9	%93,5
Toplam	%100,0	%89,9
Ortalama	%16,7	%89,8

Çizelge 5.10'da verilen eşleşen kodlar içerisindeki pay dağılımında, eşleşen kodların toplamı 100 kabul edildiğinde tabloların bu dağılıma yaptığı katkı gösterilmiştir. Tüm kodlar içerisindeki pay dağılımında ise, her bir tablodaki kayıt sayısı kendi içerisinde 100 kabul edildiğinde tabloda yer alan eşleşen kodların payları verilmiştir. Çizelgeye göre; 2018 yılı kodlarının %93,5'i, 2017 yılı kodlarının %91,9'u ve 2014 yılı kodlarının %90,4'ü eşleşen olarak sınıflanmıştır. Tüm kodlar içerisindeki payların ortalaması %89,9'dur. 2015 yılı %84,7 ile 2016 yılı %88,9 ve 2013 yılı %89,3 ile ortalama eşleşme oranının altında yer almıştır. Eşleşen kodlar içerisinde, 2018 yılı kodları %18,9, 2017 yılı kodları %17,9 ve 2016 yılı kodları %17,1'lik kısmı oluşturmaktadır. Eşleşen kodlar içerisindeki payların ortalaması %16,7'dir. 2013 yılı %15,2, 2015 yılı %15,3 ve 2014 yılı %15,5 ile bu oranın altında yer almıştır.

Yıllara göre kabul kodların dağılımı Çizelge 5.11’de verilmiştir.

Çizelge 5.11. Yıllara göre kabul kodlar

Yıl	Kabul kodlar içerisindeki pay (%)	Tüm kodlar içerisindeki pay (%)
2013	%11,7	%3,9
2014	%11,5	%3,8
2015	%20,0	%6,2
2016	%19,1	%5,5
2017	%17,3	%4,9
2018	%20,3	%5,6
Toplam	%100,0	%5,0
Ortalama	%16,7	%5,0

Çizelge 5.11’de verilen kabul kodlar içerisindeki pay dağılımında, kabul kodların toplamı 100 kabul edildiğinde tabloların bu dağılıma yaptığı katkı gösterilmiştir. Tüm kodlar içerisindeki pay dağılımında ise, her bir tablodaki kayıt sayısı kendi içerisinde 100 kabul edildiğinde tabloda yer alan kabul kodların payları verilmiştir. Çizelgeye göre; 2015 yılı kodlarının %6,2’si, 2018 yılı kodlarının %5,6’sı ve 2016 yılı kodlarının %5,5’i kabul olarak sınıflanmıştır. Tüm kodlar içerisindeki payların ortalaması %5’tir. 2014 yılı %3,8 ile 2013 yılı %3,9 ve 2017 yılı %4,9 ile bu oranın altında yer almıştır. Kabul kodlar içerisinde, 2018 yılı kodları %20,3, 2015 yılı kodları %20, 2016 yılı kodları %19,1 ve 2017 yılı kodları %17,3’lük kısmı oluşturmaktadır. Kabul kodlar içerisindeki payların ortalaması %16,7’dir. 2014 yılı %11,5, 2013 yılı %11,7 ile bu oranın altında yer almıştır.

5.1.5. Anketör ile KASİS kodu arasında aylara göre farklılık var mıdır?

KASİS programı aracılığıyla anketörün verdiği kodlar kontrol edilmiş ve aylara göre elde edilen sonuçlara Çizelge 5.12’de yer verilmiştir.

Çizelge 5.12. Aylara göre KASİS ile kontrol sonuçlarının dağılımı

Ay	YETERSİZ	SUPHELI	SOZLUK ESLESEN	EŞLEŞEN			KABUL	
				YENİ ESLESEN	BULANIK ESLESEN		KABUL	BULANIK KABUL
Ocak	5,4	0,3	17,8	71,6	0,1	5,2	0,1	0,1
Şubat	5,3	0,3	17,2	71,9	0,1	5,1	0,1	0,1
Mart	5,2	0,3	16,6	72,7	0,1	5,1	0,1	0,1
Nisan	5,3	0,3	16,4	72,9	0,1	5,0	0,1	0,1
Mayıs	5,1	0,3	16,6	72,7	0,1	5,3	0,1	0,1
Haziran	4,8	0,3	17,4	72,2	0,1	5,1	0,1	0,1
Temmuz	4,8	0,3	18,1	71,9	0,1	4,8	0,1	0,1
Ağustos	4,9	0,3	18,4	71,6	0,1	4,7	0,1	0,1
Eylül	4,6	0,2	18,4	72,0	0,1	4,6	0,1	0,1
Ekim	4,5	0,2	18,4	72,0	0,1	4,7	0,1	0,1
Kasım	4,3	0,2	18,3	72,1	0,1	4,9	0,1	0,1
Aralık	4,0	0,2	18,5	72,2	0,1	4,9	0,1	0,1
Genel Toplam	4,8	0,3	17,7	72,1	0,1	4,9	0,1	0,1

Çizelge 5.12'e göre; yetersiz kod oranı, Ocak ayında %5,4, Şubat ve Nisan ayında %5,3 ve Mart ayında %5,2'dir. Yetersiz kodların yüzdesel olarak en az olduğu aylar sırasıyla, %4 ile Aralık ayı, %4,3 ile Kasım ayı ve %4,5 ile Ekim ayıdır. Şüpheli kod olarak sınıflanmış kayıt oranı Ocak ve Ağustos arasındaki aylarda %0,3 ve diğer aylarda %0,2'dir. Aylara göre şüpheli kod dağılımı bakıldığında yılın ikinci yarısında kod atamaların daha fazla hatalı yapıldığı söylenebilir. Yetersiz ve şüpheli kodların toplamına bakıldığında en başarılı ay %4,3 ile Aralık ayı olurken en başarısız ay %5,8 ile Ocak ayıdır.

Aylara göre eşleşen kodların dağılımı Çizelge 5.13'te verilmiştir.

Çizelge 5.13. Aylara göre eşleşen kodlar

Ay	Eşleşen kodlar içerisindeki pay (%)	Tüm kodlar içerisindeki pay (%)
Ocak	%8,2	%89,0
Şubat	%8,0	%89,2
Mart	%8,8	%89,4
Nisan	%8,6	%89,4
Mayıs	%8,6	%89,3
Haziran	%8,1	%89,7
Temmuz	%7,9	%90,1
Ağustos	%7,7	%90,1
Eylül	%8,0	%90,5
Ekim	%8,6	%90,5
Kasım	%8,7	%90,5
Aralık	%8,8	%90,8
Toplam	%100,0	%89,9
Ortalama	%8,3	%89,9

Çizelge 5.13'te verilen eşleşen kodlar içerisindeki pay dağılımında, eşleşen kodların toplamı 100 kabul edildiğinde tabloların bu dağılıma yaptığı katkı gösterilmiştir. Tüm kodlar içerisindeki pay dağılımında ise, her bir tablodaki kayıt sayısı kendi içerisinde 100 kabul edildiğinde tabloda yer alan eşleşen kodların payları verilmiştir. Çizelgeye göre; Aralık ayı kodlarının %90,8'i, Eylül, Ekim ve Kasım ayı kodlarının %90,5'i eşleşen olarak sınıflanmıştır. Tüm kodlar içerisindeki payların ortalaması %89,9'dur. Ocak ile Haziran arasındaki aylar bu oranın altında yer almıştır. Eşleşen kodlar içerisinde, Mart ve Aralık ayı kodları %8,8, Kasım ayı kodları %8,7 ve Nisan, Mayıs ve Ekim ayı kodları %8,6'lık kısmı oluşturmaktadır. Eşleşen kodlar içerisindeki payların ortalaması %8,3'tür. Ocak, Şubat ve Haziran ile Eylül arasındaki aylar bu oranın altında yer almıştır.

Aylara göre kabul kodların dağılımı Çizelge 5.14'te verilmiştir.

Çizelge 5.14. Aylara göre kabul kodlar

Ay	Kabul kodlar içerisindeki pay (%)	Tüm kodlar içerisindeki pay (%)
Ocak	%8,8	%5,3
Şubat	%8,3	%5,2
Mart	%9,1	%5,2
Nisan	%8,7	%5,1
Mayıs	%9,2	%5,3
Haziran	%8,4	%5,2
Temmuz	%7,7	%4,9
Ağustos	%7,3	%4,8
Eylül	%7,4	%4,7
Ekim	%8,1	%4,8
Kasım	%8,5	%5,0
Aralık	%8,5	%4,9
Toplam	%100,0	%5,0
Ortalama	%8,3	%5,0

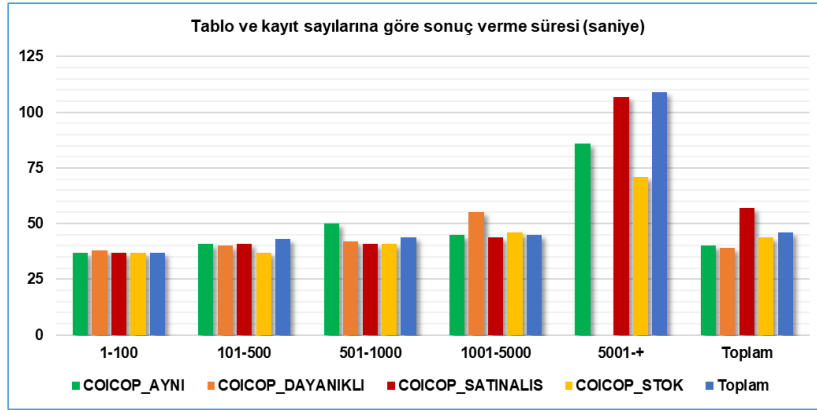
Çizelge 5.14'te verilen kabul kodlar içerisindeki pay dağılımında, kabul kodların toplamı 100 kabul edildiğinde tabloların bu dağılıma yaptığı katkı gösterilmiştir. Tüm kodlar içerisindeki pay dağılımında ise, her bir tablodaki kayıt sayısı kendi içerisinde 100 kabul edildiğinde tabloda yer alan kabul kodların payları verilmiştir. Çizelgeye göre; Ocak ve Mayıs ayı kodlarının %5,3'ü, Şubat, Mart ve Haziran ayı kodlarının %5,2'si kabul olarak sınıflanmıştır. Tüm kodlar içerisindeki payların ortalaması %5'tir. Kasım ayı hariç Temmuz ile Aralık arasındaki aylar bu oranın altında yer almıştır. Kabul kodlar içerisinde, Mayıs ayı kodları %9,2, Ocak ayı kodları %8,8 ve Nisan ayı kodları %8,7'lik kısmı oluşturmaktadır. Kabul kodlar içerisindeki payların ortalaması %8,3'tür. Temmuz ile Ekim arasındaki aylar bu oranın altında yer almıştır.

5.1.6. KASİS sonuç verme süresi tablo ve kayıt sayısına göre farklı mıdır?

Program çalışma performansının önceden tahmin edilmesi, hem kullanıcılar hem de kaynak yönetimi açısından önemlidir. Bu yüzden, kullanıcılar ve uygulama geliştiricileri, programın çalıştıktan sonra sonuç vermesi için ne kadar süre gerektiğini önceden bilmek isterler [96]. Çalışma performansı, kullanılan işletim sistemi, işlemcinin kapasitesi ve kontrol edilecek kayıt sayısı ile doğrudan ilişkilidir. Toplam sonuç verme süresi yanında kayıt başına düşen sonuç verme süresinin değerlendirme yapmak için önemli bir gösterge olduğu düşünülmektedir. Bu nedenle grafiklerde iki gösterge bir arada verilmektedir.

Sonuç verme süreleri, KASİS'in parametreleri kullanılarak olası tüm kombinasyonlara göre sistemin çalıştırılması sonucunda hesaplanmıştır.

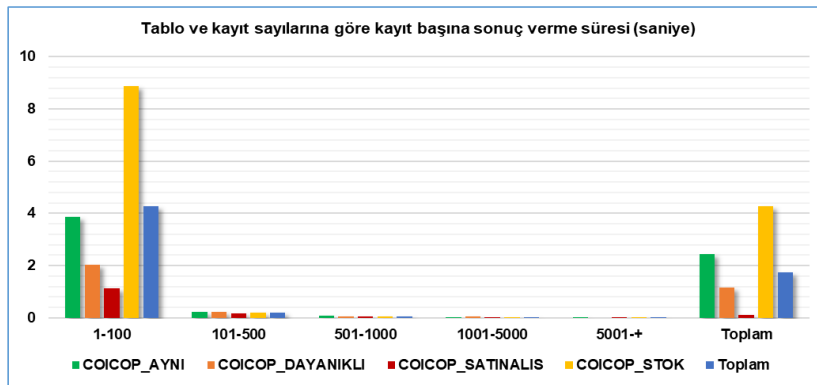
KASİS'in tablo ve kayıt sayılarına göre sonuç verme süresi Şekil 5.3'te verilmiştir.



Şekil 5.3. Tablo ve kayıt sayılarına göre sonuç verme süresi

Şekil 5.3'e göre; aynı tablosu için ortalama sonuç verme süresi 40 saniye, dayanıklı tablosu için ortalama sonuç verme süresi 39 saniye, satın alış tablosu için ortalama sonuç verme süresi 57 saniye ve stok tablosu için ortalama sonuç verme süresi 44 saniye olurken genel ortalama sonuç verme süresi 46 saniye olmuştur. Kayıt sayısı arttıkça sonuç verme süresi de artmaktadır. 5001 ve üzeri kaydın yer aldığı çalıştırmalarda sonuç verme süresi tüm tablolarda diğer kayıt sayılarına göre belirgin biçimde artmaktadır.

Tablo ve kayıt sayılarına göre kayıt başına ortalama sonuç verme süresi Şekil 5.4'te verilmiştir.

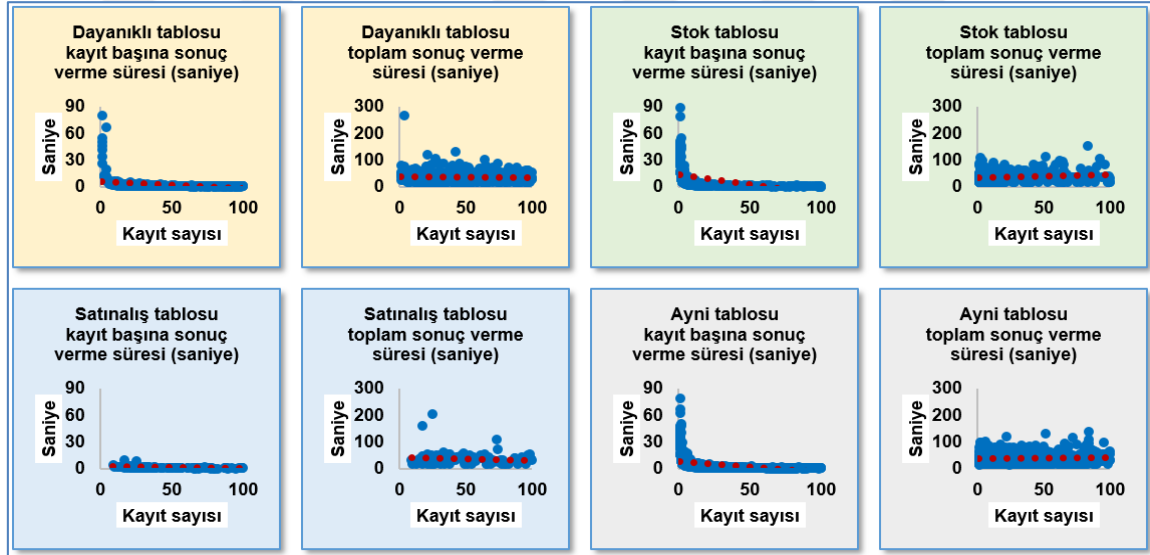


Şekil 5.4. Tablolara ve kayıt başına göre ortalama sonuç verme süresi

Şekil 5.4'te; aynı tablosu için kayıt başına ortalama sonuç verme süresi 2,4 saniye, dayanıklı tablosu için ortalama sonuç verme süresi 1,2 saniye, satın alış tablosu için ortalama sonuç verme süresi 0,1 saniye ve stok tablosu için ortalama sonuç verme süresi 4,3 saniye olurken kayıt başına genel ortalama sonuç verme süresi 1,8 saniye olmuştur. Kayıt sayısı arttıkça kayıt başına ortalama sonuç verme süresi düşmektedir.

Kayıt sayısı 1 - 100 arasında olan çalışmaların sonuç verme süresi

Kayıt sayısı 1 - 100 arasında olan çalışmaların tablo ayrımı temelinde kayıt başına düşen ve toplam sonuç verme süresi Şekil 5.5'te verilmiştir.

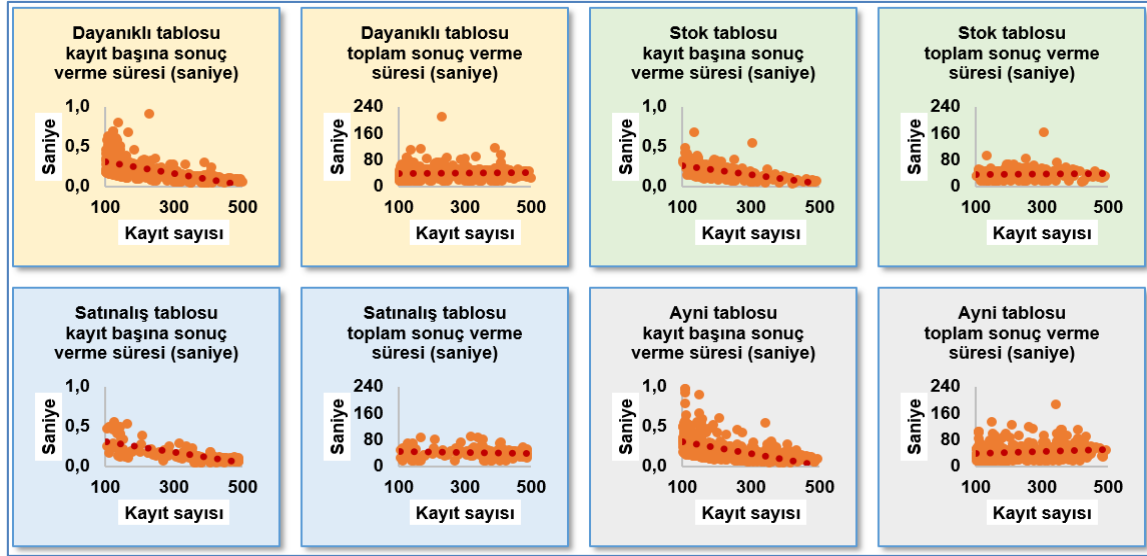


Şekil 5.5. Kayıt sayısı 1 - 100 arasında olan çalışmaların sonuç verme süresi

Şekil 5.5 incelendiğinde; kayıt sayısı arttıkça, tüm tablolarda genel sonuç verme süresinin aynı kalma eğiliminde olduğu ve kayıt başına ortalama sonuç verme süresinin azalma eğiliminde olduğu görülmektedir.

Kayıt sayısı 101 - 500 arasında olan çalışmaların sonuç verme süresi

Kayıt sayısı 101 - 500 arasında olan çalışmaların tablo ayrımı temelinde kayıt başına düşen ve toplam sonuç verme süresi Şekil 5.6'da verilmiştir.

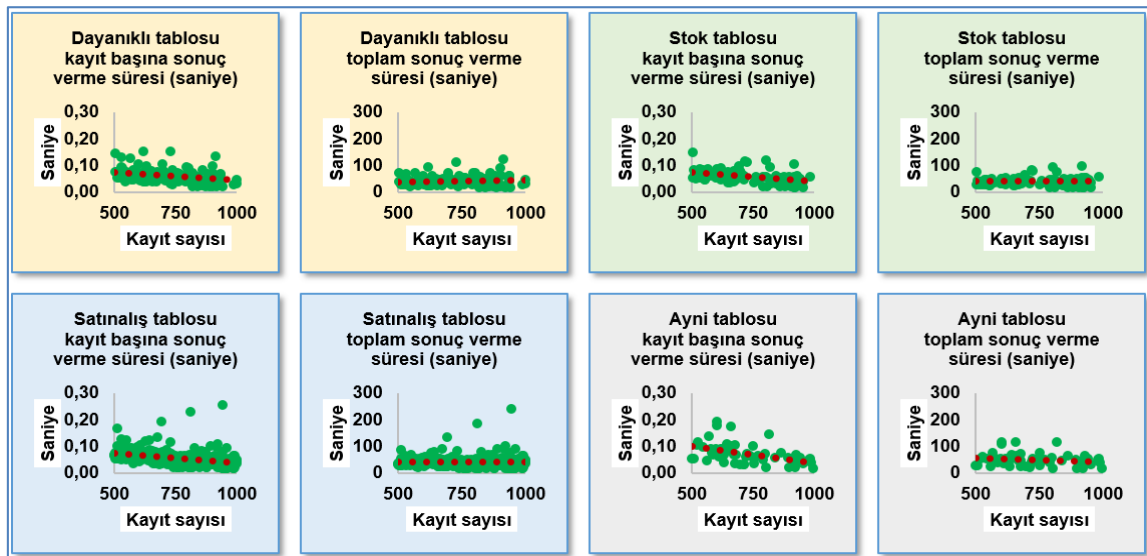


Şekil 5.6. Kayıt sayısı 101 - 500 arasında olan çalışmaların sonuç verme süresi

Şekil 5.6 incelendiğinde; kayıt sayısı arttıkça, genel sonuç verme süresinin aynı tablosu hariç aynı kalma eğiliminde olduğu ve kayıt başına ortalama sonuç verme süresinin azalma eğiliminde olduğu görülmektedir.

