

**METİN MADENCİLİĞİ İLE E-TİCARET SİTELERİNİN
BELİRLENMESİ**

Tuğba KAŞIKÇI

**YÜKSEK LİSANS TEZİ
YÖNETİM BİLİŞİM SİSTEMLERİ**

**GAZİ ÜNİVERSİTESİ
BİLİŞİM ENSTİTÜSÜ**

**OCAK 2014
ANKARA**

Tuğba KAŞIKÇI tarafından hazırlanan METİN MADENCİLİĞİ İLE E-TİCARET SİTELERİNİN BELİRLENMESİ adlı bu tezin Yüksek Lisans tezi olarak uygun olduğunu onaylarım.



Prof. Dr. Hadi Gökçen

Bu çalışma, jürimiz tarafından oy birliği / oy çokluğu ile Yönetim Bilişim Sistemleri Anabilim Dalında Yüksek lisans tezi olarak kabul edilmiştir.

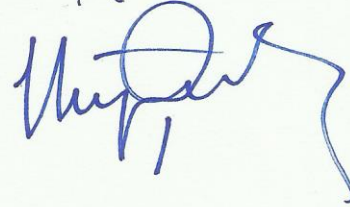
Başkan: : Prof. Dr. Hadi GÖKÇEN



Üye : Doç. Dr. Metehan TOLON



Üye : Yrd. Doç. Dr. Hüseyin ÇAKIR



Tarih : 06/01/2014

Bu tez, Gazi Üniversitesi Bilişim Enstitüsü tez yazım kurallarına uygundur.

TEZ BİLDİRİMİ

Tez içindeki bütün bilgilerin etik davranış ve akademik kurallar çerçevesinde elde edilerek sunulduğunu, ayrıca tez yazım kurallarına uygun olarak hazırlanan bu çalışmada orijinal olmayan her türlü kaynağa eksiksiz atıf yapıldığını bildiririm.

Tuğba KAŞIKÇI

METİN MADENCİLİĞİ İLE E-TİCARET SİTELERİNİN BELİRLENMESİ

(Yüksek Lisans Tezi)

Tuğba KAŞIKÇI

**GAZİ ÜNİVERSİTESİ
BİLİŞİM ENSTİTÜSÜ**

Ocak 2014

ÖZET

Bu çalışmada kullanıcı tarafından belirtilen internet sitelerinin içeriğini analiz ederek metin madenciliği yöntemleri ile bu sayfaların elektronik ticaret (e-ticaret) sitesi olup olmadığına karar veren bir uygulama geliştirilmiştir. Uygulama özel çalışmalarda kullanılmak üzere hazırlanmıştır ve kullanıcılara elektronik ticaret sitelerinin bulunmasını kolaylaştırmayı amaçlamaktadır. Metin sınıflandırmada kullanılmak üzere farklı kaynaklardan veriler toplanmış ve kullanıma hazırlanmıştır. Biri statik diğeri ise dinamik olmak üzere iki çalışma yapılmıştır. Bu çalışmalara bağlı olarak farklı eğitim kümeleri oluşturulmuş ve bunlar üzerinde K-En-Yakın-Komşu (k-NN) ve Naïve Bayes sınıflandırma algoritmaları kullanılarak elde edilen sonuçlar karşılaştırılmıştır. Ardından seçilen algoritma Java programlama dili ile masaüstü uygulaması olarak hazırlanan ara yüze aktararak kullanıma sunulmuştur.

Bilim Kodu	: 902.1.071
Anahtar Kelime	: metin madenciliği, metin sınıflandırma, elektronik ticaret, e-ticaret
Sayfa Adedi	: 140
Tez Yöneticisi	: Prof. Dr. Hadi Gökçen

DETERMINATION OF E-COMMERCE SITES BY TEXT MINING

(M.Sc. Thesis)

Tuğba KAŞIKÇI

**GAZİ UNIVERSITY
INFORMATICS INSTITUTE**

January 2014

ABSTRACT

In this thesis, an application which decide whether or not the pages specified by the user are electronic commerce (e-commerce) sites by analyzing the contents of web sites and by using text mining techniques is developed. The application is designed to be used in special studies and aims to facilitate users to find e-commerce sites. In this study, the data was collected from different sources to be used in text classification and prepared for useage. A static and a dynamic study were performed. Different training sets are prepared based on these studies and k-NN Nearest Neighbour and Naïve Bayes classification algorithms are used on these sets. And then obtained results are compared. Algorithm that gives better results is selected and a desktop application with the Java programming language is developed for usage.

Science Code	: 902.1.071
Key Words	: text mining, text classification, electronic commerce, e-commerce
Page Number	: 140
Adviser	: Prof. Dr. Hadi Gökçen

TEŞEKKÜR

Çalışmalarım süresince yardım ve katkılarıyla ben yönlendiren değerli Hocam Prof. Dr. Hadi GÖKÇEN'e, desteğini esirgemeyen değerli eşime ve arkadaşlarıma teşekkür ederim.

İÇİNDEKİLER

Sayfa

ÖZET	iv
ABSTRACT	v
TEŞEKKÜR	vi
İÇİNDEKİLER.....	vii
ÇİZELGELERİN LİSTESİ	x
ŞEKİLLERİN LİSTESİ	xv
SİMGELER VE KISALTMALAR	xx
1. GİRİŞ.....	1
2. METİN SINIFLANDIRMA	5
2.1. Metin Gösterimi	7
2.1.1. Doküman ön işleme	7
2.1.2. Anahtar kelime sözlüğünün oluşturulması.....	9
2.1.3. Vektör uzay modelinin oluşturulması.....	10
2.1.4. Vektör ağırlıklandırma	12
2.1.5. Dokümanlar arası benzerlik	13
2.1.6. Sonuçların değerlendirilmesi.....	14
2.2. Metin Sınıflandırma Algoritmaları.....	15
2.2.1. k-NN algoritması	15
2.2.2. Naïve Bayes algoritması	23
2.3. Metin Sınıflandırma Aracı WEKA.....	27
3. UYGULAMA	33
3.1. Veri Elde Etme	36

3.2. Doküman Ön İşleme	38
3.3. Anahtar Kelime Sözlüğünün Oluşturulması	39
3.3.1. Statik çalışma	39
3.3.2. Dinamik çalışma	40
3.3.3. Dinamik çalışma için farklı yüzde değerlerinin kullanılması	40
3.4. Vektör Uzay Modelinin Oluşturulması	43
3.4.1. Statik çalışma	44
3.4.2. Dinamik çalışma	46
3.5. Algoritmaların Kullanılması	61
3.5.1. Statik çalışma	61
3.5.2. Dinamik çalışma	67
3.6. Statik ve Dinamik Çalışma Sonuçlarının Karşılaştırılması	98
3.7. Uygulama Ara Yüzü	101
3.7.1. Eğitim ekranı	101
3.7.2. Sorgulama ekranı	103
4. SONUÇLAR VE ÖNERİLER	108
KAYNAKLAR	112
EKLER	116
EK-1. WEKA işlemlerini yapan Java kodu	117
EK-2. Dizin sitelerinden bağlantı toplama Java kodu	118
EK-3. Dizin sitelerinden bağlantı toplama Java kodundaki değişiklikler	119
EK-4. E-ticaret sınıfını oluşturan internet siteleri	120
EK-4. E-ticaret sınıfını oluşturan internet siteleri (Devam)	121
EK-4. E-ticaret sınıfını oluşturan internet siteleri (Devam)	122
EK-4. E-ticaret sınıfını oluşturan internet siteleri (Devam)	123
EK-4. E-ticaret sınıfını oluşturan internet siteleri (Devam)	124
EK-5. E-ticaret olmayanlar sınıfını oluşturan internet siteleri	125
EK-5. E-ticaret olmayanlar sınıfını oluşturan internet siteleri (Devam)	126
EK-5. E-ticaret olmayanlar sınıfını oluşturan internet siteleri (Devam)	127

EK-5. E-ticaret olmayanlar sınıfını oluşturan internet siteleri (Devam).....	128
EK-5. E-ticaret olmayanlar sınıfını oluşturan internet siteleri (Devam).....	129
EK-6. Durak ve gereksiz kelimelerin listesi.....	130
EK-7. %0 oranı ile elde edilen veri kümesi için anahtar kelimeler	131
EK-7. %0 oranı ile elde edilen veri kümesi için anahtar kelimeler (Devam)	132
EK-7. %0 oranı ile elde edilen veri kümesi için anahtar kelimeler (Devam)	133
EK-7. %0 oranı ile elde edilen veri kümesi için anahtar kelimeler (Devam)	134
EK-7. %0 oranı ile elde edilen veri kümesi için anahtar kelimeler (Devam)	135
EK-7. %0 oranı ile elde edilen veri kümesi için anahtar kelimeler (Devam)	136
EK-7. %0 oranı ile elde edilen veri kümesi için anahtar kelimeler (Devam)	137
EK-7. %0 oranı ile elde edilen veri kümesi için anahtar kelimeler (Devam)	138
EK-8. %5 oranı ile elde edilen veri kümesi için anahtar kelimeler	139
ÖZGEÇMİŞ.....	140

ÇİZELGELERİN LİSTESİ

Çizelge	Sayfa
Çizelge 2.1. Örnek doküman - kelime çizelgesi.....	11
Çizelge 2.2. Örnek uygulama terim sıklığı doküman - kelime çizelgesi	11
Çizelge 2.3. Örnek uygulama ikili doküman - kelime çizelgesi.....	12
Çizelge 2.4. Örnek uygulama eğitim vektörlerinin sorgu vektörüne uzaklığı (İkili) .	17
Çizelge 2.5. Örnek uygulama diğer işlemler (İkili)	17
Çizelge 2.6. Örnek uygulama benzerlik ilişkileri (İkili)	18
Çizelge 2.7. Örnek uygulama eğitim vektörlerinin sorgu vektörüne uzaklığı (TF) ...	19
Çizelge 2.8. Örnek uygulama diğer işlemler (TF)	19
Çizelge 2.9. Örnek uygulama benzerlik ilişkileri (TF)	20
Çizelge 2.10. Örnek uygulama IDF Değerleri	21
Çizelge 2.11. Örnek uygulama TF-IDF değerleri.....	21
Çizelge 2.12. Örnek - eğitim vektörlerinin sorgu vektörüne uzaklığı (TF-IDF).....	22
Çizelge 2.13. Örnek uygulama diğer işlemler (TF-IDF)	22
Çizelge 2.14. Örnek uygulama benzerlik ilişkileri (TF-IDF).....	23
Çizelge 2.15. Örnek uygulama kelimenin her sınıftaki ikili sıklığı.....	24
Çizelge 2.16. Örnek uygulama olumlu ve olumsuz sınıf sayısı	24
Çizelge 2.17. Örnek uygulama Naïve Bayes ikili diğer işlemler	24
Çizelge 2.18. Örnek uygulama Naïve Bayes ikili sonuç.....	25
Çizelge 2.19. Örnek uygulama kelimelerin her sınıftaki TF sıklığı	26
Çizelge 2.20. Örnek uygulama sınıf ve sözlükteki kelime sayısı	26

Çizelge	Sayfa
Çizelge 2.21. Örnek uygulama Naïve Bayes TF diğer işlemler	26
Çizelge 2.22. Örnek uygulama Naïve Bayes TF sonuç	26
Çizelge 2.23. Hata matrisi örneği	31
Çizelge 3.1. Kullanılan dizin siteleri listesi.....	36
Çizelge 3.2. Statik çalışma anahtar kelime sözlüğü	39
Çizelge 3.3. Yüzde değerine bağlı değişim (%0-%45).....	41
Çizelge 3.4. Yüzde değerine bağlı değişim (%50-%95).....	42
Çizelge 3.5. Dinamik çalışma k-NN - ikili ağırlıklandırma yöntemi sonucu	62
Çizelge 3.6. Dinamik çalışma k-NN - TF ağırlıklandırma yöntemi sonucu	64
Çizelge 3.7. Dinamik çalışma k-NN - TF-IDF ağırlıklandırma yöntemi sonucu.....	65
Çizelge 3.8. Dinamik çalışma Naïve Bayes ikili ağırlıklandırma yöntemi sonucu	66
Çizelge 3.9. Dinamik çalışma Naïve Bayes TF ağırlıklandırma yöntemi sonucu.....	67
Çizelge 3.10. Dinamik çalışma k-NN ikili ağırlıklandırma sonucu (%0) (k=1...8) ...	68
Çizelge 3.11. Dinamik çalışma k-NN ikili ağırlıklandırma sonucu (%0) (k=9...15) .	68
Çizelge 3.12. Dinamik çalışma k-NN ikili ağırlıklandırma sonucu (%5) (k=1...8) ...	69
Çizelge 3.13. Dinamik çalışma k-NN ikili ağırlıklandırma sonucu (%5) (k=9...15) .	69
Çizelge 3.14. Dinamik çalışma k-NN ikili ağırlıklandırma sonucu (%10) (k=1...8) .	70
Çizelge 3.15. Dinamik çalışma k-NN ikili ağırlıklandırma sonucu (%10) (k=9...15)	70
Çizelge 3.16. Dinamik çalışma k-NN ikili ağırlıklandırma sonucu (%15) (k=1...8) .	71
Çizelge 3.17. Dinamik çalışma k-NN ikili ağırlıklandırma sonucu (%15) (k=9...15)	71
Çizelge 3.18. Dinamik çalışma k-NN ikili ağırlıklandırma sonucu (%20) (k=1...8) .	72

Çizelge	Sayfa
Çizelge 3.19. Dinamik çalışma k-NN ikili ağırlıklandırma sonucu (%20) (k=9...15)	72
Çizelge 3.20. Dinamik çalışma k-NN ikili ağırlıklandırma sonucu (%25) (k=1...8)	.73
Çizelge 3.21. Dinamik çalışma k-NN ikili ağırlıklandırma sonucu (%25) (k=9...15)	73
Çizelge 3.22. Dinamik çalışma k-NN ikili ağırlıklandırma sonucu (%30) (k=1...8)	.74
Çizelge 3.23. Dinamik çalışma k-NN ikili ağırlıklandırma sonucu (%30) (k=9...15)	74
Çizelge 3.24. Dinamik çalışma k-NN ikili ağırlıklandırma sonucu (%35) (k=1...8)	.75
Çizelge 3.25. Dinamik çalışma k-NN ikili ağırlıklandırma sonucu (%35) (k=9...15)	75
Çizelge 3.26. Dinamik çalışma k-NN ikili ağırlıklandırma sonucu (%40) (k=1...8)	.76
Çizelge 3.27. Dinamik çalışma k-NN ikili ağırlıklandırma sonucu (%40) (k=9...15)	76
Çizelge 3.28. k-NN ikili için sonuçların karşılaştırılması.....	77
Çizelge 3.29. Dinamik çalışma k-NN TF ağırlıklandırma sonucu (%0) (k=1...8)	78
Çizelge 3.30. Dinamik çalışma k-NN TF ağırlıklandırma sonucu (%0) (k=9...15) ...	78
Çizelge 3.31. Dinamik çalışma k-NN TF ağırlıklandırma sonucu (%5) (k=1...8)	79
Çizelge 3.32. Dinamik çalışma k-NN TF ağırlıklandırma sonucu (%5) (k=9...15) ...	79
Çizelge 3.33. Dinamik çalışma k-NN TF ağırlıklandırma sonucu (%10) (k=1...8) ...	80
Çizelge 3.34. Dinamik çalışma k-NN TF ağırlıklandırma sonucu (%10) (k=9...15) .	80
Çizelge 3.35. Dinamik çalışma k-NN TF ağırlıklandırma sonucu (%15) (k=1...8) ...	81
Çizelge 3.36. Dinamik çalışma k-NN TF ağırlıklandırma sonucu (%15) (k=9...15) .	81
Çizelge 3.37. Dinamik çalışma k-NN TF ağırlıklandırma sonucu (%20) (k=1...8) ...	82
Çizelge 3.38. Dinamik çalışma k-NN TF ağırlıklandırma sonucu (%20) (k=9...15)	82
Çizelge 3.39. Dinamik çalışma k-NN TF ağırlıklandırma sonucu (%25) (k=1...8) ...	83

Çizelge	Sayfa
Çizelge 3.40. Dinamik çalışma k-NN TF ağırlıklandırma sonucu (%25) (k=9...15) .83	
Çizelge 3.41. Dinamik çalışma k-NN TF ağırlıklandırma sonucu (%30) (k=1...8) ...84	
Çizelge 3.42. Dinamik çalışma k-NN TF ağırlıklandırma sonucu (%30) (k=9...15) .84	
Çizelge 3.43. Dinamik çalışma k-NN TF ağırlıklandırma sonucu (%35) (k=1...8) ...85	
Çizelge 3.44. Dinamik çalışma k-NN TF ağırlıklandırma sonucu (%35) (k=9...15) .85	
Çizelge 3.45. Dinamik çalışma k-NN TF ağırlıklandırma sonucu (%40) (k=1...8) ...86	
Çizelge 3.46. Dinamik çalışma k-NN TF ağırlıklandırma sonucu (%40) (k=9...15) .86	
Çizelge 3.47. k-NN TF ağırlıklandırma için sonuçların karşılaştırılması.....87	
Çizelge 3.48. Dinamik çalışma k-NN TF-IDF ağırlıklandırma (%0) (k=1...8)87	
Çizelge 3.49. Dinamik çalışma k-NN TF-IDF ağırlıklandırma (%0) (k=9...15).....88	
Çizelge 3.50. Dinamik çalışma k-NN TF-IDF ağırlıklandırma (%5) (k=1...8)89	
Çizelge 3.51. Dinamik çalışma k-NN TF-IDF ağırlıklandırma (%5) (k=9...15).....89	
Çizelge 3.52. Dinamik çalışma k-NN TF-IDF ağırlıklandırma (%10) (k=1...8).....90	
Çizelge 3.53. Dinamik çalışma k-NN TF-IDF ağırlıklandırma (%10) (k=9...15).....90	
Çizelge 3.54. Dinamik çalışma k-NN TF-IDF ağırlıklandırma (%15) (k=1...8).....91	
Çizelge 3.55. Dinamik çalışma k-NN TF-IDF ağırlıklandırma (%15) (k=9...15).....91	
Çizelge 3.56. Dinamik çalışma k-NN TF-IDF ağırlıklandırma (%20) (k=1...8).....92	
Çizelge 3.57. Dinamik çalışma k-NN TF-IDF ağırlıklandırma (%20) (k=9...15).....92	
Çizelge 3.58. Dinamik çalışma k-NN TF-IDF ağırlıklandırma (%25) (k=1...8).....93	
Çizelge 3.59. Dinamik çalışma k-NN TF-IDF ağırlıklandırma (%25) (k=9...15).....93	
Çizelge 3.60. Dinamik çalışma k-NN TF-IDF ağırlıklandırma (%30) (k=1...8).....94	

Çizelge	Sayfa
Çizelge 3.61. Dinamik çalışma k-NN TF-IDF ağırlıklandırma (%30) (k=9...15).....	94
Çizelge 3.62. Dinamik çalışma k-NN TF-IDF ağırlıklandırma (%35) (k=1...8).....	95
Çizelge 3.63. Dinamik çalışma k-NN TF-IDF ağırlıklandırma (%35) (k=9...15).....	95
Çizelge 3.64. Dinamik çalışma k-NN TF-IDF ağırlıklandırma (%40) (k=1...8).....	96
Çizelge 3.65. Dinamik çalışma k-NN TF-IDF ağırlıklandırma (%40) (k=9...15).....	96
Çizelge 3.66. k-NN TF-IDF ağırlıklandırma için sonuçların karşılaştırılması	97
Çizelge 3.67. Dinamik çalışma Naïve Bayes ikili ağırlıklandırma yöntemi sonucu ..	97
Çizelge 3.68. Dinamik çalışma Naïve Bayes TF ağırlıklandırma yöntemi sonucu	98
Çizelge 3.69. Statik çalışma algoritma sonuçlarının karşılaştırılması	99
Çizelge 3.70. Dinamik çalışma algoritma sonuçlarının karşılaştırılması	99

ŞEKİLLERİN LİSTESİ

Şekil	Sayfa
Şekil 1.1. Web 1.0, web 2.0 ve web 3.0 gelişimi	3
Şekil 2.1. Metin sınıflandırma aşamaları	6
Şekil 2.2. Veri ön işleme adımları	7
Şekil 2.3. Vektör uzay modeli	10
Şekil 2.4. Örnek uygulama sorgu “.ARFF” dosyası	27
Şekil 2.5. Örnek uygulama eğitim “.ARFF” dosyası.....	28
Şekil 2.6. WEKA ara yüzü ekran görüntüsü	29
Şekil 2.7. WEKA ara yüzü dosya açma ekran görüntüsü	30
Şekil 2.8. Örnek uygulama eğitim ve sınıflandırma sonucu	31
Şekil 3.1. Uygulama iş akışı.....	35
Şekil 3.2. Dizin sitelerinden bağlantı toplama Java kodu	37
Şekil 3.3. Bağlantıları kontrol etme Java kodu.....	37
Şekil 3.4. Bağlantı ana sayfa ve META bilgilerini alma Java kodu.....	38
Şekil 3.5. Statik çalışma ikili doküman-kelime çizelgesi	44
Şekil 3.6. Statik çalışma TF doküman-kelime çizelgesi	45
Şekil 3.7. Statik çalışma TF-IDF doküman-kelime çizelgesi.....	46
Şekil 3.8. Dinamik çalışma ikili doküman-kelime çizelgesi (%0)	47
Şekil 3.9. Dinamik çalışma ikili doküman-kelime çizelgesi (%5)	48
Şekil 3.10. Dinamik çalışma ikili doküman-kelime çizelgesi (%10)	48
Şekil 3.11. Dinamik çalışma ikili doküman-kelime çizelgesi (%15)	49

Şekil	Sayfa
Şekil 3.12. Dinamik çalışma ikili doküman-kelime çizelgesi (%20)	49
Şekil 3.13. Dinamik çalışma ikili doküman-kelime çizelgesi (%25)	50
Şekil 3.14. Dinamik çalışma ikili doküman-kelime çizelgesi (%30)	50
Şekil 3.15. Dinamik çalışma ikili doküman-kelime çizelgesi (%35)	51
Şekil 3.16. Dinamik çalışma ikili doküman-kelime çizelgesi (%40)	51
Şekil 3.17. Dinamik çalışma TF doküman-kelime çizelgesi (%0)	52
Şekil 3.18. Dinamik çalışma TF doküman-kelime çizelgesi (%5)	52
Şekil 3.19. Dinamik çalışma TF doküman-kelime çizelgesi (%10)	53
Şekil 3.20. Dinamik çalışma TF doküman-kelime çizelgesi (%15)	53
Şekil 3.21. Dinamik çalışma TF doküman-kelime çizelgesi (%20)	54
Şekil 3.22. Dinamik çalışma TF doküman-kelime çizelgesi (%25)	54
Şekil 3.23. Dinamik çalışma TF doküman-kelime çizelgesi (%30)	55
Şekil 3.24. Dinamik çalışma TF doküman-kelime çizelgesi (%35)	55
Şekil 3.25. Dinamik çalışma TF doküman-kelime çizelgesi (%40)	56
Şekil 3.26. Dinamik çalışma TF-IDF doküman-kelime çizelgesi (%0).....	57
Şekil 3.27. Dinamik çalışma TF-IDF doküman-kelime çizelgesi (%5).....	57
Şekil 3.28. Dinamik çalışma TF-IDF doküman-kelime çizelgesi (%10).....	58
Şekil 3.29. Dinamik çalışma TF-IDF doküman-kelime çizelgesi (%15).....	58
Şekil 3.30. Dinamik çalışma TF-IDF doküman-kelime çizelgesi (%20).....	59
Şekil 3.31. Dinamik çalışma TF-IDF doküman-kelime çizelgesi (%25).....	59
Şekil 3.32. Dinamik çalışma TF-IDF doküman-kelime çizelgesi (%30).....	60

Şekil	Sayfa
Şekil 3.33. Dinamik çalışma TF-IDF doküman-kelime çizelgesi (%35).....	60
Şekil 3.34. Dinamik çalışma TF-IDF doküman-kelime çizelgesi (%40).....	61
Şekil 3.35. “k” değerine bağlı başarı yüzdesi değişimi (İkili).....	63
Şekil 3.36. “k” değerine bağlı başarı yüzdesi değişimi (TF)	64
Şekil 3.37. “k” değerine bağlı başarı yüzdesi değişimi (TF-IDF)	66
Şekil 3.38. “k” değerine bağlı başarı yüzdesi değişimi (%0) (İkili)	68
Şekil 3.39. “k” değerine bağlı başarı yüzdesi değişimi (%5) (İkili)	69
Şekil 3.40. “k” değerine bağlı başarı yüzdesi değişimi (%10) (İkili)	70
Şekil 3.41. “k” değerine bağlı başarı yüzdesi değişimi (%15) (İkili)	71
Şekil 3.42. “k” değerine bağlı başarı yüzdesi değişimi (%20) (İkili)	72
Şekil 3.43. “k” değerine bağlı başarı yüzdesi değişimi (%25) (İkili)	73
Şekil 3.44. “k” değerine bağlı başarı yüzdesi değişimi (%30) (İkili)	74
Şekil 3.45. “k” değerine bağlı başarı yüzdesi değişimi (%35) (İkili)	75
Şekil 3.46. “k” değerine bağlı başarı yüzdesi değişimi (%40) (İkili)	76
Şekil 3.47. “k” değerine bağlı başarı yüzdesi değişimi (%0) (TF).....	78
Şekil 3.48. “k” değerine bağlı başarı yüzdesi değişimi (%5) (TF).....	79
Şekil 3.49. “k” değerine bağlı başarı yüzdesi değişimi (%10) (TF).....	80
Şekil 3.50. “k” değerine bağlı başarı yüzdesi değişimi (%15) (TF).....	81
Şekil 3.51. “k” değerine bağlı başarı yüzdesi değişimi (%20) (TF).....	82
Şekil 3.52. “k” değerine bağlı başarı yüzdesi değişimi (%25) (TF).....	83
Şekil 3.53. “k” değerine bağlı başarı yüzdesi değişimi (%30) (TF).....	84

Şekil	Sayfa
Şekil 3.54. “k” değerine bağlı başarı yüzdesi değişimi (%35) (TF).....	85
Şekil 3.55. “k” değerine bağlı başarı yüzdesi değişimi (%40) (TF).....	86
Şekil 3.56. “k” değerine bağlı başarı yüzdesi değişimi (%0) (TF-IDF)	88
Şekil 3.57. “k” değerine bağlı başarı yüzdesi değişimi (%5) (TF-IDF)	89
Şekil 3.58. “k” değerine bağlı başarı yüzdesi değişimi (%10) (TF-IDF)	90
Şekil 3.59. “k” değerine bağlı başarı yüzdesi değişimi (%15) (TF-IDF)	91
Şekil 3.60. “k” değerine bağlı başarı yüzdesi değişimi (%20) (TF-IDF)	92
Şekil 3.61. “k” değerine bağlı başarı yüzdesi değişimi (%25) (TF-IDF)	93
Şekil 3.62. “k” değerine bağlı başarı yüzdesi değişimi (%30) (TF-IDF)	94
Şekil 3.63. “k” değerine bağlı başarı yüzdesi değişimi (%35) (TF-IDF)	95
Şekil 3.64. “k” değerine bağlı başarı yüzdesi değişimi (%40) (TF-IDF)	96
Şekil 3.65. Uygulama eğitim ekranı görüntüsü	101
Şekil 3.66. Satır ekle düğmesi	102
Şekil 3.67. Ortak kelime yüzdesi girişi	102
Şekil 3.68. Modeli yeniden eğit düğmesi	102
Şekil 3.69. Eğitim modeli sonuçları ekranı	103
Şekil 3.70. Sorgulama ekranı	104
Şekil 3.71. Sorgulama ekranı ve sonuçları	104
Şekil 3.72. Site ismi giriş kısmı	105
Şekil 3.73. E-ticaret sitesi mi? düğmesi	105
Şekil 3.74. Bağlantı sağlanamayan siteler listesi	105

Şekil	Sayfa
Şekil 3.75. Sonuçları temizle düğmesi.....	105
Şekil 3.76. Uygulama sonuç ekranı (e-ticaret sitesi)	106
Şekil 3.77. Uygulama sonuç ekranı (e-ticaret sitesi değil).....	107

SİMGELER VE KISALTMALAR

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

Simge	Açıklama
c	Sınıf
d	Doküman
w	Kelime
n	Kelimenin dokümanda geçme sayısı
tf	Kelimenin terim sıklığı
idf	Ters doküman sıklığı
dtf	Doküman sıklığı
D	Toplam doküman sıklığı
x	Kelimenin dokümandaki TF-IDF ağırlığı
q	Sınıfının bulunması istenen vektör
$ V $	Sözlükteki kelime sayısı
B	c sınıfında w kelimesini içeren doküman sayısı
$ C $	c sınıfındaki eğitim dokümanı sayısı
N	Sınıftaki toplam kelime sayısı
$P(d)$	Sınıf olasılığı
TP	True positive
TN	True negative
FP	False positive
FN	False negative

Kısaltmalar	Açıklamalar
ARFF	Attribute Relation File Format Özellik İlişki Dosya Biçimi
E-TİCARET	Elektronik Ticaret
HTML	Hyper Text Markup Language Zengin Metin İşaret Dili
IDF	Inverce Document Frequency Ters Doküman Sıklığı
K-NN	K Nearest Neighbour K En Yakın Komşu
SVM	Supported Vector Machines Destek Vektörleri
TF	Term Frequency Terim Sıklığı
TF-IDF	Term Frequency - Inverce Document Frequency Terim Sıklığı – Ters Doküman Sıklığı
URL	Uniform Resource Locator Tekdüzen Kaynak Bulucu
WEKA	Waikato Environment for Knowledge Analysis Bilgi Analizi için Waikato Ortamı
XML	Extensible Markup Language Genişletilebilir İşaretleme Dili

1. GİRİŞ

İnternetin hızla gelişmesiyle birlikte artık internet hayatın her alanında kullanılmaktadır. İnternetin getirdiği kolaylıklardan bazıları şunlardır: bilgiye kolay erişim, haberleşme (internetten konuşma, e-posta, çevirim içi gazete okuma vs.), kamusal olanaklardan faydalanma, para kazanma, elektronik ticaret (e-ticaret). E-ticaret ticaretin elektronik ortamda yapılmasıdır. E-ticaret ile müşteri için saat sınırı olmaksızın, hızlı, çok çeşitli ürünlerin bulunduğu, ürünlerin karşılaştırılabildiği, rekabetçi ortamda düşük fiyatlı ürünlerin bulunabildiği alışveriş yapma imkânı sağlamaktadır. Bu nedenle e-ticaret hacmi giderek artmaktadır [1].