Kayıt sayısı 501 – 1 000 arasında olan çalışmaların sonuç verme süresi

Kayıt sayısı 501 – 1 000 arasında olan çalışmaların tablo ayrımı temelinde kayıt başına düşen ve toplam sonuç verme süresi Şekil 5.7’de verilmiştir.

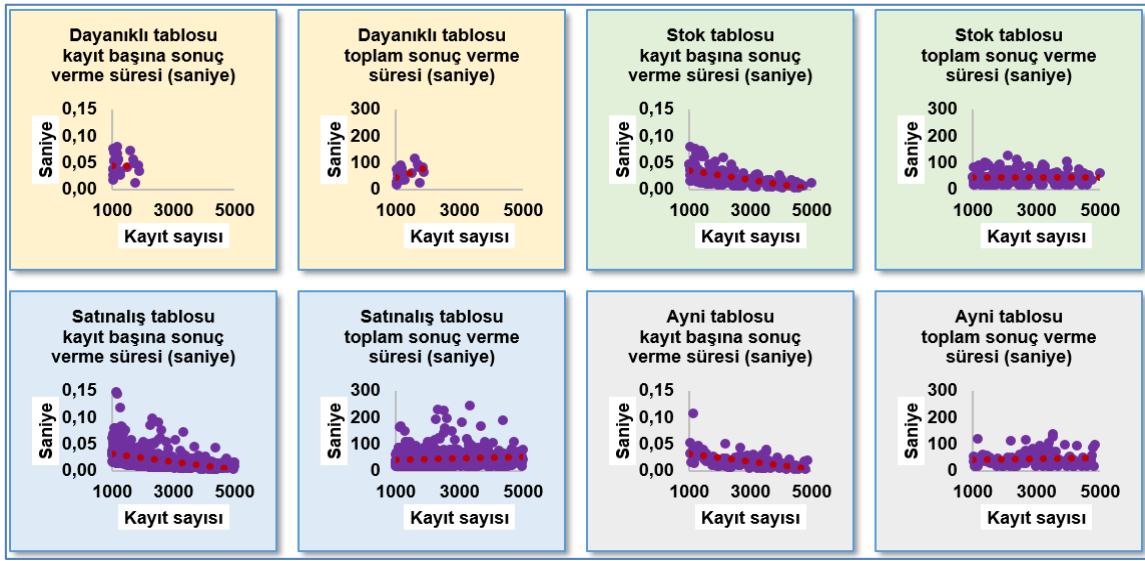


Şekil 5.7. Kayıt sayısı 501 – 1 000 arasında olan çalışmaların sonuç verme süresi

Şekil 5.7 incelendiğinde; kayıt sayısı arttıkça, genel sonuç verme süresinin aynı tablosu hariç aynı kalma eğiliminde olduğu ve kayıt başına ortalama sonuç verme süresinin azalma eğiliminde olduğu görülmektedir.

Kayıt sayısı 1 001 – 5 000 arasında olan çalışmaların sonuç verme süresi

Kayıt sayısı 1 001 – 5 000 arasında olan çalışmaların tablo ayrımı temelinde kayıt başına düşen ve toplam sonuç verme süresi Şekil 5.8’de verilmiştir.

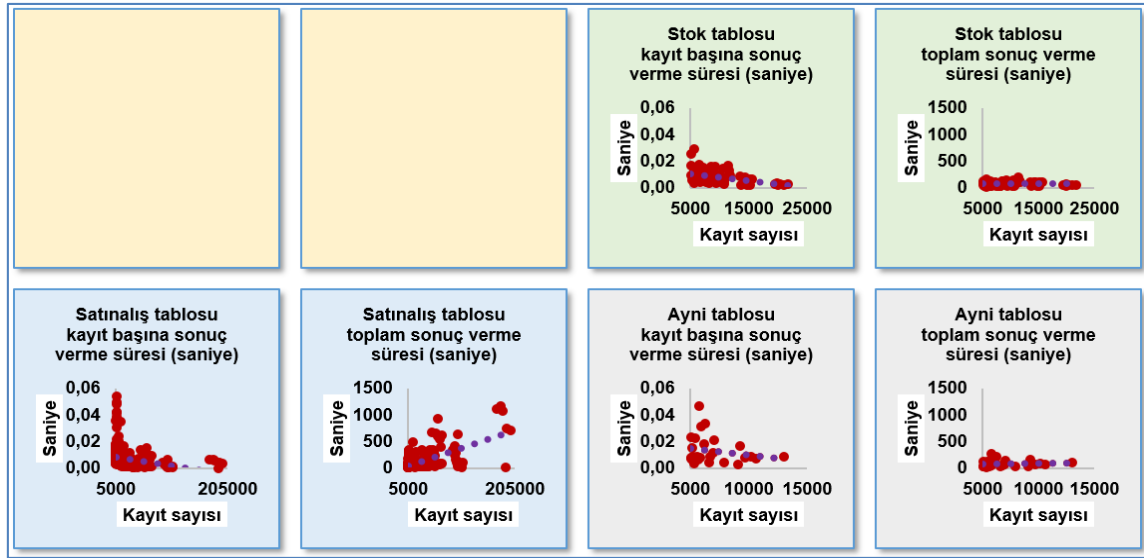


Şekil 5.8. Kayıt sayısı 1 001 – 5 000 arasında olan çalışmaların sonuç verme süresi

Şekil 5.8 incelendiğinde; kayıt sayısı arttıkça, genel sonuç verme süresinin dayanıklı tablosu hariç aynı kalma eğiliminde olduğu ve kayıt başına ortalama sonuç verme süresinin dayanıklı tablosu hariç azalma eğiliminde olduğu görülmektedir.

Kayıt sayısı 5 001 ve üzeri olan çalışmaların sonuç verme süresi

Kayıt sayısı 5 001 ve üzeri olan çalışmaların tablo ayrımı temelinde kayıt başına düşen ve toplam sonuç verme süresi Şekil 5.9’da verilmiştir. Dayanıklı tablosunda 5 001 ve üzeri kayıt bulunmadığı için grafik alanı boş bırakılmıştır.



Şekil 5.9. Kayıt sayısı 5 001 ve üzeri olan çalışmaların sonuç verme süresi

Şekil 5.9 incelendiğinde; kayıt sayısı arttıkça, genel sonuç verme süresinin satın alıs tablosu hariç aynı kalma eğiliminde olduğu ve kayıt başına ortalama sonuç verme süresinin azalma eğiliminde olduğu görülmektedir.

5.2. Kod Atama Katmanında Elde Edilen Bulgular

Anketör tarafından kod ataması yapılmış kayıtlar KASİS aracılığıyla kontrol edilmiş ve kayıtların eşleşen, yetersiz ve şüpheli olarak sınıflamaları yapılmıştır. Bu bölümde, sınıflaması yetersiz ve şüpheli olarak yapılmış kayıtlara yönelik anketörün yaptığı kod atamasının doğruluğunu teyit edebilmek veya yeni bir kod önerebilmek amacıyla anketörün yazdığı tanımdan yola çıkarak yeniden kod ataması yapılmıştır.

Kod ataması yapılırken iki ayrı yöntem kullanılmıştır. Birinci yöntem, daraltılmış listeden kod ataması yapmaktır. Bu yöntemde, sınıflama katmanında yetersiz veya şüpheli olarak sınıflanmış kayıtlar için eşleştirme katmanında kaç tane alternatif kodun ve tanımın karşılık geldiği ve bunların neler olduğu sistem tarafından bilindiği için Bölüm 2.6.2’de yer alan bulanık eşleştirme algoritmaları eşleştirme katmanında belirlenen bu kod ve tanımlar ile sınırlı kalmak koşuluyla uygulanmıştır. İkinci yöntem, geniş listeden kod ataması yapmaktır. Bu yöntemde, ilk yöntemin aksine bulanık eşleştirme algoritmaları, listede daraltma uygulanmadan sözlükte bulunan tüm kod ve tanımlar için uygulanmıştır.

İki farklı yöntemle göre yapılan KASİS kod atama sonuçlarına Çizelge 5.15'te yer verilmiştir.

Çizelge 5.15. İki farklı yöntemle göre yapılan KASİS kod atama sonuçları

Açıklama	Daraltılmış listeden kod atama	%	Geniş listeden kod atama	%
Kod ataması yapılan kayıt sayısı	647 256	100,0	647 256	100,0
Anketörün atadığı 10 basamaklı kodun doğru olduğu kayıt sayısı	163 358	25,2	98 732	15,3
Anketörün atadığı 10 basamaklı kodun yanlış olduğu kayıt sayısı	483 898	74,8	548 524	84,7
KASİS tarafından 1 kod önerisi yapılan kayıt sayısı	401 722	62,1	273 857	42,3
KASİS'in önerdiği kodun ilk beş basamağı ile anketörün atadığı kodun ilk beş basamağı aynı	231 766	35,8	149 025	23,0
KASİS'in önerdiği kodun ilk beş basamağı ile anketörün atadığı kodun ilk beş basamağı farklı	169 956	26,3	124 832	19,3
KASİS tarafından 3 kod önerisi yapılan kayıt sayısı	62 249	9,6	105 684	16,3
KASİS tarafından 4 kod önerisi yapılan kayıt sayısı	18 583	2,9	149 217	23,1
KASİS tarafından 5 kod önerisi yapılan kayıt sayısı	1 344	0,2	19 766	3,1

Çizelge 5.15'e göre; daraltılmış listeden kod atama yönteminde, kayıtların %25,2'sinde, KASİS ile anketör aynı sonuca ulaşmıştır. Bu oran, geniş listeden kod atama yönteminde %15,3'te kalmıştır. Bu kayıtlarda anketörün atadığı 10 basamaklı COICOP kodlarının doğru olduğu sonucuna ulaşılmıştır. Anketörün atadığı 10 basamaklı COICOP kodlarının hatalı olduğu kayıtların oranı, ilk yöntem için %74,8 iken ikinci yöntem için %84,7'dir. Bu kayıtlar için, KASİS'in tek kod önerdiği kayıt oranı ilk yöntemde %62,1 iken ikinci yöntemde bu oran %42,3'te kalmıştır.

Anketörün atadığı 10 basamaklı kodun hatalı olduğu kayıtlar incelendiğinde; ilk yöntemde kayıtların %35,8'inde KASİS'in önerdiği kodun ilk beş basamağı ile anketörün atadığı kodun ilk beş basamağının aynı olduğu görülmüştür. İkinci yöntemde, bu oran %23 seviyesinde kalmıştır. Bu kayıtlarda anketörün atadığı kodun doğru olarak kabul edilebileceği sonucuna ulaşılmıştır. Bu sonuçlar neticesinde, anketörün atadığı kodun doğruluğu sistem tarafından teyit edilen kayıtların oranı ilk yöntem için %61 (%25,2+%35,8), ikinci yöntem için %38,3 (%15,3+%23) olmuştur.

Sistemin önerdiği kodun ilk beş basamağı ile anketörün atadığı kodun ilk beş basamağının farklı olduğu kayıt sayısı ilk yöntemde 169 956 ve ikinci yöntemde 124 832'dir. Bu kayıtlarda KASİS'in önerdiği kodun kullanılması önerilmektedir. İlk yöntemde kayıtların %12,7'sine ve ikinci yöntemde %42,4'üne sistem tarafından üç, dört veya beş farklı kod önerisi yapılmıştır. Bu kayıtlarda, KASİS'in beş farklı bulanık eşleştirme algoritması ile

önerdiği kodlara bakılarak anketör kod atamasının doğruluğunun uzman görüşü ile onaylanması veya KASİS tarafından önerilen kodlardan birisinin seçilmesi gerekmektedir. KASİS ile kontrol edilen 2013-2018 yıllarındaki toplam kayıt sayısı 12 676 591'dir. İlk yöntemde bu kayıtların %06'sında ve ikinci yöntemde %2,2'sinde uzman görüşünün devreye girmesi gerektiği belirlenmiştir.

5.2.1 İki yöntem arasında tablolara göre farklılık var mıdır?

Tablolara göre anketör atadığı ile sistemin önerdiği kodların dağılımı Çizelge 5.16'da verilmiştir.

Çizelge 5.16. Tablolara göre sistemin önerdiği kodların dağılımı

Ana grup	Daraltılmış listeden kod atama yöntemi		Geniş listeden kod atama yöntemi	
	Anketörün atadığı ile sistemin önerdiği kod aynı	Anketörün atadığı ile sistemin önerdiği kod farklı	Anketörün atadığı ile sistemin önerdiği kod aynı	Anketörün atadığı ile sistemin önerdiği kod farklı
COICOP_AYNI	24,7	75,3	15,4	84,6
COICOP_DAYANIKLI	16,9	83,1	11,3	88,7
COICOP_SATINALIS	26,7	73,3	15,9	84,1
COICOP_STOK	11,4	88,6	8,2	91,8

Çizelge 5.16'daki yüzdesel dağılımlara bakıldığında; daraltılmış listeden kod atama yöntemi lehine yöntemler arasında farklılık bulunmaktadır. Daraltılmış listeden kod atama yönteminde anketörün atadığı kod ile uyum oranı en yüksek tablo %26,7 ile satın alış tablosu olmuştur. Uyum oranı bakımından ikinci sırada %24,7 ile aynı tablosu, üçüncü sırada %16,9 ile dayanıklı tablosu ve dördüncü sırada %11,4 ile stok tablosu kayıtları yer almaktadır. Geniş listeden kod atama yönteminde, başarılı tablo sıralaması değişmemiştir. %15,9 ile satın alış tablosu birinci, %15,4 ile aynı tablosu ikinci, %11,3 ile dayanıklı tablosu üçüncü ve %8,2 ile stok tablosu kayıtları dördüncü olmuştur.

Tablolara göre anketörün atadığı kodun hatalı olduğu kayıtlara sistemin önerdiği kod sayılarına göre dağılımı Çizelge 5.17'de verilmiştir.

Çizelge 5.17. Tablolara göre sistemin önerdiği kod sayılarına göre dağılımı

Tablo	Daraltılmış listeden kod atama yöntemi				Geniş listeden kod atama yöntemi			
	KASIS tarafından 1 kod önerisi yapılan kayıt	KASIS tarafından 3 kod önerisi yapılan kayıt	KASIS tarafından 4 kod önerisi yapılan kayıt	KASIS tarafından 5 kod önerisi yapılan kayıt	KASIS tarafından 1 kod önerisi yapılan kayıt	KASIS tarafından 3 kod önerisi yapılan kayıt	KASIS tarafından 4 kod önerisi yapılan kayıt	KASIS tarafından 5 kod önerisi yapılan kayıt
COICOP_AYNI	83,4	12,7	3,6	0,3	50,9	17,4	27,9	3,8
COICOP_DAYANIKLI	86,2	9,3	4,2	0,2	51,8	21,8	21,5	4,9
COICOP_SATINALIS	83,0	12,8	3,9	0,3	49,8	19,2	27,5	3,5
COICOP_STOK	81,0	15,0	3,7	0,4	50,0	20,6	25,8	3,6

Çizelge 5.17'ye göre, daraltılmış listeden kod atama yönteminde bir adet kod önerisi yapılan oransal olarak en yüksek tablo %86,2 ile dayanıklı tablosu olmuştur. %83,4 ile aynı ikinci, %83 ile satın alış üçüncü ve %81 ile stok tablosu son sırada yer almıştır. Geniş listeden kod atama yönteminde, %51,8 ile dayanıklı birinci, %50,9 ile aynı ikinci, %50 ile stok üçüncü ve %49,8 ile satın alış dördüncü tablo olmuştur. Sistemin üç farklı kod önerisi yaptığı yüzdesel dağılımlara bakıldığında; daraltılmış listeden kod atama yöntemi lehine yöntemler arasında farklılık bulunmaktadır. Aynı farklılık, sistemin dört farklı ve beş farklı kod önerisi yaptığı tablolarda da bulunmaktadır.

Tablolara göre sistemin önerdiği kod ile anketörün atadığı kodun ilk beş basamağına göre dağılımı Çizelge 5.18'de yer verilmiştir. Sistemin önerdiği kodun ilk beş basamağı ile anketörün atadığı kodun ilk beş basamağının uyum oranının yüksek olması anketörün atadığı kodun doğru olarak kabul edilmesini sağlamaktadır.

Çizelge 5.18. Tablolara göre kodların ilk beş basamağına göre dağılımı

Tablo	Daraltılmış listeden kod atama yöntemi		Geniş listeden kod atama yöntemi	
	KASIS'in önerdiği kodun ilk beş basamağı ile anketörün atadığı kodun ilk beş basamağı aynı	KASIS'in önerdiği kodun ilk beş basamağı ile anketörün atadığı kodun ilk beş basamağı farklı	KASIS'in önerdiği kodun ilk beş basamağı ile anketörün atadığı kodun ilk beş basamağı aynı	KASIS'in önerdiği kodun ilk beş basamağı ile anketörün atadığı kodun ilk beş basamağı farklı
COICOP_AYNI	48,4	51,6	44,3	55,7
COICOP_DAYANIKLI	69,6	30,4	60,9	39,1
COICOP_SATINALIS	56,1	43,9	53,6	46,4
COICOP_STOK	77,2	22,8	70,5	29,5

Çizelge 5.18'e göre, daraltılmış listeden kod atama yönteminde sistemin önerdiği kodun ilk beş basamağı ile anketörün atadığı kodun ilk beş basamağının aynı olduğu en yüksek tablo %77,2 ile stok tablosu olmuştur. İkinci tablo %69,6 ile dayanıklı, üçüncü tablo %56,1 ile

satın alış ve son tablo %48,4 ile aynı tablosu olmuştur. Geniş listeden kod atama yönteminde, %70,5 ile stok tablosu birinci, %60,9 ile dayanıklı tablosu ikinci, %53,6 ile satın alış üçüncü ve %44,3 ile aynı tablosu dördüncü tablo olmuştur.

Tablolara göre yetersiz ve şüpheli kodlar içerisinde kod atama sonuçları Çizelge 5.19’da verilmiştir.

Çizelge 5.19. Tablolara göre kod atama sonuçları 1

Tablo	Daraltılmış listeden kod atama yöntemi				Geniş listeden kod atama yöntemi			
	Doğrulanamadı - Hatalı Kodlama	Doğrulanamadı - Uzman Görüşü ile Seçim Gerekli	Doğrulandı - Onlu Kod Aynı	Doğrulandı - Beşli Kod Aynı	Doğrulanamadı - Hatalı Kodlama	Doğrulanamadı - Uzman Görüşü ile Seçim Gerekli	Doğrulandı - Onlu Kod Aynı	Doğrulandı - Beşli Kod Aynı
COICOP_AYNI	32,4	12,5	24,7	30,4	24,0	41,5	15,4	19,1
COICOP_DAYANIKLI	21,8	11,5	16,9	49,9	18,0	42,8	11,3	28,0
COICOP_SATINALIS	26,7	12,4	26,7	34,2	19,4	42,2	15,9	22,4
COICOP_STOK	16,4	16,9	11,4	55,4	13,5	45,9	8,2	32,3

Çizelge 5.19 incelendiğinde; daraltılmış listeden kod atama yöntemine göre stok tablosunda şüpheli ve yetersiz kodların %16,9’luk kısmı için uzman görüşünün gerekli olduğu sonucuna ulaşılmıştır. Bu oran; aynı tablosunda %12,5, satın alış tablosunda %12,4 ve dayanıklı tablosunda %11,5 olmuştur. Bu kayıtlar için, KASİS’in beş farklı bulanık eşleştirme algoritması ile önerdiği kodlara bakılarak uzman görüşü ile anketör kod atamasının doğruluğunun onaylanması veya KASİS tarafından önerilen kodlardan birisinin seçilmesi gerektiği belirlenmiştir.

Tablolara göre tüm kodlar içerisinde kod atama sonuçları Çizelge 5.20’de verilmiştir.

Çizelge 5.20. Tablolara göre kod atama sonuçları 2

Tablo	Daraltılmış listeden kod atama yöntemi				Geniş listeden kod atama yöntemi			
	Doğrulanamadı - Hatalı Kodlama	Doğrulanamadı - Uzman Görüşü ile Seçim Gerekli	Doğrulandı - Onlu Kod Aynı	Doğrulandı - Beşli Kod Aynı	Doğrulanamadı - Hatalı Kodlama	Doğrulanamadı - Uzman Görüşü ile Seçim Gerekli	Doğrulandı - Onlu Kod Aynı	Doğrulandı - Beşli Kod Aynı
COICOP_AYNI	1,8	0,7	1,4	1,7	1,3	2,3	0,9	1,1
COICOP_DAYANIKLI	1,9	1,0	1,5	4,3	1,6	3,7	1,0	2,4
COICOP_SATINALIS	1,4	0,6	1,4	1,8	1,0	2,2	0,8	1,2
COICOP_STOK	0,6	0,6	0,4	1,9	0,5	1,6	0,3	1,1

Çizelge 5.20’de daraltılmış listeden kod atama yöntemine göre, dayanıklı tablosunda tüm kayıtların %1’lik kısmı için uzman görüşüne başvurulması gerektiği sonucuna ulaşılmıştır. Bu oran; aynı tablosunda %0,7, satın alış ve stok tablolarında %0,6 olmuştur. Geniş listeden

kod atama yönteminin seçilmesi durumunda ise, ilk yöntemden biraz daha fazla kayıтта uzman görüşüne başvurulması gerektiği ortaya çıkmıştır.

5.2.2. İki yöntem arasında ana gruplara göre farklılık var mıdır?

Ana gruplara göre anketör atadığı ile sistemin önerdiği kodların dağılımı Çizelge 5.21’de verilmiştir.

Çizelge 5.21. Ana gruplara göre sistemin önerdiği kodların dağılımı

Ana grup	Daraltılmış listeden kod atama yöntemi		Geniş listeden kod atama yöntemi	
	Anketörün atadığı ile sistemin önerdiği kod aynı	Anketörün atadığı ile sistemin önerdiği kod farklı	Anketörün atadığı ile sistemin önerdiği kod aynı	Anketörün atadığı ile sistemin önerdiği kod farklı
Gıda ve alkolsüz içecekler	28,3	71,7	18,1	81,9
Alkollü içecek, sigara ve tütün	3,4	96,6	3,7	96,3
Giyim ve ayakkabı	15,6	84,4	6,1	93,9
Konut ve kira	43,7	56,3	37,5	62,5
Mobilya ve ev eşyaları	17,8	82,2	9,1	90,9
Sağlık	16,1	83,9	5,8	94,2
Ulaştırma	20,5	79,5	12,7	87,3
Haberleşme	62,4	37,6	57,6	42,4
Eğlence ve kültür	13,8	86,2	7,2	92,8
Eğitim hizmetleri	30,8	69,2	10,9	89,1
Lokanta ve oteller	23,0	77,0	10,5	89,5
Çeşitli mal ve hizmetler	13,8	86,2	11,6	88,4

Çizelge 5.21’de, daraltılmış listeden kod atama yönteminde anketörün atadığı kod ile uyum oranı en yüksek ana grup %62,4 ile haberleşme olmuştur. Uyum oranı bakımından ikinci sırada %43,7 ile kira ve üçüncü sırada %30,8 ile eğitim hizmetleri ana grubu yer almaktadır. Geniş listeden kod atama yönteminde, %57,6 ile haberleşme ana grubu birinci, %37,5 ile konut ve kira ana grubu ikinci ve %18,1 ile gıda ve alkolsüz içecekler ana grubu kayıtları üçüncü olmuştur.

Ana gruplara göre anketörün atadığı kodun hatalı olduğu kayıtlara sistemin önerdiği kod sayılarına göre dağılımı Çizelge 5.22’de verilmiştir.

Çizelge 5.22. Ana gruplara göre sistemin önerdiği kod sayılarına göre dağılımı

Ana grup	Daraltılmış listeden kod atama yöntemi				Geniş listeden kod atama yöntemi			
	KASIS tarafından 1 kod önerisi yapılan kayıt	KASIS tarafından 3 kod önerisi yapılan kayıt	KASIS tarafından 4 kod önerisi yapılan kayıt	KASIS tarafından 5 kod önerisi yapılan kayıt	KASIS tarafından 1 kod önerisi yapılan kayıt	KASIS tarafından 3 kod önerisi yapılan kayıt	KASIS tarafından 4 kod önerisi yapılan kayıt	KASIS tarafından 5 kod önerisi yapılan kayıt
Gıda ve alkolsüz içecekler	78,4	16,5	4,8	0,4	43,0	19,7	33,7	3,5
Alkollü içecek, sigara ve tütün	96,1	3,4	0,5	0,0	89,7	7,6	2,6	0,1
Giyim ve ayakkabı	72,7	17,9	8,7	0,6	31,6	20,9	39,8	7,7
Konut ve kira	80,4	15,0	4,2	0,4	49,1	28,8	20,0	2,2
Mobilya ve ev eşyaları	84,9	10,8	4,0	0,3	40,5	28,6	26,0	4,9
Sağlık	70,7	17,6	11,2	0,5	28,8	49,4	21,3	0,6
Ulaştırma	89,0	6,5	4,6	0,0	49,0	12,9	34,3	3,8
Haberleşme	95,8	3,6	1,1	0,1	60,4	22,6	16,7	0,3
Eğlence ve kültür	85,5	8,9	5,3	0,4	56,9	17,2	23,0	2,9
Eğitim hizmetleri	88,2	9,9	1,9	0,1	41,5	40,2	17,4	0,9
Lokanta ve oteller	88,3	10,2	1,4	0,1	60,6	16,4	19,4	3,6
Çeşitli mal ve hizmetler	84,2	11,2	4,4	0,1	58,0	16,3	22,5	3,1

Çizelge 5.22'ye göre, daraltılmış listeden kod atama yönteminde bir adet kod önerisi yapılan oransal olarak en yüksek ana grup %96,1 ile alkollü içecek, sigara ve tütün olmuştur. %95,3 ile haberleşme ikinci ve %88,3 ile lokanta ve oteller üçüncü sırada yer almıştır. Geniş listeden kod atama yönteminde, %89,7 ile alkollü içecek, sigara ve tütün birinci, %60,6 ile lokanta ve oteller ikinci ve %60,4 ile haberleşme üçüncü ana grup olmuştur. Sistemin üç farklı kod önerisi yaptığı yüzdesel dağılımlara bakıldığında; daraltılmış listeden kod atama yöntemi lehine yöntemler arasında farklılık bulunmaktadır. Aynı farklılık, sistemin dört farklı ve beş farklı kod önerisi yaptığı tablolarda da bulunmaktadır.

Ana gruplara göre sistemin önerdiği kod ile anketörün atadığı kodun ilk beş basamağına göre dağılımı Çizelge 5.23'te yer verilmiştir.

Çizelge 5.23. Ana gruplara göre kodların ilk beş basamağına göre dağılımı

Ana grup	Daraltılmış listeden kod atama yöntemi		Geniş listeden kod atama yöntemi	
	KASIS'in önerdiği kodun ilk beş basamağı ile anketörün atadığı kodun ilk beş basamağı aynı	KASIS'in önerdiği kodun ilk beş basamağı ile anketörün atadığı kodun ilk beş basamağı farklı	KASIS'in önerdiği kodun ilk beş basamağı ile anketörün atadığı kodun ilk beş basamağı aynı	KASIS'in önerdiği kodun ilk beş basamağı ile anketörün atadığı kodun ilk beş basamağı farklı
Gıda ve alkolsüz içecekler	68,9	31,1	72,9	27,1
Alkollü içecek, sigara ve tütün	96,4	3,6	19,0	81,0
Giyim ve ayakkabı	60,8	39,2	56,4	43,6
Konut ve kira	65,6	34,4	56,4	43,6
Mobilya ve ev eşyaları	76,7	23,3	68,7	31,3
Sağlık	58,0	42,0	60,0	40,0
Ulaştırma	91,9	8,1	84,3	15,7
Haberleşme	40,3	59,7	13,6	86,4
Eğlence ve kültür	73,5	26,5	65,1	34,9
Eğitim hizmetleri	19,9	80,1	35,1	64,9
Lokanta ve oteller	29,5	70,5	33,6	66,4
Çeşitli mal ve hizmetler	77,5	22,5	89,3	10,7

Çizelge 5.23'e göre, daraltılmış listeden kod atama yönteminde sistemin önerdiği kodun ilk beş basamağı ile anketörün atadığı kodun ilk beş basamağının aynı olduğu en yüksek ana grup %96,4 ile alkollü içecek, sigara ve tütün ana grubu olmuştur. İkinci ana grup %91,9 ile ulaştırma ve üçüncü ana grup %77,5 ile çeşitli mal ve hizmetler ana grubu olmuştur. Alkollü içecek, sigara ve tütün ana grubunda yer alan kodların %96,4'ünde anketörün atadığı kodun ilk beş basamağı ile sistemin atadığı kodun ilk beş basamağı aynı olduğu için anketör kod atamasının doğru olduğu kabul edilebilmektedir. Geniş listeden kod atama yönteminde, %89,3 ile çeşitli mal ve hizmetler ana grubu birinci, %84,3 ile ulaştırma ana grubu ikinci ve %72,9 ile alkollü içecek, sigara ve tütün üçüncü ana grup olmuştur.

Ana gruplara göre yetersiz ve şüpheli kodlar içerisinde kod atama sonuçları Çizelge 5.24'te verilmiştir.

Çizelge 5.24. Ana gruplara göre kod atama sonuçları 1

Ana grup	Daraltılmış listeden kod atama yöntemi				Geniş listeden kod atama yöntemi			
	Doğrulanamadı - Hatalı Kodlama	Doğrulanamadı - Uzman Görüşü ile Seçim Gerekli	Doğrulandı - Onlu Kod Aynı	Doğrulandı - Beşli Kod Aynı	Doğrulanamadı - Hatalı Kodlama	Doğrulanamadı - Uzman Görüşü ile Seçim Gerekli	Doğrulandı - Onlu Kod Aynı	Doğrulandı - Beşli Kod Aynı
Gıda ve alkolsüz içecekler	17,5	15,5	28,3	38,7	9,6	46,7	18,1	25,7
Alkollü içecek, sigara ve tütün	3,4	3,7	3,4	89,5	70,0	9,9	3,7	16,5
Giyim ve ayakkabı	24,1	23,0	15,6	37,3	12,9	64,3	6,1	16,7
Konut ve kira	15,6	11,0	43,7	29,7	13,4	31,8	37,5	17,3
Mobilya ve ev eşyaları	16,3	12,4	17,8	53,6	11,5	54,1	9,1	25,3
Sağlık	24,9	24,6	16,1	34,4	10,8	67,2	5,8	16,3
Ulaştırma	5,7	8,8	20,5	65,0	6,7	44,5	12,7	36,1
Haberleşme	21,4	1,8	62,4	14,5	22,1	16,8	57,6	3,5
Eğlence ve kültür	19,5	12,5	13,8	54,1	18,4	40,0	7,2	34,4
Eğitim hizmetleri	48,9	8,2	30,8	12,2	24,0	52,1	10,9	13,0
Lokanta ve oteller	47,9	9,0	23,0	20,1	36,0	35,3	10,5	18,2
Çeşitli mal ve hizmetler	16,4	13,6	13,8	56,2	5,5	37,1	11,6	45,8

Çizelge 5.24 incelendiğinde; daraltılmış listeden kod atama yöntemine göre sağlık ana grubunda şüpheli ve yetersiz kodların %24,6'lık kısmı için uzman görüşünün gerekli olduğu sonucuna ulaşılmıştır. Bu oran; giyim ve ayakkabı ana grubunda %23,0 ve alkollü içecek, sigara ve tütün ana grubunda %15,5 olmuştur. Bu kayıtlar için, KASİS'in beş farklı bulanık eşleştirme algoritması ile önerdiği kodlara bakılarak uzman görüşü ile anketör kod atamasının doğruluğunun onaylanması veya KASİS tarafından önerilen kodlardan birisinin seçilmesi gerektiği ortaya çıkmıştır.

Ana gruplara göre tüm kodlar içerisinde kod atama sonuçları Çizelge 5.25'te verilmiştir.

Çizelge 5.25. Ana gruplara göre kod atama sonuçları 2

Ana grup	Daraltılmış listeden kod atama yöntemi				Geniş listeden kod atama yöntemi			
	Doğrulanamadı - Hatalı Kodlama	Doğrulanamadı - Uzman Görüşü ile Seçim Gerekli	Doğrulandı - Onlu Kod Aynı	Doğrulandı - Beşli Kod Aynı	Doğrulanamadı - Hatalı Kodlama	Doğrulanamadı - Uzman Görüşü ile Seçim Gerekli	Doğrulandı - Onlu Kod Aynı	Doğrulandı - Beşli Kod Aynı
Gıda ve alkolsüz içecekler	0,6	0,5	0,9	1,3	0,3	1,6	0,6	0,9
Alkollü içecek, sigara ve tütün	0,1	0,1	0,1	1,5	1,2	0,2	0,1	0,3
Giyim ve ayakkabı	2,4	2,3	1,6	3,7	1,3	6,4	0,6	1,7
Konut ve kira	0,6	0,4	1,8	1,2	0,5	1,3	1,5	0,7
Mobilya ve ev eşyaları	0,8	0,6	0,8	2,5	0,5	2,5	0,4	1,2
Sağlık	1,1	1,1	0,7	1,6	0,5	3,1	0,3	0,7
Ulaştırma	0,4	0,6	1,3	4,3	0,4	2,9	0,8	2,4
Haberleşme	2,4	0,2	7,0	1,6	2,5	1,9	6,5	0,4
Eğlence ve kültür	1,3	0,8	0,9	3,6	1,2	2,6	0,5	2,3
Eğitim hizmetleri	7,2	1,2	4,5	1,8	3,5	7,7	1,6	1,9
Lokanta ve oteller	10,6	2,0	5,1	4,5	8,0	7,8	2,3	4,1
Çeşitli mal ve hizmetler	0,4	0,3	0,3	1,4	0,1	0,9	0,3	1,1

Çizelge 5.25'te daraltılmış listeden kod atama yöntemine göre, giyim ve ayakkabı ana grubunda tüm kayıtların %2,3'lük kısmı için uzman görüşüne başvurulması gerektiği sonucuna ulaşılmıştır. Bu oran; lokanta ve oteller ana grubunda %2 ve eğitim hizmetleri ana grubunda %1,2 olmuştur. Geniş listeden kod atama yönteminin seçilmesi durumunda ise, ilk yöntemden biraz daha fazla kayıta uzman görüşüne başvurulması gerektiği ortaya çıkmıştır.

5.2.3. İki yöntem arasında bölgelere göre farklılık var mıdır?

Bölgesel dağılım, sistemin üretim sürecine dahil edilmesi durumunda hangi bölgedeki anketörün sınıflama ve kod atama konusunda eksikliği olduğunun belirlenerek giderilmesi için önemli bir parametre konumuna gelecektir. Bu yüzden bölgesel sonuçların incelenerek analiz edilmesi büyük önem arz etmektedir.

Bölgelere göre anketör atadığı ile sistemin önerdiği kodların dağılımı Çizelge 5.26'da verilmiştir.

Çizelge 5.26. Bölgelere göre sistemin önerdiği kodların dağılımı

Bölge	Daraltılmış listeden kod atama yöntemi		Geniş listeden kod atama yöntemi	
	Anketörün atadığı ile sistemin önerdiği kod aynı	Anketörün atadığı ile sistemin önerdiği kod farklı	Anketörün atadığı ile sistemin önerdiği kod aynı	Anketörün atadığı ile sistemin önerdiği kod farklı
Adana	28,4	71,6	14,9	85,1
Ankara	52,3	47,7	31,1	68,9
Antalya	36,9	63,1	23,9	76,1
Balıkesir	22,0	78,0	13,1	86,9
Bursa	22,5	77,5	11,5	88,5
Denizli	21,8	78,2	12,0	88,0
Diyarbakır	22,7	77,3	13,8	86,2
Edirne	24,3	75,7	18,3	81,7
Erzurum	21,0	79,0	14,6	85,4
Gaziantep	16,8	83,2	10,5	89,5
Hatay	47,2	52,8	19,7	80,3
İstanbul	33,0	67,0	18,2	81,8
İzmir	26,5	73,5	16,2	83,8
Kars	23,5	76,5	13,8	86,2
Kastamonu	30,4	69,6	23,6	76,4
Kayseri	30,5	69,5	18,5	81,5
Kocaeli	27,9	72,1	15,7	84,3
Konya	20,8	79,2	11,9	88,1
Malatya	24,8	75,2	15,6	84,4
Manisa	25,0	75,0	17,2	82,8
Nevşehir	30,0	70,0	20,8	79,2
Samsun	40,3	59,7	30,1	69,9
Siirt	19,9	80,1	15,2	84,8
Trabzon	47,6	52,4	31,7	68,3
Van	14,6	85,4	11,3	88,7
Zonguldak	21,4	78,6	9,4	90,6

Çizelge 5.26’da, bölgesel ve yöntemler arasında oransal dağılımda farklılıklar bulunmaktadır. Daraltılmış listeden kod atama yönteminde, anketörün atadığı kod ile uyum oranı en yüksek bölge %52,3 ile Ankara olmuştur. Uyum oranı bakımından ikinci sırada %47,6 ile Trabzon ve üçüncü sırada %47,2 ile Hatay bölgesi yer almaktadır. Geniş listeden kod atama yönteminde, %31,7 ile Trabzon bölgesi birinci, %31,1 ile Ankara bölgesi ikinci ve %30,1 ile Samsun bölgesi kayıtları üçüncü olmuştur.

Bölgelere göre anketörün atadığı kodun hatalı olduğu kayıtlara sistemin önerdiği kod sayılarına göre dağılımı Çizelge 5.27’de verilmiştir.

Çizelge 5.27. Bölgelere göre sistemin önerdiği kod sayılarına göre dağılımı

Bölge	Daraltılmış listeden kod atama yöntemi				Geniş listeden kod atama yöntemi			
	KASIS tarafından 1 kod önerisi yapılan kayıt	KASIS tarafından 3 kod önerisi yapılan kayıt	KASIS tarafından 4 kod önerisi yapılan kayıt	KASIS tarafından 5 kod önerisi yapılan kayıt	KASIS tarafından 1 kod önerisi yapılan kayıt	KASIS tarafından 3 kod önerisi yapılan kayıt	KASIS tarafından 4 kod önerisi yapılan kayıt	KASIS tarafından 5 kod önerisi yapılan kayıt
Adana	84,6	13,4	1,9	0,1	49,0	16,8	31,2	3,0
Ankara	84,9	11,0	3,8	0,3	39,2	27,1	30,0	3,6
Antalya	77,4	19,0	3,4	0,2	38,1	21,9	35,6	4,4
Balıkesir	81,3	14,9	3,6	0,2	52,3	13,1	32,0	2,6
Bursa	84,4	11,2	4,0	0,4	51,7	19,9	24,4	4,0
Denizli	86,5	10,3	3,0	0,2	49,2	20,8	26,3	3,6
Diyarbakır	80,6	15,5	3,7	0,2	48,0	20,7	29,0	2,4
Edirne	94,2	4,5	1,3	0,1	67,5	19,4	12,3	0,9
Erzurum	81,3	13,8	4,7	0,2	53,4	16,0	27,9	2,7
Gaziantep	83,5	13,4	3,0	0,2	56,6	14,5	26,8	2,1
Hatay	83,5	12,1	4,1	0,3	43,8	30,8	22,3	3,2
İstanbul	82,3	13,1	4,2	0,4	44,2	20,4	29,7	5,6
İzmir	84,0	11,7	4,0	0,3	54,0	17,4	24,1	4,5
Kars	87,1	10,4	2,5	0,1	43,5	25,2	28,4	2,8
Kastamonu	84,3	12,7	2,8	0,2	56,2	16,6	24,2	2,9
Kayseri	80,8	11,7	7,1	0,5	44,3	23,8	27,5	4,5
Kocaeli	82,8	12,2	4,5	0,5	50,8	23,1	22,3	3,9
Konya	83,8	10,8	4,8	0,6	53,2	18,0	24,8	3,9
Malatya	78,7	10,3	10,6	0,3	42,0	24,7	29,5	3,9
Manisa	87,2	9,7	2,8	0,3	51,9	24,6	20,4	3,1
Nevşehir	75,6	17,4	6,8	0,2	35,7	31,8	25,2	7,3
Samsun	84,4	11,7	3,4	0,4	39,5	35,8	22,5	2,2
Siirt	82,1	14,9	2,9	0,1	46,2	19,4	32,0	2,4
Trabzon	80,6	15,1	4,0	0,3	41,8	29,3	23,9	4,9
Van	79,1	18,1	2,7	0,1	54,5	12,6	31,4	1,5
Zonguldak	83,5	12,2	3,7	0,6	51,8	20,0	24,3	3,8

Çizelge 5.27'ye göre, daraltılmış listeden kod atama yönteminde bir adet kod önerisi yapılan oransal olarak en yüksek bölge %94,2 ile Edirne olmuştur. %87,2 ile Manisa ikinci ve %87,1 ile Kars üçüncü sırada yer almıştır. Geniş listeden kod atama yönteminde, %67,5 ile Edirne birinci, %56,6 ile Gaziantep ikinci ve %56,2 ile Kastamonu üçüncü bölge olmuştur. Sistemin üç farklı kod önerisi yaptığı yüzdesel dağılımlara bakıldığında; daraltılmış listeden kod atama yöntemi lehine yöntemler arasında farklılık bulunmaktadır. Aynı farklılık, sistemin dört farklı ve beş farklı kod önerisi yaptığı tablolarda da bulunmaktadır.