Ülkelerin ekonomik devamlılığının sağlanmasında vergi çok önemli bir yer tutmaktadır. Bu bağlamda internetin gelişmesiyle e-ticaretin vergilendirilmesi günümüzde giderek daha önemli hale gelmektedir. Türkiye’de e-ticareti her yönüyle belirleyen doğrudan bir kanun bulunmamaktadır. E-ticaret ile ilgili mevcut düzenlemelerden bazıları listelenmiştir:

- Elektronik Ticaretin Düzenlenmesi Hakkında Kanun Tasarısı [2],
- İnternet Ortamında Yapılan Yayınların Düzenlenmesine Dair Usul ve Esaslar Hakkında Yönetmelik [3],
- Kişisel Verilerin Korunması Kanunu Tasarısı [4],
- Elektronik Haberleşme Sektöründe Kişisel Verilerin İşlenmesi ve Gizliliğinin Korunması Hakkında Yönetmelik [5],
- Elektronik İmza Kanunu [6] ve
- Tüketicinin Korunması Hakkında Kanun [7].

Elektronik Ticaretin Düzenlenmesi Hakkında Kanun Tasarısı elektronik araçlarla yapılan sözleşmeler ile elektronik ticarete ilişkin bilgi verme hizmeti sunanlara yönelik bazı yükümlülükler getirmektedir. E-ticaret sitesi sahiplerinin vergisel yükümlülüklerini yerine getirip getirmediikleri ülkelerin ilgili vergilendirme birimlerince önem taşımaktadır. Ülkemizde vergilendirme beyan usulüne dayalı

olduğunda inceleme yapılmaksızın kaçakların bulunması zordur. Bu nedenle bu çalışmada incelemelerde fayda sağlayacağı düşünülen bir uygulama geliştirilmiştir.

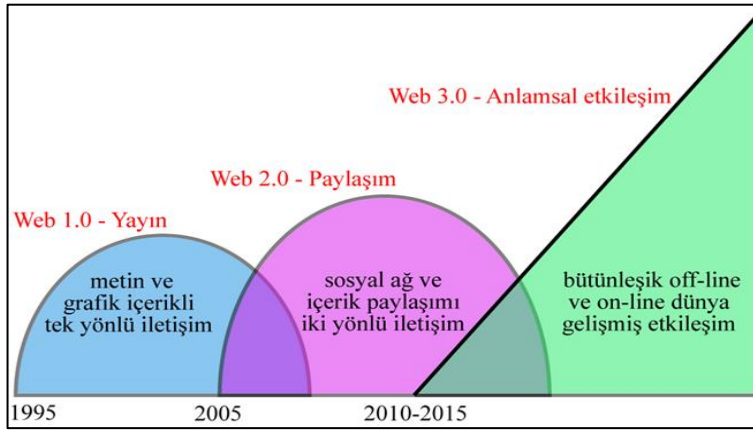
Bu tez çalışmasında kullanıcı tarafından belirtilen internet sitelerine bağlı olarak bu sitelerin içeriğini analiz eden, ardından metin madenciliği yöntemleri ile bu sitelerin e-ticaret sitesi olup olmadığına karar veren bir uygulama ve uygulama ara yüzü geliştirilmiştir. Tezin amacı, özel çalışmalar için kullanıcılara elektronik ticaret sitelerinin bulunmasını kolaylaştırmaktır. Çalışma sonucunda kullanıma sunulan uygulama mali denetim amaçlı olarak kullanılabilir. Vergilendirmeden sorumlu birimler bu uygulamayı kullanarak e-ticaret sitelerini tespit edebilir, ardından bu site sahiplerinin vergisel yükümlülüklerini yerine getirip getirmediğini araştırabilir.

Uygulamanın kullanım alanları şunlar olabilir:

- Ülkelerin vergi denetiminden sorumlu birimlerinin e-ticaret sitelerinin vergi mükellefiyeti tespiti için ön hazırlık,
- Dizin sitelerinde ilgili kategori altına yapılacak ekleme sırasında kategori önerme (mevcut durumda e-ticaret kategorisi),
- Endüstriyel pazarlamada pazar bölümlendirmesi yapmak için hedef kitlesi olarak kullanılacak e-ticaret sitelerinin tespiti:
 - bankalar tarafından düşük faizli kredi verilmek üzere e-ticaret sitelerinin tespiti
 - GSM operatörleri tarafından uygun tarife kampanyası yapılmak üzere e-ticaret sitelerinin tespiti

Bu tez kapsamında metin madenciliği ve e-ticaret konularıyla ilgili yapılan birçok çalışma incelenmiştir [8-23]. Metin madenciliği ile e-ticaretin bir arada incelendiği çalışmalara rastlanmıştır ancak bu çalışmalarda e-ticaret faaliyetleri ve geniş müşteri kitlesinden kaynaklanan veri hacminin artık stoklanmadan ziyade rekabet artırıcı yol oynaması amacıyla veri ve metin madenciliği yöntemleri ile incelenerek birbirleriyle ilişkilendirilen veriler içerisindeki desenleri ve düzensizlikleri tespit edilebilmenin mümkün olduğuna değinilmiştir. Ancak e-ticaret sitelerinin metin madenciliği yöntemleri ile tespitine dair bir çalışmaya rastlanmamıştır.

İnternetin gelişiminde şu üç kavram görülmektedir: Web 1.0, Web 2.0 ve Web 3.0. Web 1.0’de internet kullanıcısıyla internet sitesi arasında statik bir ilişki vardı. Kullanıcı sitenin pasif okuyucusuydu. Burada tek taraflı bir iletişim söz konusuydu. Web 2.0’de kullanıcı aktiftir. Kullanıcı ile internet sitesi arasında bir etkileşim vardır. Kullanıcı artık internet sayfasına müdahale edebilmektedir. Burada iki yönlü bir iletişimden söz konusudur. Web 1.0 dijital “monolog”, Web 2.0 ise “mobil tabanlı diyalog” olarak adlandırılmaktadır. Web 3.0 ise yorumlamanın temel olduğu bir sistemdir. Web 3.0 “semantik” (anlamsal web) olarak tanımlanmaktadır [24]. İnternetin yıllara göre gelişimi Şekil 1.1’de [24] verilmiştir.



Şekil 1.1. Web 1.0, web 2.0 ve web 3.0 gelişimi

Web 3.0 ortamında arama motorları “kişinin daha önceki internet davranışları” bilgisini algoritmalar kullanarak işleyerek ve yorum katarak bilgiyi sunacaktır. Web 3.0’ın amacı, internetteki bilgi arama ve düzenleme eksikliğini gidermektir [24]. İnternetin hızla gelişmesiyle birlikte artık bilgiye ulaşmadaki engeller büyük ölçüde ortadan kalmıştır. Sadece 5 yıl içinde 50 milyon kullanıcıya ulaşan internet 7’den 70’e herkesin kullandığı teknoloji ürünüdür ve her an yenilenmektedir [25]. Ancak bu durum beraberinde farklı bir zorluk getirmiştir. Bu, istenilen bilgiyi elde etmenin zorlaşmış olmasıdır. Bilgi kirliliği ve kullanıcı hata ya da eksikliklerinden kaynaklanan, özensiz bilgi paylaşımı bilgiye erişimi zorlaştırmıştır. Bu bilgi karmaşasının içerisinde kaybolmadan istenileni elde etmek için çözüm metin madenciliği ve metin sınıflandırmasıdır.

Metin madenciliği doğal dil işleme ile veri madenciliğinin bir arada kullanılması demektir [26]. Doğal dil işleme, insana özgü tüm dillerin işlenmesi ve kullanılması amacı ile araştırma yapan bilim dalıdır [27]. Veri madenciliği büyük veri yığınları içerisinde gelecekle ilgili tahminde bulunabilmeyi sağlayabilecek bağıntıların bilgisayar programı kullanarak aranmasıdır [9]. Metin madenciliğinde veri madenciliği yöntemleri ile eldeki veriler sınıflandırılarak, gruplandırılarak ya da veriler arasında ilişkiler, bağıntılar, istatistiksel sonuçlar oluşturularak modeller oluşturulur. Oluşturulan model, oluşturulduğu veri kümesinde olmayan yeni bir kayıt geldiğinde bu kayıt hakkında tahminleme yapma imkânı verir [28]. Metin sınıflandırma ise metin madenciliğinin yan uygulama alanı olup [29] dokümanları bilinen sınıflara atamaktadır. Amaç dokümanları otomatik sınıflandırmadır [10].

Tezin ikinci bölümünde metin sınıflandırma konusundan bahsedilmiştir. Ardından metin sınıflandırmada kullanılan bazı algoritmalara değinilmiştir. Daha sonra bu algoritmalar bir örnek üzerinde uygulanmıştır. Tezin üçüncü bölümünde metin sınıflandırmada kullanılmak üzere farklı kaynaklardan veriler toplanmış ve kullanıma hazır hale getirilmiştir. Biri statik diğeri ise dinamik olmak üzere iki çalışma yapılmıştır. Bu çalışmalara bağlı olarak farklı eğitim kümeleri oluşturulmuş ve bunlar üzerinde metin madenciliğinde sıklıkla kullanılan ve başarılı oldukları görülen K-En-Yakın-Komşu (k-NN) ve Naïve Bayes sınıflandırma algoritmaları kullanılmıştır. Ardından elde edilen sonuçlar içlerinden en iyi sonucu veren algoritmayı bulmak ve hazırlanan uygulamada kullanılmak üzere karşılaştırılmıştır. Daha sonra seçilen algoritma Java programlama dili ile masaüstü uygulaması olarak hazırlanan uygulama ara yüzüne aktarılmış ve kullanıma sunulmuştur. Tezin dördüncü bölümünde ise çalışmaya ait sonuçlara yer verilmiş ve gelecek çalışmalara dair önerilerde bulunulmuştur.

2. METİN SINIFLANDIRMA

Sınıflandırmanın amacı yeni bir nesnenin özelliklerini açıklamak ve bu yeni nesnenin daha önceden tanımlanmış sınıf setlerinden birine atamasını yapmaktır [30]. Sınıflandırma “denetimli makine öğrenimi” tekniklerinden birisidir [11]. Makine öğrenimi, bilgisayarların veri türlerine dayalı öğrenimini olanaklı kılan algoritmaların tasarım ve geliştirme süreçlerini konu edinen bir bilim dalıdır. Makine öğreniminde öğrenme çeşitleri denetimli öğrenme denetimsiz öğrenme olarak ikiye ayrılır [31].

Son zamanlarda metin sınıflandırma çok çeşitli alanlarda uygulanmaktadır. Bunlardan bazıları şunlardır:

- Metin bulup getirme ve kütüphane organizasyonu gibi alanlarda destek sağlayan dokümanlara sınıf ataması yapmak,
- Bu sınıfları insanların atadığı uygulamalarda sınıf atamasında rol oynamak,
- Mesajları, haberleri ve diğer metin dizisi halindeki bilgileri alıcılara ulaştırmak,
- Doğal dil işleme sistemlerinin bir parçası olarak
 - ilgisiz metinleri ve metin parçalarını filtrelemek,
 - metinleri, sınıf tabanlı işleme mekanizmalarına yönlendirmek,
 - sınırlı şekillerde bilgi edinimini sağlamak,
 - sözcük analizi işlerinde yardımcı olmak (sözcük belirsizliğini giderme) [12].

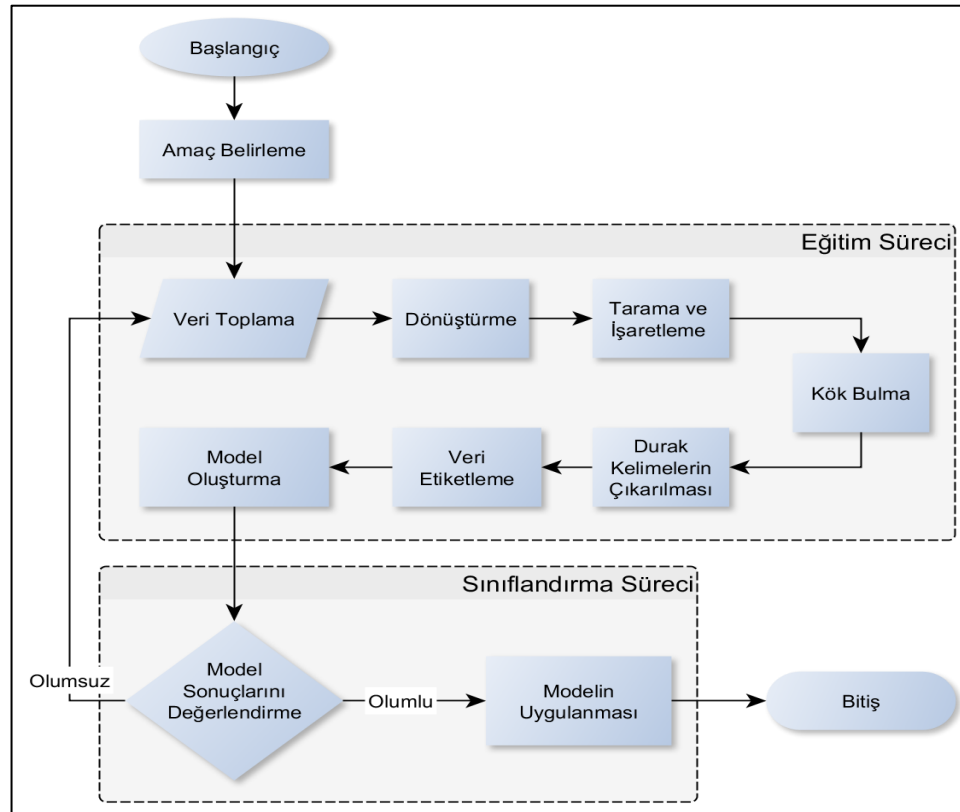
Bugüne kadar metin sınıflandırmada birçok yöntem kullanılmıştır:

- Naïve Bayes olasılıklı sınıflandırma [10, 13],
- K-En-Yakın-Komşu sınıflandırması (k-NN) [10, 13, 32],
- Karar Ağaçları [10, 13],
- Yapay Sinir Ağları [10, 13],
- Destek Vektörleri (SVM) [10].

Bunların yanında birden fazla yöntemin bir arada kullanılıp sonuçlarının karşılaştırmasının yapıldığı çalışmalar da olmuştur. Örneğin; [14-16, 33, 34]. Bahsi geçen çalışmalarda k-NN ve Naïve Bayes algoritmaları iyi sonuçları verdiği için bu çalışmada bu iki sınıflandırma yöntemi kullanılmıştır.

Denetimli öğrenmede sonuçlar eldeki verinin eğitilmesiyle elde edilir. Metin sınıflandırma iki aşamadan oluşur (Bkz. Şekil 2.1):

1. Eğitim süreci (eğitim kümesi - training) ve
2. Sınıflandırma süreci (sınama kümesi – test) [13].



Şekil 2.1. Metin sınıflandırma aşamaları

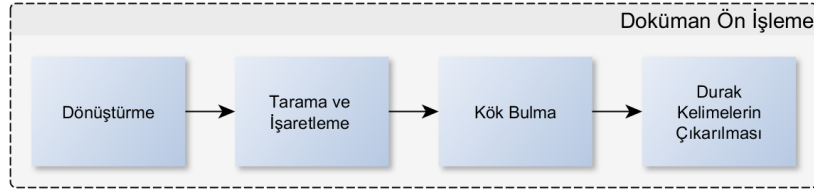
Eğitim sürecinde hangi sınıfa ait olduğu bilinen ve verilerin içerisinde rastgele seçilen bir kısım veri (eğitim kümesi - training set) eğitilerek bir model oluşturulur. Sınıflandırma süreciyse yine veri tabanından seçilen ve öğrenme kümesinde yer almayan (sınama kümesi - test set) veriler ile uygulanır [13].

2.1. Metin Gösterimi

Dokümanları sınıflandırmaya başlamadan önce, metin formatının düzgün ve kullanılmaya hazır olduğundan emin olmak için dokümanların bazı işlemlerden geçmesi gerekmektedir. Bu işlemler sırasıyla küçük bir örnek ile birlikte açıklanmıştır.

2.1.1. Doküman ön işleme

Veri ön işlemede başlangıç dokümanının eğitim ve sınıflandırma süreçlerine hazırlanması sağlanır. Veri ön işleme veri üzerinde bulunabilen problemleri çözmek, verinin doğal yapısını öğrenerek daha anlamlı ve kaliteli analiz yapabilmek ve veriden daha anlamlı bilgi üretebilmek gibi amaçlar için yapılır [17]. Doküman ön işleme adımlarından sonra metin bir diziye aktarılır. Veri ön işleme adımları Şekil 2.2’de gösterilmiştir.



Şekil 2.2. Veri ön işleme adımları

Dönüştürme: İnternette dokümanlar genellikle HTML (Hyper Text Markup Language) ve XML (Extensible Markup Language) gibi çeşitli tiplerde tutulduğundan, bunları düz metin haline dönüştürmek gerekmektedir. Bu adım HTML gibi formatlarla oluşturulmuş metinlerin etiketlerden kurtulması için gerekir.

Tarama ve işaretleme: Bu adım, metin içindeki terimleri ayıklamak için yapılır. İşaretlemede metin, simgeler ya da noktalama işaretleri ile kelime ya da ifadelere ayrılır [35]. Büyük / küçük harf duyarlılığını ortadan kaldırmak için metnin tamamı küçük harfe çevrilir.

Kök bulma: Aynı kökten gelen farklı ek almış kelimelerin doküman içerisindeki kelime sıklıklarına bakılırken aynı kelime olarak algılanması için köklerinin bulunması gerekmektedir.

Durak kelimelerin çıkarılması: Metin içerisinde çok sık geçen fakat sınıflandırmada bir anlam ifade etmeyen edat, bağlaç ve zamir gibi kelimelerin metinden çıkartılması gerekir. Bu kelimeler ayıklamada kullanılmak üzere bir tabloda ya da dosyada tutulabilir.

Örnek uygulamada eğitim kümesi olması istenen 3 tanesi e-ticaret sitesi sınıfından, diğer 3 tanesi ise e-ticaret sitesi olmayan siteler sınıfından olacak şekilde 6 internet sitesi seçilmiştir:

- www.gittigidiyor.com (Doküman 1) (e-ticaret sitesi),
- www.hepsiburada.com (Doküman 2) (e-ticaret sitesi),
- www.sahibinden.com (Doküman 3) (e-ticaret sitesi),
- www.berkehan.com (Doküman 4) (e-ticaret sitesi değil),
- www.btattoos.com (Doküman 5) (e-ticaret sitesi değil),
- www.mafm.boun.edu.tr (Doküman 6) (e-ticaret sitesi değil).

E-ticaret sitesi olup olmadığı sorgulanmak istenen site olarak ise “www.herkesemobilya.com” adresi seçilmiştir. Bu sitelerin içerikleri alınmış ve öncelikle HTML etiketlerinden temizlenmiştir. Ardından metindeki simge ve noktalama işaretleri çıkarılmıştır. Metin kelimelere ayrılarak küçük harfe çevrilmiştir. Az boyutlu bir örnek çalışma yapıldığı için durak kelimelerin çıkarılması ve kök bulma adımları göz kontrolü ile yapılmıştır. Bu tezdeki asıl uygulamalarda ise bu işlemi otomatik olarak yapan Zemberek Projesinin Java kütüphanesi kullanılmıştır.

2.1.2. Anahtar kelime sözlüğünün oluşturulması

Dokümanlar arasında kıyaslama yapabilmek için anahtar kelimelerden oluşan bir sözlüğe ihtiyaç duyulmaktadır. Bu adımda e-ticaret sınıfı için anahtar kelimeler belirlenerek anahtar kelime sözlüğü oluşturulmuştur. Anahtar kelime sözlüğü iki şekilde oluşturulabilir. Birinci seçenek belirli kelimelerin seçilmesidir [12]. Burada dikkat edilmesi gereken, anahtar kelimelerin sınıflara özel olarak, kendisini diğer sınıflardan ayıracak şekilde seçilmeye çalışılmasıdır. Anahtar kelimeler o konunun uzmanları ile görüşülerek belirlenirse nispeten daha sağlıklı sonuçlar elde edilebilir. İkinci seçenek ise metinde geçen tüm kelimelerin seçilmesidir [18].

İki yöntemde de kelimeler çalışmanın en başında statik olarak seçilmekte ve anahtar kelimelerin seçimi kullanıcıya bağlı olmaktadır. Farklı kullanıcıların aynı çalışmayı yaptığı bir durumda kullanıcıya bağlı seçimden dolayı aynı verilerle çok farklı sonuçlara ulaşılabilmektedir. Öte yandan kelimeleri dinamik olarak seçmek kullanıcının yanlış / eksik kelime seçmesini ve metinde yer alan tüm kelimeleri hesaba katarak çok boyutlu bir uzayda çalışmayı engellemektedir.

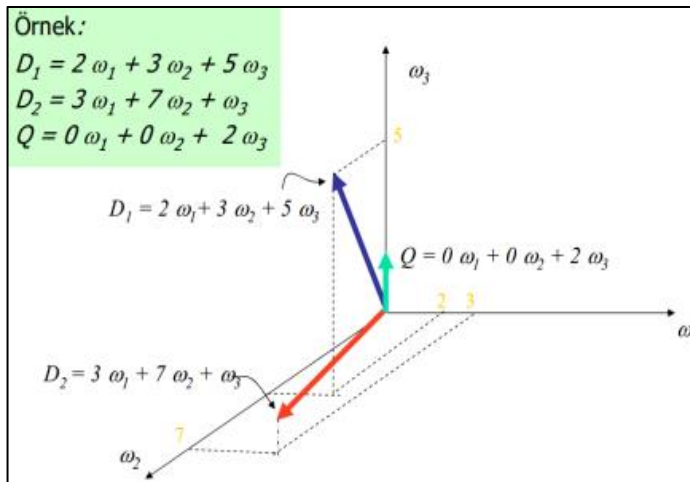
Bu tezde gerçekleştirilen iki çalışmadan birisinde eğitim kümesini oluştururken anahtar kelime sözlüğündeki kelimeler statik olarak seçilmiş, diğerinde ise dinamik olarak seçilmesi amacıyla farklı bir yöntem gerçekleştirilmiştir. Örnek uygulamada alışveriş sitelerinde ortak olarak yer aldığı düşünülen 7 kelime ele alındı. Anahtar kelime sözlüğünü oluşturan bu kelimeler şunlardır:

- uygun (Kelime 1),
- indirim (Kelime 2),
- ürün (Kelime 3),
- sepet (Kelime 4),
- stok (Kelime 5),
- kampanya (Kelime 6) ve
- alışveriş (Kelime 7).

2.1.3. Vektör uzay modelinin oluşturulması

Vektör uzay modeli bilgi çıkarımı, bilgi filtreleme, indeksleme gibi alanlarda kullanılan matematiksel bir modeldir ve doğal dil belgelerinin çok boyutlu uzayda özel bir anlamını simgelemektedir [12]. Metinleri vektörlerle ifade edebilmek için öncelikle uzay eksenlerinin belirlenmesi gerekir. Uzay eksenleri, sınıflandırmada o sınıfın belirleyici kelimelerinden (anahtar kelimeler) oluşur.

Dokümanlar Şekil 2.3'te [12] gösterildiği gibi kelimelerin vektörleri olarak ifade edilir (D_i : eğitim dokümanları, Q : sorgu dokümanı). Eksenler kelimeleri (w_i) ifade etmektedirler [12].



Şekil 2.3. Vektör uzay modeli

Vektör uzay modelinde kelimeler ve dokümanlar $i - j$ boyutlu bir uzayda gösterilir. Bu şekilde her doküman bir vektör olarak gösterilebilir. i 'nin değeri, incelenen kelime sayısıdır. j 'nin değeri ise incelenen doküman sayısıdır. Her bir boyut bir kelimeye karşılık gelmektedir. Örnek bir doküman - kelime çizelgesi Çizelge 2.1'de gösterilmiştir.

Çizelge 2.1. Örnek doküman - kelime çizelgesi

Kelime Doküman	Kelime 1	Kelime 2	Kelime 3	Kelime i
Doküman 1	Kelime Sıklığı	Kelime Sıklığı	Kelime Sıklığı	Kelime Sıklığı
Doküman 2	Kelime Sıklığı	Kelime Sıklığı	Kelime Sıklığı	Kelime Sıklığı
Doküman 3	Kelime Sıklığı	Kelime Sıklığı	Kelime Sıklığı	Kelime Sıklığı
Doküman j	Kelime Sıklığı	Kelime Sıklığı	Kelime Sıklığı	Kelime Sıklığı

Doküman – kelime çizelgesi oluşturulurken terim sıklığı ve kelimenin dokümanda geçip geçmediği bilgisi dikkate alınır.

Örnek uygulamada vektör uzay modeli -7 kelime olduğu için- 7 boyutludur. Doküman 1 (D1), Doküman 2 (D2) ve Doküman 3 (D3) olumlu sınıfta yer alırken, Doküman 4 (D4), Doküman 5 (D5) ve Doküman 6 (D6) olumsuz sınıfta yer almaktadır. Bu dokümanlar ile Sorgu Dokümanına (q) ait terim sıklığı doküman-kelime çizelgesi Çizelge 2.2’de gösterilmiştir. Çizelge 2.3’te ise doküman-kelime çizelgesi ikili olarak verilmiştir.

Çizelge 2.2. Örnek uygulama terim sıklığı doküman - kelime çizelgesi

Kelime Doküman	Kelime 1	Kelime 2	Kelime 3	Kelime 4	Kelime 5	Kelime 6	Kelime 7
	uygun	indirim	ürün	sepet	stok	kampanya	alışveriş
Doküman 1	6	0	18	18	18	9	5
Doküman 2	8	11	21	21	21	11	2
Doküman 3	3	0	14	14	14	7	3
Doküman 4	1	0	0	0	1	0	0
Doküman 5	2	0	0	0	1	0	0
Doküman 6	2	1	0	0	0	0	0
Sorgu Dokümanı	7	5	12	12	12	0	1

Çizelge 2.3. Örnek uygulama ikili doküman - kelime çizelgesi

Kelime Doküman	Kelime 1	Kelime 2	Kelime 3	Kelime 4	Kelime 5	Kelime 6	Kelime 7
	uygun	indirim	ürün	sepet	stok	kampanya	alışveriş
Doküman 1	1	0	1	1	1	1	1
Doküman 2	1	1	1	1	1	1	1
Doküman 3	1	0	1	1	1	1	1
Doküman 4	1	0	0	0	1	0	0
Doküman 5	1	0	0	0	1	0	0
Doküman 6	1	1	0	0	0	0	0
Sorgu Dokümanı	1	1	1	1	1	0	1

Vektör uzay modelinin oluşturulmasında dokümanlardaki her bir kelimenin ağırlıklandırma önemlidir. Vektör uzay modeli ağırlıklandırma sonucu oluşturulur.

2.1.4. Vektör ağırlıklandırma

Vektörlerin ağırlıklandırılması sınıflandırmanın başarısını etkilemektedir. Ağırlıklandırmada terim sıklığı (TF), ters doküman sıklığı (IDF), terim sıklığı - ters doküman sıklığı (TF-IDF), terim ayrıştırma değeri, olasılıksal terim ağırlıklandırma, tek terim doğruluğu, genetik algoritmalar gibi teknikler kullanılmaktadır. Bunlardan yaygın olarak kullanılan [11-13, 36] üç tanesi şu şekilde açıklanmıştır:

TF: Bu yöntemde metinde daha sık geçen kelimelerin, diğerlerine göre daha önemli olduğu düşünülür. TF ağırlıklandırma yönteminin hesaplanması Eş. 2.1'dedir [11].

$$TF_{ij} = \frac{n_{ij}}{|d_i|} \quad (2.1)$$

d_i : i. doküman

n_{ij} : d_i dokümanında w_j kelimesinin geçme sayısı

TF_{ij} : terim sıklığı

IDF: Bu yöntemde metinde daha az sayıda geçen kelimelerin ayırt edici özelliğinin diğerlerine göre daha fazla olduğu düşünülür. IDF ağırlıklandırma yönteminin hesaplanması Eş. 2.2’de verilmiştir [12].

$$IDF_j = \log \frac{n}{n_j} \quad (2.2)$$

n_j : w_j kelimesinin geçtiği doküman sayısı (doküman sıklığı)

n : dokümanlar kümesindeki doküman sayısı

IDF_j : w_j kelimesinin devrik doküman sıklığı

TF-IDF: Bu yöntemde bir dokümanda bulunan ancak diğer dokümanlarda daha az görülen bir terimin ağırlığının daha fazla olduğu düşünülür. TF-IDF ağırlıklandırma yönteminin hesaplanması Eş. 2.3’te verilmiştir [12].

$$x_{ij} = TF_{ij} \times IDF_j \quad (2.3)$$

x_{ij} : w_j kelimesinin d_i dokümanındaki TF-IDF ağırlığı

2.1.5. Dokümanlar arası benzerlik

Dokümanlar arasındaki ilişki bulunmak istendiğinde, her bir eğitim vektörü ile sorgu vektörü arasındaki açının *kosinüs değeri* hesaplanır ve sıralanır. İki belge arasındaki benzerlik şu şekilde ifade edilir: $s(d, d') \in R$. İki vektör arasındaki açının kosinüsü Eş. 2.4’te verilmiştir [36].

$$\cos(d_1, d_2) = d_1 \bullet d_2 / \|d_1\| \|d_2\| \quad (2.4)$$

$d_1 \bullet d_2$: d_i dokümanlarının uzunlukları çarpımı

$\|d_i\|$: d_i dokümanının uzunluğu

2.1.6. Sonuçların değerlendirilmesi

Eğitim dokümanlarının w ve sorgu dokümanının (q) vektör uzay modeli bulunur. Kullanıcı sorgusunun, her bir eğitim dokümanı ile olan benzerliği hesaplanır. Tüm dokümanlar vektörel olarak Eş. 2.5'teki gibi temsil edilir [12]. Sorgu dokümanı ile diğer dokümanlar arasındaki kosinüs benzerliği Eş. 2.6 ve 2.7'deki verilmiştir [36].

$$d_i = (wd_{i1}, wd_{i2}, \dots, wd_{ij}, \dots) \quad (2.5)$$

d_i : terimin doküman içerisindeki ağırlığı

w_{ij} : eğitim dokümanı vektörü

$$\text{sim}(d_i, q) = \cos \theta \quad (2.6)$$

$$\text{sim}(d_i, q) = \frac{d_i * q}{|d_i| |q|} = \frac{\sum_j (w_{i,j} * w_{q,j})}{\sqrt{\sum_j w_{i,j}^2} \sqrt{\sum_j w_{q,j}^2}} \quad (2.7)$$

d_i : terimin doküman içerisindeki ağırlığı

w_{ij} : eğitim dokümanı vektörü

q : sınıfının bulunması istenen vektör

Görülebileceği gibi (Bkz. Eş. 2.6 ve 2.7), eğer iki dokümanın hiç bir ortak kelimesi yoksa sonuç değeri 0 olur ($\text{sim}(d_i, q) = 0$). Öte yandan iki dokümanın tüm kelimeleri ortaksa (iki dokümanda da anahtar kelimelerin tamamı geçiyor ise) sonuç değeri 1 olur ($\text{sim}(d_i, q) = 1$). İşlem sonucunda çıkan değerler 0 ile 1 arasında olur ve yüksek olan değerler aranır [19]. Benzerlik oranları karşılaştırılır. En büyük benzerlik oranı dikkate alınarak sorgu vektörü ilgili sınıfa atanır. Örnek uygulama ile ilgili sonuçlara her bir algoritma içinde ayrı ayrı yer verilmiştir.

2.2. Metin Sınıflandırma Algoritmaları

Bu bölümde, yaygın olarak kullanılan [10, 13, 32] mesafeye dayalı sınıflandırma yöntemlerinden k-NN algoritması ve istatistiğe dayalı sınıflandırma yöntemlerinden Naïve Bayes algoritmasına değinilmiştir.

2.2.1. k-NN algoritması

k-NN algoritması içinde farklı sınıftan elemanların bulunduğu bir gruptan sorgulanmak istenilen dokümanın sınıfına benzeyen sınıfları bulur [20]. En yakın komşu araması, uzaklık bakımından kendisine en yakın olanı noktayı/noktaları bulmaya yarar. Aynı zamanda “benzerlik araması” ya da “en yakın nokta araması” olarak bilinir [37]. k-NN sınıflandırması sınıflandırmada yaygın olarak kullanılan denetimli makine öğrenimi tekniklerindendir. Yeni bir dokümanı sınıflandırmak için, eğitim kümesinde onun “komşu”su olarak adlandırılan dokümanlar kontrol edilir. Bu karşılaştırma yeni doküman ile her bir doküman arasındaki uzaklığa göre yapılır. Karşılaştırma sonucunda yeni doküman kendisine en yakın olan k sayıdaki (önceden eğitilmiş) dokümandan sınıfça çoğunlukta olanın sınıfına atanır. k sayısı dikkate alınacak en yakın komşu sayısını belirtir [19].

W uzayında S sayıdaki noktanın q dokümanı için sorgulandığı bir durumda S ile q arasındaki en yakın uzaklık bulunmaya çalışılır [37]. Benzerlik karşılaştırması için farklı yöntemler kullanılmaktadır. Bunlardan birisi, iki dokümanda da geçen ortak kelimelerin sayısıdır. İkinci ve yaygın olarak kullanılan yöntem ise iki nokta arasındaki açının kosinüs değerini bulmaktır [20].

k-NN algoritmasının avantajları şunlardır; uygulanabilirliği basittir, gürültülü eğitim dokümanlarına karşı dirençlidir ve eğitim sayısının fazla olduğu durumda etkilidir. k-NN algoritmasının dezavantajları şunlardır; sayısı önceden net olarak belirlenemeyen k parametreye ihtiyaç duyar ve uygulanması zaman gerektirmektedir.

Tüm eğitim dokümanları ve sınıfının bulunması istenen sorgu dokümanı vektörel olarak ifade edilirler. k-NN algoritmasında sorgu dokümanı ile diğer dokümanlar arasındaki kosinüs benzerliği hesaplanır (Bkz. Eş. 2.6 ve 2.7) [36]. Eğer iki dokümanın hiç ortak kelimesi yoksa sonuç değeri 0 ($sim(d_i, q) = 0$), tüm kelimeleri ortaksa sonuç değeri 1 olur ($sim(d_i, q) = 1$) [19].

Benzerlik hesaplamalarından sonra sınıfça üstün sayıda olan sınıfa atama yapılacağından bahsedilmişti. Eğer komşu olan iki sınıfın da eleman sayısı eşitse, bu durumda iki seçenek uygulanabilir. Birincisi, doküman her iki sınıfa da atanabilir. İkincisi ise k-NN algoritması için ağırlıklandırma yöntemleri kullanılabilir. k-NN algoritması terim ağırlıklandırmada kullanılan yöntemlerden bazıları şunlardır [11]:

1. İkili Ağırlıklandırma,
2. TF Ağırlıklandırma (Term Frequency - Terim Sıklığı) ,
3. TF-IDF Ağırlıklandırma (Term Frequency – Inverse Document Frequency).

k-NN algoritması ikili ağırlıklandırma yöntemi

İkili yöntemde vektörlerin ağırlıkları dokümanda bulunma veya bulunmamasına göre belirlenir. Yani kelimenin o dokümanda 3 kere bulunması ağırlığını değiştirmeyecektir. Kelimenin dokümanda olması 1, olmaması 0 ile ifade edilmektedir. Sözlükteki her bir kelimenin dokümanda geçip geçmediği kontrol edilir, 0 ya da 1 olarak ağırlıklandırılır. Bu şekilde eğitim dokümanlarını (internet sitelerini) kullanarak ikili ağırlıklandırma yöntemi ile kelimelerin dokümanda geçip geçmediğine göre vektörler oluşturulur.

Her bir eğitim vektörünün sorgu vektörüne olan uzaklığı hesaplanır. Ardından benzerlik oranları bulunur (Bkz. Eş. 2.7). Ardından bunlardan en yüksek değere sahip k tane en yakın olan dokümandan sınıf sayısı çoğunlukta olana sorgu dokümanının ataması yapılır. Örnek uygulama ile ilgili olarak doküman ön işleme aşamasında eğitim vektörleri ve sorgu vektörünü ağırlıklandırılmıştı (Bkz. Çizelge 2.3).

İkili olarak ifade edilen her eğitim vektörünün sorgu vektörüne (q) olan uzaklığı hesaplanarak Çizelge 2.4'te verilmiştir. Eş. 2.7'nin uygulanması için gereken diğer işlemler Çizelge 2.5'te verilmiştir.

Çizelge 2.4. Örnek uygulama eğitim vektörlerinin sorgu vektörüne uzaklığı (İkili)

Sorgu Kelime Çarpımları	1	2	3	4	5	6	7	$\sum_j (w_{i,j} * w_{q,j})$
D1 * q	1	0	1	1	1	0	1	5
D2 * q	1	1	1	1	1	0	1	6
D3 * q	1	0	1	1	1	0	1	5
D4 * q	1	0	0	0	1	0	0	2
D5 * q	1	0	0	0	1	0	0	2
D6 * q	1	1	0	0	0	0	0	2

Çizelge 2.5. Örnek uygulama diğer işlemler (İkili)

Sorgu Kelime Kareleri	1	2	3	4	5	6	7	$\sum_j w_i^2$	$\sqrt{\sum_j w_i^2}$
D1^2	1	0	1	1	1	1	1	6	2,4495
D2^2	1	1	1	1	1	1	1	7	2,6458
D3^2	1	0	1	1	1	1	1	6	2,4495
D4^2	1	0	0	0	1	0	0	2	1,4142
D5^2	1	0	0	0	1	0	0	2	1,4142
D6^2	1	1	0	0	0	0	0	2	1,4142
q^2	1	1	1	1	1	0	1	6	2,4495

Dokümanlar arasındaki benzerlik ilişkisini gösteren benzerlik hesaplamasının (Bkz. Eş. 2.7) sonucu Çizelge 2.6'da verilmiştir. Benzerlik oranının 1'e yakın olması, iki dokümanın aynı sınıfta olması anlamına gelmektedir. k-NN algoritması için gereken k parametresinin seçilmesinin gerekmektedir. k = 2 varsayımında bulunarak, oranı 1'e en yakın olan 2 değerin ait olduğu sınıflar, sorgu dokümanının komşuları demektir. Yani D1 ve D2 (ya da D2 ve D3) en yüksek değere sahip oldukları için sorgu dokümanının komşularıdır. Bu iki komşunun ait olduğu sınıf "olumlu" sınıf olduğundan, q vektörü de bu sınıfa aittir. Bu durumda "www.herkesemobilya.com" sitesi bir e-ticaret sitesidir.

Çizelge 2.6. Örnek uygulama benzerlik ilişkileri (İkili)

$\text{sim}(D_i, q)$	Olasılık
$\text{sim}(D1, q) =$	0,8333
$\text{sim}(D2, q) =$	0,9258
$\text{sim}(D3, q) =$	0,8333
$\text{sim}(D4, q) =$	0,5774
$\text{sim}(D5, q) =$	0,5774
$\text{sim}(D6, q) =$	0,5774

k-NN algoritması TF ağırlıklıklandırma yöntemi

TF yöntemde vektörlerin ağırlıkları dokümanda bulunma sayısına göre belirlenir. Sözlükteki her bir kelimenin dokümanda kaçar kez geçtiği kontrol edilerek vektörler ağırlıklandırılır.

Daha sonra her bir eğitim vektörünün sorgu vektörüne olan uzaklığı hesaplanır. Ardından benzerlik oranları bulunur (Bkz. Eş. 2.7). Ardından bunlardan en yüksek değere sahip k tane en yakın olan dokümandan sınıf sayısı çoğunlukta olana sorgu dokümanının ataması yapılır.

Örnek uygulama ile ilgili olarak doküman ön işleme aşamasında eğitim vektörleri ve sorgu vektörünü ağırlıklandırılmıştı (Bkz. Çizelge 2.2). Bu algoritmanın bir önceki algoritmadan tek farkı, ağırlıklandırmanın ikili yerine terim sıklığı dikkate alınarak yapılmış olmasıdır. Her bir eğitim vektörünün sorgu vektörüne (q) olan uzaklığı Çizelge 2.7’de hesaplanmıştır. Eş. 2.7’nin uygulanması için gereken diğer işlemler Çizelge 2.8’de verilmiştir.

Çizelge 2.7. Örnek uygulama eğitim vektörlerinin sorgu vektörüne uzaklığı (TF)

Sorgu Kelime Çarpımları	1	2	3	4	5	6	7	$\sum_j (w_{i,j} * w_{q,j})$
D1 * q	42	0	216	216	216	0	5	695
D2 * q	56	55	252	252	252	0	2	869
D3 * q	21	0	168	168	168	0	3	528
D4 * q	7	0	0	0	12	0	0	19
D5 * q	14	0	0	0	12	0	0	26
D6 * q	14	5	0	0	0	0	0	19

Çizelge 2.8. Örnek uygulama diğer işlemler (TF)

Sorgu Kelime Kareleri	1	2	3	4	5	6	7	$\sum_j w_i^2$	$\sqrt{\sum_j w_i^2}$
D1^2	36	0	324	324	324	81	25	1114	33,3766
D2^2	64	121	441	441	441	121	4	1633	40,4104
D3^2	9	0	196	196	196	49	9	655	25,5930
D4^2	1	0	0	0	1	0	0	2	1,4142
D5^2	4	0	0	0	1	0	0	5	2,2361
D6^2	4	1	0	0	0	0	0	5	2,2361
q^2	49	25	144	144	144	0	1	507	22,5167

Dokümanlar arasındaki benzerlik ilişkisini gösteren benzerlik hesaplamasının (Bkz. Eş. 2.7) sonucu Çizelge 2.9’da verilmiştir. Yine $k = 2$ varsayımında bulunarak, oranı 1’e en yakın olan 2 değerin ait olduğu sınıflar, sorgu dokümanının komşuları demektir. Yani D1 ve D2 en yüksek değere sahip oldukları için q dokümanının komşularıdır. Bu iki komşunun ait olduğu sınıf “olumlu” sınıf olduğundan, q vektörü de bu sınıfa aittir. Bu durumda “www.herkesemobilya.com” sitesi bir e-ticaret sitesidir.

Çizelge 2.9. Örnek uygulama benzerlik ilişkileri (TF)

$\text{sim}(D_i, q)$	Olasılık
$\text{sim}(D1, q) =$	0,92478
$\text{sim}(D2, q) =$	0,9550
$\text{sim}(D3, q) =$	0,9162
$\text{sim}(D4, q) =$	0,5967
$\text{sim}(D5, q) =$	0,51640
$\text{sim}(D6, q) =$	0,3774

k-NN – TF-IDF ağırlıklandırma yöntemi

TF-IDF ağırlıklandırmasında her bir dokümandaki kelimelerin sıklığı rol oynamaktadır. Böylece dokümanda daha fazla görülen kelimeler varsa (terim sıklığı yüksek) o doküman için daha değerli olduğu anlaşılır. Ayrıca IDF tüm dokümanlarda seyrek görülen kelimeler ile ilgili bir ölçü verir. Bu değer tüm eğitim dokümanlarından hesaplanır. Bu yüzden eğer bir kelime dokümanlarda sık geçiyorsa doküman için belirleyici olmadığı düşünülebilir (durak kelimeler). Eğer kelime dokümanlarda çok sık geçmiyorsa o kelimenin o doküman için belirleyici özelliği vardır denebilir. TF-IDF genel olarak sorgu vektörü ile eğitim dokümanı vektörü arasındaki benzerlik oranını bulmak için kullanılır. TF-IDF fonksiyonunun çeşitli versiyonları mevcuttur [21]. Bu algoritmanın uygulanması sözlükteki her bir kelime için Eş. 2.8, 2.9'daki formüller uygulanarak vektörler ağırlıklandırılır.

$$w_{d,t} = tf_{d,t} \bullet idf_t \quad (2.8)$$

$$idf_t = \log\left(\frac{D}{dft}\right) \quad (2.9)$$

tf : Dokümanda bulunan kelimenin dokümandaki görülme sıklığı

D : Toplam doküman sayısı

dft : Doküman sıklığı (Kaç eğitim dokümanında ilgili kelime geçmiş)

Daha sonra her bir eğitim vektörünün sorgu vektörüne olan uzaklığı hesaplanır. Benzerlik oranları bulunur (Bkz. Eş. 2.7). Ardından bunlardan en yüksek değere sahip k tane en yakın olan dokümandan sınıf sayısı çoğunlukta olana sorgu dokümanının ataması yapılır.

Örnek uygulama ile ilgili olarak doküman ön işleme aşamasında eğitim vektörleri ve sorgu vektörünü ağırlıklandırılmıştı (Bkz. Çizelge 2.2). IDF değerlerinin hesaplanması Çizelge 2.10’da gösterilmiştir. Eş. 2.8 yardımı ile TF-IDF değerleri Çizelge 2.11’de hesaplanmıştır.

Çizelge 2.10. Örnek uygulama IDF Değerleri

Kelimeler Hesaplamala	1	2	3	4	5	6	7
D	6						
dft	6	2	3	3	4	4	3
IDF=log(D/dft)	0	0,4771	0,3010	0,3010	0,1760	0,1760	0,3010

Çizelge 2.11. Örnek uygulama TF-IDF değerleri

Kelimeler Hesaplamala	1	2	3	4	5	6	7
TF1 * IDF	0	0	5,4185	5,4185	3,1696	1,5848	1,5056
TF2 * IDF	0	5,2483	6,3216	6,3216	3,6979	1,9370	0,6021
TF3 * IDF	0	0	4,2144	4,2144	2,4653	1,2326	0,9031
TF4 * IDF	0	0	0	0	0,1761	0	0
TF5 * IDF	0	0	0	0	0,1761	0	0
TF6 * IDF	0	0,4771	0	0	0	0	0
TFS * IDF	0	2,3856	3,6124	3,6124	2,1131	0	0,3010

Eğitim vektörlerinin sorgu vektörüne olan uzaklığı Çizelge 2.12’de hesaplanmıştır. Eş. 2.7’nin uygulanması için gereken diğer işlemler Çizelge 2.13’te verilmiştir.

Çizelge 2.12. Örnek - eğitim vektörlerinin sorgu vektörüne uzaklığı (TF-IDF)

Kelimeler Sorgu Çarpımları	1	2	3	4	5	6	7	$\sum_j (w_{i,j} * w_{q,j})$
D1 * q	0	0	19,5700	19,5700	6,6980	0	0,4530	46,2983
D2 * q	0	12,5200	22,8400	22,8400	7,8140	0	0,1810	66,1878
D3 * q	0	0	15,2200	15,2200	5,2090	0	0,2720	35,9292
D4 * q	0	0	0	0	0,3720	0	0	0,3721
D5 * q	0	0	0	0	0,3720	0	0	0,3721
D6 * q	0	1,1380	0	0	0	0	0	1,1382

Çizelge 2.13. Örnek uygulama diğer işlemler (TF-IDF)

Kelimeler Sorgu Kareleri	1	2	3	4	5	6	7	$\sum_j w_i^2$	$\sqrt{\sum_j w_i^2}$
D1^2	0	0	383,1000	383,1000	44,8600	0	0,2050	811,3260	28,4800
D2^2	0	156,8000	521,5000	521,5000	61,0600	0	0,0330	1260,8201	35,5100
D3^2	0	0	231,8000	231,8000	27,1400	0	0,0740	490,7519	22,1500
D4^2	0	0	0	0	0,1380	0	0	0,1385	0,3720
D5^2	0	0	0	0	0,1380	0	0	0,1385	0,3720
D6^2	0	1,2960	0	0	0	0	0	1,2956	1,1380
q^2	0	1,1380	1,0870	1,0870	0,3720	0	0,0910	3,7758	1,0670

Dokümanlar arasındaki benzerlik ilişkisini gösteren benzerlik hesaplamasının (Bkz. Eş. 2.7) sonucu Çizelge 2.14’te verilmiştir. Yine $k = 2$ varsayımında bulunarak, oranı 1’e en yakın olan 2 değerin ait olduğu sınıflar, sorgu dokümanının komşuları demektir. Yani D1 ve D2 en yüksek değere sahip oldukları için q dokümanının komşularıdır. Bu iki komşunun ait olduğu sınıf “olumlu” sınıf olduğundan, q vektörü de bu sınıfa aittir. Bu durumda “www.herkesemobilya.com” sitesi bir e-ticaret sitesidir.

Çizelge 2.14. Örnek uygulama benzerlik ilişkileri (TF-IDF)

$\text{sim}(D_i, q)$	Olasılık
$\text{sim}(D1, q) =$	0,4687
$\text{sim}(D2, q) =$	0,5376
$\text{sim}(D3, q) =$	0,4678
$\text{sim}(D4, q) =$	0,2884
$\text{sim}(D5, q) =$	0,2884
$\text{sim}(D6, q) =$	0,2884

2.2.2. Naïve Bayes algoritması

Naïve (saf) Bayes Algoritması istatistiğe dayalı, Bayes Sınıflandırma Algoritmalarından birisidir. Olasılık tabanlıdır. Bu algoritmaya göre kelimeler sınıftan bağımsızdır [38]. Uygulanabilirliği kolay bir algoritma olup performans bakımından da başarılıdır. Naïve Bayes sınıflandırıcısı için ikili (Multi-variate) ve TF (Multinomial) ağırlıklandırma modeli olmak üzere 2 model mevcuttur [22].

Naïve Bayes algoritması ikili ağırlıklandırma (Multi-variate model)

Bu algoritma ile dokümanlarda kelimelerin olup olmasına bakılarak ağırlıklandırma yapılır. İkili ağırlıklandırma metodu kullanılarak dokümanların ağırlıkları bulunur. Doküman 1'ler ve 0'lardan oluşacak şekilde ifade edilir. İki olasılıklı durumlarda ikili sınıflandırma kullanılabilir. Örneğin; e-ticaret sitesi olan siteler sınıfı ve e-ticaret sitesi olmayan siteler sınıfı.

Sonuçta sözlük boyutu kadar 1'ler ve 0'lardan oluşan vektörler elde edilir. Bu vektörün herhangi bir sınıfta olma olasılığını hesaplamak için kullanılır. d vektörünün c_j sınıfında olma olasılığının hesaplanması Eş. 2.10, 2.11 ve 2.12'de gösterilmiştir.

$$p(d | c_j) = \prod_{t=1}^{|V|} p(w_t | c_j)^{x_t} (1 - p(w_t | c_j))^{(1-x_t)} \quad (2.10)$$

$$p(w_t | c_j) = \frac{1 + B_{jt}}{2 + |C_j|} \quad (2.11)$$

$$M(C) = P(word1 | C)^{n1} P(word2 | C)^{n2} \dots P(wordv | C)^{nv} P(|C) \quad (2.12)$$

$|V|$: sözlükteki kelime sayısı

B_{jt} : c_j sınıfında bulunan ve w_t kelimesini içeren eğitim dokümanı sayısı

$|C_j|$: c_j sınıfında bulunan eğitim dokümanı sayısı

x_t : Kelimenin ağırlığı (1 ya da 0)

Sorgu dokümanı $M(C)$ değeri en büyük olan sınıfa atanır. Örnek uygulama için olumlu ve olumsuz dokümanlarda geçen kelimeleri kendi içlerinde toplanarak hangi kelimenin hangi sınıfta (ikili olarak) kaç kez geçtiği Çizelge 2.15'te bulunmuştur. Olumlu ve olumsuz sınıf sayısı Çizelge 2.16'da verilmiştir. Eş. 2.10'un uygulanması için gereken diğer bilgiler Çizelge 2.17'de, sonuçları ise Çizelge 2.18'de verilmiştir.

Çizelge 2.15. Örnek uygulama kelimenin her sınıftaki ikili sıklığı

Sınıflar \ Kelimeler	1	2	3	4	5	6	7
D_olumlu	3	1	3	3	3	3	3
D_olumsuz	3	1	0	0	2	0	0

Çizelge 2.16. Örnek uygulama olumlu ve olumsuz sınıf sayısı

Sınıf	Sayı
cj_olumlu	3
cj_olumsuz	3

Çizelge 2.17. Örnek uygulama Naïve Bayes ikili diğer işlemler

Sınıflar \ Kelimeler	1	2	3	4	5	6	7
Olumlu	0,8	0,4	0,8	0,8	0,8	0,8	0,8
Olumsuz	0,8	0,4	0,2	0,2	0,6	0,2	0,2

Çizelge 2.18. Örnek uygulama Naïve Bayes ikili sonuç

Kelimeler Sınıflar	1	2	3	4	5	6	7	Sonuç
M(Olumlu)	0,8	0,4	0,8	0,8	0,8	0,2	0,8	0,02621
M(Olumsuz)	0,8	0,4	0,2	0,2	0,6	0,8	0,2	0,0012

En büyük $M(C)$ değeri olumlu sınıfa ait olduğundan, sorgu dokümanı olumlu sınıfa atanır. Bu durumda “www.herkesemobilya.com” sitesi bir e-ticaret sitesidir.

Naïve Bayes algoritması TF ağırlıklandırma (Multinomial model)

Doküman içindeki kelimelerin tekrar sayılarını dikkate alınır. Bu algoritmadaki varsayım, her bir kelimenin tekrar etme sayısının diğer kelimelerin tekrar etme sayılarından bağımsız olmasıdır [23]. Sonuçta sözlük boyutu kadar, kelimelerin tekrar sayılarından oluşan vektörler elde edilir. d vektörünün c_j sınıfında olma olasılığı Eş. 2.13, 2.14 ve 2.15’te gösterilmiştir.

$$p(d | c_j) = p(|d|) |d|! \prod_{t=1}^{|V|} \frac{p(w_t | c_j)^{x_t}}{x_t!} \quad (2.13)$$

$$p(w_t | c_j) = \frac{1 + N_{jt}}{|V| + V_j} \quad (2.14)$$

$$M(C) = P(word1 | C)^{n1} P(word2 | C)^{n2} \dots P(wordv | C)^{nv} P(|C|) \quad (2.15)$$

d : sınıf sayısı

N_{jt} : j sınıfındaki dokümanlar içinde t kelimesinin görülme sıklığı

N_j : j sınıfındaki toplam kelime sayısı

$P | d |$: sınıf olasılığı

x_t : kelime sıklığı

$|V|$: sözlükteki kelime sayısı

Sorgu dokümanı $M(C)$ değeri en büyük olan sınıfa atanır. Örnek uygulama için olumlu ve olumsuz dokümanlarda geçen kelimeler kendi içlerinde toplanarak hangi

kelimenin hangi sınıfta kaç kez geçtiği bilgisi bulunarak Çizelge 2.19’da gösterilmiştir. Sınıf sayısı ve sözlükteki kelime sayısı ise Çizelge 2.20’de verilmiştir. Eş. 2.15’in uygulanması için gereken diğer işlemler Çizelge 2.21’de, Eş. 2.15’in uygulanmasıyla elde edilen sonuçlar ise Çizelge 2.22’de verilmiştir.

Çizelge 2.19. Örnek uygulama kelimelerin her sınıftaki TF sıklığı

Kelimeler Sınıflar	1	2	3	4	5	6	7	Toplam
D_olumlu	17	11	53	53	53	27	10	224
D_olumsuz	5	1	0	0	2	0	0	8

Çizelge 2.20. Örnek uygulama sınıf ve sözlükteki kelime sayısı

Hesaplama	Değerler
P(d)	0,5
 d !	2
 V 	7

Çizelge 2.21. Örnek uygulama Naïve Bayes TF diğer işlemler

Kelimeler Sınıflar	1	2	3	4	5	6	7	Çarpım
D_olumlu	3E-12	3E-09	6E-17	6E-17	6E-17	1	0,048	8,9E-71
D_olumsuz	3E-07	4E-07	2E-23	2E-23	9E-18	1	0,067	1,7E-77

Çizelge 2.22. Örnek uygulama Naïve Bayes TF sonuç

Sınıflar	Sonuç
M(Olumlu)	8,9E-71
M(Olumsuz)	1,7E-77

En büyük M(C) değeri olumlu sınıfa ait olduğundan, sorgu dokümanı olumlu sınıfa atanır. Bu durumda “www.herkesemobilya.com” sitesi bir e-ticaret sitesidir.

2.3. Metin Sınıflandırma Aracı WEKA

WEKA (Waikato Environment for Knowledge Analysis) veri madenciliği ve metin madenciliğinde kullanılan ücretsiz bir programdır. WEKA makine öğrenimi amacıyla Waikato Üniversitesinde Java programlama dili ile geliştirilmiştir.

WEKA, geliştirilen ara yüz aracılığıyla kullanılabildiği gibi Java kütüphaneleri kullanılarak geliştirilen yeni programların içerisinde de kolaylıkla kullanılabilmektedir [39]. Bu şekilde platformdan bağımsız olarak çalışabilmektedir. WEKA bünyesinde 49 doküman ön işleme aracı, 76 sınıflandırma/regresyon algoritması ve 8 kümeleme algoritması barındırmaktadır.

WEKA’da “Attribute-Relation File Format” (ARFF) ve “Comma-Separated Variables” (CSV) gibi formatlardaki veriler kullanılır. “ARFF” dosyalarında verilerin özelliklerini (attributes) içeren başlık bilgisi ve birbirinden virgül karakteri ile ayrılmış olan veri bulunur. Verilerin özellikleri; nominal, sayısal (numeric), metin (string), tarih (date) veya ilişkisel (relational) olarak olabilir. Örnek uygulamanın sorgu ve eğitim verilerini içeren “ARFF” formatı Şekil 2.4 ve Şekil 2.5’te verilmiştir.

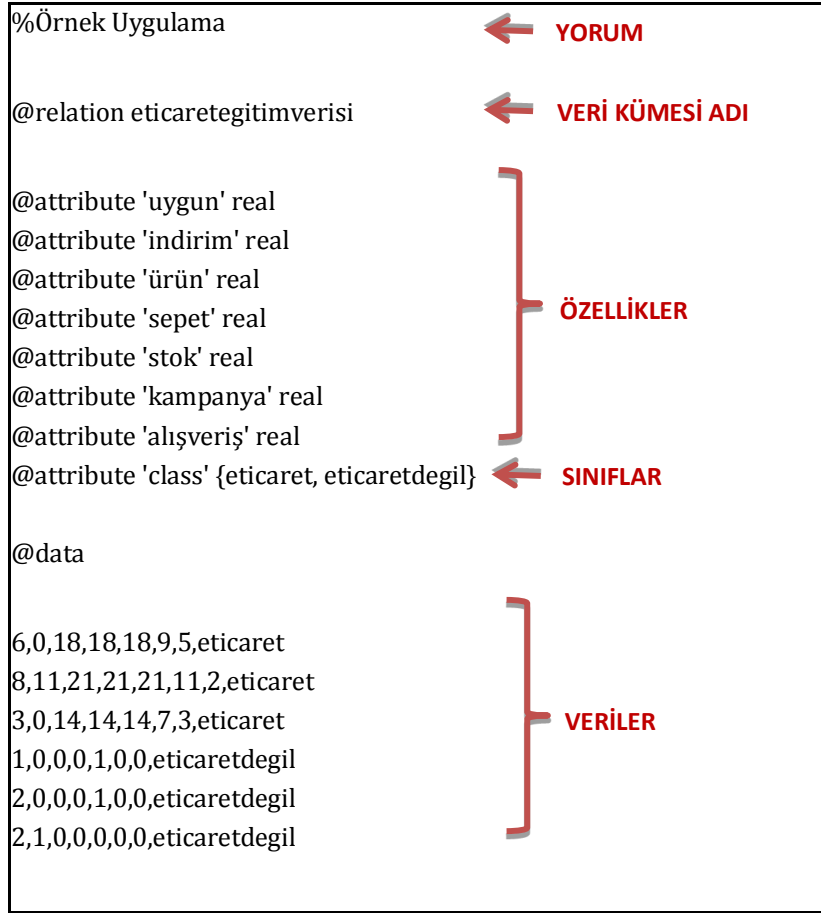
```
@relation eticarettestverisi

@attribute 'uygun' real
@attribute 'indirim' real
@attribute 'ürün' real
@attribute 'sepet' real
@attribute 'stok' real
@attribute 'kampanya' real
@attribute 'alışveriş' real
@attribute 'class' {eticaret, eticaretdegil}

@data

7,5,12,12,12,0,1,?
```

Şekil 2.4. Örnek uygulama sorgu “.ARFF” dosyası

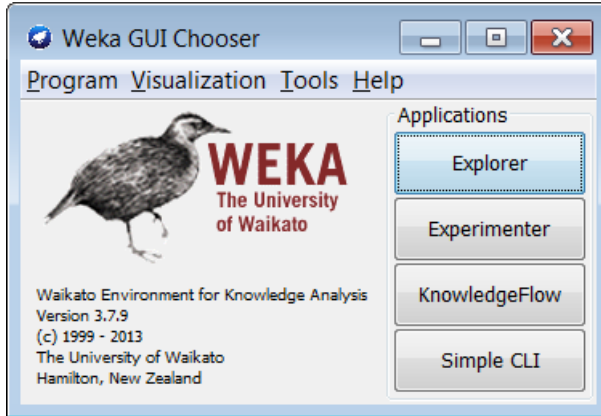


Şekil 2.5. Örnek uygulama eğitim “.ARFF” dosyası

WEKA programının ara yüz ekranı Şekil 2.6’da verilmiştir. WEKA ara yüzünde aşağıda belirtilen dört bölüm bulunmaktadır:

- Explorer,
- Experimenter,
- KnowledgeFlow ve
- Simple CLI (Command Line Interface)

“Explorer” bölümünde veri üzerinde araştırma yapılır, “Experimenter” bölümünde veri üzerinde deneyler yapılarak istatistiksel sonuçlar çıkarılmaya çalışılır, “KnowledgeFlow” bölümünde “Explorer” bölümünde yapılanlar işlemleri görselleştirerek yapılır ve bilgi akışının modellenmesi sağlanır, “Simple CLI” bölümündeyse işlemler ara yüz yerine komut satırı kullanılarak gerçekleştirilir [40].