Bölgelere göre sistemin önerdiği kod ile anketörün atadığı kodun ilk beş basamağına göre dağılımı Çizelge 5.28'da yer verilmiştir.

Çizelge 5.28. Bölgelere göre kodların ilk beş basamağına göre dağılımı

Bölge	Daraltılmış listeden kod atama yöntemi		Geniş listeden kod atama yöntemi	
	KASIS'in önerdiği kodun ilk beş basamağı ile anketörün atadığı kodun ilk beş basamağı aynı	KASIS'in önerdiği kodun ilk beş basamağı ile anketörün atadığı kodun ilk beş basamağı farklı	KASIS'in önerdiği kodun ilk beş basamağı ile anketörün atadığı kodun ilk beş basamağı aynı	KASIS'in önerdiği kodun ilk beş basamağı ile anketörün atadığı kodun ilk beş basamağı farklı
Adana	74,0	26,0	66,4	33,6
Ankara	60,7	39,3	55,4	44,6
Antalya	51,9	48,1	48,4	51,6
Balıkesir	70,5	29,5	70,7	29,3
Bursa	43,7	56,3	35,4	64,6
Denizli	36,3	63,7	24,9	75,1
Diyarbakır	70,2	29,8	72,3	27,7
Edirne	76,1	23,9	72,3	27,7
Erzurum	74,1	25,9	66,5	33,5
Gaziantep	70,6	29,4	60,0	40,0
Hatay	56,3	43,7	61,2	38,8
İstanbul	48,4	51,6	47,3	52,7
İzmir	55,9	44,1	61,0	39,0
Kars	71,7	28,3	69,0	31,0
Kastamonu	71,5	28,5	71,3	28,7
Kayseri	60,2	39,8	54,8	45,2
Kocaeli	53,2	46,8	49,3	50,7
Konya	58,7	41,3	58,8	41,2
Malatya	54,5	45,5	46,6	53,4
Manisa	38,2	61,8	27,3	72,7
Nevşehir	50,1	49,9	47,2	52,8
Samsun	54,5	45,5	37,0	63,0
Siirt	64,9	35,1	69,7	30,3
Trabzon	70,7	29,3	56,2	43,8
Van	76,0	24,0	66,7	33,3
Zonguldak	53,9	46,1	64,3	35,7

Çizelge 5.28'e göre, daraltılmış listeden kod atama yönteminde sistemin önerdiği kodun ilk beş basamağı ile anketörün atadığı kodun ilk beş basamağının aynı olduğu en yüksek bölge %76,1 ile Edirne bölgesi olmuştur. İkinci bölge %76 ile Van bölgesi ve üçüncü bölge %74,1 ile Erzurum bölgesi olmuştur. Geniş listeden kod atama yönteminde, %72,3 ile Diyarbakır ve Edirne bölgesi birinci, %71,3 ile Kastamonu bölgesi ikinci ve %70,7 ile Balıkesir bölgesi üçüncü olmuştur.

Bölgelere göre yetersiz ve şüpheli kodlar içerisinde kod atama sonuçları Çizelge 5.29'da verilmiştir.

Çizelge 5.29. Bölgelere göre kod atama sonuçları 1

Bölge	Daraltılmış listeden kod atama yöntemi				Geniş listeden kod atama yöntemi			
	Doğrulanamadı - Hatalı Kodlama	Doğrulanamadı - Uzman Görüşü ile Seçim Gerekli	Doğrulandı - Onlu Kod Aynı	Doğrulandı - Beşli Kod Aynı	Doğrulanamadı - Hatalı Kodlama	Doğrulanamadı - Uzman Görüşü ile Seçim Gerekli	Doğrulandı - Onlu Kod Aynı	Doğrulandı - Beşli Kod Aynı
Adana	15,7	11,0	28,4	44,8	14,0	43,4	14,9	27,7
Ankara	15,9	7,2	52,3	24,6	12,1	41,9	31,1	15,0
Antalya	23,5	14,3	36,9	25,4	15,0	47,1	23,9	14,1
Balıkesir	18,7	14,6	22,0	44,7	13,3	41,5	13,1	32,1
Bursa	36,8	12,1	22,5	28,6	29,6	42,7	11,5	16,2
Denizli	43,1	10,6	21,8	24,6	32,5	44,7	12,0	10,8
Diyarbakır	18,5	15,0	22,7	43,8	11,5	44,8	13,8	29,9
Edirne	17,0	4,4	24,3	54,3	15,2	26,6	18,3	39,9
Erzurum	16,7	14,8	21,0	47,6	15,3	39,8	14,6	30,3
Gaziantep	20,4	13,8	16,8	49,0	20,3	38,8	10,5	30,4
Hatay	19,3	8,7	47,2	24,8	13,6	45,1	19,7	21,5
İstanbul	28,5	11,9	33,0	26,6	19,1	45,6	18,2	17,1
İzmir	27,2	11,8	26,5	34,5	17,7	38,5	16,2	27,6
Kars	18,9	9,9	23,5	47,8	11,7	48,7	13,8	25,9
Kastamonu	16,7	10,9	30,4	42,0	12,3	33,4	23,6	30,6
Kayseri	22,3	13,4	30,5	33,8	16,3	45,4	18,5	19,8
Kocaeli	27,9	12,4	27,9	31,7	21,7	41,5	15,7	21,1
Konya	27,5	12,8	20,8	39,0	19,3	41,2	11,9	27,6
Malatya	26,9	16,0	24,8	32,3	18,9	49,0	15,6	16,5
Manisa	40,4	9,6	25,0	25,0	31,3	39,8	17,2	11,7
Nevşehir	26,4	17,1	30,0	26,5	14,9	51,0	20,8	13,4
Samsun	22,9	9,3	40,3	27,5	17,4	42,3	30,1	10,2
Siirt	23,1	14,3	19,9	42,7	11,9	45,7	15,2	27,3
Trabzon	12,4	10,2	47,6	29,8	12,5	39,7	31,7	16,0
Van	16,2	17,9	14,6	51,3	16,1	40,3	11,3	32,3
Zonguldak	30,3	13,0	21,4	35,4	16,8	43,6	9,4	30,2

Çizelge 5.29 incelendiğinde; daraltılmış listeden kod atama yöntemine göre Van bölgesinde şüpheli ve yetersiz kodların %17,9'luk kısmı için uzman görüşünün gerekli olduğu sonucuna ulaşılmıştır. Bu oran; Nevşehir bölgesinde %17,1 ve Malatya bölgesinde %16 olmuştur. Bu kayıtlar için, KASİS'in beş farklı bulanık eşleştirme algoritması ile önerdiği kodlara bakılarak uzman görüşü ile anketör kod atamasının doğruluğunun onaylanması veya KASİS tarafından önerilen kodlardan birisinin seçilmesi gerektiği belirlenmiştir.

Bölgelere göre tüm kodlar içerisinde kod atama sonuçları Çizelge 5.30'da verilmiştir.

Çizelge 5.30. Bölgelere göre kod atama sonuçları 2

Bölge	Daraltılmış listeden kod atama yöntemi				Geniş listeden kod atama yöntemi			
	Doğrulanamadı - Hatalı Kodlama	Doğrulanamadı - Uzman Görüşü ile Seçim Gerekli	Doğrulandı - Onlu Kod Aynı	Doğrulandı - Beşli Kod Aynı	Doğrulanamadı - Hatalı Kodlama	Doğrulanamadı - Uzman Görüşü ile Seçim Gerekli	Doğrulandı - Onlu Kod Aynı	Doğrulandı - Beşli Kod Aynı
Adana	0,7	0,5	1,2	1,9	0,6	1,8	0,6	1,2
Ankara	0,1	0,0	0,3	0,1	0,1	0,2	0,2	0,1
Antalya	1,3	0,8	2,0	1,4	0,8	2,6	1,3	0,8
Balıkesir	1,7	1,3	2,0	4,0	1,2	3,7	1,2	2,9
Bursa	2,4	0,8	1,5	1,9	2,0	2,8	0,8	1,1
Denizli	3,2	0,8	1,6	1,8	2,4	3,3	0,9	0,8
Diyarbakır	1,4	1,2	1,8	3,4	0,9	3,5	1,1	2,3
Edirne	0,4	0,1	0,6	1,3	0,4	0,6	0,4	1,0
Erzurum	1,8	1,6	2,3	5,2	1,7	4,4	1,6	3,3
Gaziantep	1,1	0,7	0,9	2,5	1,1	2,0	0,5	1,6
Hatay	0,2	0,1	0,4	0,2	0,1	0,4	0,2	0,2
İstanbul	1,0	0,4	1,2	1,0	0,7	1,7	0,7	0,6
İzmir	2,3	1,0	2,2	2,9	1,5	3,2	1,3	2,3
Kars	1,0	0,5	1,2	2,5	0,6	2,5	0,7	1,3
Kastamonu	0,5	0,3	0,9	1,3	0,4	1,0	0,7	0,9
Kayseri	0,6	0,4	0,9	1,0	0,5	1,3	0,5	0,6
Kocaeli	1,4	0,6	1,4	1,6	1,1	2,1	0,8	1,0
Konya	2,9	1,4	2,2	4,2	2,1	4,4	1,3	3,0
Malatya	1,4	0,8	1,3	1,7	1,0	2,6	0,8	0,9
Manisa	1,8	0,4	1,1	1,1	1,4	1,8	0,8	0,5
Nevşehir	1,2	0,7	1,3	1,2	0,7	2,2	0,9	0,6
Samsun	0,2	0,1	0,4	0,3	0,2	0,5	0,3	0,1
Siirt	4,0	2,5	3,4	7,4	2,0	7,9	2,6	4,7
Trabzon	0,2	0,1	0,6	0,4	0,2	0,5	0,4	0,2
Van	3,5	3,9	3,2	11,2	3,5	8,8	2,5	7,0
Zonguldak	0,9	0,4	0,7	1,1	0,5	1,4	0,3	0,9

Çizelge 5.30’da daraltılmış listeden kod atama yöntemine göre, Van bölgesinde tüm kayıtların %3,9’luk kısmı için uzman görüşüne başvurulması gerektiği sonucuna ulaşılmıştır. Bu oran; Siirt bölgesinde %2,5 ve Erzurum bölgesinde %1,6 olmuştur. Geniş listeden kod atama yönteminin seçilmesi durumunda ise, ilk yöntemden biraz daha fazla kayıtta uzman görüşüne başvurulması gerektiği ortaya çıkmıştır.

5.2.4. İki yöntem arasında yıllara göre farklılık var mıdır?

Yıllara göre anketör atadığı ile sistemin önerdiği kodların dağılımı Çizelge 5.31’de verilmiştir.

Çizelge 5.31. Yıllara göre sistemin önerdiği kodların dağılımı

Yıl	Daraltılmış listeden kod atama yöntemi		Geniş listeden kod atama yöntemi	
	Anketörün atadığı ile sistemin önerdiği kod aynı	Anketörün atadığı ile sistemin önerdiği kod farklı	Anketörün atadığı ile sistemin önerdiği kod aynı	Anketörün atadığı ile sistemin önerdiği kod farklı
2013	23,2	76,8	15,3	84,7
2014	24,1	75,9	15,2	84,8
2015	23,2	76,8	14,9	85,1
2016	23,8	76,2	14,7	85,3
2017	30,7	69,3	16,2	83,8
2018	53,7	46,3	19,0	81,0

Çizelge 5.31’de, daraltılmış listeden kod atama yönteminde anketörün atadığı kod ile uyum oranı en yüksek yıl %53,7 ile 2018 yılı olmuştur. Uyum oranı bakımından ikinci sırada %30,7 ile 2017 ve üçüncü sırada %24,1 ile 2014 yılı yer almaktadır. Geniş listeden kod atama yönteminde, %19 ile 2018 yılı birinci, %16,2 ile 2017 yılı ikinci ve %15,3 ile 2013 yılı kayıtları üçüncü olmuştur.

Yıllara göre anketörün atadığı kodun hatalı olduğu kayıtlara sistemin önerdiği kod sayılarına göre dağılımı Çizelge 5.32’de verilmiştir.

Çizelge 5.32. Yıllara göre sistemin önerdiği kod sayılarına göre dağılımı

Yıl	Daraltılmış listeden kod atama yöntemi				Geniş listeden kod atama yöntemi			
	KASİS tarafından 1 kod önerisi yapılan kayıt	KASİS tarafından 3 kod önerisi yapılan kayıt	KASİS tarafından 4 kod önerisi yapılan kayıt	KASİS tarafından 5 kod önerisi yapılan kayıt	KASİS tarafından 1 kod önerisi yapılan kayıt	KASİS tarafından 3 kod önerisi yapılan kayıt	KASİS tarafından 4 kod önerisi yapılan kayıt	KASİS tarafından 5 kod önerisi yapılan kayıt
2013	84,0	10,8	4,9	0,4	53,7	17,8	24,4	4,1
2014	82,6	12,3	4,8	0,3	50,9	18,2	26,5	4,5
2015	82,6	13,6	3,6	0,2	50,7	17,8	28,1	3,4
2016	84,5	12,3	2,9	0,2	52,5	15,9	28,6	3,0
2017	83,3	13,5	2,9	0,3	38,9	30,2	27,7	3,2
2018	66,8	31,2	1,9	0,1	34,1	32,8	31,0	2,1

Çizelge 5.32’ye göre, daraltılmış listeden kod atama yönteminde bir adet kod önerisi yapılan oransal olarak en yüksek yıl %84,5 ile 2016 olmuştur. %84 ile 2013 yılı ikinci ve %83,3 ile 2017 yılı üçüncü sırada yer almıştır. Geniş listeden kod atama yönteminde, %53,7 ile 2013 birinci, %52,5 ile 2016 ikinci ve %50,9 ile 2014 üçüncü yıl olmuştur. Sistemin üç farklı kod önerisi yaptığı yüzdesel dağılımlara bakıldığında; daraltılmış listeden kod atama yöntemi lehine yöntemler arasında farklılık bulunmaktadır. Aynı farklılık, sistemin dört farklı ve beş farklı kod önerisi yaptığı tablolarda da bulunmaktadır.

Yıllara göre sistemin önerdiği kod ile anketörün atadığı kodun ilk beş basamağına göre dağılımı Çizelge 5.33'te yer verilmiştir.

Çizelge 5.33. Yıllara göre kodların ilk beş basamağına göre dağılımı

Yıl	Daraltılmış listeden kod atama yöntemi				Geniş listeden kod atama yöntemi			
	KASIS'in önerdiği kodun ilk beş basamağı ile anketörün atadığı kodun ilk beş basamağı aynı		KASIS'in önerdiği kodun ilk beş basamağı ile anketörün atadığı kodun ilk beş basamağı farklı		KASIS'in önerdiği kodun ilk beş basamağı ile anketörün atadığı kodun ilk beş basamağı aynı		KASIS'in önerdiği kodun ilk beş basamağı ile anketörün atadığı kodun ilk beş basamağı farklı	
2013		46,7		53,3		51,4		48,6
2014		45,5		54,5		48,3		51,7
2015		63,9		36,1		56,5		43,5
2016		64,1		35,9		55,0		45,0
2017		70,3		29,7		63,5		36,5
2018		67,2		32,8		67,9		32,1

Çizelge 5.33'e göre, daraltılmış listeden kod atama yönteminde sistemin önerdiği kodun ilk beş basamağı ile anketörün atadığı kodun ilk beş basamağının aynı olduğu en yüksek yıl %70,3 ile 2017 yılı olmuştur. İkinci yıl %67,2 ile 2018 ve üçüncü yıl %64,1 ile 2016 olmuştur. Geniş listeden kod atama yönteminde, %67,9 ile 2018 yılı birinci, %63,5 ile 2017 yılı ikinci ve %56,5 ile 2015 yılı üçüncü olmuştur.

Yıllara göre yetersiz ve şüpheli kodlar içerisinde kod atama sonuçları Çizelge 5.34'te verilmiştir.

Çizelge 5.34. Yıllara göre kod atama sonuçları 1

Yıl	Daraltılmış listeden kod atama yöntemi				Geniş listeden kod atama yöntemi			
	Doğrulanamadı - Hatalı Kodlama	Doğrulanamadı - Uzman Görüşü ile Seçim Gerekli	Doğrulandı - Onlu Kod Aynı	Doğrulandı - Beşli Kod Aynı	Doğrulanamadı - Hatalı Kodlama	Doğrulanamadı - Uzman Görüşü ile Seçim Gerekli	Doğrulandı - Onlu Kod Aynı	Doğrulandı - Beşli Kod Aynı
2013		34,4		12,3		23,2		30,1
2014		34,2		13,2		24,1		28,5
2015		22,9		13,4		23,2		40,6
2016		23,1		11,8		23,8		41,2
2017		17,1		11,6		30,7		40,6
2018		10,2		15,4		53,7		20,8

Çizelge 5.34 incelendiğinde; daraltılmış listeden kod atama yöntemine göre 2018 yılında şüpheli ve yetersiz kodların %15,4'lük kısmı için uzman görüşünün gerekli olduğu sonucuna ulaşılmıştır. Bu oran; 2015 yılında %13,4 ve 2014 yılında %13,2 olmuştur. Bu kayıtlar için, KASIS'in beş farklı bulanık eşleştirme algoritması ile önerdiği kodlara bakılarak uzman

görüşü ile anketör kod atamasının doğruluğunun onaylanması veya KASİS tarafından önerilen kodlardan birisinin seçilmesi gerekmektedir.

Yıllara göre tüm kodlar içerisinde kod atama sonuçları Çizelge 5.35’te verilmiştir.

Çizelge 5.35. Yıllara göre kod atama sonuçları 2

Yıl	Daraltılmış listeden kod atama yöntemi				Geniş listeden kod atama yöntemi			
	Doğrulanamadı - Hatalı Kodlama	Doğrulanamadı - Uzman Görüşü ile Seçim Gerekli	Doğrulandı - Onlu Kod Aynı	Doğrulandı - Beşli Kod Aynı	Doğrulanamadı - Hatalı Kodlama	Doğrulanamadı - Uzman Görüşü ile Seçim Gerekli	Doğrulandı - Onlu Kod Aynı	Doğrulandı - Beşli Kod Aynı
2013	2,4	0,8	1,6	2,1	1,5	2,7	1,0	1,6
2014	2,0	0,8	1,4	1,7	1,3	2,4	0,9	1,2
2015	2,1	1,2	2,1	3,7	1,7	3,8	1,4	2,2
2016	1,3	0,7	1,3	2,3	1,1	2,2	0,8	1,4
2017	0,5	0,4	1,0	1,3	0,4	1,6	0,5	0,7
2018	0,1	0,1	0,5	0,2	0,1	0,5	0,2	0,2

Çizelge 5.35’te daraltılmış listeden kod atama yöntemine göre, 2015 yılında tüm kayıtların %1,2’lik kısmı için uzman görüşüne başvurulması gerektiği sonucuna ulaşılmıştır. Bu oran; 2013 ve 2014 yılında %08 olmuştur. Geniş listeden kod atama yönteminin seçilmesi durumunda ise, ilk yöntemden biraz daha fazla kayıta uzman görüşüne başvurulması gerektiği ortaya çıkmıştır.

5.2.5. İki yöntem arasında aylara göre farklılık var mıdır?