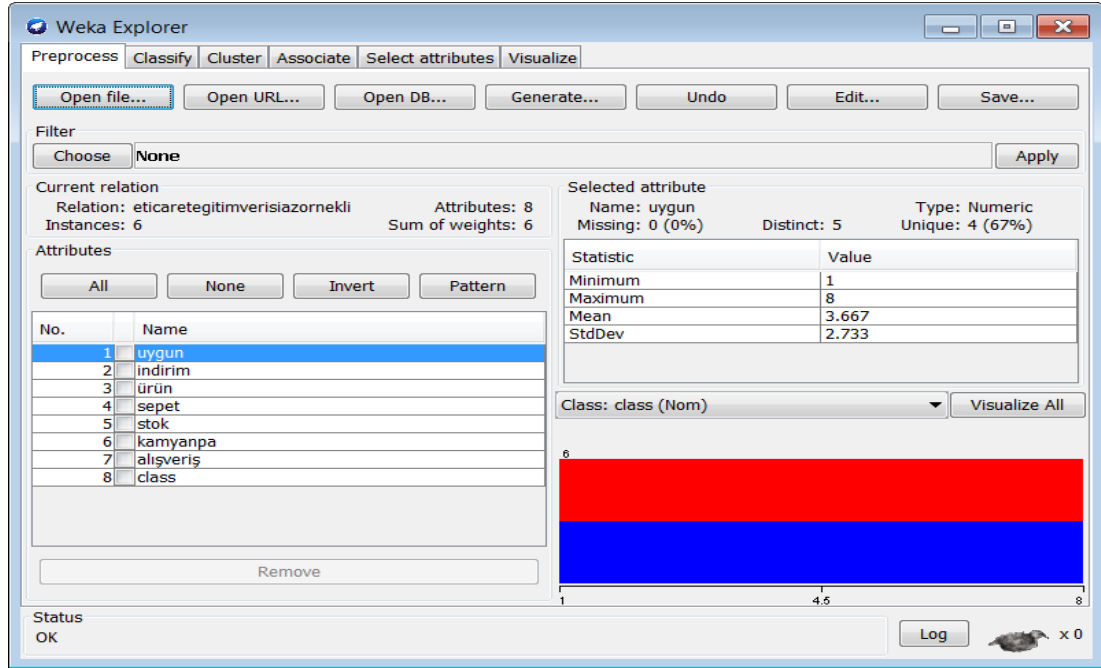
Şekil 2.6. WEKA ara yüzü ekran görüntüsü

“Explorer” bölümünde doküman ön işleme işlemleri yapılabilir, sınıflandırma, kümeleme ve birliktelik algoritmaları kullanılabilir, özellik seçimi yapılabilir ve farklı nitelikler arasındaki ilişkilerin görselleştirilmesi sağlanabilir. Veriler metin dosyasından, internette dosyanın bulunduğu konumdan ya da veri tabanından alınabileceği gibi yapay olarak da oluşturulabilmektedir. Örnek eğitim veri dosyasının WEKA’da açıldıktan sonraki ekran görüntüsü Şekil 2.7’de verilmiştir.

Veriler seçildikten sonra gerekirse veri ön işleme aşamaları gerçekleştirilir. Ardından sınıflandırma aşamasına geçilir. Bu aşamada istenilen sınıflandırma algoritması seçilerek model eğitim kümesindeki veriler ile eğitilir.

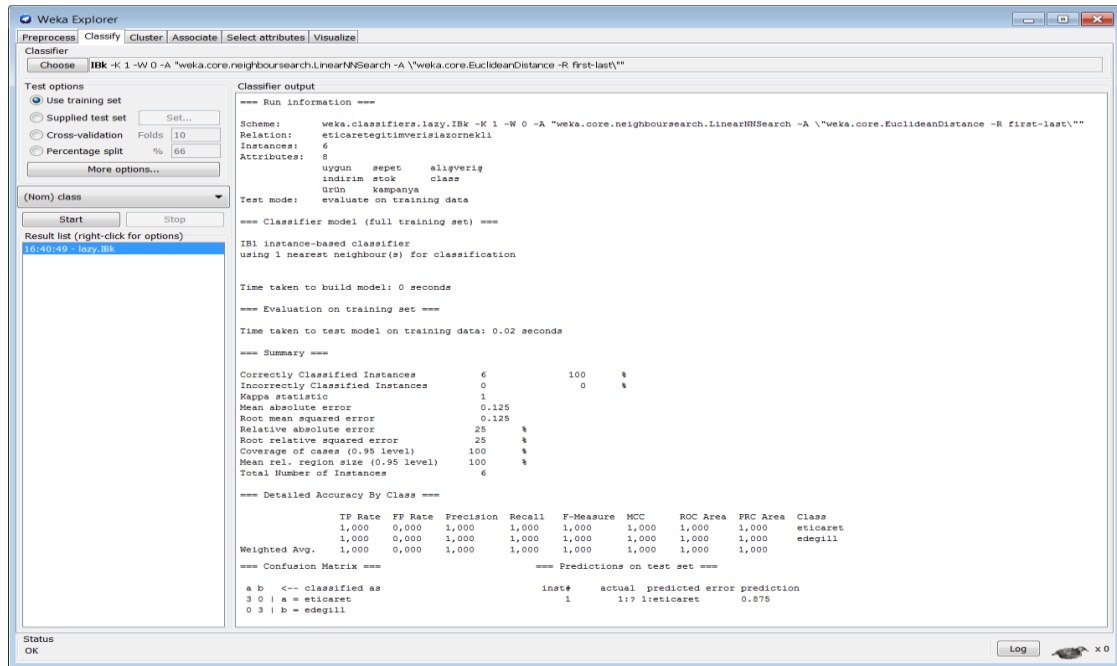
Eğitim aşamasında dört seçenek kullanılabilir:

1. *Eğitim kümesini kullanma (Use training set)*: Veri kümesi sadece modeli eğitmek için kullanılır.
2. *Sınama kümesi ile sınama (Supplied test set)*: Veri kümesi “Use training set” seçeneğiyle eğitildikten sonra modelin doğruluğu ikinci bir küme ile sınanır.
3. *Çapraz geçerleme (Cross-validation)*: Veri kümesi n eşit kümeye ayrılır ve eğitim için $n-1$ küme (validating set), sınama içinse 1 küme (test set) kullanılır. Bu işlem belirtilen küme sayısı kadar tekrarlanarak sistem iyileştirilmeye çalışılır.
4. *Yüzde ayırma (Percentage split)*: Veri kümesini ikiye böler. Veri kümesinin belirli bir yüzdesi modelin eğitilmesi, kalan kısmı ise modelin sınanması için kullanılır [41].



Şekil 2.7. WEKA ara yüzü dosya açma ekran görüntüsü

Modelin eğitilmesinden sonra sonuçlar ekrana gelmektedir. “Run Information” kısmında algoritma seçenekleri, veri kümesi adı, veri sayısı, özellikler, ve test seçeneği yer alır. “Classifier model” kısmında sınıflandırma modelinin metinsel gösterimi ve modeli oluşturmak için geçen süre gösterilir. “Summary” kısmında sınıflandırıcının tahmin etme başarısı ile ilgili istatistiki bilgileri yer alır. “Detailed Accuracy By Class” kısmında sınıflandırıcının tahmin doğruluğu daha detaylı olarak sınıf bazında gösterilir. “Confusion Matrix” (hata matrisi) kısmındaysa her sınıfa kaç adet verinin atandığı bilgisi matris şeklinde verilir [40]. Örnek uygulamanın k-NN ($k = 1$) algoritmasıyla eğitilmesiyle oluşan modelin görüntüsü Şekil 2.8’de verilmiştir.



Şekil 2.8. Örnek uygulama eğitim ve sınıflandırma sonucu

Sınıflandırma modelinin başarısının değerlendirilmesinde hata matrisi, doğruluk (accuracy), kesinlik (precision), hassasiyet (recall) ve F-Ölçütü kullanılır. Bu bilgilere ulaşmak için hata matrisi kullanılır. İki sınıftan oluşan bir model için hata matrisi Çizelge 2.23'te verilmiştir.

Çizelge 2.23. Hata matrisi örneği

Doğru Sınıf \ Tahmin Edilen Sınıf	A Sınıfı	B Sınıfı
	A Sınıfı	B Sınıfı
A Sınıfı	TP	FN
B Sınıfı	FP	TN

Hata matrisinde TP (True Positive) gerçekte A sınıfında olup model tarafından da A sınıfına atanan veri sayısını, FN (False Negative) gerçekte A sınıfında olup model tarafından B sınıfına atanan veri sayısını, TN (True Negative) gerçekte B sınıfında olup model tarafından da B sınıfına atanan veri sayısını ve FP (False Positive) gerçekte B sınıfında olup model tarafından A sınıfına atanan veri sayısını ifade etmektedir [28].

Doğruluk tüm verinin yüzde kaçının doğru olarak sınıflandırıldığı bilgisini verir. Hesaplanışı Eş. 2.16'da verilmiştir. Kesinlik gerçekte ilgili sınıfın yüzde kaçının doğru olarak sınıflandırıldığı bilgisini verir. Hesaplanışı Eş. 2.17'de verilmiştir. Hassasiyet tahmin edilen ilgili sınıfın yüzde kaçının doğru olarak sınıflandırıldığı bilgisini verir. Hesaplanışı Eş. 2.18'de verilmiştir. F-Ölçütü ise duyarlılık ve kesinliğin harmonik ortalamasıdır. Hesaplanışı Eş. 2.19'da verilmiştir [28].

$$Doğruluk = \frac{TP + TN}{TP + FP + FN + TN} \quad (2.16)$$

$$Kesinlik = \frac{TP}{TP + FN} \quad (2.17)$$

$$Duyarlılık = \frac{TP}{TP + FP} \quad (2.18)$$

$$F - \text{Ölçütü} = \frac{2 \times \text{Duyarlılık} \times \text{Kesinlik}}{\text{Duyarlılık} + \text{Kesinlik}} \quad (2.19)$$

3. UYGULAMA

Bu tezde özel araştırmalar için kullanıcılara e-ticaret sitelerinin tespitini kolaylaştırmak amacıyla, kullanıcı tarafından belirtilen internet sitelerinin içeriğini analiz ederek metin madenciliği yöntemleri ile bu sitelerinin e-ticaret sitesi olup olmadığına karar veren bir uygulama yapılmıştır. Çalışmada öncelikle farklı yöntemlerle elde edilen veriler üzerinde farklı metin madenciliği algoritmaları kullanılmış ve karşılaştırılmıştır. Ardından seçilen algoritma kullanılarak Java programlama dili ile bir ara yüz oluşturulmuştur.

Bu tezde Java programlama dilinin kullanılmasının sebebi Java'nın platformdan bağımsız, yüksek performanslı, nesneye yönelik ve çok işlevli bir dil olmasıdır. Java, özellikle ağ ve Internet uygulamaları için geliştirilmiş üst düzey bir programlama dilidir [42]. Java ile herhangi bir ortamda oluşturulan bir yazılım başka bir ortamda da kullanılabilir

Geliştirilen uygulamanın oluşturulmasında kullanılan yazılımlar şunlardır:

- İşletim sistemi : Windows
- Programlama dili : Java
- Kullanıcı ara yüzü dili : Swing
- Kullanılan kütüphaneler : WEKA, Zemberek ve JSoup
- Programlama dili geliştirme yazılımı : Eclipse ve Netbeans

Bu tezde hem WEKA'nın "Explorer" ara yüzü hem de Java ile yazılan kod parçalarında WEKA'ya ait Java kütüphaneleri kullanılmıştır. Geliştirilen kod EK-1'de verilmiştir. Sonuçların değerlendirilmesinde doğru sınıflandırma yüzdesi, hata matrisi, doğruluk, kesinlik, duyarlılık ve F-Ölçütü bilgileri dikkate alınmıştır.

İnternet sayfalarını sınıflandırma metin, resim, doküman yapısı gibi birçok parametreye bağlı olarak gerçekleştirilebilmektedir. İnternet sayfalarının sınıflandırması şu şekilde kategorilere ayrılmıştır:

1. Alan adı uzmanlarınca yapılan elle sınıflandırma,
2. Kümeleme yaklaşımları ile yapılan sınıflandırma,
3. Meta etiketleri¹ ile yapılan sınıflandırma,
4. Doküman içeriği ve meta etiketleri ile yapılan sınıflandırma,
5. Sadece doküman içeriğine göre ile yapılan sınıflandırma ve
6. Bağlantı ve içerik analizi ile yapılan sınıflandırma [43].

Bu çalışma 4. madde olan doküman içeriği ve META etiketlerini dikkate alarak yapılmıştır. Çalışmada iki sınıf bulunmaktadır:

- “elektronik ticaret sitesi olan siteler” sınıfı (olumlu - eticaret) ve
- “elektronik ticaret sitesi olmayan siteler” sınıfı (olumsuz - edegill).

Çalışmada yapılan işlemler sırasıyla şunlardır:

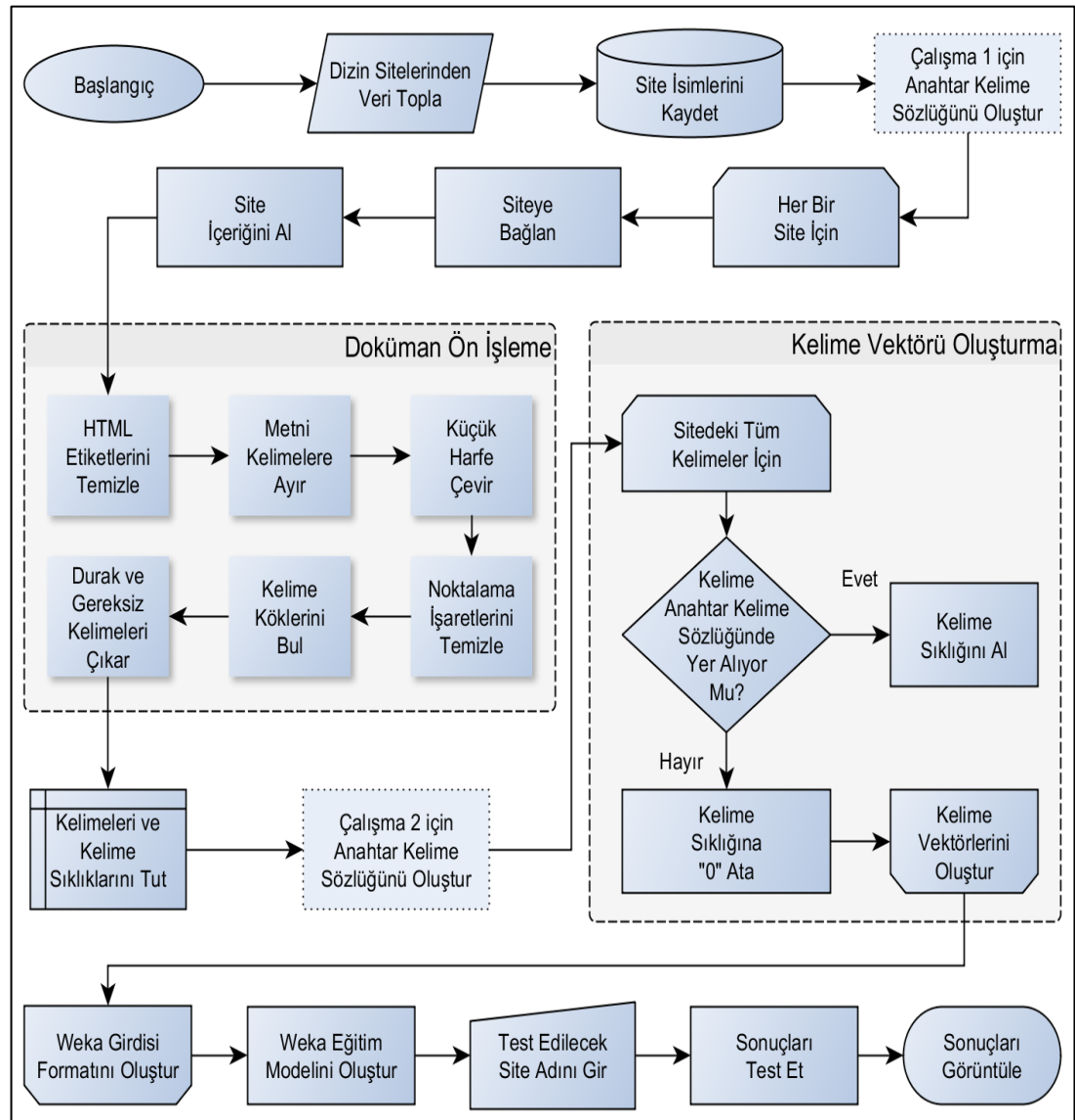
- Dizin sitelerinden² eğitim verisi olarak kullanılmak üzere e-ticaret sitesi olan ve olmayan site isimleri toplanmıştır.
- Toplanan bu sitelere Java programlama dili ile yazılan kod parçacıklarıyla bağlanılıp site içerikleri elde edilmiştir.
- Veri ön işleme aşamaları gerçekleştirilerek veriler etiketlenmiştir (sınıf belirleme).
- Anahtar kelimelerin oluşturulması aşamasında biri statik diğeri dinamik olmak üzere iki yöntem kullanılmıştır. Statik çalışma ile 1, dinamik çalışma ile 9 farklı veri kümesi elde edilmiştir.
- Her bir veri kümesi için vektör uzay modelleri oluşturulmuştur.
- Denetimli öğrenme tekniklerinden k-NN ve Naïve Bayes algoritmaları kullanılmıştır.
- Algoritmaların eğitim verisi üzerindeki başarıları karşılaştırılmıştır.
- Algoritmalarından birisi seçilerek eğitim modeli oluşturulmuştur.

¹ Meta etiketleri internet sitesi hakkında sayfa bilgisi, anahtar kelimeler, dil bilgisi, telif hakkı gibi yapısal bilgilerin tutulduğu kısımdır [44].

² Dizin siteleri kullanıcıların aradıkları bilgileri kolayca bulmasını amaçlayan sitelerdir ve siteleri elle veri tabanına eklenir [45].

- Java ile modelin girdi olarak kullanılacağı bir ara yüz hazırlanmıştır.
- Bu ara yüz yardımıyla kullanıcıların istedikleri siteleri test ederek ilgili sitenin e-ticaret sitesi olup olmadığının bildirilmesi sağlanmıştır.

Çalışmada yapılan işlemlere ait uygulama iş akışı Şekil 3.1’de akış şeması şeklinde görselleştirilerek gösterilmiştir.



Şekil 3.1. Uygulama iş akışı

Bu tezde iki çalışma yapılmıştır. Her iki çalışmada da metin sınıflandırma aşamaları olan veri elde etme, doküman ön işleme, anahtar kelimelerin oluşturulması, vektör

uzay modelinin oluşturulması, vektör ağırlıklandırma, dokümanlar arasındaki benzerliklerin bulunması ve sonuçların değerlendirilmesi işlemleri yapılmıştır. İki çalışma arasındaki fark anahtar kelime sözlüğünün oluşturulması aşamasında farklı yöntemlerin izlenmiş olmasıdır. Anahtar kelime sözlüğü statik çalışmada statik olarak seçilen 30 kelime ile oluşturulurken dinamik çalışma e-ticaret sitelerinin belirlenen sayıdaki yüzdesinde ortak olarak geçen kelimeler ele alınarak oluşturulmuştur. Ayrıca dinamik çalışmada anahtar kelimelerin oluşturulması aşamasında farklı yüzde değerleri kullanılmıştır. Ardından bu çalışmaların sonuçları karşılaştırılmıştır.

3.1. Veri Elde Etme

Bu çalışmada eğitim verisi olarak kullanılmak üzere 13 dizin sitesinin “alışveriş” kategorisinde yer alan 1893 site bağlantısı toplanmıştır. Bu dizin siteleri Çizelge 3.1’de listelenmiştir.

Çizelge 3.1. Kullanılan dizin siteleri listesi

www.alisverisrehberi.com	www.dmoz.org	www.sitetanitimi.com
www.benimsayfam.org	www.linkcenneti.net	www.trdizin.org
www.bilgirehberi.com	www.mavinokta.com	www.turkarama.com
www.buneki.org	www.netbul.gen.tr	
www.dizin.sitesi.web.tr	www.sektoreldizin.com	

Eğitim seti verisini elde etmede Java programla dili kullanılmıştır. Dizin sitelerinde yer alan sitelerin bağlantılarını toplamak etmek için Java programlama dili ile kod yazılmıştır. İlgili kod parçacığı Şekil 3.2’de verilmiştir.

```

public static List<String> extractLinks(String url) throws IOException {
    final ArrayList<String> result = new ArrayList<String>();
    Document doc = Jsoup.connect(url).get(); //Siteye bağlan
    //Site bilgisini içeren HTML etiketlerini seç
    Elements dizinsitesindenlinks = doc.select("span.urlimg");
    for (Element link : dizinsitesindenlinks) { //Her etiketin içindeki url'leri al
        result.add(link.text().replace("www.", "")); //başındaki "www."yi sil
    }
    return result;
}

```

Şekil 3.2. Dizin sitelerinden bağlantı toplama Java kodu

Her dizin sitesinin HTML kod yazımı aynı olmadığı için site bağlantılarını elde ederken farklı dizin siteleri için farklı kodlar yazıldı. Bu farklılık EK-2'deki kodun “dizinsitesindenlinks” adlı elemanında gerçekleştirilmiştir. Bu eleman üzerinde yapılan değişiklikler EK-3'te verilmiştir.

Dizin sitelerinden toplanan 1893 bağlantı içerisinde 339 adet tekrarlı kayıtn varlığı tespit edilmiştir. Tekrarlı kayıtların tekilleştirilmesi ile birlikte bağlantı sayısı 1554'e düşmüştür. Bu sitelerin hala kullanımda olup olmadığı yazılan Java kodu ile kontrol edilmiştir. İlgili kod Şekil 3.30'te verilmiştir. Sitelerin güncelliğinin sorgulanmasından sonra 93 siteye sayfanın kullanımda olmaması nedeniyle bağlantı sağlanamadığı görülmüştür.

```

BufferedReader siteleriDosyadanOku = new BufferedReader(new
FileReader("C:/siteler.txt"));
String satir = "";
String text = "";
//tüm sitelerin tutulduğu dosyada satır olduğu sürece yap:
while ((satir = siteleriDosyadanOku.readLine()) != null) {
    String baglantiadi = satir;
    URL url = new URL("http://www." + baglantiadi);
    URLConnection urlcon = url.openConnection();
    urlcon.connect(); //siteye bağlan
}

```

Şekil 3.3. Bağlantıları kontrol etme Java kodu

Kalan 1461 site incelenmiş ve dizin sitesine e-ticaret sitesi olarak eklenmiş olan bazı sitelerin gerçekte sipariş verme, ödeme yapma gibi e-ticaret sitesi özelliklerini [46] taşımadığı, bazılarının ise sadece reklam amaçlı olduğu gözlemlenmiştir. Toplanan sitelerden büyük bir kısmının gerçekten de kaydedildikleri gibi e-ticaret kategorisine ait olmadıkları gözlemlenmiştir.

Doğru veriler ile yola çıkmanın veri madenciliğinin en önemli aşamalarından biri olduğu [47] göz önünde bulundurularak 1461 adet site tek tek incelenmiştir. Her bir siteye tarayıcı aracılığıyla bağlanılmış ve o sitenin gerçekten e-ticaret sitesi olup olmadığına bakılmıştır. 686 tanesinin gerçekten e-ticaret sitesi olduğu, 775 tanesinin ise e-ticaret sitesi olmadığı görülmüştür. E-ticaret sitesi olan ve olmayan sitelerin adları sırasıyla EK-4 ve EK-5’te verilmiştir. Bu veriler doğrultusunda 686 siteden “e-ticaret sitesi” sınıfı, 775 siteden “e-ticaret sitesi değil” sınıfı oluşturulmuştur. Bu site isimleri daha sonra kullanılmak üzere bir metin dosyasına kaydedilmiştir.

Toplanan 1461 adet sitenin ana sayfa içerikleri veri olarak kullanılmıştır. JSoup¹ adlı Java kütüphanesini kullanarak yazılan Java kodu ile metin dosyasındaki site adları okunarak siteye bağlanılarak ana sayfa ve META bilgileri alınmıştır. İlgili kod parçacığı Şekil 3.4’te verilmiştir.

```
BufferedReader siteleridosyadanoku = new BufferedReader(new FileReader("C:/siteler.txt"));
String satir = "";
String text = "";
int tm = 20000; /* bağlantı zaman aşımı süresi */
//tüm sitelerin tutulduğu dosyada satır olduğu sürece yap:
while ((satir = siteleridosyadanoku.readLine()) != null) {
    String baglantiadi = satir;
    System.out.print("\n" + baglantiadi + " adresi inceleniyor...\n");
    URL url = new URL("http://www." + baglantiadi);
    try {
        URLConnection urlcon = url.openConnection();
        urlcon.connect(); // siteye bağlan
        //bağlantı zaman aşımı süresi boyunca bağlanmayı dene
        Document doc = Jsoup.parse(url, tm);
        text = doc.text();
        System.out.println(baglantiadi + ";" + text);
    } catch (Exception ex) {
        System.err.println("\nURL adresi bulunamadı: " + ex.getMessage());
        ex.printStackTrace();
    }
}
```

Şekil 3.4. Bağlantı ana sayfa ve META bilgilerini alma Java kodu

3.2. Doküman Ön İşleme

Sitelerin (dokümanların) metin şeklinde gösterilebilmesi için Jsoup Java kütüphanesinden ve Zemberek [48] adlı açık kaynak kodlu Türkçe Doğal Dil İşleme Java Kütüphanesinden yararlanılmıştır. Zemberek yazım denetimi, hatalı kelimeler

¹ JSoup internet sitelerinin kaynak kodlarını indirmeyi, bu kodları ayrıştırıp içerisinden istenilen verileri çekmeyi sağlayan bir Java kütüphanesidir [50]

için öneri, heceleme, hatalı kodlama temizleme gibi işlevlere sahiptir [49]. Bu kütüphaneler kullanılarak yazılan Java kodu ile sitelere bağlanıp site içerikleri alınmıştır. Ardından doküman ön işleme aşamaları olan *dönüştürme, tarama ve işaretleme, kök bulma ve durak ve gereksiz kelimelerin çıkarılması* işlemleri yapılmıştır. Bu bağlamda metin içerisindeki noktalama işaretleri kaldırılmış, metnin tamamı küçük harfe çevrilerek kelimelere ayrılmış ve içerisindeki HTML etiketleri temizlenmiştir. Metin içerisinde çok sık geçen fakat sınıflandırmada bir anlam ifade etmeyen edat, bağlaç ve zamir gibi kelimeler ile Türkiye'ye ait şehir isimleri metinden çıkartılmıştır. Durak ve gereksiz kelimelerin listesi EK-6'da verilmiştir. Ardından kelime kökleri bulunmuştur.

3.3. Anahtar Kelime Sözlüğünün Oluşturulması

Anahtar kelime sözlüğünün oluşturulmasında biri *statik* diğeri *dinamik* olmak üzere iki çalışma yapılmıştır. Bu çalışmalar Bölüm 3.3.1., 3.3.2. ve 3.3.3.'te anlatılmıştır.

3.3.1. Statik çalışma

Bu çalışmada e-ticaret sitelerinde sıkça geçtiğini gözlemlenen 30 adet anahtar kelime ele alınmıştır. Bu kelimeler bulunurken tüm sitelerdeki kelimelerin toplamda kaçar kez geçtiği bulunmuştur. Ardından e-ticaret ile ilgili olduğu düşünülenler seçilmiştir. Anahtar kelime sözlüğünü oluşturan bu kelimeler Çizelge 3.2'de verilmiştir.

Çizelge 3.2. Statik çalışma anahtar kelime sözlüğü

alışveriş	banka	ekle	fırsat	fiyat
gönder	haberdar	hesap	hizmet	iade
incele	kampanya	kargo	katalog	kdv
marka	müşteri	öde	promosyon	sat
sepet	sipariş	sözleşme	teslim	ticaret
tl	toptan	ucuz	ücret	ürün

3.3.2. Dinamik çalışma

Yapılan incelemeler sonucunda statik çalışmada seçilmesi gerektiği halde seçilmemiş olan, gözden kaçan ya da seçilmemesi gerektiği halde seçilen kelimelerin olabileceği görüldü. Bu nedenle dinamik çalışmanın yapılmasına karar verildi ve anahtar kelimelerin geliştirilen program tarafından kullanılan eğitim verisinde geçen kelimelere bağlı olarak otomatik olarak seçilmesine karar verildi. Bu çalışmada eğitim kümesinde ele alınan e-ticaret sitelerinin tamamında geçen kelimeler ele alınarak anahtar kelime sözlüğü oluşturuldu. Ancak bu şekilde kelime boyutunun çoğaldığı gözlemlenmiştir ve bir sitede çok sayıda geçen fakat farklı bir sitede hiç geçmeyen ya da yazım hatası yapılarak farklı şekilde algılanan kelimeler de sözlüğe alınmıştır.