Aylara göre anketör atadığı ile sistemin önerdiği kodların dağılımı Çizelge 5.36’da verilmiştir.

Çizelge 5.36. Aylara göre sistemin önerdiği kodların dağılımı

Ay	Daraltılmış listeden kod atama yöntemi		Geniş listeden kod atama yöntemi	
	Anketörün atadığı ile sistemin önerdiği kod aynı	Anketörün atadığı ile sistemin önerdiği kod farklı	Anketörün atadığı ile sistemin önerdiği kod aynı	Anketörün atadığı ile sistemin önerdiği kod farklı
Ocak	25,1	74,9	16,0	84,0
Şubat	25,1	74,9	15,8	84,2
Mart	25,9	74,1	15,4	84,6
Nisan	25,5	74,5	15,3	84,7
Mayıs	25,9	74,1	15,6	84,4
Haziran	25,1	74,9	14,8	85,2
Temmuz	23,6	76,4	14,4	85,6
Ağustos	22,5	77,5	13,1	86,9
Eylül	23,8	76,2	13,5	86,5
Ekim	26,5	73,5	15,9	84,1
Kasım	26,8	73,2	16,3	83,7
Aralık	26,9	73,1	16,7	83,3

Çizelge 5.36’da, daraltılmış listeden kod atama yönteminde anketörün atadığı kod ile uyum oranı en yüksek ay %26,9 ile Aralık olmuştur. Uyum oranı bakımından ikinci sırada %26,8 ile Kasım ve üçüncü sırada %26,5 ile Ekim ayı yer almaktadır. Geniş listeden kod atama yönteminde, %16,7 ile Aralık ayı birinci, %16,3 ile Kasım ayı ikinci ve %16 ile Ocak ayı kayıtları üçüncü olmuştur.

Aylara göre anketörün atadığı kodun hatalı olduğu kayıtlara sistemin önerdiği kod sayılarına göre dağılımı Çizelge 5.37’de verilmiştir.

Çizelge 5.37. Aylara göre sistemin önerdiği kod sayılarına göre dağılımı

Ay	Daraltılmış listeden kod atama yöntemi				Geniş listeden kod atama yöntemi			
	KASIS tarafından 1 kod önerisi yapılan kayıt	KASIS tarafından 3 kod önerisi yapılan kayıt	KASIS tarafından 4 kod önerisi yapılan kayıt	KASIS tarafından 5 kod önerisi yapılan kayıt	KASIS tarafından 1 kod önerisi yapılan kayıt	KASIS tarafından 3 kod önerisi yapılan kayıt	KASIS tarafından 4 kod önerisi yapılan kayıt	KASIS tarafından 5 kod önerisi yapılan kayıt
Ocak	82,2	12,8	4,8	0,3	49,3	19,8	27,2	3,7
Şubat	82,5	13,3	3,9	0,3	49,9	18,7	27,6	3,8
Mart	83,3	12,7	3,8	0,3	51,1	18,8	26,2	3,9
Nisan	83,1	12,8	3,9	0,3	49,8	19,5	27,0	3,7
Mayıs	82,7	13,4	3,7	0,3	50,4	19,4	26,8	3,3
Haziran	83,4	12,8	3,6	0,2	50,1	18,2	28,2	3,4
Temmuz	82,1	13,7	4,0	0,2	49,5	18,4	28,5	3,6
Ağustos	83,7	12,8	3,3	0,3	50,0	18,9	27,9	3,2
Eylül	84,0	12,2	3,5	0,3	49,2	19,4	27,9	3,5
Ekim	82,9	12,8	4,0	0,2	48,4	20,6	27,4	3,6
Kasım	83,2	12,7	3,8	0,3	50,5	19,6	26,4	3,5
Aralık	83,6	12,4	3,7	0,4	50,8	19,9	25,3	4,0

Çizelge 5.37’ye göre, daraltılmış listeden kod atama yönteminde bir adet kod önerisi yapılan aylar %82,1 ile %84 arasında yer almaktadır. Geniş listeden kod atama yönteminde, bir adet kod önerisi yapılan aylar %48,4 ile %51,1 arasında yer almaktadır. Sistemin üç farklı kod önerisi yaptığı yüzdesel dağılımlara bakıldığında; daraltılmış listeden kod atama yöntemi lehine yöntemler arasında farklılık bulunmaktadır. Aynı farklılık, sistemin dört farklı ve beş farklı kod önerisi yaptığı tablolarda da bulunmaktadır.

Aylara göre sistemin önerdiği kod ile anketörün atadığı kodun ilk beş basamağına göre dağılımı Çizelge 5.38’de yer verilmiştir.

Çizelge 5.38. Aylara göre kodların ilk beş basamağına göre dağılımı

Ay	Daraltılmış listeden kod atama yöntemi				Geniş listeden kod atama yöntemi			
	KASIS'in önerdiği kodun ilk beş basamağı ile anketörün atadığı kodun ilk beş basamağı aynı		KASIS'in önerdiği kodun ilk beş basamağı ile anketörün atadığı kodun ilk beş basamağı farklı		KASIS'in önerdiği kodun ilk beş basamağı ile anketörün atadığı kodun ilk beş basamağı aynı		KASIS'in önerdiği kodun ilk beş basamağı ile anketörün atadığı kodun ilk beş basamağı farklı	
Ocak		60,3		39,7		55,7		44,3
Şubat		59,6		40,4		53,5		46,5
Mart		57,9		42,1		55,9		44,1
Nisan		58,7		41,3		55,2		44,8
Mayıs		58,4		41,6		55,5		44,5
Haziran		60,0		40,0		56,7		43,3
Temmuz		60,2		39,8		57,9		42,1
Ağustos		57,6		42,4		55,2		44,8
Eylül		56,4		43,6		54,1		45,9
Ekim		53,1		46,9		50,6		49,4
Kasım		53,9		46,1		50,5		49,5
Aralık		54,7		45,3		51,1		48,9

Çizelge 5.38'e göre, daraltılmış listeden kod atama yönteminde sistemin önerdiği kodun ilk beş basamağı ile anketörün atadığı kodun ilk beş basamağının aynı olduğu en yüksek ay %60,3 ile Ocak ayı olmuştur. İkinci ay %60,2 ile Temmuz ayı ve üçüncü ay %60 ile Haziran ayı olmuştur. Geniş listeden kod atama yönteminde, %57,9 ile Temmuz ayı birinci, %56,7 ile Haziran ayı ikinci ve %55,9 ile Mart ayı üçüncü olmuştur.

Aylara göre yetersiz ve şüpheli kodlar içerisinde kod atama sonuçları Çizelge 5.39'da verilmiştir.

Çizelge 5.39. Aylara göre kod atama sonuçları 1

Ay	Daraltılmış listeden kod atama yöntemi				Geniş listeden kod atama yöntemi			
	Doğrulanamadı - Hatalı Kodlama	Doğrulanamadı - Uzman Görüşü ile Seçim Gerekli	Doğrulandı - Onlu Kod Aynı	Doğrulandı - Beşli Kod Aynı	Doğrulanamadı - Hatalı Kodlama	Doğrulanamadı - Uzman Görüşü ile Seçim Gerekli	Doğrulandı - Onlu Kod Aynı	Doğrulandı - Beşli Kod Aynı
Ocak	24,5	13,4	25,1	37,1	18,4	42,6	16,0	23,1
Şubat	25,0	13,1	25,1	36,8	19,5	42,2	15,8	22,4
Mart	26,0	12,4	25,9	35,8	19,1	41,4	15,4	24,2
Nisan	25,6	12,6	25,5	36,3	18,9	42,5	15,3	23,3
Mayıs	25,5	12,8	25,9	35,7	18,9	41,8	15,6	23,6
Haziran	25,0	12,5	25,1	37,5	18,5	42,5	14,8	24,2
Temmuz	24,9	13,7	23,6	37,8	17,9	43,2	14,4	24,5
Ağustos	27,5	12,6	22,5	37,3	19,5	43,4	13,1	24,0
Eylül	27,9	12,2	23,8	36,1	19,5	43,9	13,5	23,1
Ekim	28,6	12,5	26,5	32,4	20,1	43,4	15,9	20,6
Kasım	28,0	12,3	26,8	32,8	20,9	41,4	16,3	21,3
Aralık	27,7	12,0	26,9	33,5	20,7	41,0	16,7	21,6

Çizelge 5.39 incelendiğinde; daraltılmış listeden kod atama yöntemine göre Temmuz ayında şüpheli ve yetersiz kodların %13,7’lik kısmı için uzman görüşünün gerekli olduğu sonucuna ulaşılmıştır. Bu oran; Ocak ayında %13,4 ve Şubat ayında %13,1 olmuştur. Bu kayıtlar için, KASİS’in beş farklı bulanık eşleştirme algoritması ile önerdiği kodlara bakılarak uzman görüşü ile anketör kod atamasının doğruluğunun onaylanması veya KASİS tarafından önerilen kodlardan birisinin seçilmesi gerektiği belirlenmiştir.

Aylara göre tüm kodlar içerisinde kod atama sonuçları Çizelge 5.40’ta verilmiştir.

Çizelge 5.40. Aylara göre kod atama sonuçları 2

Ay	Daraltılmış listeden kod atama yöntemi				Geniş listeden kod atama yöntemi			
	Doğrulanamadı - Hatalı Kodlama	Doğrulanamadı - Uzman Görüşü ile Seçim Gerekli	Doğrulandı - Onlu Kod Aynı	Doğrulandı - Beşli Kod Aynı	Doğrulanamadı - Hatalı Kodlama	Doğrulanamadı - Uzman Görüşü ile Seçim Gerekli	Doğrulandı - Onlu Kod Aynı	Doğrulandı - Beşli Kod Aynı
Ocak	1,4	0,8	1,4	2,1	1,1	2,5	0,9	1,3
Şubat	1,4	0,7	1,4	2,1	1,1	2,4	0,9	1,3
Mart	1,4	0,7	1,4	2,0	1,0	2,3	0,8	1,3
Nisan	1,4	0,7	1,4	2,0	1,1	2,4	0,9	1,3
Mayıs	1,4	0,7	1,4	1,9	1,0	2,2	0,8	1,3
Haziran	1,3	0,6	1,3	1,9	0,9	2,2	0,8	1,2
Temmuz	1,2	0,7	1,2	1,9	0,9	2,2	0,7	1,2
Ağustos	1,4	0,6	1,2	1,9	1,0	2,2	0,7	1,2
Eylül	1,3	0,6	1,1	1,7	0,9	2,1	0,6	1,1
Ekim	1,4	0,6	1,3	1,5	1,0	2,1	0,8	1,0
Kasım	1,3	0,6	1,2	1,5	0,9	1,9	0,7	1,0
Aralık	1,2	0,5	1,1	1,4	0,9	1,8	0,7	0,9

Çizelge 5.40’ta daraltılmış listeden kod atama yöntemine göre, tüm kayıtların %05 ile %08’lik kısmı için uzman görüşüne başvurulması gerektiği sonucuna ulaşılmıştır. Geniş listeden kod atama yönteminin seçilmesi durumunda ise, ilk yöntemden biraz daha fazla olarak tüm kayıtların yaklaşık %2’lik kısmı için uzman görüşüne başvurulması gerektiği ortaya çıkmıştır.

6. SONUÇ VE ÖNERİLER

Araştırmanın amacı, kullanılan sınıflamaya göre en uygun kod atamasını yapabilecek bir bilgisayar programının tasarımını, geliştirmesini ve değerlendirmesini yapmaktır. Program SAS tabanlı olarak geliştirilmiştir. Sistemin etkinliği 2013-2018 yılları arasında TÜİK'in uygulamış olduğu HBA verisi üzerinde test edilmiştir. Geliştirilen sistem COICOP kodları üzerinde denenmiş olmasına rağmen tanım sözlüğü olan ve geliştirilecek yeni tüm sınıflamalarda kullanılma özelliğine sahiptir. Veriler üzerinden yapılan analizler neticesinde elde edilen bulgulara göre ulaşılan sonuçlar iki başlık altında verilmiştir.

Dilbilgisi, Eşleştirme ve Sınıflama Katmanında Elde Edilen Sonuçlar

Çalışma kapsamında, yaklaşık 13 milyon kayıt için anketörün atadığı kodların doğru olup olmadığı KASİS tarafından kontrol edilmiştir. Kontrol sonucunda kodlarla ilgili olarak temel olarak dört olası durum ortaya çıkmaktadır.

1. *Eşleşen kod:* Anketörün atadığı kodun doğru olduğu KASİS tarafından teyit edilmiştir. Eşleşen kodda üç farklı tür bulunmaktadır.

Sözlük eşleşen kod: Anketörün yazdığı harcama tanımı sözlükte birebir yer almaktadır ve anketörün atadığı kodun doğru olduğu KASİS tarafından teyit edilmiştir.

Yeni eşleşen kod: Anketörün yazdığı harcama tanımı sözlükte birebir yer almamakla birlikte dilbilgisi, eşleştirme ve sınıflama katmanlarından geçerek anketörün atadığı kodun doğru olduğu KASİS tarafından teyit edilmiştir.

Bulanık eşleşen kod: Anketörün yazdığı harcama tanımındaki yazım yanlışı KASİS tarafından düzeltilerek anketörün atadığı kodun doğru olduğu KASİS tarafından teyit edilmiştir.

2. *Şüpheli kod:* Anketörün atadığı kodun hatalı olduğu KASİS tarafından teyit edilmiştir.

3. *Yetersiz kod:* Anketörün yazdığı harcama tanımı ile atanabilecek birden fazla kodun olduğu KASİS tarafından teyit edilmiştir.

4. *Kabul kod:* Anketörün atadığı kodun sistem tarafından önerilen kodların arasında olduğu, anketörün atadığı kod ile önerilen kodların ilk beş basamağının aynı olduğu ve önerilen kod sayısının beşten az olduğu KASİS tarafından teyit edilmiştir. Kabul kod ve bulanık kabul kod olmak üzere iki farklı türü bulunmaktadır.

Bulanık kabul kod: Anketörün yazdığı harcama tanımındaki yazım yanlışları KASİS tarafından düzeltilerek kabul kod olduğu KASİS tarafından teyit edilmiştir.

Sistemin belirlemiş olduğu şüpheli kodların kesin olarak düzeltilmesi ve yetersiz kodların tanımlarında daha fazla ayrıntıya yer verilmesi gerekmektedir. Sistemin çalıştırılması sonucunda oluşan bulguların sistematik bir şekilde sunulmasını sağlamak amacıyla tablo, ana grup, bölge, yıl ve ay olmak üzere 5 adet sistem parametresi belirlenmiştir.

Bu parametrelerden ilki tablo parametresidir. Dayanıklı, stok, satın alış ve aynı olmak üzere dört farklı tablo türü bulunmaktadır. Stok ve aynı kodlarında yüksek eşleşme oranları belirlenmiştir. Bu durumun bu tablolardaki kayıt çeşitliliğinin sınırlı olmasından kaynaklandığı düşünülmektedir. Dayanıklı ve satın alış tablosundaki kayıtlar daha çeşitli olduğundan eşleşme oranları nispeten daha düşük seviyede kalmıştır. Dayanıklı tablosundaki kayıtların diğer tablodaki kayıtlardan daha yüksek kabul oranına sahip olması bu tabloda yer alan tanımların uygun kodu verebilecek şekilde daha ayrıntılı olarak yazılması gerektiğine işaret etmektedir. Dayanıklı tablosu kayıtlarında kod ataması yaparken anketörlerin daha kontrollü ve dikkatli olmaları gerektiği ortaya çıkmıştır. Dayanıklı tablosunda, anketörlerin harcama tanımı ile kod tanımı arasındaki uyumu yakalamaları gerekmektedir. İlk etapta bu kayıtların kabul kod olarak sınıflanmaları için anketörlerin yazdığı tanımlarda iyileşme sağlanmalı ikinci etapta bu kayıtların eşleşen kod olarak sınıflanmalarını sağlayacak ayrıntıda tanım yazmaları konusunda anketörler eğitilmelidir.

İkinci olarak belirlenen ana grup parametresinde 12 farklı ana grup bulunmaktadır. Elde edilen bulgulardan, kod çeşitliliğinin fazla olduğu ana gruplarda eşleşme oranının düşük olduğu, tersi durumda ise eşleşme oranının yüksek olduğu belirlenmiştir. Sistemin yaptığı kontroller sonucunda, lokanta ve oteller, eğitim ve haberleşme ana grubu kodlarında diğer ana grup kodlarından daha yüksek düzeyde yetersiz ve şüpheli kod oranı belirlenmiştir. Bu gruplarla ilgili kod ataması yaparken anketörlerin daha kontrollü ve dikkatli olmaları gerektiği ortaya çıkmıştır. Ayrıca, sözlükte bu gruplarla ilgili tanımlarda anketörlerin yüksek

eşleşme oranlarını yakalayabilmeleri için gereken metinsel düzeltmeler yapılmalıdır.

Üçüncü olarak belirlenen bölge parametresinde 26 bölge bulunmaktadır. Bölgesel olarak bulgulara bakıldığında, yüksek oranlarda yetersiz ve şüpheli oranına sahip Van, Siirt, Erzurum ve Konya bölgelerindeki anketörlere yönelik kod atama ve sınıflama eğitimlerinin verilmesi gerektiği ortaya çıkmıştır.

Yıl dördüncü parametre olarak belirlenmiştir. 2015 yılını dışarıda bırakırsak anketörler tarafından tanımların her geçen yıl daha iyi yazılarak doğru kod atamasının yapıldığı belirlenmiştir. 2015 yılında yapılan kod atamalarına ise daha detaylı bakılması gerektiği sonucuna ulaşılmıştır.

Son parametre olarak ay belirlenmiştir. Yılın ilk aylarında yüksek olan yetersiz kayıt oranları, yılın son aylarına doğru düşmüştür. Kahvehanede içilen çay, taze üzüm, elma, çilek şeftali ve portakal gibi mevsimsel özellikli ürün ve hizmetlerin bu konuda etkili olduğu sonucuna ulaşılmıştır. 2017 yılına kadar bölgelerde görevli anketörler her yılın son haftasında işten çıkarılır yeni yılın ilk ayı ortasında tekrar işe alınırdı. Özellikle yılın son ayı ile ilk ayında anketör performansındaki düşüş nedeniyle kod atamalarında daha çok hata olabileceği beklentisi yapılan analizlerde ortaya çıkmamıştır.

Smith [14] ICD kodları ile ilgili yaptığı araştırmada kodlama hataları yapıldığını belirlemiş ve Rossi ve diğ. [15] kodların elle atanması esnasında atamayı yapan personel yanlışlığının kodlama hatalarına sebebiyet verdiğini ortaya koymuştur. Appel ve Hellermann [16] ise kodların elle atanmasının maliyetli olmasına vurgu yapmıştır. Benzer sonuçlar bu çalışma sonucunda da ortaya çıkmıştır. Anketörün atadığı kodların gözle kontrol edilirken gösterilen çaba ve harcanan emeğe rağmen hala kodlarda eksikliklerin kaldığı belirlenmiştir. KASİS bize hangi tabloda, hangi ana grupta, hangi bölgede, hangi yılda ve hangi ayda bu eksikliğin bulunduğunu listeleyebilmektedir. Sistem tarafından hangi kaydın doğru sınıflandığı, hangi kaydın yanlış sınıflandığı ve yanlış sınıflanan bu kayıtlara nasıl bir işlem yapılacağı bilgisinin sistem tarafından verildiği düşünüldüğünde; gözle veya elle yapılan yöntemlerle yapılabilmesi imkânsız bir görev sistem tarafından yerine getirilmektedir. Ayrıca, kodlama hataları barındıran veri ile yapılacak istatistiksel analizler ve kurulacak modellerde dikkatli olunması gerektiği Belloni ve diğ. [19] tarafından ortaya koyulmuştur. Çıkan bu listeye göre gereken önlemlerin alınarak kodlardaki eksikliklerin gelecekte

oluşmaması sağlanmalı ve sonucunda personel sayısı ile diğer maliyet kalemlerinde tasarruflar sağlanmalıdır.

Kod Atama Katmanında Elde Edilen Bulgular

Kod ataması yapılırken, daraltılmış listeden ve geniş listeden olmak üzere iki farklı yöntem kullanılmıştır. Anketör tarafından yapılan kod atamasının doğruluğunun sistem tarafından teyit edilemediği kayıtlara bu iki farklı yöntem kullanılarak tekrar kod ataması yapılarak bulgular ortaya koyulmuştur. Sistem tarafından kod ataması şüpheli bulunan veya tanımı kod ataması yapabilecek kadar ayrıntı içermeyen kayıtları veri setinden dışlamadan önce araştırmacılara alternatif bir yöntem sunmak ve bulunan bu eksiklikleri sonuçları resmi olarak açıklanmış verileri düzeltebilmek için bir yöntem önermektir. Bu yöntemlerden daraltılmış listeye uygulanan ilk yöntemin daha iyi sonuç verdiği ortaya konulmuştur. Ancak sözlükte yer alan tanımlara göre listeyi daraltamadığınız durumlarda bulanık eşleştirme algoritmalarını kullanarak kod doğrulaması yapmak her zaman başvurulması gereken bir yöntem olmak zorundadır.

Otomatik kod atama programlarının performansını karşılaştırmak için genellikle kullanılan kriterler; verimlilik, güvenilirlik ve hız olmaktadır. Ancak, bu kriterler mutlak değildir [97]. KASİS ile kayıtların %94,9'una anketör ile aynı kod ataması yapılarak bu kayıtlarda yapılan kod atamanın doğru olduğu sonucuna ulaşılmıştır.

Otomatik kod atama sisteminin doğru kodu atama performansı, kullanılan sınıflamanın ve verinin karmaşıklığına bağlı olarak değişebilmektedir. Ancak genel bir perspektif sunabilmesi açısından, KASİS'in kod atama performansı benzer sistemler ile karşılaştırıldığında başarılı olduğu sonucuna ulaşılmış ve KASİS'in atadığı kodlar ile anketörün atadığı kodlar karşılaştırıldığında aralarında bir fark olmadığı belirlenmiştir. Roessingh ve Bethlehem [98], aile harcama anketinde üç farklı yöntemle otomatik kod ataması gerçekleştirerek %94, %85, %78 oranlarında doğru kod ataması yapmışlardır. Yeni Zelanda İstatistik Ofisi, Census 2013 verilerinde meslek ve okul sonrası yeterlilik değişkenlerini kod atamak için SVM algoritmasını kullanmışlardır. Her iki değişken için test verilerinde %50 doğruluk oranına ulaşmışlardır [46]. Tourigny ve Moloney [99], 1991 Kanada Nüfus Sayımı'nda yer alan yedi farklı değişken için yapılan otomatik kod atama sonucunda %92'lik doğruluk oranına ulaşmışlardır. Haslinger [100], Avusturya Nüfus

Sayımı'ndaki çalışılan yer değişkenine %96, eğitim değişkenine %92 ve iktisadi faaliyet değişkenine %50 oranında otomatik kod atama sistemi ile kod ataması yapmıştır.

Öneriler

Çalışma kapsamında oluşturulan sistemin etkinliği yayınlanmış veri üzerinde test edilerek bulgu ve sonuçlar ortaya konulmuştur. Bu sistemin, istatistik üretim aşamasında canlı veri üzerinde uygulanabilirliği sağlanmalıdır. Canlı veri üzerinde uygulanabilirliği iki farklı yöntemle sağlanabilir. Birinci yöntem olarak veri giriş programına ilave edilecek dinamik kontroller ile bu işlem gerçekleştirilebilir. Bu sayede anketörün yapacağı kodlama hatası anketöre dinamik olarak iletilerek hata oluşmadan kaynağında engellenmesi sağlanabilir. İkinci yöntem olarak anketör veri girişi yaptıktan ve veriyi veri tabanına gönderdikten sonra buradaki veri üzerinde kodlama hatalarını bulma amaçlı kontroller gerçekleştirilebilir. Daha sonra anketöre hangi kayıta hata yaptığı bilgisi iletilerek hatanın düzeltilmesi sağlanabilir. Sonuç olarak, hangi yöntem seçilirse seçilsin sonuçlar kamuoyu ile paylaşılmadan önce kod atamasında yapılan hata ve eksikliklerin giderilmesi ve kod atama kalitesinin artırılması sağlanmalıdır.

Bu sistemin çıktıları hangi anketörün veya bölgenin sınıflama konusunda eğitime ihtiyacı olduğunun belirlenmesi için eğitim ve takip amaçlı olarak kullanılabilir. Bu kapsamda, bölgesel ve anketör bazlı bir performans takip sistemi kurulmalıdır.

Ulaşılan bulgu ve sonuçlardan anketörün kod ataması ile KASİS'in kod ataması arasında önemli farklar olmadığı belirlenmiştir. Bu kapsamda, anketörün kod ataması yapması yerine otomatik yöntemlerle kod ataması yapılması sonucunda anketörlerin üzerindeki iş yükü azaltılmış ve atanan kodlarda doğruluk sağlanmış olacaktır. Ayrıca, anketör sayısında azalma ve diğer maliyetlerde azalma gerçekleşecektir. Bu sistemin, TÜFE'de kullanılacak olan modern fiyat toplama yöntemleri için önemli olan kod atama faaliyetini yerine getirebileceği de ortaya konulmuştur. Sistem bu amaçla da kullanılmalıdır.

Geliştirilen sistem şu an SAS tabanlı olarak çalışabilmektedir. SAS tabanlı programlar görselliğe pek fazla müsaade etmediği için kullanıcı dostu değildir. Sistemin gelecek versiyonlarının python, Java ve C programlama dilleri kullanılarak oluşturulmalı ve bu

sayede hem sistemin kullanıcı dostu bir hale gelmesi hem de sistemin hızı ve verimliliğinin artması sağlanmalıdır.



KAYNAKLAR

1. İnternet: Türkiye İstatistik Kurumu Kalite Güven Çerçevesi. (2015). URL: http://web.archive.org/web/20190818103144/http://www.tuik.gov.tr/jsp/duyuru/upload/TUIK_Kalite_Guvence_Cercevesi.pdf, Son Erişim Tarihi: 18.08.2019.
2. Akdeniz, F. (2013). İstatistikte Yeni Eğilimler ve Yöntemler. *Journal of Statistical Research*, 10(3), 35-48.
3. İnternet: Türkiye İstatistik Kurumu Sınıflama Sunucusu. (2006). URL: <http://web.archive.org/web/20161016152805/https://biruni.tuik.gov.tr/DIESS/>, Son Erişim Tarihi: 18.08.2019.
4. Simões, M.d.G., Freitas, M. C. V. d., Rodríguez-Bravo, B. (2016). Theory of classification and classification in libraries and archives: Convergences and divergences. *Knowledge Organization*, 43(7), 530-538.
5. International Organization for Standardization. (1988). ISO 5127-6-1988. Documentation et information – Vocabulaire: partie 6: langages documentaires. In Documentation et information: recueil de normes ISO I. Genève: ISO, 87-111.
6. İnternet: Türkiye İstatistik Kurumu Sınıflama Sunucusu. (2006). URL: <http://web.archive.org/web/20160414175213/https://biruni.tuik.gov.tr/DIESS/SiniflamaTurListeAction.do>, Son Erişim Tarihi: 31.08.2019.
7. İnternet: Türkiye Odalar ve Borsalar Birliği NACE Rev.2 Altılı Ekonomik Faaliyet Sınıflaması Hakkında Duyuru. (2011). URL: http://web.archive.org/web/20180713151926/http://gen.tobb.org.tr/ggnot/images/bilgi_notu/258_DUYURU-GD1.pdf, Son Erişim Tarihi: 31.08.2019.
8. İnternet: 5429 sayılı Türkiye İstatistik Kanunu. (2005). URL: <http://www.mevzuat.gov.tr/MevzuatMetin/1.5.5429.pdf>, Son Erişim Tarihi: 31.08.2019.
9. Schierholz, M. (2014). *Automating Survey Coding for Occupation*, Yüksek Lisans Tezi. Ludwig Maximilians Universität Institut für Statistik, München.
10. Statistics Netherlands. (2012). *Theme: Coding; Interpreting Short Descriptions Using a Classification*. The Hague/Heerlen: Statistics Netherlands.
11. Aggarwal, C. C., Zhai, C. A. (2012). *Mining Text Data*. Boston: Springer, 163-222.
12. Groves, R. M., Fowler, F. J., Couper, M. P., Lepkowski, J. M., Singer, E., Tourangeau, R. (2009). *Survey Methodology*, New Jersey: John Wiley & Sons, 329-367.
13. Esuli, A., Sebastiani, F. (2010). Machines That Learn How To Code Open-Ended Survey Data. *International Journal of Market Research*, 52(6), 775-800.

14. Smith, M. (1989). Hospital discharge diagnoses: How accurate are they and their International Classification of Diseases (ICD) Codes? *The New Zealand medical journal*. 102(1), 507-8.
15. Rossi, C., Alberti, V., Mancino, G., Flor, L., Martello, T., Poeta, L., Lise, M., Diana, L., Puppini, G. (1993). Comparison between manual and automatic coding of medical record statistical cards at a university hospital. *Medical informatics*, 18(1), 53-9.
16. Appel, M. V., Hellerman, E. (1983). Census Bureau Experiments with Automated Industry and Occupation Coding, Paper presented at Proceedings of the American Statistical Association, 32-40.
17. Altunışık, R. (2008). Anketlerde Veri Kalitesinin İyileştirilmesi İçin Öntest (Pilot Test) Yöntemler. *Pazarlama ve Pazarlama Araştırmaları Dergisi*, 2(1), 1-17.
18. Meade, A. W., Craig, S. B. (2012). Identifying careless responses in survey data. *Psychological Methods*, 17(1), 437–455.
19. Belloni, M., Brugiavini, A., Meschi, E., Tijdens, K. (2016). Measuring and detecting errors in occupational coding: an analysis of share data. *Journal of Official Statistics*, 32(4), 917-945.
20. İnternet: United Nations Statistical Commission and Economic Commission for Europe Conference of European Statisticians Statistical Standards And Studies - No. 48 Statistical Data Editing Volume No. 2 Methods and Techniques. (1997). URL: <http://web.archive.org/web/20190903100906/http://www.unece.org/fileadmin/DAM/stats/publications/editing/SDE2.pdf>, Son Erişim Tarihi: 03.09.2019.
21. Eurostat. (1997). *New Techniques and Technologies for Statistics II Proceedings of the Second Bonn Seminar*, Netherlands: IOS Press, 33-44.
22. Hilden, J., Habbema, J.D.F, Bjerregaard, B. (1978a). The measurement of performance in probabilistic diagnosis, I. The problem, descriptive tools, and measures based on classification matrices. *Methods of information in medicine*, 17(1), 217-226.
23. Hilden, J., Habbema, J.D.F, Bjerregaard, B. (1978b). The measurement of performance in probabilistic diagnosis, II. Trustworthiness of the exact values of the diagnostic probabilities. *Methods of information in medicine*, 17(1), 227- 237.
24. Hilden, J., Habbema, J.D.F, Bjerregaard, B. (1978c). The measurement of performance in probabilistic diagnosis, III. Methods based on continuous 54 functions of the diagnostic probabilities, *Methods of information in medicine*, 17(1), 238-246.
25. Campanelli, P., Thomson, K., Moon, N., Staples, T. (1997). The quality of occupational coding in the United Kingdom. In L. Lyberg, P. Biemer, M. Collins, E. De Leeuw, C. Dippo, N. Schwarz and D. Trewin (Eds.), *Survey Measurement and Process Quality*, New York: Wiley, pp. 437–453.

26. İnternet: Eurostat Household Budget Surveys in the EU, Methodology and recommendations for harmonisation. (2003). <http://web.archive.org/web/20190713043155/https://ec.europa.eu/eurostat/web/products-manuals-and-guidelines/-/KS-BF-03-003?inheritRedirect=true>, 06.09.2019.
27. Oğur, R., Tekbaş, Ö. F. (2003). Anket Nasıl Hazırlanır?. *Sürekli Tıp Eğitimi Dergisi*, 12(9), 336-341.
28. Oğuzlar, A., Kızılkaya, Y. M. (2018). Bazı denetimli öğrenme algoritmalarının R programlama dili ile kıyaslanması. *Karadeniz Uluslararası Bilimsel Dergi*, 37(37), 90-98.
29. İnternet: Eurostat Harmonised Index of Consumer Prices Practical Guide for Processing Supermarket Scanner Data. (2017). URL: <http://web.archive.org/web/20190623220837/https://circabc.europa.eu/sd/a/8e1333df-ca16-40fc-bc6a-1ce1be37247c/Practical-Guide-Supermarket-Scanner-Data-September-2017.pdf>, Son Erişim Tarihi: 02.09.2019.
30. Breton, R., Flower, T., Mayhew, M., Metcalfe, E., Milliken, N., Payne, C., Smith, T., Winton, J., Woods, A. (2016). Research indices using web-scraped data: May 2016 update. Technical Report. Office for National Statistics, Newport.
31. İnternet: Türkiye İstatistik Kurumu Hanehalkı Bütçe Anketi Genel Açıklamalar. (2017). URL: <http://web.archive.org/web/20190903075621/http://www.tuik.gov.tr/PreHaberBultenleri.do?id=27840>, Son Erişim Tarihi: 10.09.2019.
32. İnternet: Türkiye Odalar ve Borsalar Birliği Sorularla Örnekleme, Araştırma ve Veri Derleme Sorularla Merak Ettikleriniz Yayın Serisi-2. (2013). URL: <http://web.archive.org/web/20180729130254/https://www.tobb.org.tr/Documents/yayinlar/seri-yayinlar/SY-Sorular-ornekler.pdf>, Son Erişim Tarihi: 02.09.2019.
33. İnternet: The World Bank DIME Wiki. (2019). URL: [http://web.archive.org/web/20190902151422/https://dimewiki.worldbank.org/wiki/Pen-and-Paper_Personal_Interviews_\(PAPI\)](http://web.archive.org/web/20190902151422/https://dimewiki.worldbank.org/wiki/Pen-and-Paper_Personal_Interviews_(PAPI)), Son Erişim Tarihi: 02.09.2019.
34. İnternet: Türkiye İstatistik Kurumu Kamuoyu Duyurusu. (2019). URL: https://web.archive.org/web/20190902123241/http://www.tuik.gov.tr/duyurular/duyuru_4054.pdf, Son Erişim Tarihi: 02.09.2019.
35. United Nations. (2018). *Classification of Individual Consumption According to Purpose (COICOP) 2018*. New York: United Nations.
36. International Labour Organization. (1926). *Methods of Conducting Family Budget Enquiries*. Geneva: International Labour Organization.
37. İnternet: Eurostat Household Budget Survey (HBS). URL: <http://web.archive.org/web/20190623214549/https://ec.europa.eu/eurostat/web/household-budget-surveys/policy-context>, 07.09.2019.

38. Conrad, F. G. (1997). Using Expert Systems to Model and Improve Survey Classification Processes. In L. Lyberg, P. Biemer, M. Collins, E. De Leeuw, C. Dippo, N. Schwarz and D. Trewin (Eds.), *Survey Measurement and Process Quality*, New York: Wiley, pp. 393–414.
39. Bushnell, D. (1998, August). *An Evaluation of Computer-assisted Occupation Coding*. Paper presented at Proceedings of the International Conference, Southampton.
40. Fielding, J., Fielding, N., Hughes G. (2013). Opening Up Open-Ended Survey Data Using Qualitative Software, *Quality & Quantity*, 47(6), 3261–3276.
41. Clarke, F. R., Brooker S. (2011, August). *Use of Machine Learning for Automated Survey Coding*. Paper presented at International Statistical Institute Proceedings of the 58th World Statistics Congress, Dublin.
42. Geis, A. (2009). Handbuch für die Berufsvercodung. *Center for Survey Design and Methodology*, GESIS, Mannheim.
43. Drasch, K., Matthes, B., Munz, M., Paulus, W., Valentin, M. A. (2012). Arbeiten und Lernen im Wandel: Teil V: Die Codierung der offenen Angaben zur beruflichen Tätigkeit, Ausbildung und Branche.
44. Jung Y., Yoo J., Myaeng SH., Han DC. (2008). A Web-Based Automated System for Industry and Occupation Coding. In: Bailey J., Maier D., Schewe KD., Thalheim B., Wang X.S. (Eds.) *Web Information Systems Engineering - WISE 2008*. WISE 2008. Lecture Notes in Computer Science, vol 5175. Springer, Berlin, Heidelberg.
45. Internet: Die Klassifizierung der Berufe und der Wirtschaftszweige im Sozio-oekonomischen Panel. (2002). URL: http://web.archive.org/web/20180420083205/https://www.diw.de/documents/dokumente/17/diw_01.c.40132.de/vercodung.pdf, Son Erişim Tarihi: 04.09.2019.
46. Chu, K., Poirier, C. (2015, April). *Machine learning documentation initiative*. Paper presented at Workshop on the Modernisation of Statistical Production, Switzerland.
47. Bethmann, A., Schierholz, M., Wenzig, K., Zielonka M. (2014, August). *Automatic Coding of Occupations Using Machine Learning Algorithms for Occupation Coding in Several German Panel Surveys*. Paper presented at Proceedings of Statistics Canada Symposium, Canada.
48. Internet: Beck, M., Dumpert, F., Feuerhake J. (2018, December). Machine Learning in Official Statistics, URL: <https://arxiv.org/abs/1812.10422>, Son Erişim Tarihi: 04.09.2019.
49. De Matteis, S., Jarvis, D., Young, H., Young, A., Allen, N., Potts, J., Darnton, A., Rushton, L., Cullinan, P. (2017). Occupational self-coding and automatic recording (OSCAR): a novel web-based tool to collect and code lifetime job histories in large population-based studies. *Scandinavian Journal of Work, Environment & Health*, 43(2), 181-186.

50. Wright, G. (2011, October). *Probabilistic Record Linkage in SAS*. Paper presented at Proceedings of Western Users of SAS Software, California.
51. İnternet: Statistics Canada Directive on Microdata Linkage. (2017). URL: <http://web.archive.org/web/20190418100828/https://www.statcan.gc.ca/eng/record/policy4-1>, Son Erişim Tarihi: 05.09.2019.
52. Winkler, W. E. (1999, June). *The state of record linkage and current research problems*. Paper presented at Proceedings of the Survey Methods Section of SSC Annual Meeting, Regina.
53. Tejada, S., Knoblock, C., Minton, S. (2001). Learning Object Identification Rules for Information Extraction, *Information Systems*, 26 (8), 607-633.
54. Do, H.H., Rahm, E. (2001). *COMA – A system for flexible combination of schema matching approaches*. Paper presented at Proceedings of the Very Large Databases Conference.
55. Gravano, L., Ipeirotis, P. G., Jagadish, H. V., Koudas, N., Muthukrishnan, S., Srivastava, D. (2001). *Approximate String Joins in a Database (Almost) for Free*. Paper presented at Proceedings of the Very Large Databases Conference.
56. Ananthakrishna, R., Chaudhuri, S., Ganti, V. (2002). *Eliminating Fuzzy Duplicates in Data Warehouses*. Paper presented at Proceedings of the Very Large Databases Conference.
57. Benjelloun, O., Garcia-Molina, H., Su, Q., Widom, J. (2005). Swoosh: A Generic Approach to Entity Resolution, Stanford University technical report, Stanford.
58. İnternet: Statistics Canada Reports on Special Business Projects an Overview of Selected International Business Record Linkage Programs. (2016). URL: <http://web.archive.org/web/20190702090512/https://www150.statcan.gc.ca/n1/pub/18-001-x/18-001-x2016001-eng.htm>, Son Erişim Tarihi: 05.09.2019.
59. Hernandez, M. A., Stolfo, S. J. (1995). The Merge/Purge Problem for Large Databases, *ACM SIGMOD Record*, 24(1), 127–138.
60. Elfeky, M. G., Verykios, V. S., Elmagarmid, A. K. (2002). *TAILOR: A Record Linkage Toolbox*. Paper presented at Proc. 18th IEEE Int'l Conf. Data Eng.
61. Verykios, V.S., Moustakides, G. V., Elfeky, M. G. (2002). A Bayesian Decision Model for Cost Optimal Record Matching, *The VLDB Journal*, 12(1), 28–40.
62. Gu, L., Baxter, R., Vickers, D., Rainsford, C. (2003). Record linkage: Current practice and future directions, *Commonwealth Scientific and Industrial Research Organisation Mathematical and Information Science*, 3.
63. van Herk-Sukel, M.P., Lemmens, V. E., van de Poll-Franse, L., Herings, R. M., Coebergh, J. W. (2012). Record linkage for pharmacoepidemiological studies in cancer patients, *Pharmacoepidemiology and Drug Safety*, 21(1), 94–103.