Anahtar kelime sözlüğünü oluşturmak amacıyla elde edilen kelime kökleri bir araya getirildi. Ardından tüm e-ticaret sitelerinin en az belirlenen yüzdesinde ortak olarak geçen kelimeler dikkate alınarak vektör oluşturma aşamasında kullanılmak üzere anahtar kelime sözlüğü oluşturuldu. Bu şekilde yazım yanlışlarının, Türkçe harici dillerdeki kelimelerin ve özel isimlerin sözlüğe alınmasının önüne geçildi. Ayrıca vektör uzay modelini oluşturan kelime boyutu sayısı da azaltılmış oldu. Örneğin eğitim verisi olarak kullanılan 1461 siteden e-ticaret sitelerinde geçen tekil kelime sayısı 8605 iken, en az $147(>1461*0,10)$ e-ticaret sitesinde birden geçen kelime sayısı 144 olarak bulunmuştur. Anahtar kelime sözlüğü bu kelimeler ile oluşturulmuştur. Böylece anahtar kelimeler statik çalışmadaki gibi önceden belirlenmemiş, eğitim verisindeki e-ticaret sitelerindeki ortak kelimeler ele alınmıştır.

3.3.3. Dinamik çalışma için farklı yüzde değerlerinin kullanılması

Dinamik çalışma için e-ticaret sitelerinin en az belirlenen yüzdesinde ortak olarak geçen kelimeler dikkate alınarak anahtar kelime sözlüğü oluşturuldu. Bu yüzde oranının değişmesinin modelin başarısını nasıl etkilediğini görmek amacıyla %0, %5, %10, %15, %20, %25, %30, %35, %40, %45, %50, %55, %60, %65, %70, %75,

%80, %85, %90 ve %95 değerleri çalışmada kullanılmıştır. Örneğin e-ticaret sitelerinin en az %5’inde ortak olarak geçen kelimeler ele alınarak toplanan tüm siteler tarandı ve anahtar kelime sözlüğü oluşturuldu. Ardından elde edilen eğitim kümesi için bir sonraki bölümde anlatılacağı üzere vektör uzay modeli oluşturuldu ve eğitim verisi üzerinde metin madenciliği algoritmaları uygulandı. Bu işlem diğer yüzde değerleri için de tekrarlandı.

Eğitim verisi olarak 1461 site kullanılmasına rağmen bağlantı sırasında oluşan bazı problemlerden dolayı her çalıştırmada tam olarak 1461 siteye erişim sağlanamamıştır. Bağlantı sağlanarak eğitimde hesaba katılan site sayısına bağlı olarak tüm sitelerde geçen toplam kelime sayısı, ortak kelimeleri oluştururken hesaba katılacak en az site sayısı ve anahtar kelime sayısı da değişiklik göstermiştir. Bu değişiklikler Çizelge 3.3 ve Çizelge 3.4’te verilmiştir.

Çizelge 3.3. Yüzde değerine bağlı değişim (%0-%45)

Sayı \ Yüzde Değeri	0	5	10	15	20	25	30	35	40	45
Bakılacak Site Sayısı	1461	1461	1461	1461	1461	1461	1461	1461	1461	1461
Bakılan Site Sayısı	1399	1435	1433	1431	1426	1433	1424	1430	1426	1429
Toplam Tekil Kelime Sayısı	8605	8623	8622	8617	8620	8621	8608	8607	8619	8595
Ortak Kelimeleri Oluştururken Hesaba Katılacak En Az Site Sayısı	0	76	147	220	293	366	439	512	585	658
Anahtar Kelime Sayısı	8605	349	144	65	39	19	10	5	1	0

Anahtar kelime sözlüğünü oluştururken e-ticaret sitelerinde ortak olarak geçen kelimeleri alınmasaydı %0 oranı ile elde edilen verideki 8605 kelime ile uğraşılması gerekecekti. Çizelge 3.4 ve Çizelge 3.5’te görüldüğü üzere sitelerin %5’inde geçen ortak kelime sayısı 349 iken, %10’unda geçen ortak kelime sayısı 144, %15’inde geçen ortak kelime sayısı 65, %20’sinde geçen ortak kelime sayısı 39, %25’inde geçen ortak kelime sayısı 19, %30’unda geçen ortak kelime sayısı 10, %35’inde geçen ortak kelime sayısı 5 ve %40’ında geçen ortak kelime sayısı 1’dir.

Çizelge 3.4. Yüzde değerine bağlı değişim (%50-%95)

Sayı \ Yüzde Değeri	50	55	60	65	70	75	80	85	90	95
Bakılacak Site Sayısı	1461	1461	1461	1461	1461	1461	1461	1461	1461	1461
Bakılan Site Sayısı	1427	1432	1430	1431	1429	1436	1426	1430	1434	1434
Toplam Tekil Kelime Sayısı	8614	8634	8645	8611	8649	8629	8628	8627	8612	8618
Ortak Kelimeleri Oluştururken Hesaba Katılacak En Az Site Sayısı	731	804	877	950	1023	1096	1169	1242	1315	1388
Anahtar Kelime Sayısı	0	0	0	0	0	0	0	0	0	0

Daha çok sitede ortak kelime aradıkça anahtar kelime sözlüğünü oluşturacak olan ortak kelime sayısı düşmüştür. Sitelerin %45 ve daha fazlasında ise ortak olarak geçen kelimeye ulaşamamıştır. Bu nedenle vektör uzay modeli ve algoritmaların uygulanması aşamaları 9 yüzde değeri için yapılmıştır: %0, %5, %10, %15, %20, %25, %30, %35 ve %40.

Söz konusu oranlar ile elde edilen veri kümeleri için anahtar kelimeler şunlardır:

- %0 oranı ile elde edilen veri kümesi için anahtar kelimeler: Bkz. EK-7,
- %5 oranı ile elde edilen veri kümesi için anahtar kelimeler: Bkz. EK-8,
- %10 oranı ile elde edilen veri kümesi için anahtar kelimeler: “*ad, adet, adres, aksesuar, al, alt, alışveriş, an, ana, anne, ara, ay, aç, bak, bakı, bakım, bank, banka, bebek, beyaz, bil, bilgi, bilgisayar, bul, bulun, cam, dahil, destek, değ, ek, ekle, el, elektronik, erkek, et, ev, fiyat, form, fırsat, garanti, gel, gir, giriş, giz, gizli, grup, gönder, gör, gün, günü, güven, haber, hak, hazırla, hediye, hesap, hizmet, hız, iade, il, ilet, in, indir, iste, iş, kampanya, kargo, kart, kategori, kayıt, kdv, kişi, kullan, kur, kurum, kutu, kırmızı, liste, malzeme, mark, marka, markala, mağaza, model, müşteri, ol, renk, saat, sak, saklı, sat, satış, sayfa, sepet, set, seç, sipariş, sistem, site, sor, soru, su, söz, sözleş, sık, tak, takip, takı, takım, tek, tel, telefon, teslimat, tl, türkiye, ula, ulaş, unut, ver, yap, yar, yardım, yaz, yazı, ye, yeni, yer, yıl, çanta, çiçek, çocuk, öde, ödem, özel, ürün, üye, şart, şifre*”,

- %15 oranı ile elde edilen veri kümesi için anahtar kelimeler: “aksesuar, al, alışveriş, an, ana, ara, bak, bebek, beyaz, bilgi, el, et, ev, fiyat, garanti, gel, gir, giriş, giz, gizli, gör, gün, güven, hak, hediye, hesap, hizmet, iade, il, ilet, in, indir, kampanya, kargo, kategori, kullan, liste, marka, müşteri, ol, sak, sat, satış, sayfa, sepet, set, seç, sipariş, sistem, site, söz, sözleş, takip, teslimat, tl, unut, yap, yaz, yeni, öde, ödem, özel, ürün, üye, şifre”,
- %20 oranı ile elde edilen veri kümesi için anahtar kelimeler: “al, alışveriş, an, ana, ara, bilgi, et, fiyat, gir, giriş, giz, gizli, gün, güven, hak, hizmet, il, ilet, in, indir, kategori, kullan, müşteri, ol, sat, satış, sayfa, sepet, sipariş, site, teslimat, tl, yap, yeni, öde, ödem, özel, ürün, üye”,
- %25 oranı ile elde edilen veri kümesi için anahtar kelimeler: “alışveriş, ara, gir, giriş, giz, gizli, güven, hak, ilet, in, ol, sat, sepet, sipariş, site, tl, yeni, ürün, üye”
- %30 oranı ile elde edilen veri kümesi için anahtar kelimeler: “gir, hak, ilet, ol, sat, sepet, tl, yeni, ürün, üye”,
- %35 oranı ile elde edilen veri kümesi için anahtar kelimeler: “ilet, ol, sepet, tl, ürün”,
- %40 oranı ile elde edilen veri kümesi için anahtar kelime: “ürün”,
- %45, %50, %55, %60, %65, %70, %75, %80, %85, %90 ve %95 değerleri için anahtar kelime bulunamamıştır. Yani e-ticaret sitelerinin %45 ve yukarısında, bu sitelerin tamamında birden geçen ortak kelime bulunamamıştır.

3.4. Vektör Uzay Modelinin Oluşturulması

Statik ve dinamik çalışmada kullanılan eğitim seti için farklı vektör uzay modelleri oluşturulmuştur. Vektör uzay modeli statik çalışma için 30 boyut ile oluşturulmuştur. Dinamik çalışma için ise sırasıyla 8605 (%0), 349 (%5), 144 (%10), 65 (%15), 39 (%20), 19 (%25), 10 (%30), 5 (%35) ve 1 (%40) boyut ile oluşturulmuştur.

İkili ağırlıklandırma yöntemi

Bu yöntem e-ticaret sitelerinin sırasıyla en az %0, %5, %10, %15, %20, %25, %30, %35 ve %40’ında ortak olarak geçen kelimelerin anahtar sözlüğü oluşturmada kullanılması ile elde edilen eğitim kümesi için vektör uzay modelleri oluşturulmuştur.

%0 oranı ile elde edilen veri kümesi

E-ticaret sitelerinin en az %0’ında ortak olarak geçen kelimelerin yani tüm kelimelerin anahtar sözlüğü oluşturmada kullanılması ile elde edilen eğitim kümesi için ikili ağırlıklandırma yöntemi kullanılarak oluşturulan vektör uzay modelinin bir kesiti Şekil 3.8’de gösterilmiştir.

Viewer

Relation: eticaretgetimverisiCalismas2025-weka.filters.unsupervised.attribute.StringToWordVector-R1-W1000-prune-ratio-1.0-10-stemmerweka.com.stemmers.NullStemmer-M1-tokenizerweka.com.tokenizers.WordTokenizer-delimiters "\\\n,\\\\.\\

Şekil 3.8. Dinamik çalışma ikili doküman-kelime çizelgesi (%0)

%5 oranı ile elde edilen veri kümesi

E-ticaret sitelerinin en az %5’inde ortak olarak geçen kelimelerin anahtar sözlüğü oluşturmada kullanılması ile elde edilen eğitim kümesi için ikili ağırlıklandırma yöntemi kullanılarak oluşturulan vektör uzay modelinin bir kesiti Şekil 3.9’da gösterilmiştir.

Viewer

Relation: eticaretigitimverisiCalsma22005-weka.filters.unsupervised.attribute.StringToWordVector-R1-W1000-prune-rate-1.0-N0-stemmerweka.core.stemmers.NullStemmer-M1-tokenizerweka.core.tokenizers.WordTokenizer-delimiters "\r\n\t,;:\'\"

No.	1: class	2: ad	3: ada	4: adet	5: adres	6: alış	7: aksesuar	8: al	9: ala	10: alan	11: alet	12: alt	13: altın	14: alin	15: alışveriş	16: amaç	17: an	18: ana	19: anahtar	20: anne	21: ar	22: ara	23: araba	24: aranjman	25: araç
1	edegill	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0
2	edegill	0.0	0.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0	1.0	0.0	0.0	0.0	0.0	1.0	1.0	0.0	0.0	0.0	1.0	0.0	0.0
3	eticaret	0.0	0.0	0.0	0.0	0.0	1.0	1.0	1.0	1.0	1.0	0.0	0.0	1.0	1.0	0.0	0.0	0.0	0.0	1.0	1.0	1.0	0.0	0.0	1.0
4	eticaret	0.0	0.0	0.0	1.0	0.0	0.0	1.0	1.0	1.0	1.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0	1.0	1.0	0.0	0.0	1.0	0.0	0.0
5	eticaret	0.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0
6	edegill	0.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
7	eticaret	0.0	0.0	0.0	1.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0	1.0	0.0	0.0
8	eticaret	0.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0	0.0	1.0	1.0	0.0	1.0	0.0	0.0	0.0	1.0	1.0	0.0	1.0	0.0	0.0	0.0
9	edegill	0.0	0.0	0.0	0.0	0.0	0.0	1.0	1.0	1.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0	1.0	0.0	0.0	1.0	0.0	0.0	0.0	1.0
10	edegill	1.0	0.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0	1.0	0.0	0.0	1.0	0.0	0.0	0.0

«

!!!

»

Undo

OK

Cancel

Şekil 3.9. Dinamik çalışma ikili doküman-kelime çizelgesi (%5)

%10 oranı ile elde edilen veri kümesi

E-ticaret sitelerinin en az %10’inde ortak olarak geçen kelimelerin anahtar sözlüğü oluşturmada kullanılması ile elde edilen eğitim kümesi için ikili ağırlıklandırma yöntemi kullanılarak oluşturulan vektör uzay modelinin bir kesiti Şekil 3.10’da gösterilmiştir.

No.	1: class	2: adet	3: adres	4: aksesuar	5: al	6: alt	7: alışveriş	8: an	9: ana	10: anne	11: ara	12: ay	13: aç	14: bak	15: baki	16: bakım	17: banka	18: bebek	19: beyaz	20: bil	21: bilgi	22: bul	23: bulun	24: cam	25: dahil
1	edegill	0.0	0.0	0.0	0.0	0.0	0.0	1.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0	0.0
2	edegill	0.0	1.0	0.0	0.0	1.0	0.0	1.0	1.0	0.0	1.0	1.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
3	eticaret	0.0	0.0	1.0	1.0	0.0	1.0	0.0	0.0	1.0	1.0	0.0	1.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0	1.0	0.0	0.0	0.0	1.0
4	eticaret	0.0	1.0	0.0	1.0	0.0	1.0	1.0	1.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
5	eticaret	0.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
6	edegill	0.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
7	eticaret	0.0	0.0	1.0	0.0	1.0	1.0	0.0	0.0	1.0	1.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0	1.0	0.0	1.0
8	edegill	0.0	0.0	0.0	1.0	0.0	1.0	1.0	0.0	0.0	0.0	0.0	1.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0	1.0	1.0	0.0	0.0
9	edegill	0.0	0.0	0.0	1.0	0.0	0.0	1.0	1.0	0.0	1.0	1.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
10	edegill	0.0	0.0	1.0	1.0	0.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0

Şekil 3.10. Dinamik çalışma ikili doküman-kelime çizelgesi (%10)

%15 oranı ile elde edilen veri kümesi

E-ticaret sitelerinin en az %15’inde ortak olarak geçen kelimelerin anahtar sözlüğü oluşturmada kullanılması ile elde edilen eğitim kümesi için ikili ağırlıklandırma yöntemi kullanılarak oluşturulan vektör uzay modelinin bir kesiti Şekil 3.11’de gösterilmiştir.

Relation: eticaretigitimverisiCalsma22015-weka.filters.unsupervised.attribute.StringToWordVector-R1-W1000-prune-rate-1.0-N0-stemmerveka.core.stemmers.NullStemmer-M1-tokenizerveka.core.tokenizers.WordTokenizer-delimiters "\r\n\t,;:\\"/>

No.	1: class	2: al	3: al	4: algi	5: an	6: ana	7: ara	8: bak	9: bebek	10: beyaz	11: bilgi	12: el	13: et	14: ev	15: fiyat	16: garanti	17: gel	18: gir	19: giriş	20: giz	21: gizli	22: gör	23: gün	24: güven	25: hak
1	edegill	0.0	0.0	0.0	1.0	1.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0
2	edegill	0.0	0.0	0.0	1.0	1.0	1.0	0.0	0.0	0.0	0.0	0.0	1.0	1.0	0.0	0.0	1.0	0.0	0.0	0.0	0.0	1.0	1.0	0.0	1.0
3	eticaret	1.0	1.0	1.0	0.0	0.0	1.0	0.0	1.0	0.0	1.0	1.0	1.0	1.0	0.0	0.0	1.0	1.0	1.0	1.0	1.0	0.0	1.0	0.0	1.0
4	eticaret	0.0	1.0	1.0	1.0	1.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0	1.0	1.0	1.0	1.0	1.0	0.0	0.0	1.0
5	eticaret	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0	1.0	0.0	1.0	1.0	0.0	0.0	0.0	1.0	0.0	1.0
6	edegill	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0	1.0
7	eticaret	1.0	0.0	0.0	0.0	0.0	1.0	1.0	0.0	0.0	1.0	0.0	1.0	1.0	0.0	0.0	0.0	1.0	1.0	0.0	0.0	0.0	0.0	1.0	1.0
8	eticaret	1.0	0.0	1.0	0.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0	1.0	0.0	1.0	1.0	1.0	0.0	0.0	0.0	1.0	1.0	1.0
9	edegill	0.0	1.0	1.0	1.0	0.0	0.0	1.0	0.0	0.0	1.0	0.0	1.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0
10	edegill	0.0	1.0	0.0	1.0	1.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0	0.0	1.0	1.0

Şekil 3.11. Dinamik çalışma ikili doküman-kelime çizelgesi (%15)

%20 oranı ile elde edilen veri kümesi

E-ticaret sitelerinin en az %20'sinde ortak olarak geçen kelimelerin anahtar sözlüğü oluşturmada kullanılması ile elde edilen eğitim kümesi için ikili ağırlıklandırma yöntemi kullanılarak oluşturulan vektör uzay modelinin bir kesiti Şekil 3.12'de gösterilmiştir.

Relation: eticaretigitimverisiCalsma22020-weka.filters.unsupervised.attribute.StringToWordVector-R1-W1000-prune-rate-1.0-N0-stemmerveka.core.stemmers.NullStemmer-M1-tokenizerveka.core.tokenizers.WordTokenizer-delimiters "\r\n\t,;:\\"/>

No.	1: class	2: al	3: algi	4: an	5: ana	6: ara	7: bilgi	8: et	9: fiyat	10: gir	11: giriş	12: giz	13: gizli	14: gün	15: güven	16: hak	17: hizmet	18: il	19: ile	20: in	21: indir	22: kategori	23: kullan	24: müşteri	25: ol
1	edegill	0.0	0.0	1.0	1.0	1.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0	1.0	0.0	1.0	0.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0	1.0
2	eticaret	1.0	1.0	0.0	0.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	0.0	1.0	1.0	0.0	1.0	0.0	0.0	0.0	1.0	0.0	0.0	1.0
3	eticaret	1.0	1.0	1.0	1.0	1.0	0.0	0.0	0.0	1.0	1.0	1.0	1.0	0.0	0.0	1.0	1.0	0.0	1.0	0.0	0.0	0.0	1.0	1.0	1.0
4	eticaret	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0	1.0	1.0	1.0	0.0	1.0	0.0	1.0	0.0	0.0	0.0	1.0	1.0	0.0	0.0	0.0	1.0
5	edegill	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0	1.0	0.0	1.0	0.0	0.0	0.0	0.0	1.0	0.0
6	eticaret	0.0	0.0	0.0	1.0	1.0	1.0	0.0	1.0	1.0	0.0	0.0	0.0	0.0	1.0	1.0	0.0	0.0	1.0	1.0	0.0	0.0	1.0	0.0	1.0
7	eticaret	0.0	1.0	0.0	0.0	1.0	0.0	0.0	1.0	1.0	1.0	0.0	0.0	1.0	1.0	1.0	0.0	0.0	1.0	1.0	0.0	1.0	1.0	0.0	1.0
8	edegill	1.0	1.0	0.0	0.0	1.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0	1.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0	1.0
9	edegill	1.0	0.0	1.0	1.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0	0.0	1.0	0.0	1.0	1.0	0.0	0.0	1.0	1.0	0.0	1.0
10	edegill	1.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0	1.0	0.0	1.0

Şekil 3.12. Dinamik çalışma ikili doküman-kelime çizelgesi (%20)

%25 oranı ile elde edilen veri kümesi

E-ticaret sitelerinin en az %25'inde ortak olarak geçen kelimelerin anahtar sözlüğü oluşturmada kullanılması ile elde edilen eğitim kümesi için ikili ağırlıklandırma yöntemi kullanılarak oluşturulan vektör uzay modelinin bir kesiti Şekil 3.13'te gösterilmiştir.

No.	1: class	2: alışveriş	3: ara	4: gir	5: giriş	6: giz	7: gizli	8: güven	9: hak	10: ilet	11: in	12: ol	13: sat	14: sepet	15: sipariş	16: site	17: tl	18: yeni	19: ürün	20: üye
	Nominal	Numeric	Numeric	Numeric	Numeric	Numeric	Numeric	Numeric	Numeric	Numeric	Numeric	Numeric	Numeric	Numeric	Numeric	Numeric	Numeric	Numeric	Numeric	Numeric
1	edegill	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0	1.0	0.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0	1.0	0.0
2	edegill	0.0	1.0	0.0	0.0	0.0	0.0	0.0	1.0	1.0	0.0	1.0	0.0	0.0	1.0	1.0	0.0	1.0	1.0	0.0
3	eticaret	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	0.0	0.0	1.0	1.0	0.0	0.0	1.0	1.0	1.0	1.0	1.0
4	eticaret	1.0	1.0	1.0	1.0	1.0	1.0	0.0	1.0	1.0	0.0	1.0	0.0	1.0	1.0	1.0	1.0	0.0	0.0	0.0
5	eticaret	0.0	0.0	1.0	1.0	0.0	0.0	0.0	1.0	0.0	1.0	1.0	1.0	1.0	0.0	0.0	1.0	1.0	1.0	1.0
6	edegill	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0	0.0
7	eticaret	0.0	1.0	1.0	1.0	0.0	0.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	0.0
8	eticaret	1.0	1.0	1.0	1.0	0.0	0.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	0.0	1.0	0.0	1.0	1.0
9	edegill	1.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0	1.0	0.0	1.0	0.0	0.0	0.0	1.0	0.0	1.0	0.0	0.0
10	edegill	0.0	1.0	0.0	0.0	0.0	0.0	0.0	1.0	1.0	0.0	1.0	1.0	0.0	0.0	0.0	1.0	0.0	0.0	1.0

Şekil 3.13. Dinamik çalışma ikili doküman-kelime çizelgesi (%25)

%30 oranı ile elde edilen veri kümesi

E-ticaret sitelerinin en az %30’unda ortak olarak geçen kelimelerin anahtar sözlüğü oluşturmada kullanılması ile elde edilen eğitim kümesi için ikili ağırlıklandırma yöntemi kullanılarak oluşturulan vektör uzay modelinin bir kesiti Şekil 3.14’te gösterilmiştir.

No.	1: class	2: gir	3: hak	4: ilet	5: ol	6: sat	7: sepet	8: tl	9: yeni	10: ürün	11: ürün	12: üye
	Nominal	Numeric	Numeric	Numeric	Numeric	Numeric	Numeric	Numeric	Numeric	Numeric	Numeric	Numeric
1	edegill	0.0	1.0	1.0	1.0	0.0	0.0	0.0	1.0	1.0	0.0	0.0
2	eticaret	1.0	1.0	0.0	1.0	1.0	0.0	1.0	1.0	1.0	1.0	1.0
3	eticaret	1.0	1.0	1.0	1.0	0.0	1.0	1.0	0.0	0.0	0.0	0.0
4	eticaret	1.0	1.0	0.0	1.0	1.0	1.0	1.0	1.0	1.0	0.0	1.0
5	edegill	0.0	1.0	1.0	0.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0
6	eticaret	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	0.0	0.0
7	eticaret	1.0	1.0	1.0	1.0	1.0	1.0	0.0	0.0	1.0	0.0	1.0
8	edegill	0.0	1.0	1.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
9	edegill	0.0	1.0	1.0	1.0	1.0	0.0	1.0	0.0	0.0	0.0	1.0
10	edegill	1.0	0.0	0.0	1.0	0.0	0.0	0.0	1.0	1.0	0.0	1.0

Şekil 3.14. Dinamik çalışma ikili doküman-kelime çizelgesi (%30)

%35 oranı ile elde edilen veri kümesi

E-ticaret sitelerinin en az %35’inde ortak olarak geçen kelimelerin anahtar sözlüğü oluşturmada kullanılması ile elde edilen eğitim kümesi için ikili ağırlıklandırma yöntemi kullanılarak oluşturulan vektör uzay modelinin bir kesiti Şekil 3.15’te gösterilmiştir.

No.	1: class Nominal	2: ilet Numeric	3: ol Numeric	4: sepet Numeric	5: tl Numeric	6: ürün Numeric
1	edegill	1.0	0.0	0.0	0.0	1.0
2	edegill	1.0	1.0	0.0	0.0	1.0
3	eticaret	0.0	1.0	0.0	1.0	1.0
4	eticaret	1.0	1.0	1.0	1.0	0.0
5	eticaret	0.0	1.0	1.0	1.0	1.0
6	edegill	1.0	0.0	0.0	0.0	1.0
7	eticaret	1.0	1.0	1.0	1.0	1.0
8	eticaret	1.0	1.0	1.0	0.0	1.0
9	edegill	1.0	1.0	0.0	0.0	0.0
10	edegill	1.0	1.0	0.0	1.0	0.0

Şekil 3.15. Dinamik çalışma ikili doküman-kelime çizelgesi (%35)

%40 oranı ile elde edilen veri kümesi

E-ticaret sitelerinin en az %40'ında ortak olarak geçen kelimelerin anahtar sözlüğü oluşturmada kullanılması ile elde edilen eğitim kümesi için ikili ağırlıklandırma yöntemi kullanılarak oluşturulan vektör uzay modelinin bir kesiti Şekil 3.16'da gösterilmiştir.

No.	1: class Nominal	2: ürün Numeric
1	edegill	1.0
2	eticaret	1.0
3	eticaret	0.0
4	eticaret	1.0
5	edegill	1.0
6	eticaret	1.0
7	eticaret	1.0
8	edegill	0.0
9	edegill	0.0
10	edegill	1.0

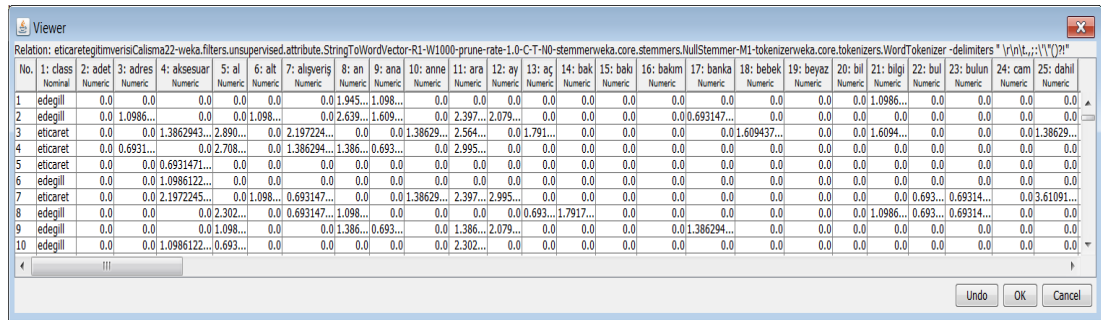
Şekil 3.16. Dinamik çalışma ikili doküman-kelime çizelgesi (%40)

TF ağırlıklandırma yöntemi

Bu yöntem e-ticaret sitelerinin sırasıyla en az %0, %5, %10, %15, %20, %25, %30, %35 ve %40'ında ortak olarak geçen kelimelerin anahtar sözlüğü oluşturmada kullanılması ile elde edilen eğitim kümesi için vektör uzay modelleri oluşturulmuştur.

%10 oranı ile elde edilen veri kümesi

E-ticaret sitelerinin en az %10’inde ortak olarak geçen kelimelerin anahtar sözlüğü oluşturmada kullanılması ile elde edilen eğitim kümesi için TF ağırlıklandırma yöntemi kullanılarak oluşturulan vektör uzay modelinin bir kesiti Şekil 3.19’da gösterilmiştir.

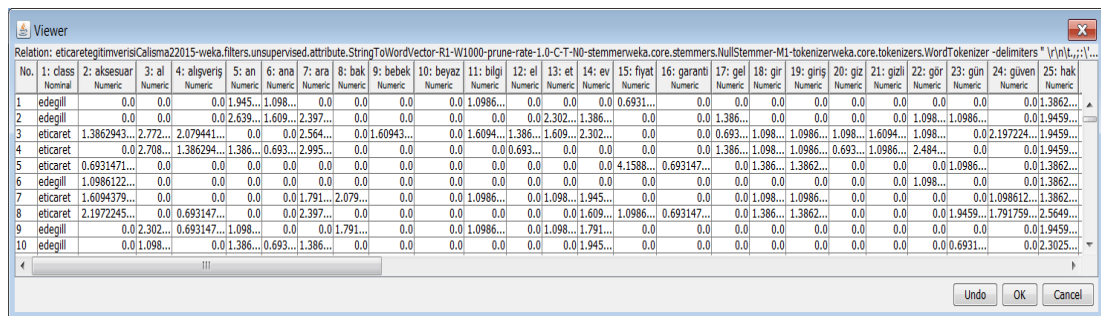


No.	1: class	2: adet	3: adres	4: aksesuar	5: al	6: alt	7: alışveriş	8: an	9: ana	10: anne	11: ara	12: ay	13: aç	14: bak	15: baki	16: bakım	17: banka	18: bebek	19: beyaz	20: bil	21: bilgi	22: bul	23: bulun	24: cam	25: dahil
1	edegill	0.0	0.0	0.0	0.0	0.0	0.0	1.945...	1.098...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0986...	0.0	0.0	0.0	0.0
2	edegill	0.0	1.0986...	0.0	0.0	1.098...	0.0	2.639...	1.609...	0.0	2.397...	2.079...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
3	eticaret	0.0	0.0	1.3862943...	2.890...	0.0	2.197224...	0.0	0.0	1.38629...	2.564...	0.0	1.791...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.6094...	0.0	0.0	0.0	0.0
4	eticaret	0.0	0.6931...	0.0	2.708...	0.0	1.386294...	1.386...	0.693...	0.0	2.995...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
5	eticaret	0.0	0.0	0.6931471...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
6	edegill	0.0	0.0	1.0986122...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
7	eticaret	0.0	0.0	2.1972245...	0.0	1.098...	0.693147...	0.0	0.0	1.38629...	2.397...	2.995...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.693...	0.69314...	0.0	0.0
8	edegill	0.0	0.0	0.0	2.302...	0.0	0.693147...	1.098...	0.0	0.0	0.0	0.0	0.0	0.693...	1.7917...	0.0	0.0	0.0	0.0	0.0	1.0986...	0.693...	0.69314...	0.0	0.0
9	edegill	0.0	0.0	0.0	1.098...	0.0	0.0	1.386...	0.693...	0.0	1.386...	2.079...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
10	edegill	0.0	0.0	1.0986122...	0.693...	0.0	0.0	0.0	0.0	0.0	2.302...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0

Şekil 3.19. Dinamik çalışma TF doküman-kelime çizelgesi (%10)

%15 oranı ile elde edilen veri kümesi

E-ticaret sitelerinin en az %15’inde ortak olarak geçen kelimelerin anahtar sözlüğü oluşturmada kullanılması ile elde edilen eğitim kümesi için TF ağırlıklandırma yöntemi kullanılarak oluşturulan vektör uzay modelinin bir kesiti Şekil 3.20’de gösterilmiştir.



No.	1: class	2: aksesuar	3: al	4: alışveriş	5: an	6: ana	7: ara	8: bak	9: bebek	10: beyaz	11: bilgi	12: el	13: et	14: ev	15: fiyat	16: garanti	17: gel	18: gir	19: giriş	20: giz	21: gizli	22: gör	23: gün	24: güven	25: hak
1	edegill	0.0	0.0	0.0	1.945...	1.098...	0.0	0.0	0.0	0.0	1.0986...	0.0	0.0	0.0	0.6931...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
2	edegill	0.0	0.0	0.0	2.639...	1.609...	2.397...	0.0	0.0	0.0	0.0	2.302...	1.386...	0.0	0.0	1.386...	0.0	1.386...	0.0	0.0	0.0	1.098...	1.0986...	0.0	1.3862...
3	eticaret	1.3862943...	2.772...	2.079441...	0.0	0.0	2.564...	0.0	1.60943...	0.0	1.6094...	1.386...	1.609...	2.302...	0.0	0.0	0.693...	1.098...	1.0986...	1.098...	1.6094...	1.098...	0.0	2.197224...	1.9459...
4	eticaret	0.0	2.708...	1.386294...	1.386...	0.693...	2.995...	0.0	0.0	0.0	0.0	0.693...	0.0	0.0	0.0	0.0	1.386...	1.098...	1.0986...	0.693...	1.0986...	2.484...	0.0	0.0	1.9459...
5	eticaret	0.6931471...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	4.1588...	0.693147...	0.0	1.386...	1.3862...	0.0	0.0	0.0	1.0986...	0.0	1.3862...
6	edegill	1.0986122...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.098...	0.0	0.0	1.3862...
7	eticaret	1.6094379...	0.0	0.0	0.0	0.0	1.791...	2.079...	0.0	0.0	0.0	1.0986...	0.0	1.098...	1.945...	0.0	0.0	0.0	1.098...	1.0986...	0.0	0.0	0.0	0.0	1.098612...
8	eticaret	2.1972245...	0.0	0.693147...	0.0	0.0	2.397...	0.0	0.0	0.0	0.0	0.0	0.0	1.609...	1.0986...	0.693147...	0.0	1.386...	1.3862...	0.0	0.0	0.0	1.9459...	1.791759...	2.5649...
9	edegill	0.0	2.302...	0.693147...	1.098...	0.0	0.0	1.791...	0.0	0.0	1.0986...	0.0	1.098...	1.791...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
10	edegill	0.0	1.098...	0.0	1.386...	0.693...	1.386...	0.0	0.0	0.0	0.0	0.0	0.0	1.945...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.6931...	0.0

Şekil 3.20. Dinamik çalışma TF doküman-kelime çizelgesi (%15)

%20 oranı ile elde edilen veri kümesi

E-ticaret sitelerinin en az %20'sinde ortak olarak geçen kelimelerin anahtar sözlüğü oluşturmada kullanılması ile elde edilen eğitim kümesi için TF ağırlıklandırma yöntemi kullanılarak oluşturulan vektör uzay modelinin bir kesiti Şekil 3.21'de gösterilmiştir.

Relation: eticaretgizliimersCalisma22020-weka.filters.unsupervised.attribute.StringToWordVector-R1-W1000-prune-rate-1.0-C-T-W0-stemmerweka.core.stemmers.NullStemmer-M1-tokenizerweka.core.tokenizers.WordTokenizers-delimiters "\\\n\\\\t\\\\s\\\\.\\\\|\\

Şekil 3.21. Dinamik çalışma TF doküman-kelime çizelgesi (%20)

%25 oranı ile elde edilen veri kümesi

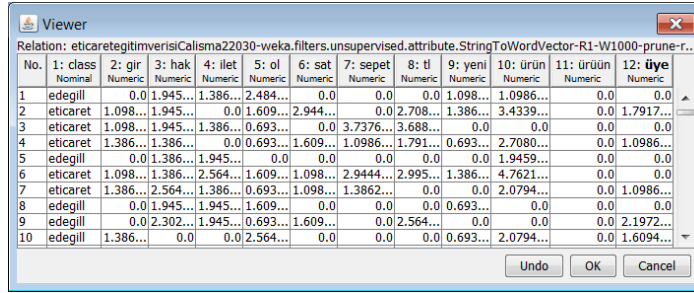
E-ticaret sitelerinin en az %25'inde ortak olarak geçen kelimelerin anahtar sözlüğü oluşturmada kullanılması ile elde edilen eğitim kümesi için TF ağırlıklandırma yöntemi kullanılarak oluşturulan vektör uzay modelinin bir kesiti Şekil 3.22'de gösterilmiştir.

No.	1: class	2: alisveris	3: ara	4: gir	5: giriş	6: giz	7: gizli	8: güven	9: hak	10: ilet	11: in	12: ol	13: sat	14: sepet	15: sipariş	16: site	17: ti	18: yeni	19: ürün	20: üye
1	edegill	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.386...	1.386...	0.0	0.0	1.098...	0.0	0.0	0.0	0.0	0.0	0.0
2	edegill	0.0	2.397...	0.0	0.0	0.0	0.0	0.0	0.0	1.945...	1.386...	0.0	2.484...	0.0	0.0	0.693147...	1.3862...	0.0	1.0986...	1.0986...
3	eticaret	2.079441...	2.564...	1.098...	1.098...	1.098...	1.609...	2.19722...	1.945...	0.0	0.0	1.609...	2.944...	0.0	0.0	1.0986...	2.639...	1.6094...	3.6109...	1.7917...
4	eticaret	1.386294...	2.995...	1.098...	1.098...	0.693...	1.098...	0.0	0.0	1.945...	1.386...	0.0	0.693...	0.0	3.73766...	1.609437...	1.3862...	3.688...	0.0	0.0
5	eticaret	0.0	0.0	1.386...	1.386...	0.0	0.0	0.0	0.0	1.386...	0.0	0.693...	0.693...	1.609...	1.09861...	0.0	0.0	1.791...	0.6931...	2.7080...
6	edegill	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.386...	1.945...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.9459...	0.0
7	eticaret	0.0	1.791...	1.098...	1.098...	0.0	0.0	1.09861...	1.386...	2.564...	1.791...	1.609...	1.098...	2.94443...	0.693147...	1.3862...	2.995...	1.3862...	4.7621...	0.0
8	eticaret	0.693147...	2.397...	1.386...	1.386...	0.0	0.0	1.79175...	2.564...	1.386...	2.564...	0.693...	1.098...	1.38629...	0.0	0.6931...	0.0	0.0	2.0794...	1.0986...
9	edegill	0.693147...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.945...	1.945...	0.0	1.609...	0.0	0.0	0.0	0.0	1.0986...	0.6931...	0.0
10	edegill	0.0	1.386...	0.0	0.0	0.0	0.0	0.0	0.0	2.302...	1.945...	0.0	0.693...	1.609...	0.0	0.0	1.6094...	2.564...	0.0	0.0

Şekil 3.22. Dinamik çalışma TF doküman-kelime çizelgesi (%25)

%30 oranı ile elde edilen veri kümesi

E-ticaret sitelerinin en az %30’unda ortak olarak geçen kelimelerin anahtar sözlüğü oluşturmada kullanılması ile elde edilen eğitim kümesi için TF ağırlıklandırma yöntemi kullanılarak oluşturulan vektör uzay modelinin bir kesiti Şekil 3.23’te gösterilmiştir.

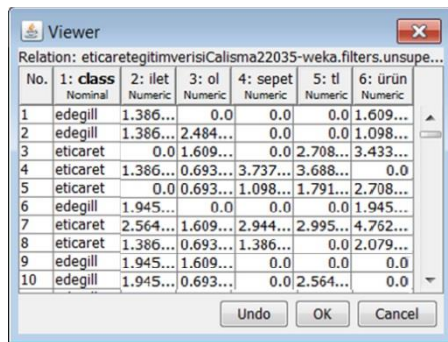


No.	1: class	2: gir	3: hak	4: ilet	5: ol	6: sat	7: sepet	8: tl	9: yeni	10: ürün	11: ürün	12: üye
	Nominal	Numeric	Numeric	Numeric	Numeric	Numeric	Numeric	Numeric	Numeric	Numeric	Numeric	Numeric
1	edegill	0.0	1.945...	1.386...	2.484...	0.0	0.0	0.0	1.098...	1.0986...	0.0	0.0
2	eticaret	1.098...	1.945...	0.0	1.609...	2.944...	0.0	2.708...	1.386...	3.4339...	0.0	1.7917...
3	eticaret	1.098...	1.945...	1.386...	0.693...	0.0	3.7376...	3.688...	0.0	0.0	0.0	0.0
4	eticaret	1.386...	1.386...	0.0	0.693...	1.609...	1.0986...	1.791...	0.693...	2.7080...	0.0	1.0986...
5	edegill	0.0	1.386...	1.945...	0.0	0.0	0.0	0.0	0.0	1.9459...	0.0	0.0
6	eticaret	1.098...	1.386...	2.564...	1.609...	1.098...	2.9444...	2.995...	1.386...	4.7621...	0.0	0.0
7	eticaret	1.386...	2.564...	1.386...	0.693...	1.098...	1.3862...	0.0	0.0	2.0794...	0.0	1.0986...
8	edegill	0.0	1.945...	1.945...	1.609...	0.0	0.0	0.0	0.693...	0.0	0.0	0.0
9	edegill	0.0	2.302...	1.945...	0.693...	1.609...	0.0	2.564...	0.0	0.0	0.0	2.1972...
10	edegill	1.386...	0.0	0.0	2.564...	0.0	0.0	0.0	0.693...	2.0794...	0.0	1.6094...

Şekil 3.23. Dinamik çalışma TF doküman-kelime çizelgesi (%30)

%35 oranı ile elde edilen veri kümesi

E-ticaret sitelerinin en az %35’inde ortak olarak geçen kelimelerin anahtar sözlüğü oluşturmada kullanılması ile elde edilen eğitim kümesi için TF ağırlıklandırma yöntemi kullanılarak oluşturulan vektör uzay modelinin bir kesiti Şekil 3.24’te gösterilmiştir.

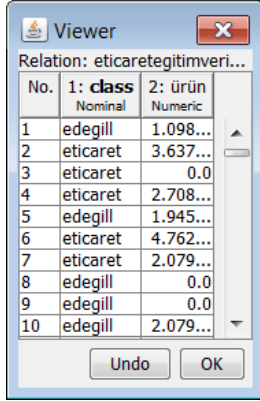


No.	1: class	2: ilet	3: ol	4: sepet	5: tl	6: ürün
	Nominal	Numeric	Numeric	Numeric	Numeric	Numeric
1	edegill	1.386...	0.0	0.0	0.0	1.609...
2	edegill	1.386...	2.484...	0.0	0.0	1.098...
3	eticaret	0.0	1.609...	0.0	2.708...	3.433...
4	eticaret	1.386...	0.693...	3.737...	3.688...	0.0
5	eticaret	0.0	0.693...	1.098...	1.791...	2.708...
6	edegill	1.945...	0.0	0.0	0.0	1.945...
7	eticaret	2.564...	1.609...	2.944...	2.995...	4.762...
8	eticaret	1.386...	0.693...	1.386...	0.0	2.079...
9	edegill	1.945...	1.609...	0.0	0.0	0.0
10	edegill	1.945...	0.693...	0.0	2.564...	0.0

Şekil 3.24. Dinamik çalışma TF doküman-kelime çizelgesi (%35)

%40 oranı ile elde edilen veri kümesi

E-ticaret sitelerinin en az %40'ında ortak olarak geçen kelimelerin anahtar sözlüğü oluşturmada kullanılması ile elde edilen eğitim kümesi için TF ağırlıklandırma yöntemi kullanılarak oluşturulan vektör uzay modelinin bir kesiti Şekil 3.25'te gösterilmiştir.



No.	1: class Nominal	2: ürün Numeric
1	edegill	1.098...
2	eticaret	3.637...
3	eticaret	0.0
4	eticaret	2.708...
5	edegill	1.945...
6	eticaret	4.762...
7	eticaret	2.079...
8	edegill	0.0
9	edegill	0.0
10	edegill	2.079...

Şekil 3.25. Dinamik çalışma TF doküman-kelime çizelgesi (%40)

TF-IDF ağırlıklandırma yöntemi

Bu yöntem e-ticaret sitelerinin sırasıyla en az %0, %5, %10, %15, %20, %25, %30, %35 ve %40'ında ortak olarak geçen kelimelerin anahtar sözlüğü oluşturmada kullanılması ile elde edilen eğitim kümesi için vektör uzay modelleri oluşturulmuştur.

%0 oranı ile elde edilen veri kümesi

E-ticaret sitelerinin en az %0'ında ortak olarak geçen kelimelerin yani tüm kelimelerin anahtar sözlüğü oluşturmada kullanılması ile elde edilen eğitim kümesi için TF-IDF ağırlıklandırma yöntemi kullanılarak oluşturulan vektör uzay modelinin bir kesiti Şekil 3.26'da gösterilmiştir.

Viewer

Relation: eticaretigitimveriCalsma22005-weka.filters.unsupervised.attribute.StringToWordVector-R1-W1000-prune-rate-1.0-C-T-F-N0-stemmerweka.core.stemmers.NullStemmer-M1-tokenizerweka.core.tokenizers.WordTokenizer-delimiters "\r\n\t,;:'\"

No.	1: class	2: abone	3: acor	4: ada	5: adres	6: adaptir	7: adet	8: adres	9: ahmet	10: ahgap	11: ale	12: ak	13: akseusur	14: alvayum	15: alu	16: alu	17: al	18: ala	19: alan	20: alarm	21: alel	22: al	23: alt	24: abn	25: aljars	26: alcu	27: alim	28: aln	29: alg	30: algeris
1	edegill	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
2	edegill	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
3	eticaret	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
4	eticaret	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
5	eticaret	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
6	edegill	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
7	eticaret	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
8	eticaret	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
9	edegill	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
10	edegill	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0

UndoOKCancel

Şekil 3.26. Dinamik çalışma TF-IDF doküman-kelime çizelgesi (%0)

%5 oranı ile elde edilen veri kümesi

E-ticaret sitelerinin en az %5’inde ortak olarak geçen kelimelerin anahtar sözlüğü oluşturmada kullanılması ile elde edilen eğitim kümesi için TF-IDF ağırlıklandırma yöntemi kullanılarak oluşturulan vektör uzay modelinin bir kesiti Şekil 3.27’de gösterilmiştir.

No.	1: class	2: ad	3: ada	4: adet	5: adres	6: ahgap	7: akseusur	8: al	9: ala	10: alan	11: alel	12: alt	13: altAtn	14: altin	15: algeris	16: amac	17: an	18: ana	19: anahtar	20: anne	21: ar	22: ara	23: araba	24: aranjman	25: arac
1	edegill	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
2	edegill	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
3	eticaret	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
4	eticaret	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
5	eticaret	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
6	edegill	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
7	eticaret	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
8	eticaret	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
9	edegill	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
10	edegill	1.872...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0

Şekil 3.27. Dinamik çalışma TF-IDF doküman-kelime çizelgesi (%5)

%10 oranı ile elde edilen veri kümesi

E-ticaret sitelerinin en az %10’inde ortak olarak geçen kelimelerin anahtar sözlüğü oluşturmada kullanılması ile elde edilen eğitim kümesi için TF-IDF ağırlıklandırma yöntemi kullanılarak oluşturulan vektör uzay modelinin bir kesiti Şekil 3.28’de gösterilmiştir.

No.	1: class	2: adet	3: adres	4: aksesuar	5: al	6: alt	7: algiyeris	8: an	9: ana	10: anne	11: ara	12: ay	13: ac	14: bak	15: bako	16: bakim	17: banka	18: bebek	19: beyaz	20: bil	21: bilgi	22: bul	23: bulun	24: cam	25: dahil
1	edegill	0.0	0.0	0.0	0.0	0.0	0.0	1.636...	1.134...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
2	edegill	0.0	1.6504...	0.0	0.0	1.717...	0.0	2.219...	1.662...	0.0	1.903...	3.614...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
3	eticaret	0.0	0.0	1.9580763...	2.316...	0.0	2.535252...	0.0	0.0	2.67553...	2.035...	0.0	2.970...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
4	eticaret	0.0	1.0413...	0.0	2.170...	0.0	1.599566...	1.165...	0.716...	0.0	2.377...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
5	eticaret	0.0	0.0	0.9790381...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
6	edegill	0.0	0.0	1.5517387...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
7	eticaret	0.0	0.0	3.1034775...	0.0	1.717...	0.799783...	0.0	0.0	2.67553...	1.903...	5.206...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
8	edegill	0.0	0.0	0.0	1.845...	0.0	0.799783...	0.923...	0.0	0.0	0.0	0.0	1.149...	2.5881...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
9	edegill	0.0	0.0	0.0	0.880...	0.0	0.0	1.165...	0.716...	0.0	1.100...	3.614...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
10	edegill	0.0	0.0	1.5517387...	0.555...	0.0	0.0	0.0	0.0	0.0	1.827...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0

Şekil 3.28. Dinamik çalışma TF-IDF doküman-kelime çizelgesi (%10)

%15 oranı ile elde edilen veri kümesi

E-ticaret sitelerinin en az %15’inde ortak olarak geçen kelimelerin anahtar sözlüğü oluşturmada kullanılması ile elde edilen eğitim kümesi için TF-IDF ağırlıklandırma yöntemi kullanılarak oluşturulan vektör uzay modelinin bir kesiti Şekil 3.29’da gösterilmiştir.

No.	1: class	2: adet	3: adres	4: aksesuar	5: al	6: alt	7: algiyeris	8: an	9: ana	10: anne	11: ara	12: ay	13: ac	14: bak	15: bako	16: bakim	17: banka	18: bebek	19: beyaz	20: bil	21: bilgi	22: bul	23: bulun	24: cam	25: dahil
1	edegill	0.0	0.0	0.0	1.643...	1.131...	0.0	0.0	0.0	0.0	0.0	0.9883...	0.0	0.0	0.0	0.7137...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
2	edegill	0.0	0.0	0.0	2.228...	1.657...	1.884...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.944...	1.712...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
3	eticaret	1.9286100...	2.200...	2.387264...	0.0	0.0	2.016...	0.0	2.72023...	0.0	1.4479...	1.956...	1.359...	2.844...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
4	eticaret	0.0	2.149...	1.591509...	1.170...	0.713...	2.354...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
5	eticaret	0.9643050...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
6	edegill	1.5283872...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
7	eticaret	2.2390468...	0.0	0.0	0.0	0.0	1.408...	2.982...	0.0	0.0	0.9883...	0.0	0.927...	2.404...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
8	eticaret	3.0567745...	0.0	0.795754...	0.0	0.0	1.884...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
9	edegill	0.0	1.827...	0.795754...	0.927...	0.0	0.0	2.569...	0.0	0.0	0.9883...	0.0	0.927...	2.213...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
10	edegill	0.0	0.872...	0.0	1.170...	0.713...	1.089...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0

Şekil 3.29. Dinamik çalışma TF-IDF doküman-kelime çizelgesi (%15)

%20 oranı ile elde edilen veri kümesi

E-ticaret sitelerinin en az %20’sinde ortak olarak geçen kelimelerin anahtar sözlüğü oluşturmada kullanılması ile elde edilen eğitim kümesi için TF-IDF ağırlıklandırma yöntemi kullanılarak oluşturulan vektör uzay modelinin bir kesiti Şekil 3.30’da gösterilmiştir.

Viewer

Relation: eticaretegitimverisiCalisma2020-weka.filters.unsupervised.attribute.StringToWordVector-R1-W1000-prune-rate-1.0-C-T-F-N0-stemmerveka.core.stemmers.NullStemmer-M1-tokenizerweka.core.tokenizers.WordTokenizer-delimiters '*\n(\n\t,...

No.	1: class	2: al	3: algiyari	4: an	5: ara	6: ara	7: bilgi	8: et	9: fiyat	10: gir	11: giris	12: giz	13: gizli	14: gün	15: güven	16: hak	17: hizmet	18: il	19: illet	20: in	21: indir	22: kategori	23: kullan	24: müşteri	25: ol	
	Nominal	Numeric	Numeric	Numeric	Numeric	Numeric	Numeric	Numeric	Numeric	Numeric	Numeric	Numeric	Numeric	Numeric	Numeric	Numeric	Numeric	Numeric	Numeric	Numeric	Numeric	Numeric	Numeric	Numeric	Numeric	
1	edegill	0.0	0.0	0.0	2.219...	1.664...	1.883...	0.0	1.951...	0.0	0.0	0.0	0.0	0.0	1.2843...	0.0	0.9459...	0.0	0.0	0.659...	0.0	0.0	0.0	0.0	0.0	1.213...
2	eticaret	2.315...	2.403014...	0.0	0.0	2.015...	1.456...	1.364...	1.428...	0.986...	1.0832...	1.319...	1.9369...	0.0	2.003303...	0.9459...	0.0	0.547...	0.0	0.0	0.0	2.0021977...	0.0	0.0	0.0	0.786...
3	eticaret	2.169...	1.602009...	1.165...	0.716...	2.353...	0.0	0.0	0.0	0.986...	1.0832...	0.832...	1.3221...	0.0	0.0	0.9459...	1.186319...	0.0	0.659...	0.0	0.0	0.0	0.61879...	1.2382060...	0.338...	0.0
4	eticaret	0.0	0.0	0.0	0.0	0.0	0.0	0.0	4.284...	1.244...	1.3668...	0.0	0.0	1.2843...	0.0	0.6739...	0.0	0.0	0.718...	0.9057...	2.2290121...	0.0	0.0	0.0	0.0	0.338...
5	edegill	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.6739...	1.377274...	0.0	0.926...	0.0	0.0	0.0	0.0	1.2382060...	0.0
6	eticaret	0.0	0.0	0.0	0.0	1.407...	0.994...	0.931...	0.0	0.986...	1.0832...	0.0	0.0	0.0	1.001651...	0.6739...	0.0	1.220...	1.856...	2.3414...	0.0	0.0	0.0	0.0	0.0	0.786...
7	eticaret	0.0	0.801004...	0.0	0.0	1.883...	0.0	0.0	1.131...	1.244...	1.3668...	0.0	0.0	0.0	2.6918...	1.633623...	1.2468...	0.0	0.659...	2.734...	0.0	2.0021977...	0.98076...	0.0	0.0	0.338...
8	edegill	1.844...	0.801004...	0.923...	0.0	0.0	0.994...	0.931...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.9459...	2.563593...	0.0	0.926...	0.0	0.0	0.0	0.0	0.0	0.0	0.786...
9	edegill	0.880...	0.0	1.165...	0.716...	1.089...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.8103...	0.0	1.1193...	0.0	0.547...	0.926...	0.0	1.3667125...	0.98076...	0.0	0.0	0.338...
10	edegill	0.555...	0.0	0.0	0.0	1.809...	0.0	0.0	0.0	1.244...	1.3668...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.3667125...	1.73716...	0.0	1.252...	

|||

UndoOKCancel

Şekil 3.30. Dinamik çalışma TF-IDF doküman-kelime çizelgesi (%20)

%25 oranı ile elde edilen veri kümesi

E-ticaret sitelerinin en az %25’inde ortak olarak geçen kelimelerin anahtar sözlüğü oluşturmada kullanılması ile elde edilen eğitim kümesi için TF-IDF ağırlıklandırma yöntemi kullanılarak oluşturulan vektör uzay modelinin bir kesiti Şekil 3.31’de gösterilmiştir.

No.	1: class	2: algiyari	3: ara	4: gir	5: giris	6: giz	7: gizli	8: güven	9: hak	10: illet	11: in	12: ol	13: sat	14: sepet	15: siparis	16: site	17: tl	18: yeni	19: ürün	20: üye
1	edegill	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.674...	0.660...	0.0	0.0	0.798...	0.0	0.0	0.0	0.0	0.0	0.7432...	0.0
2	edegill	0.0	1.888...	0.0	0.0	0.0	0.0	0.0	0.946...	0.660...	0.0	1.214...	0.0	0.0	0.755240...	0.9657...	0.0	0.7455...	0.5073...	0.0
3	eticaret	2.394754...	2.019...	0.989...	1.086...	1.327...	1.948...	2.01023...	0.946...	0.0	0.0	0.786...	2.139...	0.0	0.0	0.7653...	2.189...	1.0921...	1.6675...	1.6083...
4	eticaret	1.596502...	2.359...	0.989...	1.086...	0.837...	1.330...	0.0	0.946...	0.660...	0.0	0.338...	0.0	3.63398...	1.753613...	0.9657...	3.060...	0.0	0.0	0.0
5	eticaret	0.0	0.0	1.249...	1.371...	0.0	0.0	0.0	0.674...	0.0	0.725...	0.338...	1.169...	1.06813...	0.0	0.0	1.486...	0.4703...	1.2505...	0.9861...
6	edegill	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.674...	0.926...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.8986...	0.0
7	eticaret	0.0	1.410...	0.989...	1.086...	0.0	0.0	1.00511...	0.674...	1.221...	1.875...	0.786...	0.798...	2.86275...	0.755240...	0.9657...	2.485...	0.9407...	2.1991...	0.0
8	eticaret	0.798251...	1.888...	1.249...	1.371...	0.0	0.0	1.63927...	1.247...	0.660...	2.685...	0.338...	0.798...	1.34783...	0.0	0.4828...	0.0	0.0	0.9602...	0.9861...
9	edegill	0.798251...	0.0	0.0	0.0	0.0	0.0	0.0	0.946...	0.926...	0.0	0.786...	0.0	0.0	0.0	0.7653...	0.0	0.4703...	0.0	0.0
10	edegill	0.0	1.091...	0.0	0.0	0.0	0.0	0.0	1.120...	0.926...	0.0	0.338...	1.169...	0.0	0.0	1.1212...	2.128...	0.0	0.0	1.9722...

Şekil 3.31. Dinamik çalışma TF-IDF doküman-kelime çizelgesi (%25)

%30 oranı ile elde edilen veri kümesi

E-ticaret sitelerinin en az %30’unda ortak olarak geçen kelimelerin anahtar sözlüğü oluşturmada kullanılması ile elde edilen eğitim kümesi için TF-IDF ağırlıklandırma yöntemi kullanılarak oluşturulan vektör uzay modelinin bir kesiti Şekil 3.32’de gösterilmiştir.

No.	1: class Nominal	2: gir Numeric	3: hak Numeric	4: ilet Numeric	5: ol Numeric	6: sat Numeric	7: sepet Numeric	8: tl Numeric	9: yeni Numeric	10: ürün Numeric	11: ürünü Numeric	12: üye Numeric
1	edegill	0.0	0.953...	0.663...	1.211...	0.0	0.0	0.0	0.745...	0.5097...	0.0	0.0
2	eticaret	0.989...	0.953...	0.0	0.784...	2.169...	0.0	2.218...	0.940...	1.5933...	0.0	1.5982...
3	eticaret	0.989...	0.953...	0.663...	0.338...	0.0	3.6199...	3.022...	0.0	0.0	0.0	0.0
4	eticaret	1.248...	0.679...	0.0	0.338...	1.185...	1.0640...	1.468...	0.470...	1.2564...	0.0	0.9799...
5	edegill	0.0	0.679...	0.931...	0.0	0.0	0.0	0.0	0.9028...	0.0	0.0	0.0
6	eticaret	0.989...	0.679...	1.227...	0.784...	0.809...	2.8517...	2.454...	0.940...	2.2095...	0.0	0.0
7	eticaret	1.248...	1.256...	0.663...	0.338...	0.809...	1.3426...	0.0	0.0	0.9648...	0.0	0.9799...
8	edegill	0.0	0.953...	0.931...	0.784...	0.0	0.0	0.0	0.470...	0.0	0.0	0.0
9	edegill	0.0	1.128...	0.931...	0.338...	1.185...	0.0	2.101...	0.0	0.0	0.0	1.9599...
10	edegill	1.248...	0.0	0.0	1.250...	0.0	0.0	0.0	0.470...	0.9648...	0.0	1.4356...

Şekil 3.32. Dinamik çalışma TF-IDF doküman-kelime çizelgesi (%30)

%35 oranı ile elde edilen veri kümesi

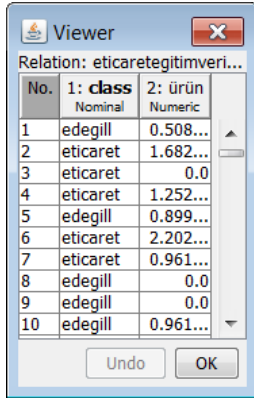
E-ticaret sitelerinin en az %35’inde ortak olarak geçen kelimelerin anahtar sözlüğü oluşturmada kullanılması ile elde edilen eğitim kümesi için TF-IDF ağırlıklandırma yöntemi kullanılarak oluşturulan vektör uzay modelinin bir kesiti Şekil 3.33’te gösterilmiştir.