64. Herings, R., M., C. (1993). *PHARMO: A record linkage system for postmarketing surveillance of prescription drugs in The Netherlands*. Utrecht University.
65. Reitsma, J., B. (1999). *Registers in cardiovascular epidemiology*. UVA University.
66. Statistics Netherlands, (2014). *Record linkage in health data: a simulation study*. Hague/Heerlen: Statistics Netherlands.
67. Porter, E. H., Winkler W. E. (1997). Approximate string comparison and its effect on an advanced record linkage system, *SRD research report*.
68. Oberaigner, W. (2007). Errors in survival rates caused by routinely used deterministic record linkage methods, *Methods of Information in Medicine*, 46(1), 420–424.
69. Shlomo N. (2019). Overview of Data Linkage Methods for Policy Design and Evaluation. In Crato N., Paruolo P. (Eds.), *Data-Driven Policy Impact Evaluation*. Springer, Cham, pp. 47-65.
70. Lafler, K. P., Sloan, S. (2018, October). *A Quick Look at Fuzzy Matching Programming Techniques Using SAS Software*. Paper presented at MWSUG 2018 Conference Proceedings.
71. Dunn, T. (2014). *Getting the Warm and Fuzzy Feeling with Inexact Matching*. Paper presented at Proceedings of the 2014 SAS Global Forum (SGF) Conference.
72. Patman, F., Thompson, P. (2003). *Names: A new frontier in text mining*. Paper presented at in ISI-2003.
73. Friedman, C., Sideli, R. (1992). Tolerating spelling errors during patient validation, *Computers and Biomedical Research*, 25(1), 486–509.
74. Christen, P. (2006). *A Comparison of Personal Name Matching: Techniques and Practical Issues*. Paper presented at Proc. IEEE Sixth Data Mining Workshop.
75. Hall, P. A., Dowling, G. R. (1980). Approximate string matching. *ACM Computing Surveys*, 12(4), 381–402.
76. Şahin, İ., Temür, G. (2016). *Levenshtein Mesafe Algoritması İçin Bir Donanım Modülü Tasarımı*. 5. Uluslararası Meslek Yüksekokulları Sempozyumu'nda sunuldu, Kosova.
77. İnternet: Jaro Winkler Distance. (2019). URL: <http://web.archive.org/web/20190228054643/http://alias-.com/lingpipe/docs/api/com/aliasi/spell/JaroWinklerDistance.html>, Son Erişim Tarihi: 09.09.2019.
78. İnternet: Levenshtein Mesafe Algoritması. (2010). URL: <http://web.archive.org/web/20181219194638/http://bilgisayarkavramlari.sadievrenseker.com/2010/12/30/levenshtein-mesafe-algoritmasi-levenshtein-distance>, Son Erişim Tarihi: 09.09.2019.

79. İnternet: SAS COMPGED Function. (2011). URL: <http://web.archive.org/web/20110214081910/http://support.sas.com/documentation/cdl/en/lrdict/64316/HTML/default/viewer.htm#a002206133.htm>, Son Erişim Tarihi: 12.09.2019.
80. İnternet: n-gram. (2011). URL: <http://web.archive.org/web/20190504163236/http://bilgisayarkavramlari.sadievrenseker.com/2011/04/23/n-gram>, Son Erişim Tarihi: 09.09.2019.
81. Richey, R. C., Klein, J. D. (2008). Research on design and development. In J. M. Spector, M. D. Merrill, J. van Merriënboer, M. P. Driscoll (Eds.), *Handbook of research for educational communications and technology*. Mahwah, New Jersey: Lawrence Erlbaum Associates, pp. 748–757.
82. Richey, R. C., Klein, J. D. (2014). Design and development research. In J. M. Spector, M. D. Merrill, J. Elen, M. J. Bishop (Eds.), *Handbook of research on educational communications and technology*. Mahwah, New York: Springer, pp. 141–150.
83. Büyüköztürk, Ş., Kılıç Çakmak, E., Akgün, Ö. E., Karadeniz, Ş., Demirel, F. (2015). *Bilimsel Araştırma Yöntemleri* (Genişletilmiş 19. Baskı). Ankara: Pegem Akademi Yayıncılık.
84. Sommerville, I. (2015). *Software Engineering (10th Edition)*. New Jersey: Pearson Prentice Hall.
85. Ersöz, S., Ersöz, O.Ö. (2015). *İşletmelerde Bilgi Sistemleri Uygulamalı Örneklerle*. Ankara: Nobel Akademik Yayıncılık.
86. Gökçen, H. (2011). *Yönetim Bilgi/Bilişim Sistemleri: Analiz ve Tasarım*. Ankara: Afşar Matbaacılık.
87. Kendall, K. E., Kendall, J. E. (2011). *Systems analysis and design*. New Jersey: Pearson Prentice Hall.
88. Türkiye İstatistik Kurumu Hanehalkı Bütçe Anketi Mikro Veri Seti, CD. (2013). Ankara.
89. Türkiye İstatistik Kurumu Hanehalkı Bütçe Anketi Mikro Veri Seti, CD. (2014). Ankara.
90. Türkiye İstatistik Kurumu Hanehalkı Bütçe Anketi Mikro Veri Seti, CD. (2015). Ankara.
91. Türkiye İstatistik Kurumu Hanehalkı Bütçe Anketi Mikro Veri Seti, CD. (2016). Ankara.
92. Türkiye İstatistik Kurumu Hanehalkı Bütçe Anketi Mikro Veri Seti, CD. (2017). Ankara.

93. Türkiye İstatistik Kurumu Hanehalkı Bütçe Anketi Mikro Veri Seti, CD. (2018). Ankara.
94. İnternet: Bakanlar Kurulu'nun 2002/4720 sayılı Kararı. (2002). URL: <http://web.archive.org/web/20190215042935/http://www.resmigazete.gov.tr/eskiler/2002/09/20020922.htm>, Son Erişim Tarihi: 11.09.2019.
95. Cadieux, R., Daniel R. B. (2014). *Matching Rules: Too Loose, Too Tight, or Just Right?*. Paper presented at Proceedings of the 2014 SAS Global Forum (SGF) Conference.
96. Popescu A. D., Ercegovac V., Balmin A., Branco M., Ailamaki A. (2012). *Same Queries, Different Data: Can We Predict Runtime Performance?*, Paper presented at IEEE 28th International Conference on Data Engineering Workshops.
97. Riviere, P. (1995). *Outline of a theory of automated coding*. Paper presented at Conference of European Statisticians, Athens.
98. Roessingh, M., Bethlehem, J. (1983). Trigram coding in the family expenditure survey in statistics, Netherlands Central Bureau of Statistics.
99. Tourigny, J. Y., Moloney, J. (1997). Statistical Data Editing. Paper presented at United Nations Statistical Commission and Economic Commission for Europe, New York and Geneva.
100. Haslinger. A. (1997). *Automatic Coding and Text Processing using N-grams*. Paper presented at Conference of European Statisticians, New York and Geneva.



EK-1. İçerisinde “ekmek” ifadesi geçen COICOP kodlarının listesi

METİNSEL İFADE	COICOP KOD	COICOP KOD TANIMI
ekmek	0111301010	Ekmek Normal Ekmek
	0111301020	Ekmek Kepekli Ekmek
	0111301030	Ekmek Belediye Ekmeği (Halk Ekmeği) (Normal)
	0111301040	Ekmek Belediye Ekmeği (Halk Ekmeği) (Kepekli)
	0111301050	Ekmek Ata Ekmek
	0111301060	Ekmek Baston Ekmek (Baget Ekmek)
	0111301070	Ekmek Bazlama (Serme, Patil Ekmek vb.)
	0111301080	Ekmek Cevizli Ekmek
	0111301090	Ekmek Çavdar Ekmeği
	0111301100	Ekmek Çiçek Ekmek
	0111301110	Ekmek Francala
	0111301120	Ekmek Hamburger Ekmeği
	0111301130	Ekmek Haşhaşlı Ekmek
	0111301140	Ekmek Isparta Ekmeği
	0111301150	Ekmek Karışık (Buğday, Çavdar vb. tahıl karışımı)
	0111301160	Ekmek Köy Ekmeği
	0111301170	Ekmek Lavaş
	0111301180	Ekmek Mısır Ekmeği
	0111301190	Ekmek Patatesli Ekmek
	0111301200	Ekmek Pide (Ramazan Pidesi, Tırnak Pide vb.)
	0111301210	Ekmek Pide Tahinli
	0111301220	Ekmek Sandviç Ekmeği
	0111301230	Ekmek Soyalı Ekmek
	0111301240	Ekmek Susamlı Ekmek
	0111301250	Ekmek Sütlü Ekmek
	0111301260	Ekmek Tahinli Ekmek
	0111301270	Ekmek Tandır Ekmeği
	0111301280	Ekmek Tost Ekmeği
	0111301290	Ekmek Trabzon Ekmeği
	0111301300	Ekmek Tuzsuz Ekmek
	0111301310	Ekmek Yufka Ekmeği
	0111301320	Ekmek Yulafli Ekmek
	0111301330	Ekmek Zeytinli Ekmek
	0111301340	Ekmek Diyet Ekmeği
	0111301350	Ekmek Katlama (Yöresel)
	0111301360	Ekmek Bebe Ekmeği
	0111301370	Ekmek Tatlı Maya Ekmeği
	0111301380	Ekmek Tam Buğday Unundan Ekmek

EK-1. (Devam) İçerisinde “ekmek” ifadesi geçen COICOP kodlarının listesi

METİNSEL İFADE	COICOP KOD	COICOP KOD TANIMI
ekmek	0111301390	Ekmek Belediye Ekmeği (Halk Ekmeği) (Köy Ekmeği)
	0111301400	Ekmek Belediye Ekmeği (Halk Ekmeği) (Cabata (Zeytinyağlı) Ekmeği)
	0111301410	Ekmek Belediye Ekmeği (Halk Ekmeği) (Çeşnibahar Ekmeği)
	0111301420	Ekmek Belediye Ekmeği (Halk Ekmeği) (Fındıklı Ekmek)
	0111301430	Ekmek Belediye Ekmeği (Halk Ekmeği) (Somun Ekmek)
	0111301440	Ekmek Belediye Ekmeği (Halk Ekmeği) (Glutensiz Ekmek)
	0111301450	Ekmek Bartın (Çöven) Ekmeği
	0111301460	Ekmek Yağlı Pide
	0111301470	Ekmek Roll Ekmek
	0111301480	Ekmek Nohutlu Ekmek
	0111301490	Ekmek Tava Ekmeği
	0111301500	Ekmek Gobit Ekmeği
	0111301510	Ekmek Belediye Ekmeği (Halk Ekmeği) (Şampiyon Ekmek)
	0111301520	Ekmek Belediye Ekmeği (Halk Ekmeği) (Çavdar Ekmeği)
	0111301530	Ekmek Belediye Ekmeği (Halk Ekmeği) (Tam Buğday Ekmeği)
	0111301540	Ekmek Belediye Ekmeği (Halk Ekmeği) (Tost Ekmeği)
	0111301550	Ekmek Alman Ekmeği
	0111301560	Ekmek Balina Ekmek
	0111301570	Ekmek Cabata (Zeytinyağlı) Ekmek
	0111301580	Ekmek Keten Tohumlu Ekmek
	0111301590	Ekmek Ruşeymli Ekmek
	0111301600	Ekmek Katıklı Ekmek (Yöresel)
	0111301610	Ekmek Organik (Ekşi Maya ile Yapılan-Her Çeşit) Ekmek
	0111301980	Ekmek Diğer

EK-2. Mikro verilere erişim izin yazısı



T.C.
TÜRKİYE İSTATİSTİK KURUMU BAŞKANLIĞI
Bilgi Dağıtım Ve İletişim Daire Başkanlığı

Sayı : 27964695-622.03-E.1668
Konu : Tüketim Harcamaları İstatistikleri

22.01.2018

Sayın Levent AHİ

İlgi : 17.01.2018 tarihli yazınız.

TÜİK, İşgücü ve Yaşam Koşulları Dairesi Başkanlığı adına, "Uluslararası İstatistiksel Sınıflamalara Yönelik Kod Atama Sistemi" konulu araştırma projesi için talep ettiğiniz B Grubu Hanehalkı Bütçe Araştırması 2012-2016 yılları mikro verileri; 11/10/2012 tarihli B.22.1.TÜİ.01.78.01.02-622.02-4253 sayılı kurum personellerine başvuru formunu doldurup göndermeleri ve çalışmalarının kayıt altına alınmasını takiben, herhangi bir ücret ödemeksizin mikro verileri temin etmesine/kullanmasına erişim hakkı sağlayan Başkanlık oluru ile verilmesine Değerlendirme Kurulunca karar verilmiştir. Ekte verilen taahhütnamenin imzalanması durumunda talep ettiğiniz mikro veriler tarafınıza elektronik ortamda iletilecektir.

Mikro Veri kullanılarak yapılan tüm araştırmalara, Kurum web sitesi ana sayfası kütüphane linki aracılığıyla veya <http://kutuphane.tuik.gov.tr/yordambt/yordam.php> linkinde arama bölümünde "TÜİK Mikro Verileri ile Hazırlanan Çalışmalar" seçilerek erişim sağlayabilirsiniz.

Bilgilerinizi rica ederim.



e-imzalıdır
Abdullah ÖZDEMİR
Daire Başkanı V.

Ek : Taahhütname (1 Sayfa)

Not: 5070 sayılı elektronik imza kanununun 5 maddesi gereği bu belge elektronik imza ile imzalanmıştır.

Devlet Mahallesi Necatibey Caddesi No:114 Çankaya/Ankara
Tel: (0312) 410 02 22 Faks: (0312) 417 04 32
KEP: tuikbilgidaagitimveiletisimdairesibaskanligi@hs01.kep.tr

Bilgi için: Ümit ÖĞÜT
Uzman
Telefon No: (312) 410 02 47

EK-3. Hanehalkı harcama kayıt defteri

FORM 2

T.C.
TÜRKİYE İSTATİSTİK KURUMU
BAŞKANLIĞI

Bülten No: _____

Ay : _____

1. HAFTA (1-5. GÜNLER)

DİKKAT: Bu anket hanehalkına gelir, sosyal-ekonomik durum ve tüketim yapısı hakkında bilgi edinmek amacı ile 10.11.2005 tarih ve 5429 sayılı Kanun uyarınca yapılmaktadır. Toplanan bilgiler *gizli* tutulacak olup, *istatistik dışındaki başka hiçbir amaçla kullanılmaması, açıklanmaması, kayıt aracı yapılmaması.* Bu soru kağıdına zamanında veya doğru cevap vermemeyişler ile toplanan kişisel bilgilerin açıklanması 5429 Sayılı Kanunun 53 ve 54 üncü maddelerine göre cezalandırılır.

HANEHALKI BÜTÇE ANKETİ HARCAMA KAYIT DEFTERİ

Hanehalkı Sorumlusunun Adı Soyadı

Anketörün Adı Soyadı

Kontrolörün Adı Soyadı

Görüşme Yapılan Hanehalkı Fertinin Adı Soyadı

Hanehalkuna yapılacak olan bir sonraki ziyaret tarihi

____ / ____ / _____

Bölge Müdürlüğü	İl	İlçe	Bucak	Köy	Mahalle	Cadde / Sokak	Dış kapı no	İç kapı no
_____	_____	_____	_____	_____	_____	_____	_____	_____
[]	[]	[]	[]	[]	[]	_____	_____	_____

T.C.
TÜRKİYE İSTATİSTİK KURUMU
BÖLGE MÜDÜRLÜKLERİ TELEFON LİSTESİ

BÖLGE MÜDÜRLÜĞÜ ADI	BAGLI OLAN İLLER	OTOMATİK KOD	TELEFON NO	GSM NO
ADANA BÖLGE MÜDÜRLÜĞÜ	Adana, Mersin	322	457 65 56	0 530 879 37 12
ANKARA BÖLGE MÜDÜRLÜĞÜ	Ankara	312	481 94 00	0 530 879 37 10
ANTALYA BÖLGE MÜDÜRLÜĞÜ	Antalya, Burdur, Isparta	242	243 45 60	0 530 879 37 14
BALIKESİR BÖLGE MÜDÜRLÜĞÜ	Balıkesir, Çanakkale	266	244 99 45	0 530 879 37 15
BURSA BÖLGE MÜDÜRLÜĞÜ	Bursa, Eskişehir, Bilecik	224	361 75 25	0 530 879 37 16
DENİZLİ BÖLGE MÜDÜRLÜĞÜ	Denizli, Muğla, Aydın	258	266 65 22	0 530 879 37 20
DIYARBAKIR BÖLGE MÜDÜRLÜĞÜ	Diyarbakir, Şanlıurfa	412	223 80 24	0 530 879 37 21
EDİRNE BÖLGE MÜDÜRLÜĞÜ	Edirne, Kırklareli, Tekirdağ	284	213 64 48	0 530 879 37 22
ERZURUM BÖLGE MÜDÜRLÜĞÜ	Erzurum, Bayburt, Erzincan	442	235 20 15	0 530 879 37 25
GAZİANTEP BÖLGE MÜDÜRLÜĞÜ	Gaziantep, Adıyaman, Kilis	342	336 94 00	0 530 879 37 27
HATAY BÖLGE MÜDÜRLÜĞÜ	Hatay, Kahramanmaraş, Osmaniye	326	225 44 00	0 530 879 37 31
İSTANBUL BÖLGE MÜDÜRLÜĞÜ - 1	İstanbul-Beylikte	212	258 92 96	0 530 879 37 34
İSTANBUL BÖLGE MÜDÜRLÜĞÜ - 2	İstanbul-Anadoluhisari	216	469 22 00	0 530 879 37 34
İSTANBUL BÖLGE MÜDÜRLÜĞÜ - 3	İstanbul-Kağıthane	212	580 07 83	0 530 879 37 34
İZMİR BÖLGE MÜDÜRLÜĞÜ	İzmir	232	483 14 54	0 530 879 37 35
KARS BÖLGE MÜDÜRLÜĞÜ	Kars, Ardahan, Şirnak, Ağrı	474	223 26 02	0 530 879 37 36
KASTAMONU BÖLGE MÜDÜRLÜĞÜ	Kastamonu, Çankırı, Sinop	366	215 50 92	0 530 879 37 37
KAYSERİ BÖLGE MÜDÜRLÜĞÜ	Kayseri, Sivas, Yozgat	352	233 42 32	0 530 879 37 38
KOCALISI BÖLGE MÜDÜRLÜĞÜ	Kocaeli, Bolu, Sakarya, Düzce, Yalova	262	300 56 00	0 530 879 37 18
KONYA BÖLGE MÜDÜRLÜĞÜ	Konya, Karaman	332	353 25 60	0 530 879 37 19
MALATYA BÖLGE MÜDÜRLÜĞÜ	Malatya, Tunceli, Bingöl, Elazığ	422	323 30 41	0 530 879 37 23
MANİSA BÖLGE MÜDÜRLÜĞÜ	Manisa, Uşak, Kütahya, Afyon	236	232 85 00	0 530 879 37 24
MEYVEHİÇ BÖLGE MÜDÜRLÜĞÜ	Nevşehir, Konya, Karaman, Niğde, Aksaray	384	212 82 23	0 530 879 37 26
SAMSUN BÖLGE MÜDÜRLÜĞÜ	Samsun, Amasya, Çorum, Tokat	362	431 25 08	0 530 879 37 28
SİRT BÖLGE MÜDÜRLÜĞÜ	Sirt, Mardin, Şırnak, Batman	484	223 49 00	0 530 879 37 29
TRABZON BÖLGE MÜDÜRLÜĞÜ	Trabzon, Artvin, Gümüşhane, Rize, Ordu, Giresun	462	321 57 49	0 530 879 37 30
VAN BÖLGE MÜDÜRLÜĞÜ	Van, Bitlis, Hakkari, Muş	432	214 25 11	0 530 879 37 32
ZONGULDAK BÖLGE MÜDÜRLÜĞÜ	Zonguldak, Bartın, Karabük	372	253 79 70	0 530 879 37 33

[illegible][illegible]

EK-4. Ferdi harcama kayıt defteri



TÜRKİYE İSTATİSTİK KURUMU

FORM 3

Bülten No: _____

HANEHALKI BÜTÇE ANKETİ

FERDİ HARCAMA KAYIT DEFTERİ



(Lütfen anket ayı içerisinde yaptığınız kişisel harcamaları
unutmamak için arka sayfa belirtilen örneğe göre kayıt ediniz.
Yaptığınız harcamalar hiç bir şekilde başka bir kişiye
açıklanmayacaktır.)

Fert Sıra No : _____

Adı ve Soyadı :

T.C. TÜRKİYE İSTATİSTİK KURUMU BAŞKANLIĞI

**BU DEFTERİ ANKET SONUNDA
ANKETÖRE İADE EDİNİZ**

SATIN ALINAN MAL VE HİZMETİN										
Sıra No.	Harcamanın adı	Markası	Cins ve ayrıntılı tanımı	Miktarı	Ölçü birimi (kg, adet, lt, metre, m ² , paket, çift, dakika, aykıl v.b.)	Toplam Ödenen (TL)	Alışverişin yapıldığı yer (Bakkal, market, Aşçımarket, Benekli İstasyonu, otobüs, baki, bilet, eczane, kafe, zincirler v.b.)	Harcanan paranın, harcanan cinsin ve harcanan miktarın toplamı (TL)	Harcanan paranın, harcanan cinsin ve harcanan miktarın toplamı (TL)	Alışverişin yapıldığı gün
01	02	03	04	05	06	07	08	09	10	11
1	Çay (çilen)		Küçük çay bardağı	5	Bardak	2,00	Gayretçi			1
2	Kola	Pepsi	330 ml kutu kola	1	Adet	3,00	İşyeri kantini			1
3	Sipara	Polo	Sipara Filtreli Silms	4	Paket	24	Bakkal	X		4
4	İlaç	Novalgın	Ağrı kesici	1	Paket	4,00	Eczane			4
5	Sakız	Falm	Tekli	1	Adet	0,50	Bakkal			5
6	Çekir		Kiçifile (küçük çin)	1	Adet	75,00	İnternet (Trendyol)	X		6
7	Gazete	SABAH	Günlük gazete	1	Adet	0,75	Gazete bayii			7
8	Fon çalıtırma		2. sınıf kadın kuarfoni	1	Defa	8,00	Kuarfoni	X		9
9	Otobüs kartı		5 TL kart, tam	1	Defa	15,00	Bilet satı yeri			15
10	Simit			1	Adet	1,00	Seyyar satıcı			24

EK-5. İstatistiki Bölge Birimleri Sınıflaması (İBBS)

Düzy 1 (12 bölge)		Düzy 2 (26 alt bölge)		Düzy 3 (81 il)
TR1	İstanbul	TR10	İstanbul alt bölgesi	İstanbul
TR2	Batı Marmara	TR21	Tekirdağ alt bölgesi	Tekirdağ, Edirne, Kırklareli
		TR22	Balıkesir alt bölgesi	Balıkesir, Çanakkale
TR3	Ege	TR31	İzmir alt bölgesi	İzmir
		TR32	Aydın alt bölgesi	Aydın, Denizli, Muğla
		TR33	Manisa alt bölgesi	Manisa, Afyonkarahisar, Kütahya, Uşak
TR4	Doğu Marmara	TR41	Bursa alt bölgesi	Bursa, Eskişehir, Bilecik
		TR42	Kocaeli alt bölgesi	Kocaeli, Sakarya, Düzce, Bolu, Yalova
TR5	Batı Anadolu	TR51	Ankara alt bölgesi	Ankara
		TR52	Konya alt bölgesi	Konya, Karaman
TR6	Akdeniz	TR61	Antalya alt bölgesi	Antalya, Isparta, Burdur
		TR62	Adana alt bölgesi	Adana, Mersin
		TR63	Hatay alt bölgesi	Hatay, Kahramanmaraş, Osmaniye
TR7	Orta Anadolu	TR71	Kırıkkale alt bölgesi	Kırıkkale, Aksaray, Niğde, Nevşehir, Kırşehir
		TR72	Kayseri alt bölgesi	Kayseri, Sivas, Yozgat
TR8	Batı Karadeniz	TR81	Zonguldak alt bölgesi	Zonguldak, Karabük, Bartın
		TR82	Kastamonu alt bölgesi	Kastamonu, Çankırı, Sinop
		TR83	Samsun alt bölgesi	Samsun, Tokat, Çorum, Amasya
TR9	Doğu Karadeniz	TR90	Trabzon alt bölgesi	Trabzon, Ordu, Giresun, Rize, Artvin, Gümüşhane
TRA	Kuzeydoğu Anadolu	TRA1	Erzurum alt bölgesi	Erzurum, Erzincan, Bayburt
		TRA2	Ağrı alt bölgesi	Ağrı, Kars, Iğdır, Ardahan
TRB	Ortadoğu Anadolu	TRB1	Malatya alt bölgesi	Malatya, Elazığ, Bingöl, Tunceli
		TRB2	Van alt bölgesi	Van, Muş, Bitlis, Hakkâri
TRC	Güneydoğu Anadolu	TRC1	Gaziantep alt bölgesi	Gaziantep, Adıyaman, Kilis
		TRC2	Şanlıurfa alt bölgesi	Şanlıurfa, Diyarbakır
		TRC3	Mardin alt bölgesi	Mardin, Batman, Şırnak, Siirt

EK-6. KASİS program menüsü

SEÇİM YAPINIZ:

1. ARŞİVLENMİŞ VERİYİ SINIFLAMA

TABLO SEÇİNİZ:
ANA GRUP SEÇİNİZ:
BÖLGE SEÇİNİZ:
YIL SEÇİNİZ:
AY SEÇİNİZ:

- 1.1. ÇALIŞTIR
- 1.2. BİR ÖNCEKİ EKRANA GERİ DÖN
- 1.3. ÇIKIŞ
- 1.4. TEKRAR SEÇİM YAP

2. ARŞİVLENMİŞ VERİYE DARALTILMIŞ LİSTEDEN KOD ATAMA

2.1. SON ÇALIŞTIRILAN KAYITLARA KOD ATA

TABLO SEÇİNİZ:
ANA GRUP SEÇİNİZ:
BÖLGE SEÇİNİZ:
YIL SEÇİNİZ:
AY SEÇİNİZ:

- 2.1.1. KAYITLARI GÖSTER
- 2.1.2. BİR ÖNCEKİ EKRANA GERİ DÖN
- 2.1.3. ÇIKIŞ
- 2.1.4. TEKRAR SEÇİM YAP

2.2. VERİ TABANINDAKİ KAYITLARA KOD ATA

TABLO SEÇİNİZ:
ANA GRUP SEÇİNİZ:
BÖLGE SEÇİNİZ:
YIL SEÇİNİZ:
AY SEÇİNİZ:

- 2.2.1. KAYITLARI GÖSTER
- 2.2.2. BİR ÖNCEKİ EKRANA GERİ DÖN
- 2.2.3. ÇIKIŞ
- 2.2.4. TEKRAR SEÇİM YAP

- 2.3. BİR ÖNCEKİ EKRANA GERİ DÖN
- 2.4. ÇIKIŞ

3. ARŞİVLENMİŞ VERİYE GENİŞ LİSTEDEN KOD ATAMA

3.1. SON ÇALIŞTIRILAN KAYITLARA KOD ATA

TABLO SEÇİNİZ:
ANA GRUP SEÇİNİZ:
BÖLGE SEÇİNİZ:
YIL SEÇİNİZ:
AY SEÇİNİZ:

- 3.1.1. KAYITLARI GÖSTER
- 3.1.2. BİR ÖNCEKİ EKRANA GERİ DÖN
- 3.1.3. ÇIKIŞ
- 3.1.4. TEKRAR SEÇİM YAP

EK-6. (Devam) KASİS program menüsü

3.2. VERİ TABANINDAKİ KAYITLARA KOD ATA

TABLO SEÇİNİZ:
ANA GRUP SEÇİNİZ:
BÖLGE SEÇİNİZ:
YIL SEÇİNİZ:
AY SEÇİNİZ:

- 3.2.1. KAYITLARI GÖSTER
- 3.2.2. BİR ÖNCEKİ EKRANA GERİ DÖN
- 3.2.3. ÇIKIŞ
- 3.2.4. TEKRAR SEÇİM YAP

- 3.3. BİR ÖNCEKİ EKRANA GERİ DÖN
- 3.4. ÇIKIŞ

4. MADDE TANIMI GİREREK KOD ATAMA

BİRİNCİ MADDE TANIMI:
İKİNCİ MADDE TANIMI:
ÜÇÜNCÜ MADDE TANIMI:
DÖRDÜNCÜ MADDE TANIMI:
BEŞİNCİ MADDE TANIMI:

4.1. SEÇİM YAPINIZ:

- 4.1.1. TANIMDA GEÇEN KELİMELERİN GEÇTİĞİ YERE GÖRE KOD ATA
- 4.1.2. TANIMDA GEÇEN KELİMELERİN BÜTÜNLÜĞÜNÜ KORUYARAK KOD ATA
- 4.1.3. BULANIK EŞLEŞTİRME YÖNTEMLERİNİ KULLANARAK KOD ATA
- 4.1.4. BİR ÖNCEKİ EKRANA GERİ DÖN
- 4.1.5. ÇIKIŞ

5. RAPORLAR

5.1. EŞLEŞEN KAYITLARI GÖR

5.1.1. SON ÇALIŞTIRMADAKİ KAYITLARI GÖR

TABLO SEÇİNİZ:
ANA GRUP SEÇİNİZ:
BÖLGE SEÇİNİZ:
YIL SEÇİNİZ:
AY SEÇİNİZ:

- 5.1.1.1. KAYITLARI GÖSTER
- 5.1.1.2. BİR ÖNCEKİ EKRANA GERİ DÖN
- 5.1.1.3. ÇIKIŞ
- 5.1.1.4. TEKRAR SEÇİM YAP

5.1.2. VERİ TABANINDAKİ KAYITLARI GÖR

TABLO SEÇİNİZ:
ANA GRUP SEÇİNİZ:
BÖLGE SEÇİNİZ:
YIL SEÇİNİZ:
AY SEÇİNİZ:

- 5.1.2.1. KAYITLARI GÖSTER
- 5.1.2.2. BİR ÖNCEKİ EKRANA GERİ DÖN
- 5.1.2.3. ÇIKIŞ
- 5.1.2.4. TEKRAR SEÇİM YAP

EK-6. (Devam) KASİS program menüsü

5.1.3. BİR ÖNCEKİ EKRANA GERİ DÖN

5.1.4. ÇIKIŞ

5.2. ŞÜPHELİ KAYITLARI GÖR

5.2.1. SON ÇALIŞTIRMADAKİ KAYITLARI GÖR

TABLO SEÇİNİZ:

ANA GRUP SEÇİNİZ:

BÖLGE SEÇİNİZ:

YIL SEÇİNİZ:

AY SEÇİNİZ:

5.2.1.1. KAYITLARI GÖSTER

5.2.1.2. BİR ÖNCEKİ EKRANA GERİ DÖN

5.2.1.3. ÇIKIŞ

5.2.1.4. TEKRAR SEÇİM YAP

5.2.2. VERİ TABANINDAKİ KAYITLARI GÖR

TABLO SEÇİNİZ:

ANA GRUP SEÇİNİZ:

BÖLGE SEÇİNİZ:

YIL SEÇİNİZ:

AY SEÇİNİZ:

5.2.2.1. KAYITLARI GÖSTER

5.2.2.2. BİR ÖNCEKİ EKRANA GERİ DÖN

5.2.2.3. ÇIKIŞ

5.2.2.4. TEKRAR SEÇİM YAP

5.2.3. BİR ÖNCEKİ EKRANA GERİ DÖN

5.2.4. ÇIKIŞ

5.3. YETERSİZ KAYITLARI GÖR

5.3.1. SON ÇALIŞTIRMADAKİ KAYITLARI GÖR

TABLO SEÇİNİZ:

ANA GRUP SEÇİNİZ:

BÖLGE SEÇİNİZ:

YIL SEÇİNİZ:

AY SEÇİNİZ:

5.3.1.1. KAYITLARI GÖSTER

5.3.1.2. BİR ÖNCEKİ EKRANA GERİ DÖN

5.3.1.3. ÇIKIŞ

5.3.1.4. TEKRAR SEÇİM YAP

5.3.2. VERİ TABANINDAKİ KAYITLARI GÖR

TABLO SEÇİNİZ:

ANA GRUP SEÇİNİZ:

BÖLGE SEÇİNİZ:

YIL SEÇİNİZ:

AY SEÇİNİZ:

5.3.2.1. KAYITLARI GÖSTER

5.3.2.2. BİR ÖNCEKİ EKRANA GERİ DÖN

5.3.2.3. ÇIKIŞ

5.3.2.4. TEKRAR SEÇİM YAP

EK-6. (Devam) KASİS program menüsü

5.3.3. BİR ÖNCEKİ EKRANA GERİ DÖN

5.3.4. ÇIKIŞ

5.4. EŞLEŞMEYEN KELİMELERİ GÖR

5.4.1. SON ÇALIŞTIRMADAKİ KAYITLARI GÖR

TABLO SEÇİNİZ:

ANA GRUP SEÇİNİZ:

BÖLGE SEÇİNİZ:

YIL SEÇİNİZ:

AY SEÇİNİZ:

5.4.1.1. KAYITLARI GÖSTER

5.4.1.2. BİR ÖNCEKİ EKRANA GERİ DÖN

5.4.1.3. ÇIKIŞ

5.4.1.4. TEKRAR SEÇİM YAP

5.4.2. VERİ TABANINDAKİ KAYITLARI GÖR

TABLO SEÇİNİZ:

ANA GRUP SEÇİNİZ:

BÖLGE SEÇİNİZ:

YIL SEÇİNİZ:

AY SEÇİNİZ:

5.4.2.1. KAYITLARI GÖSTER

5.4.2.2. BİR ÖNCEKİ EKRANA GERİ DÖN

5.4.2.3. ÇIKIŞ

5.4.2.4. TEKRAR SEÇİM YAP

5.5. COICOP SÖZLÜĞÜNÜ GÖR

5.5.1. 2013 YILI SÖZLÜĞÜ

5.5.2. 2014 YILI SÖZLÜĞÜ

5.5.3. 2015 YILI SÖZLÜĞÜ

5.5.4. 2016 YILI SÖZLÜĞÜ

5.5.5. 2017 YILI SÖZLÜĞÜ

5.5.6. 2018 YILI SÖZLÜĞÜ

5.5.7. BİR ÖNCEKİ EKRANA GERİ DÖN

5.5.8. ÇIKIŞ

5.6. KABUL EDİLEBİLİR KAYITLARI GÖR

5.6.1. SON ÇALIŞTIRMADAKİ KAYITLARI GÖR

TABLO SEÇİNİZ:

ANA GRUP SEÇİNİZ:

BÖLGE SEÇİNİZ:

YIL SEÇİNİZ:

AY SEÇİNİZ:

5.6.1.1. KAYITLARI GÖSTER

5.6.1.2. BİR ÖNCEKİ EKRANA GERİ DÖN

5.6.1.3. ÇIKIŞ

5.6.1.4. TEKRAR SEÇİM YAP

EK-6. (Devam) KASİS program menüsü

5.6.2. VERİ TABANINDAKİ KAYITLARI GÖR

TABLO SEÇİNİZ:
ANA GRUP SEÇİNİZ:
BÖLGE SEÇİNİZ:
YIL SEÇİNİZ:
AY SEÇİNİZ:

- 5.6.2.1. KAYITLARI GÖSTER
- 5.6.2.2. BİR ÖNCEKİ EKRANA GERİ DÖN
- 5.6.2.3. ÇIKIŞ
- 5.6.2.4. TEKRAR SEÇİM YAP

5.6.3. BİR ÖNCEKİ EKRANA GERİ DÖN

5.6.4. ÇIKIŞ

5.7. PROGRAMI ÇALIŞTIRMA KAYITLARINI GÖR

5.7.1. SON ÇALIŞTIRMADAKİ KAYITLARI GÖR

TABLO SEÇİNİZ:
ANA GRUP SEÇİNİZ:
BÖLGE SEÇİNİZ:
YIL SEÇİNİZ:
AY SEÇİNİZ:

- 5.7.1.1. KAYITLARI GÖSTER
- 5.7.1.2. BİR ÖNCEKİ EKRANA GERİ DÖN
- 5.7.1.3. ÇIKIŞ
- 5.7.1.4. TEKRAR SEÇİM YAP

5.7.2. VERİ TABANINDAKİ KAYITLARI GÖR

TABLO SEÇİNİZ:
ANA GRUP SEÇİNİZ:
BÖLGE SEÇİNİZ:
YIL SEÇİNİZ:
AY SEÇİNİZ:

- 5.7.2.1. KAYITLARI GÖSTER
- 5.7.2.2. BİR ÖNCEKİ EKRANA GERİ DÖN
- 5.7.2.3. ÇIKIŞ
- 5.7.2.4. TEKRAR SEÇİM YAP

5.8. DÖNÜŞÜMLERİ GÖR

5.8.1. DÖNÜŞÜM SÖZLÜĞÜNÜ GÖR

5.8.2. SON ÇALIŞTIRMADAKİ DÖNÜŞÜMLERİ GÖR

TABLO SEÇİNİZ:
ANA GRUP SEÇİNİZ:
BÖLGE SEÇİNİZ:
YIL SEÇİNİZ:
AY SEÇİNİZ:

- 5.8.2.1. KAYITLARI GÖSTER
- 5.8.2.2. BİR ÖNCEKİ EKRANA GERİ DÖN
- 5.8.2.3. ÇIKIŞ
- 5.8.2.4. TEKRAR SEÇİM YAP

EK-6. (Devam) KASİS program menüsü

5.8.3. VERİ TABANINDAKİ DÖNÜŞÜMLERİ GÖR (option96)

TABLO SEÇİNİZ:
ANA GRUP SEÇİNİZ:
BÖLGE SEÇİNİZ:
YIL SEÇİNİZ:
AY SEÇİNİZ:

- 5.8.3.1. KAYITLARI GÖSTER
- 5.8.3.2. BİR ÖNCEKİ EKRANA GERİ DÖN
- 5.8.3.3. ÇIKIŞ
- 5.8.3.4. TEKRAR SEÇİM YAP

- 5.8.4. BİR ÖNCEKİ EKRANA GERİ DÖN
- 5.8.5. ÇIKIŞ

5.9. KOD ATAMALARINI GÖR

- 5.9.1. MADDE TANIMI GİRİLEREK YAPILAN ATAMALARI GÖR
- 5.9.2. ARŞİVLENMİŞ VERİYE YAPILAN DARALTILMIŞ LİSTEDEN KOD ATAMALARI GÖR
- 5.9.3. ARŞİVLENMİŞ VERİYE GENİŞ LİSTEDEN KOD ATAMALARI GÖR
- 5.9.4. BİR ÖNCEKİ EKRANA GERİ DÖN
- 5.9.5. ÇIKIŞ

- 5.10. BİR ÖNCEKİ EKRANA GERİ DÖN
- 5.11. ÇIKIŞ

6. ÇIKIŞ

ÖZGEÇMİŞ

Kişisel Bilgiler

Soyadı, adı : AHİ, Levent

Uyruğu : T.C.

Doğum tarihi ve yeri :

Medeni hali :

Telefon :

e-mail :

Eğitim

Derece	Eğitim Birimi	Mezuniyet Tarihi
Doktora	Gazi Üniversitesi / Yönetim Bilişim Sistemleri	Devam ediyor
Yüksek lisans	Hacettepe Üniversitesi / İstatistik	2015
Lisans	Hacettepe Üniversitesi / İstatistik	1996
Lise	Ankara Ticaret Lisesi	1990

İş Deneyimi

Yıl	Yer	Görev
2006-Halen	Türkiye İstatistik Kurumu	İstatistikçi
2003-2006	Türk Telekomünikasyon A.Ş.	Telekom Uzman Yardımcısı
2003-2003	Dışbank A.Ş.	İç Kontrol Uzmanı
2001-2003	Toprakbank A.Ş.	Müfettiş Yardımcısı
2000-2000	Bayındırbank A.Ş.	Pazarlama Uzmanı
1997-1999	Sümerbank A.Ş.	Pazarlama Uzmanı

Yabancı Dil

İngilizce

Yayınlar

Yüksek Lisans Tezi:

1. Ahi, L. (2015). Veri Madenciliği Yöntemleri İle Ana Harcama Gruplarının Paylarının Tahmini. Yüksek Lisans Tezi, Hacettepe Üniversitesi Fen Bilimleri Enstitüsü İstatistik Bölümü, Ankara.

Makaleler:

1. Karabulut, T., Ahi, L., Gürsoy, Z. (2016). Kişisel Gelir Dağılımındaki Adaletsizliğin Giderilmesinde Transfer Gelirlerinin Etkisi: Türkiye Örneği. *Balkan Sosyal Bilimler Dergisi*, 732-741.
2. Ahi, L., Kılıç Çakmak, E. (2020). Kod Atama Sistemi (KASİS) ile Otomatik Kod Atama. *Bilgi ve İletişim Teknolojileri Dergisi*, 2 (1), 73-87.
3. Ahi, L., Kılıç Çakmak, E. (2020). Uluslararası İstatistiksel Sınıflamalara Yönelik Kod Atama Sistemi (KASİS). *Bilişim Teknolojileri Dergisi*, 13 (3), 313-327.



GAZİLİ OLMAK AYRICALIKTIR.