No.	1: class Nominal	2: ilet Numeric	3: ol Numeric	4: sepet Numeric	5: tl Numeric	6: ürün Numeric
1	edegill	0.663...	0.0	0.0	0.0	0.7470...
2	edegill	0.663...	1.217...	0.0	0.0	0.5099...
3	eticaret	0.0	0.788...	0.0	2.228...	1.5938...
4	eticaret	0.663...	0.339...	3.6261...	3.035...	0.0
5	eticaret	0.0	0.339...	1.0658...	1.474...	1.2569...
6	edegill	0.931...	0.0	0.0	0.0	0.9031...
7	eticaret	1.227...	0.788...	2.8565...	2.465...	2.2103...
8	eticaret	0.663...	0.339...	1.3449...	0.0	0.9651...
9	edegill	0.931...	0.788...	0.0	0.0	0.0
10	edegill	0.931...	0.339...	0.0	2.110...	0.0

Şekil 3.33. Dinamik çalışma TF-IDF doküman-kelime çizelgesi (%35)

%40 oranı ile elde edilen veri kümesi

E-ticaret sitelerinin en az %40’ında ortak olarak geçen kelimelerin anahtar sözlüğü oluşturmada kullanılması ile elde edilen eğitim kümesi için TF-IDF ağırlıklandırma yöntemi kullanılarak oluşturulan vektör uzay modelinin bir kesiti Şekil 3.34’te gösterilmiştir.



No.	1: class	2: ürün
	Nominal	Numeric
1	edegill	0.508...
2	eticaret	1.682...
3	eticaret	0.0
4	eticaret	1.252...
5	edegill	0.899...
6	eticaret	2.202...
7	eticaret	0.961...
8	edegill	0.0
9	edegill	0.0
10	edegill	0.961...

Şekil 3.34. Dinamik çalışma TF-IDF doküman-kelime çizelgesi (%40)

3.5. Algoritmaların Kullanılması

Eğitim verisi üzerinde mesafeye dayalı sınıflandırma yöntemlerinden k-NN algoritması (ikili ağırlıklandırma, TF ağırlıklandırma ve TF-IDF ağırlıklandırma) ve istatistiğe dayalı sınıflandırma yöntemlerinden Naïve Bayes (ikili ağırlıklandırma ve TF ağırlıklandırma) algoritmaları kullanılmıştır.

Algoritmaların karşılaştırılmasında doğru sınıflandırma yüzdesi (correctly classified instances %), doğru sınıflandırma sayısı (doğru sınıflandırılan doküman sayısı - correctly classified instances), yanlış sınıflandırma yüzdesi (incorrectly classified instances %), yanlış sınıflandırma sayısı (yanlış sınıflandırılan doküman sayısı - incorrectly classified instances), TP oranı (TP Rate), FP oranı (FP Rate), Kesinlik (Precision), Hassasiyet (Recall) ve F-Ölçütü (F-Measure) değerleri kullanılmıştır.

3.5.1. Statik çalışma

Statik çalışma için önceden belirlenmiş olan 30 adet anahtar kelime oluşturulmuş eğitim kümesi için k-NN (ikili, TF ve TF-IDF) ve Naïve Bayes (ikili ve TF) algoritmaları uygulanmıştır.

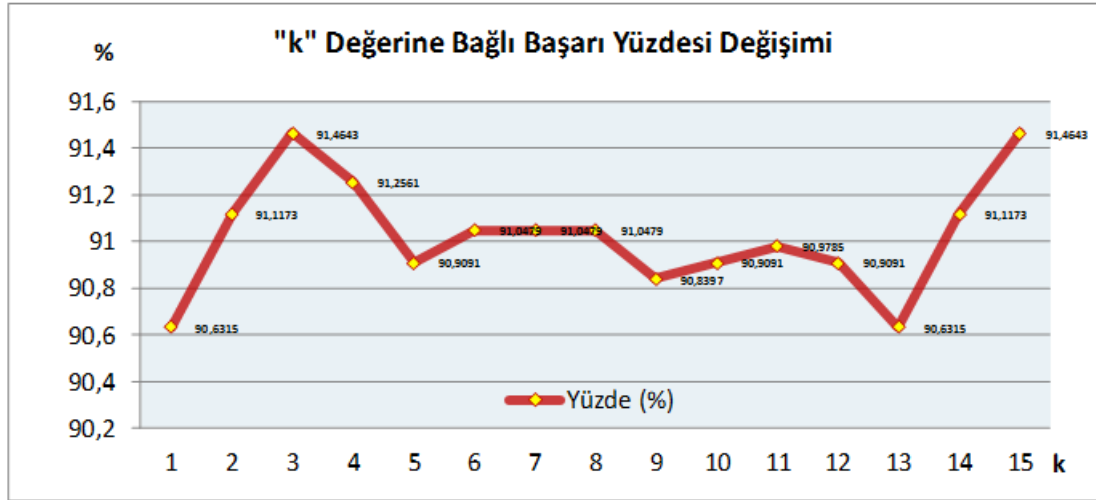
k-NN algoritması ikili ağırlıklandırma yöntemi

k-NN algoritması için ikili ağırlıklandırma yönteminde kullanılan eğitim kümesinin vektör uzay modelinin bir kesiti gösterilmişti (Bkz. Şekil 3.5). Bu eğitim verisi üzerinde k-NN algoritması ikili ağırlıklandırma yönteminin uygulanmasında “k” değeri için “1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14 ve 15” değerleri denenmiştir. Özet algoritma sonuçları Çizelge 3.5’te verilmiştir.

Çizelge 3.5. Dinamik çalışma k-NN - ikili ağırlıklandırma yöntemi sonucu

Ölçüt k	Doğru Sınıflandırma Yüzdesi (%)	Yanlış Sınıflandırma Yüzdesi (%)	Doğru Sınıflandırma Sayısı	Yanlış Sınıflandırma Sayısı	TP Oranı	FP Oranı	Kesinlik	Hassasiyet	F-Ölçütü
k=1	90,6315	9,3685	1306	135	0,809	0,177	0,823	0,809	0,809
k=2	91,1173	8,8827	1313	128	0,810	0,179	0,820	0,810	0,811
k=3	91,4643	8,5357	1318	123	0,819	0,160	0,848	0,819	0,818
k=4	91,2561	8,7439	1315	126	0,813	0,165	0,842	0,813	0,812
k=5	90,9091	9,0909	1310	131	0,808	0,168	0,844	0,808	0,806
k=6	91,0479	8,9521	1312	129	0,811	0,165	0,847	0,811	0,809
k=7	91,0479	8,9521	1312	129	0,810	0,164	0,851	0,810	0,807
k=8	91,0479	8,9521	1312	129	0,808	0,166	0,850	0,808	0,805
k=9	90,8397	9,1603	1309	132	0,804	0,168	0,849	0,804	0,801
k=10	90,9091	9,0909	1310	131	0,801	0,171	0,848	0,801	0,798
k=11	90,9785	9,0215	1311	130	0,800	0,172	0,848	0,800	0,797
k=12	90,9091	9,0909	1310	131	0,798	0,173	0,847	0,798	0,795
k=13	90,6315	9,3685	1306	135	0,809	0,177	0,823	0,809	0,809
k=14	91,1173	8,8827	1313	128	0,810	0,179	0,820	0,810	0,811
k=15	91,4643	8,5357	1318	123	0,819	0,160	0,848	0,819	0,818

En başarılı sonucun % 91,4643 ile k = 3 olduğu durumda elde edildiği görülmüştür. “k” değerine bağlı olarak algoritma başarı yüzdesinin değişimi Şekil 3.35’te verilmiştir.



Şekil 3.35. “k” değerine bağlı başarı yüzdesi değişimi (İkili)

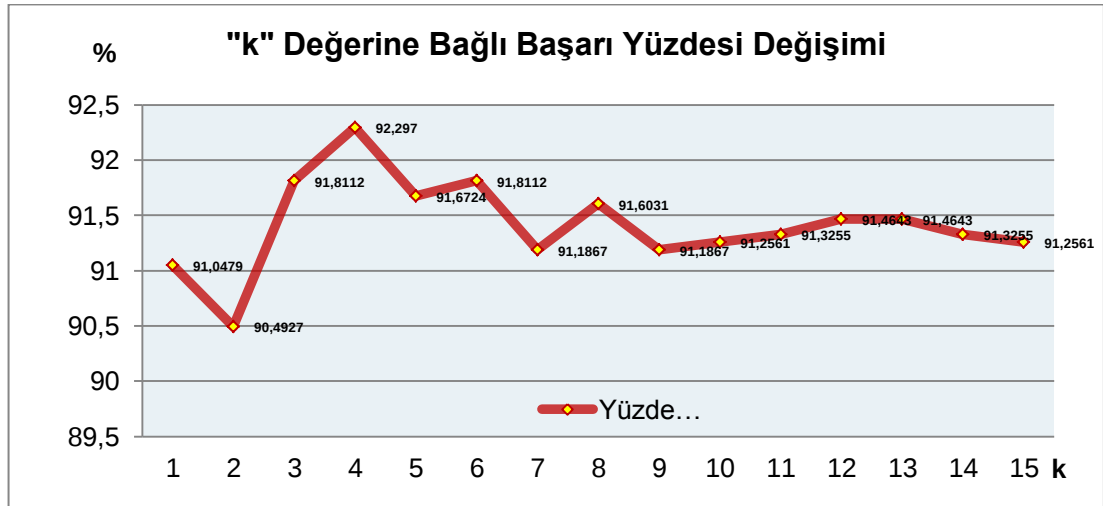
k-NN algoritması TF ağırlıklıklandırma yöntemi

k-NN algoritması için TF ağırlıklandırma yönteminde kullanılan eğitim kümesinin vektör uzay modelinin bir kesiti gösterilmişti (Bkz. Şekil 3.6). Bu eğitim verisi üzerinde k-NN algoritması TF ağırlıklandırma yönteminin uygulanmasında “k” değeri için “1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14 ve 15” değerleri denenmiştir. Özet algoritma sonuçları Çizelge 3.6’da verilmiştir.

En başarılı sonucun % 92,2970 ile $k = 4$ olduğu durumda elde edildiği görülmüştür. “k” değerine bağlı olarak algoritma başarı yüzdesinin değişimi Şekil 3.36’da verilmiştir.

Çizelge 3.6. Dinamik çalışma k-NN - TF ağırlıklandırma yöntemi sonucu

Ölçüt k	Doğru Sınıflandırma Yüzdesi (%)	Yanlış Sınıflandırma Yüzdesi (%)	Doğru Sınıflandırma Sayısı	Yanlış Sınıflandırma Sayısı	TP Oranı	FP Oranı	Kesinlik	Hassasiyet	F-Ölçütü
k=1	91,0479	8,9521	1312	129	0,910	0,094	0,912	0,910	0,910
k=2	90,4927	9,5073	1304	137	0,905	0,094	0,906	0,905	0,905
k=3	91,8112	8,1888	1323	118	0,918	0,087	0,920	0,918	0,918
k=4	92,2970	7,7030	1330	111	0,923	0,079	0,923	0,923	0,923
k=5	91,6724	8,3276	1321	120	0,917	0,089	0,919	0,917	0,916
k=6	91,8112	8,1888	1323	118	0,918	0,086	0,919	0,918	0,918
k=7	91,1867	8,8133	1314	127	0,912	0,095	0,915	0,912	0,911
k=8	91,6031	8,3969	1320	121	0,916	0,090	0,918	0,916	0,916
k=9	91,1867	8,8133	1314	127	0,912	0,095	0,915	0,912	0,911
k=10	91,2561	8,7439	1315	126	0,913	0,093	0,915	0,913	0,912
k=11	91,3255	8,6745	1316	125	0,913	0,094	0,917	0,913	0,913
k=12	91,4643	8,5357	1318	123	0,915	0,092	0,918	0,915	0,914
k=13	91,4643	8,5357	1318	123	0,915	0,093	0,919	0,915	0,914
k=14	91,3255	8,6745	1316	125	0,913	0,094	0,917	0,913	0,913
k=15	91,2561	8,7439	1315	126	0,913	0,095	0,917	0,913	0,912



Şekil 3.36. "k" değerine bağlı başarı yüzdesi değişimi (TF)

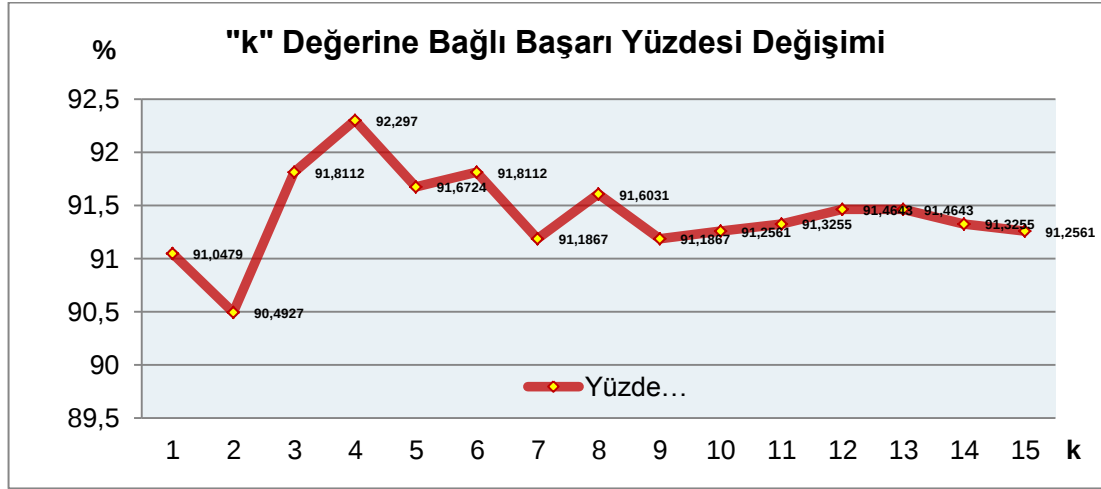
k-NN – TF-IDF ağırlıklandırma yöntemi

k-NN algoritması için TF-IDF ağırlıklandırma yönteminde kullanılan eğitim kümesinin vektör uzay modelinin bir kesiti gösterilmişti (Bkz. Şekil 3.7). Bu eğitim verisi üzerinde k-NN algoritması TF-IDF ağırlıklandırma yönteminin uygulanmasında “k” değeri için “1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14 ve 15” değerleri denenmiştir. Özet algoritma sonuçları Çizelge 3.7’de verilmiştir.

Çizelge 3.7. Dinamik çalışma k-NN - TF-IDF ağırlıklandırma yöntemi sonucu

Ölçüt k	Doğru Sınıflandırma Yüzdesi (%)	Yanlış Sınıflandırma Yüzdesi (%)	Doğru Sınıflandırma Sayısı	Yanlış Sınıflandırma Sayısı	TP Oranı	FP Oranı	Kesinlik	Hassasiyet	F-Ölçütü
k=1	91,0479	8,9521	1312	129	0,910	0,094	0,912	0,910	0,910
k=2	90,4927	9,5073	1304	137	0,905	0,094	0,906	0,905	0,905
k=3	91,8112	8,1888	1323	118	0,918	0,087	0,920	0,918	0,918
k=4	92,2970	7,7030	1330	111	0,923	0,079	0,923	0,923	0,923
k=5	91,6724	8,3276	1321	120	0,917	0,089	0,919	0,917	0,916
k=6	91,8112	8,1888	1323	118	0,918	0,086	0,919	0,918	0,918
k=7	91,1867	8,8133	1314	127	0,912	0,095	0,915	0,912	0,911
k=8	91,6031	8,3969	1320	121	0,916	0,090	0,918	0,916	0,916
k=9	91,1867	8,8133	1314	127	0,912	0,095	0,915	0,912	0,911
k=10	91,2561	8,7439	1315	126	0,913	0,093	0,915	0,913	0,912
k=11	91,3255	8,6745	1316	125	0,913	0,094	0,917	0,913	0,913
k=12	91,4643	8,5357	1318	123	0,915	0,092	0,918	0,915	0,914
k=13	91,4643	8,5357	1318	123	0,915	0,093	0,919	0,915	0,914
k=14	91,3255	8,6745	1316	125	0,913	0,094	0,917	0,913	0,913
k=15	91,2561	8,7439	1315	126	0,913	0,095	0,917	0,913	0,912

En başarılı sonucun % 92,2970 ile k = 4 olduğu durumda elde edildiği görülmüştür. “k” değerine bağlı olarak algoritma başarı yüzdesinin değişimi Şekil 3.37’de verilmiştir.



Şekil 3.37. “k” değerine bağlı başarı yüzdesi değişimi (TF-IDF)

Naïve Bayes algoritması ikili ağırlıklandırma (Multi-variate model)

Naïve Bayes algoritması için ikili ağırlıklandırma yönteminde kullanılan eğitim kümesinin vektör uzay modelinin bir kesiti gösterilmiştir (Bkz. Şekil 3.5). Naïve Bayes algoritması için ikili ağırlıklandırma yönteminin başarısı % 90,9091 olarak bulunmuştur. Özet algoritma sonuçları Çizelge 3.8’de verilmiştir.

Çizelge 3.8. Dinamik çalışma Naïve Bayes ikili ağırlıklandırma yöntemi sonucu

Ölçüt k	Doğru Sınıflandırma Yüzdesi (%)	Yanlış Sınıflandırma Yüzdesi (%)	Doğru Sınıflandırma Sayısı	Yanlış Sınıflandırma Sayısı	TP Oranı	FP Oranı	Kesinlik	Hassasiyet	F-Ölçütü
Naïve Bayes İkili	90,9091	9,0909	1310	131	0,909	0,089	0,910	0,909	0,909

Naïve Bayes algoritması TF ağırlıklandırma (Multinomial model)

Naïve Bayes algoritması için TF ağırlıklandırma yönteminde kullanılan eğitim kümesinin vektör uzay modelinin bir kesiti gösterilmiştir (Bkz. Şekil 3.6). Naïve Bayes algoritması için TF ağırlıklandırma yönteminin başarısı % 90,7009 olarak bulunmuştur. Özet algoritma sonuçları Çizelge 3.9’da verilmiştir.

Çizelge 3.9. Dinamik çalışma Naïve Bayes TF ağırlıklandırma yöntemi sonucu

Ölçüt k	Doğru Sınıflandırma Yüzdesi (%)	Yanlış Sınıflandırma Yüzdesi (%)	Doğru Sınıflandırma Sayısı	Yanlış Sınıflandırma Sayısı	TP Oranı	FP Oranı	Kesinlik	Hassasiyet	F- Ölçütü
Naïve Bayes TF	90,7009	9,2991	1307	134	0,907	0,095	0,907	0,907	0,907

3.5.2. Dinamik çalışma

Dinamik çalışma için e-ticaret sitelerinin en az belirlenen yüzdesinde ortak olarak geçen kelimeler ile anahtar kelime sözlüğü oluşturulmuş, ardından k-NN (ikili, TF ve TF-IDF) ve Naïve Bayes (ikili ve TF) algoritmaları uygulanmıştır. Dinamik çalışma belirlenen yüzde değerinin değişiminin modelin başarısını nasıl etkilediğini görmek amacıyla %0, %5, %10, %15, %20, %25, %30, %35 ve %40 değerleri ile algoritmalar üzerinde uygulanmıştır.

k-NN algoritması ikili ağırlıklandırma yöntemi

Eğitim verisi üzerinde k-NN algoritması ikili ağırlıklandırma yöntemi %0, %5, %10, %15, %20, %25, %30, %35 ve %40 değerleri ile uygulanmıştır.

%0 oranı ile elde edilen veri kümesi

E-ticaret sitelerinin en az %0'ında ortak olarak geçen kelimeler yani tüm kelimeler dikkate alınarak elde edilen eğitim verisi üzerinde k-NN algoritması ikili ağırlıklandırma yönteminin uygulanmasında “k” değeri için “1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14 ve 15” değerleri denenmiştir. Algoritma başarı yüzdeleri Çizelge 3.10 ve Çizelge 3.11’de verilmiştir.

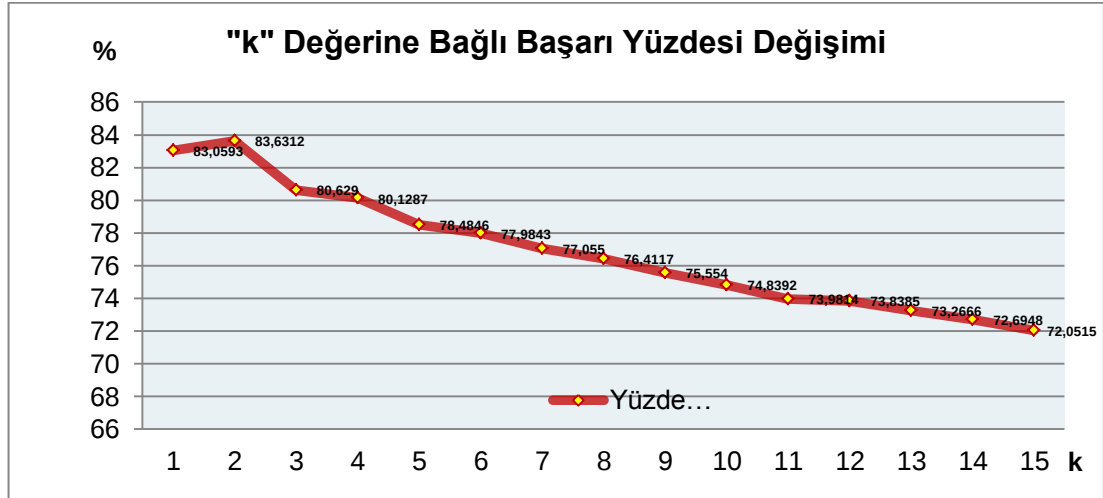
Çizelge 3.10. Dinamik çalışma k-NN ikili ağırlıklandırma sonucu (%0) (k=1...8)

Yüzde Değeri \ k	k=1	k=2	k=3	k=4	k=5	k=6	k=7	k=8
Doğru Sınıflandırma Yüzdesi (%)	83,0593	83,6312	80,6290	80,1287	78,4846	77,9843	77,0550	76,4117
Yanlış Sınıflandırma Yüzdesi (%)	16,9407	16,3688	19,3710	19,8713	21,5154	22,0157	22,9450	23,5883

Çizelge 3.11. Dinamik çalışma k-NN ikili ağırlıklandırma sonucu (%0) (k=9...15)

Yüzde Değeri \ k	k=9	k=10	k=11	k=12	k=13	k=14	k=15
Doğru Sınıflandırma Yüzdesi (%)	75,5540	74,8392	73,9814	73,8385	73,2666	72,6948	72,0515
Yanlış Sınıflandırma Yüzdesi (%)	24,4460	25,1608	26,0186	26,1615	26,7334	27,3052	27,9485

En başarılı sonucun % 83,6312 ile k = 2 olduğu durumda elde edildiği görülmüştür. “k” değerine bağlı olarak algoritma başarı yüzdesinin değişimi Şekil 3.38’de verilmiştir.



Şekil 3.38. “k” değerine bağlı başarı yüzdesi değişimi (%0) (İkili)

%5 oranı ile elde edilen veri kümesi

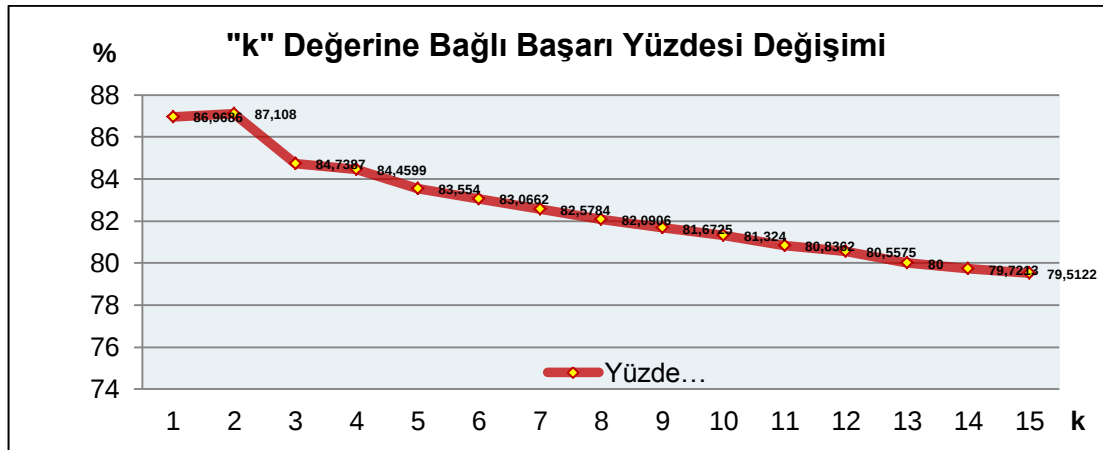
E-ticaret sitelerinin en az %5’inde ortak olarak geçen kelimeler dikkate alınarak elde edilen eğitim verisi üzerinde k-NN algoritması ikili ağırlıklandırma yönteminin uygulanmasında “k” değeri için “1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14 ve 15” değerleri denenmiştir. Algoritma başarı yüzdeleri Çizelge 3.12 ve Çizelge 3.13’te verilmiştir. En başarılı sonucun % 87,108 ile k = 2 olduğu durumda elde edildiği görülmüştür. “k” değerine bağlı algoritma başarı yüzdesi değişimi Şekil 3.39’dadır.

Çizelge 3.12. Dinamik çalışma k-NN ikili ağırlıklandırma sonucu (%5) (k=1...8)

Yüzde Değeri \ k	k=1	k=2	k=3	k=4	k=5	k=6	k=7	k=8
Doğru Sınıflandırma Yüzdesi (%)	86,9686	87,1080	84,7387	84,4599	83,5540	83,0662	82,5784	82,0906
Yanlış Sınıflandırma Yüzdesi (%)	13,0314	12,8920	15,2613	15,5401	16,4460	16,9338	17,4216	17,9094

Çizelge 3.13. Dinamik çalışma k-NN ikili ağırlıklandırma sonucu (%5) (k=9...15)

Yüzde Değeri \ k	k=9	k=10	k=11	k=12	k=13	k=14	k=15
Doğru Sınıflandırma Yüzdesi (%)	81,6725	81,3240	80,8362	80,5575	80,0000	79,7213	79,5122
Yanlış Sınıflandırma Yüzdesi (%)	18,3275	18,6760	19,1638	19,4425	20,0000	20,2787	20,4878



Şekil 3.39. “k” değerine bağlı başarı yüzdesi değişimi (%5) (İkili)

%10 oranı ile elde edilen veri kümesi

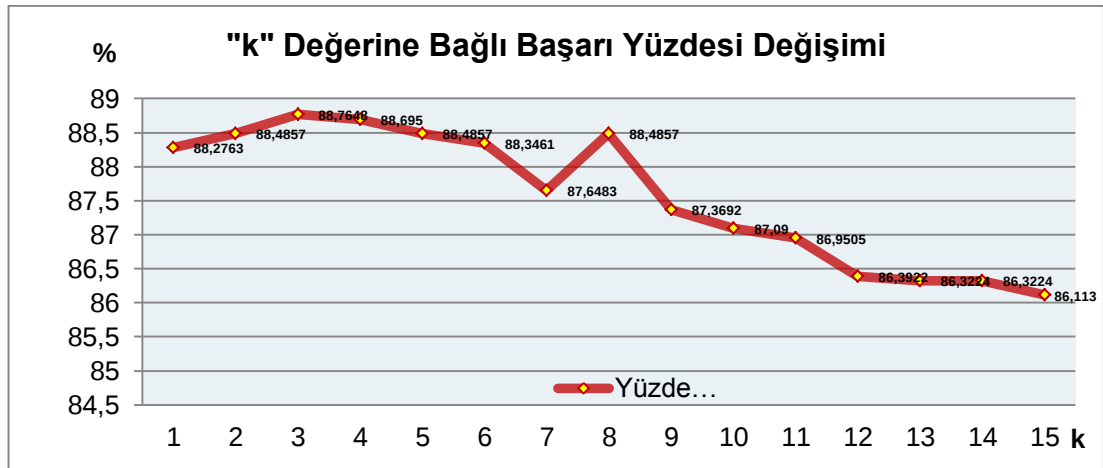
E-ticaret sitelerinin en az %10'unda ortak olarak geçen kelimeler dikkate alınarak elde edilen eğitim verisi üzerinde k-NN algoritması ikili ağırlıklandırma yönteminin uygulanmasında “k” değeri için “1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14 ve 15” değerleri denenmiştir. Algoritma başarı yüzdeleri Çizelge 3.14 ve Çizelge 3.15'tedir. En başarılı sonucun % 88,7648 ile k = 3 olduğu durumda elde edildiği görülmüştür. “k” değerine bağlı algoritma başarı yüzdesinin değişimi Şekil 3.40'ta verilmiştir.

Çizelge 3.14. Dinamik çalışma k-NN ikili ağırlıklandırma sonucu (%10) (k=1...8)

Yüzde Değeri \ k	k=1	k=2	k=3	k=4	k=5	k=6	k=7	k=8
Doğru Sınıflandırma Yüzdesi (%)	88,2763	88,4857	88,7648	88,6950	88,4857	88,3461	87,6483	88,4857
Yanlış Sınıflandırma Yüzdesi (%)	11,7237	11,5143	11,2352	11,3050	11,5143	11,6539	12,3517	12,5611

Çizelge 3.15. Dinamik çalışma k-NN ikili ağırlıklandırma sonucu (%10) (k=9...15)

Yüzde Değeri \ k	k=9	k=10	k=11	k=12	k=13	k=14	k=15
Doğru Sınıflandırma Yüzdesi (%)	87,3692	87,0900	86,9505	86,3922	86,3224	86,3224	86,1130
Yanlış Sınıflandırma Yüzdesi (%)	12,6308	12,9100	13,0495	13,6078	13,6776	13,6776	13,8870



Şekil 3.40. “k” değerine bağlı başarı yüzdesi değişimi (%10) (İkili)

%15 oranı ile elde edilen veri kümesi

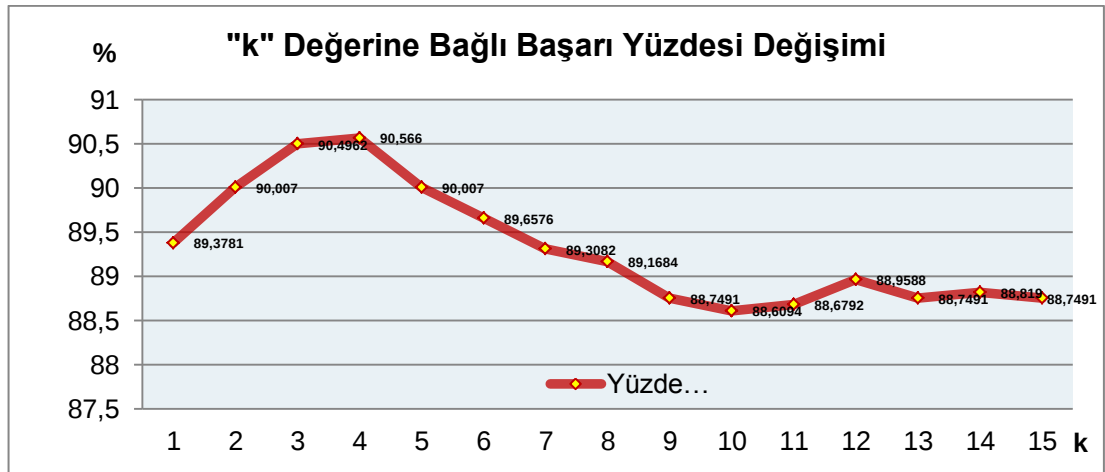
E-ticaret sitelerinin en az %15’inde ortak olarak geçen kelimeler dikkate alınarak elde edilen eğitim verisi üzerinde k-NN algoritması ikili ağırlıklandırma yönteminin uygulanmasında “k” değeri için “1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14 ve 15” değerleri denenmiştir. Algoritma başarı yüzdeleri Çizelge 3.16 ve Çizelge 3.17’dedir. En başarılı sonucun % 90,5660 ile k = 4 olduğu durumda elde edildiği görülmüştür. “k” değerine bağlı olarak algoritma başarı yüzdesinin değişimi Şekil 3.41’dedir.

Çizelge 3.16. Dinamik çalışma k-NN ikili ağırlıklandırma sonucu (%15) (k=1...8)

Yüzde Değeri \ k	k=1	k=2	k=3	k=4	k=5	k=6	k=7	k=8
Doğru Sınıflandırma Yüzdesi (%)	89,3781	90,0070	90,4962	90,5660	90,0070	89,6576	89,3082	89,1684
Yanlış Sınıflandırma Yüzdesi (%)	10,6219	9,9930	9,5038	9,4340	9,9930	10,3424	10,6918	10,8316

Çizelge 3.17. Dinamik çalışma k-NN ikili ağırlıklandırma sonucu (%15) (k=9...15)

Yüzde Değeri \ k	k=9	k=10	k=11	k=12	k=13	k=14	k=15
Doğru Sınıflandırma Yüzdesi (%)	88,7491	88,6094	88,6792	88,9588	88,7491	88,8190	88,7491
Yanlış Sınıflandırma Yüzdesi (%)	11,2509	11,3906	11,3208	11,0412	11,2509	11,1810	11,2509



Şekil 3.41. “k” değerine bağlı başarı yüzdesi değişimi (%15) (İkili)

%20 oranı ile elde edilen veri kümesi

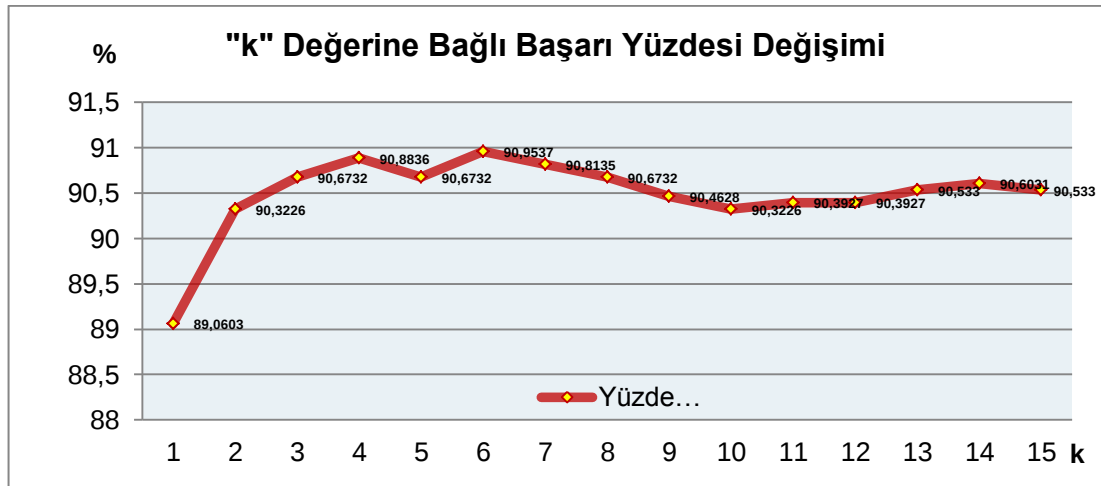
E-ticaret sitelerinin en az %20'sinde ortak olarak geçen kelimeler dikkate alınarak elde edilen eğitim verisi üzerinde k-NN algoritması ikili ağırlıklandırma yönteminin uygulanmasında “k” değeri için “1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14 ve 15” değerleri denenmiştir. Algoritma başarı yüzdeleri Çizelge 3.18 ve Çizelge 3.19’dadır. En başarılı sonucun % 90,8836 ile k = 4 olduğu durumda elde edildiği görülmüştür. “k” değerine bağlı olarak algoritma başarı yüzdesinin değişimi Şekil 3.42’dedir.

Çizelge 3.18. Dinamik çalışma k-NN ikili ağırlıklandırma sonucu (%20) (k=1...8)

k	k=1	k=2	k=3	k=4	k=5	k=6	k=7	k=8
Doğru Sınıflandırma Yüzdesi (%)	89,0603	90,3226	90,6732	90,8836	90,6732	90,9537	90,8135	90,6732
Yanlış Sınıflandırma Yüzdesi (%)	10,9397	9,6774	9,3268	9,1164	9,3268	9,0463	9,1865	9,3268

Çizelge 3.19. Dinamik çalışma k-NN ikili ağırlıklandırma sonucu (%20) (k=9...15)

k	k=9	k=10	k=11	k=12	k=13	k=14	k=15
Doğru Sınıflandırma Yüzdesi (%)	90,4628	90,3226	90,3927	90,3927	90,5330	90,6031	90,5330
Yanlış Sınıflandırma Yüzdesi (%)	9,5372	9,6774	9,6073	9,6073	9,4670	9,3969	9,4670



Şekil 3.42. “k” değerine bağlı başarı yüzdesi değişimi (%20) (İkili)

%25 oranı ile elde edilen veri kümesi

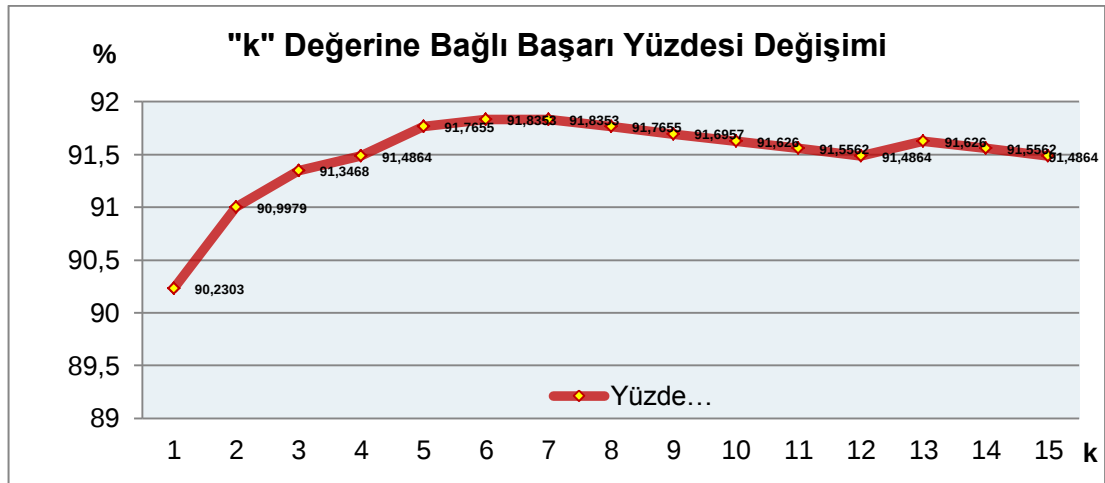
E-ticaret sitelerinin en az %25’inde ortak olarak geçen kelimeler dikkate alınarak elde edilen eğitim verisi üzerinde k-NN algoritması ikili ağırlıklandırma yönteminin uygulanmasında “k” değeri için “1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14 ve 15” değerleri denenmiştir. Algoritma başarı yüzdeleri Çizelge 3.20 ve Çizelge 3.21’dedir. En başarılı sonucun % 91,8353 ile k = 6 ve k = 7 olduğu durumda elde edilmiştir. “k” değerine bağlı olarak algoritma başarı yüzdesi değişimi Şekil 3.43’tedir.

Çizelge 3.20. Dinamik çalışma k-NN ikili ağırlıklandırma sonucu (%25) (k=1...8)

Yüzde Değeri \ k	k=1	k=2	k=3	k=4	k=5	k=6	k=7	k=8
Doğru Sınıflandırma Yüzdesi (%)	90,2303	90,9979	91,3468	91,4864	91,7655	91,8353	91,8353	91,7655
Yanlış Sınıflandırma Yüzdesi (%)	9,7697	9,0021	8,6532	8,5136	8,2345	8,1647	8,1647	8,2345

Çizelge 3.21. Dinamik çalışma k-NN ikili ağırlıklandırma sonucu (%25) (k=9...15)

Yüzde Değeri \ k	k=9	k=10	k=11	k=12	k=13	k=14	k=15
Doğru Sınıflandırma Yüzdesi (%)	91,6957	91,6260	91,5562	91,4864	91,6260	91,5562	91,4864
Yanlış Sınıflandırma Yüzdesi (%)	8,3043	8,3740	8,4438	8,5136	8,3740	8,4438	8,5136



Şekil 3.43. “k” değerine bağlı başarı yüzdesi değişimi (%25) (İkili)

%30 oranı ile elde edilen veri kümesi

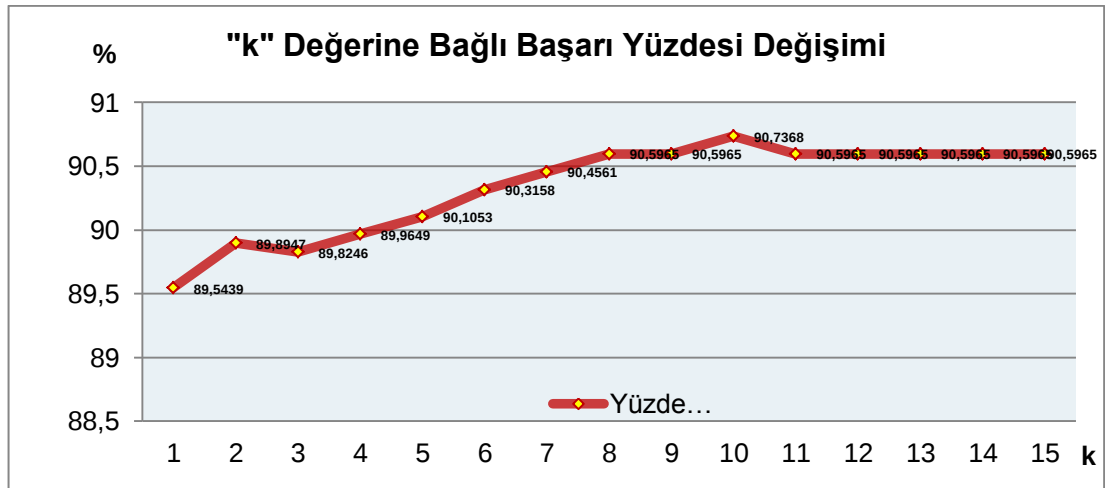
E-ticaret sitelerinin en az %30'unda ortak olarak geçen kelimeler dikkate alınarak elde edilen eğitim verisi üzerinde k-NN algoritması ikili ağırlıklandırma yönteminin uygulanmasında “k” değeri için “1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14 ve 15” değerleri denenmiştir. Algoritma başarı yüzdeleri Çizelge 3.22 ve Çizelge 3.23'tedir. En başarılı sonucun %90,7368 ile k = 10 olduğu durumda elde edildiği görülmüştür. “k” değerine bağlı olarak algoritma başarı yüzdesinin değişimi Şekil 3.44'tedir.

Çizelge 3.22. Dinamik çalışma k-NN ikili ağırlıklandırma sonucu (%30) (k=1...8)

Yüzde Değeri \ k	k=1	k=2	k=3	k=4	k=5	k=6	k=7	k=8
Doğru Sınıflandırma Yüzdesi (%)	89,5439	89,8947	89,8246	89,9649	90,1053	90,3158	90,4561	90,5965
Yanlış Sınıflandırma Yüzdesi (%)	10,4561	10,1053	10,1754	10,0351	9,8947	9,6842	9,5439	9,4035

Çizelge 3.23. Dinamik çalışma k-NN ikili ağırlıklandırma sonucu (%30) (k=9...15)

Yüzde Değeri \ k	k=9	k=10	k=11	k=12	k=13	k=14	k=15
Doğru Sınıflandırma Yüzdesi (%)	90,5965	90,7368	90,5965	90,5965	90,5965	90,5965	90,5965
Yanlış Sınıflandırma Yüzdesi (%)	9,4035	9,2632	9,4035	9,4035	9,4035	9,4035	9,4035



Şekil 3.44. “k” değerine bağlı başarı yüzdesi değişimi (%30) (İkili)

%35 oranı ile elde edilen veri kümesi

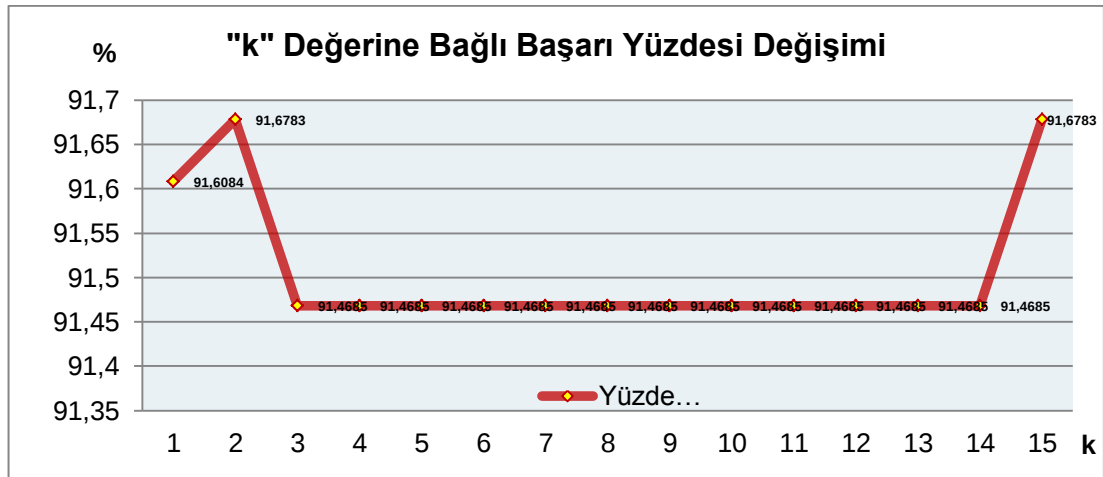
E-ticaret sitelerinin en az %35’inde ortak olarak geçen kelimeler dikkate alınarak elde edilen eğitim verisi üzerinde k-NN algoritması ikili ağırlıklandırma yönteminin uygulanmasında “k” değeri için “1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14 ve 15” değerleri denenmiştir. Algoritma başarı yüzdeleri Çizelge 24 ve Çizelge 25’te verilmiştir. En başarılı sonucun % 91,6783 ile k = 2 ve k = 15 olduğu durumlarda elde edilmiştir. “k” değerine bağlı algoritma başarı yüzdesi değişimi Şekil 3.45’tedir.

Çizelge 3.24. Dinamik çalışma k-NN ikili ağırlıklandırma sonucu (%35) (k=1...8)

Yüzde Değeri \ k	k=1	k=2	k=3	k=4	k=5	k=6	k=7	k=8
Doğru Sınıflandırma Yüzdesi (%)	91,6084	91,6783	91,4685	91,4685	91,4685	91,4685	91,4685	91,4685
Yanlış Sınıflandırma Yüzdesi (%)	8,3916	8,3217	8,5315	8,5315	8,5315	8,5315	8,5315	8,5315

Çizelge 3.25. Dinamik çalışma k-NN ikili ağırlıklandırma sonucu (%35) (k=9...15)

Yüzde Değeri \ k	k=9	k=10	k=11	k=12	k=13	k=14	k=15
Doğru Sınıflandırma Yüzdesi (%)	91,4685	91,4685	91,4685	91,4685	91,4685	91,4685	91,6783
Yanlış Sınıflandırma Yüzdesi (%)	8,5315	8,5315	8,5315	8,5315	8,5315	8,5315	8,3217



Şekil 3.45. “k” değerine bağlı başarı yüzdesi değişimi (%35) (İkili)

%40 oranı ile elde edilen veri kümesi

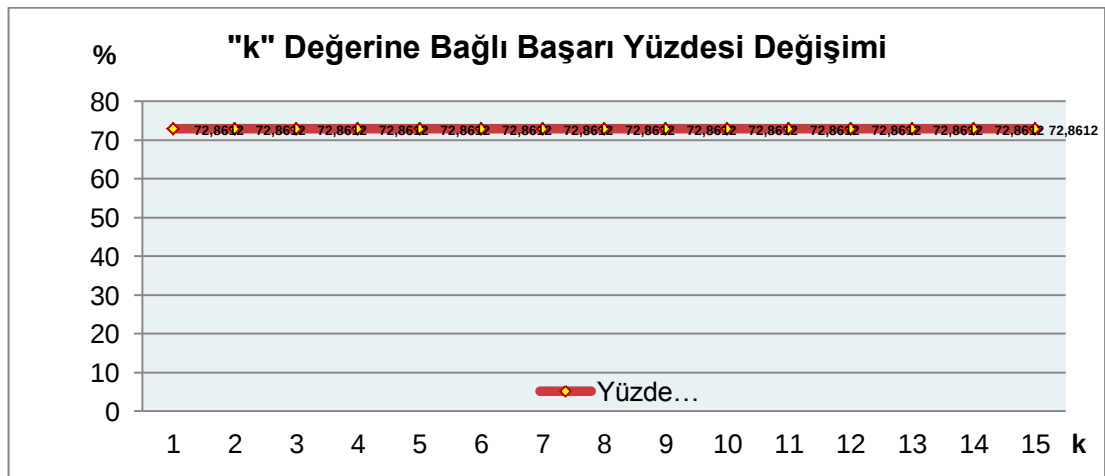
E-ticaret sitelerinin en az %40'ında ortak olarak geçen kelimeler dikkate alınarak elde edilen eğitim verisi üzerinde k-NN algoritması ikili ağırlıklandırma yönteminin uygulanmasında “k” değeri için “1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14 ve 15” değerleri denenmiştir. Algoritma başarı yüzdeleri Çizelge 3.36 ve Çizelge 3.37’de verilmiştir. Sonucun % 78,8612 ile 1-15 arasındaki tüm k değerleri için elde edildiği görülmüştür. “k” değerine bağlı algoritma başarı yüzdesi değişimi Şekil 3.46’dadır.

Çizelge 3.26. Dinamik çalışma k-NN ikili ağırlıklandırma sonucu (%40) (k=1...8)

Yüzde Değeri \ k	k=1	k=2	k=3	k=4	k=5	k=6	k=7	k=8
Doğru Sınıflandırma Yüzdesi (%)	72,8612	72,8612	72,8612	72,8612	72,8612	72,8612	72,8612	72,8612
Yanlış Sınıflandırma Yüzdesi (%)	27,1388	27,1388	27,1388	27,1388	27,1388	27,1388	27,1388	27,1388

Çizelge 3.27. Dinamik çalışma k-NN ikili ağırlıklandırma sonucu (%40) (k=9...15)

Yüzde Değeri \ k	k=9	k=10	k=11	k=12	k=13	k=14	k=15
Doğru Sınıflandırma Yüzdesi (%)	72,8612	72,8612	72,8612	72,8612	72,8612	72,8612	72,8612
Yanlış Sınıflandırma Yüzdesi (%)	27,1388	27,1388	27,1388	27,1388	27,1388	27,1388	27,1388



Şekil 3.46. “k” değerine bağlı başarı yüzdesi değişimi (%40) (İkili)

k-NN algoritması ikili ağırlıklandırma için sonuçların karşılaştırılması

Eğitim verisi üzerinde k-NN algoritması ikili ağırlıklandırma yöntemi için farklı oranların kullanılmasıyla elde edilen sonuçların karşılaştırılması Çizelge 3.28’de verilmiştir. En başarılı sonucun %91,8353 başarıyla %25 ve $k = 7$ olduğu durumda elde edildiği gözlemlenmiştir.

Çizelge 3.28. k-NN ikili için sonuçların karşılaştırılması

Ölçüt \ Yüzde Değeri	0%	5%	10%	15%	20%	25%	30%	35%	40%
Başarı Yüzdesi (%)	83,6312	87,108	88,7648	90,5660	90,8836	91,8353	90,7368	91,6783	78,8612
k Değeri	2	2	3	4	4	7	10	2,15	1-15

k-NN algoritması TF ağırlıklandırma yöntemi

Eğitim verisi üzerinde k-NN algoritması TF ağırlıklandırma yöntemi %0, %5, %10, %15, %20, %25, %30, %35 ve %40 değerleri ile uygulanmıştır.

%0 oranı ile elde edilen veri kümesi

E-ticaret sitelerinin en az %0’ında ortak olarak geçen kelimeler yani tüm kelimeler dikkate alınarak elde edilen eğitim verisi üzerinde k-NN algoritması TF ağırlıklandırma yönteminin uygulanmasında “k” değeri için “1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14 ve 15” değerleri denenmiştir. Algoritma başarı yüzdeleri Çizelge 3.29 ve Çizelge 3.30’da verilmiştir.

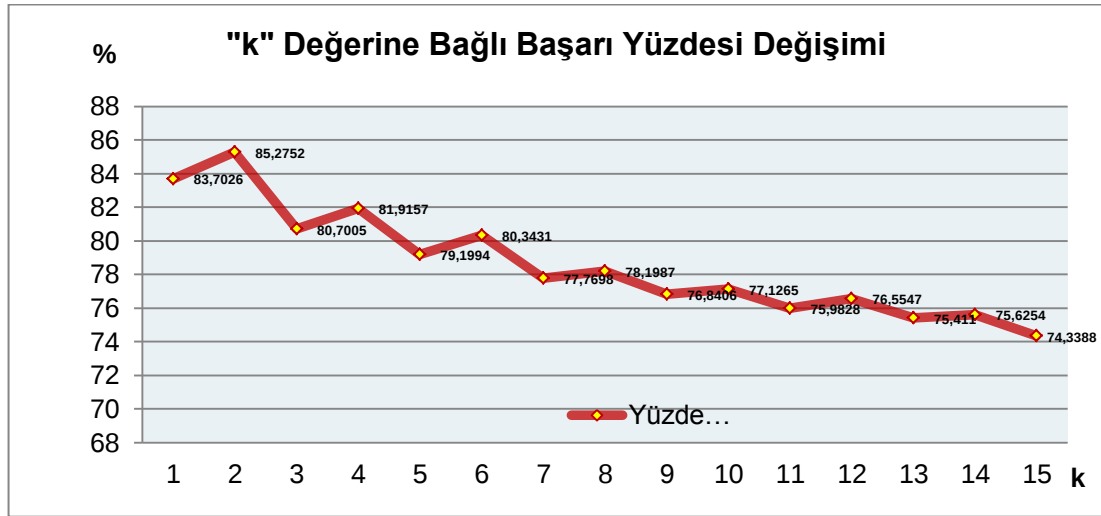
En başarılı sonucun % 85,2752 ile $k = 2$ olduğu durumda elde edildiği görülmüştür. “k” değerine bağlı olarak algoritma başarı yüzdesinin değişimi Şekil 3.47’de verilmiştir.

Çizelge 3.29. Dinamik çalışma k-NN TF ağırlıklandırma sonucu (%0) (k=1...8)

Yüzde Değeri \ k	k=1	k=2	k=3	k=4	k=5	k=6	k=7	k=8
Doğru Sınıflandırma Yüzdesi (%)	83,7026	85,2752	80,7005	81,9157	79,1994	80,3431	77,7698	78,1987
Yanlış Sınıflandırma Yüzdesi (%)	16,2974	14,7248	19,2995	18,0843	20,8006	19,6569	22,2302	21,8013

Çizelge 3.30. Dinamik çalışma k-NN TF ağırlıklandırma sonucu (%0) (k=9...15)

Yüzde Değeri \ k	k=9	k=10	k=11	k=12	k=13	k=14	k=15
Doğru Sınıflandırma Yüzdesi (%)	76,8406	77,1265	75,9828	76,5547	75,4110	75,6254	74,3388
Yanlış Sınıflandırma Yüzdesi (%)	23,1594	22,8735	24,0172	23,4453	24,5890	24,3746	25,6612



Şekil 3.47. “k” değerine bağlı başarı yüzdesi değişimi (%0) (TF)

%5 oranı ile elde edilen veri kümesi

E-ticaret sitelerinin en az %5’inde ortak olarak geçen kelimeler dikkate alınarak elde edilen eğitim verisi üzerinde k-NN algoritması TF ağırlıklandırma yönteminin uygulanmasında “k” değeri için “1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14 ve 15” değerleri denenmiştir. Algoritma başarı yüzdeleri Çizelge 3.31 ve Çizelge 3.32’de verilmiştir.

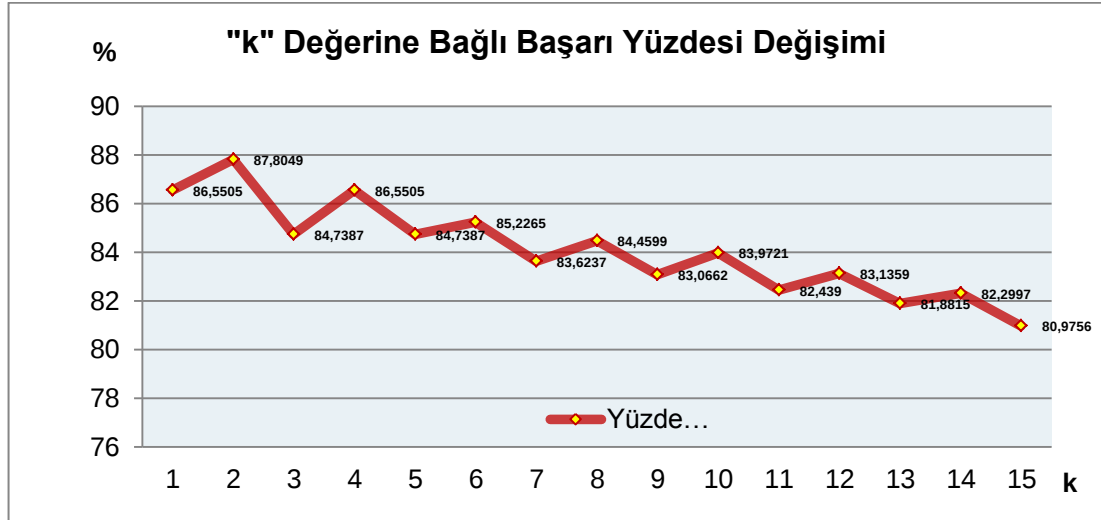
Çizelge 3.31. Dinamik çalışma k-NN TF ağırlıklandırma sonucu (%5) (k=1...8)

Yüzde Değeri \ k	k=1	k=2	k=3	k=4	k=5	k=6	k=7	k=8
Doğru Sınıflandırma Yüzdesi (%)	86,5505	87,8049	84,7387	86,5505	84,7387	85,2265	83,6237	84,4599
Yanlış Sınıflandırma Yüzdesi (%)	13,4495	12,1951	15,2613	13,4495	15,2613	14,7735	16,3763	15,5401

Çizelge 3.32. Dinamik çalışma k-NN TF ağırlıklandırma sonucu (%5) (k=9...15)

Yüzde Değeri \ k	k=9	k=10	k=11	k=12	k=13	k=14	k=15
Doğru Sınıflandırma Yüzdesi (%)	83,0662	83,9721	82,439	83,1359	81,8815	82,2997	80,9756
Yanlış Sınıflandırma Yüzdesi (%)	16,9338	16,0279	17,561	16,8641	18,1185	17,7003	19,0244

En başarılı sonucun % 87,8049 ile k = 2 olduğu durumda elde edildiği görülmüştür. “k” değerine bağlı olarak algoritma başarı yüzdesinin değişimi Şekil 3.48’de verilmiştir.



Şekil 3.48. “k” değerine bağlı başarı yüzdesi değişimi (%5) (TF)

%10 oranı ile elde edilen veri kümesi

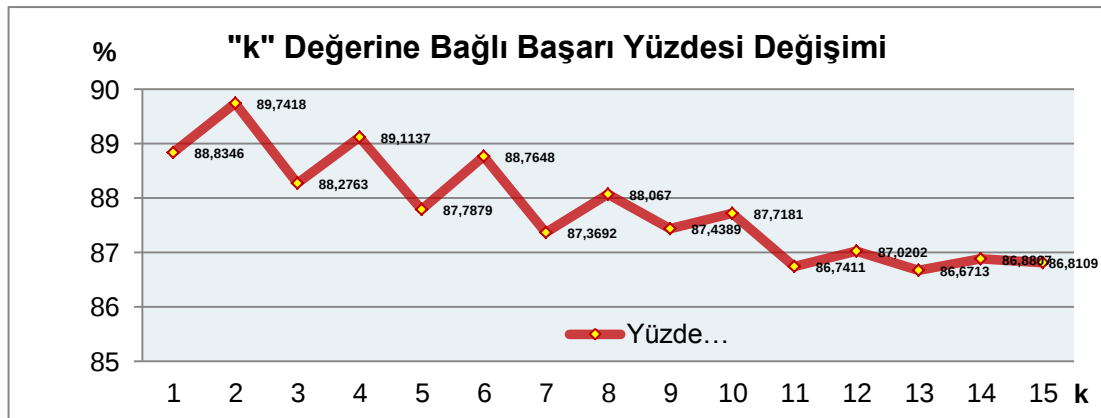
E-ticaret sitelerinin en az %10'unda ortak olarak geçen kelimeler dikkate alınarak elde edilen eğitim verisi üzerinde k-NN algoritması TF ağırlıklandırma yönteminin uygulanmasında “k” değeri için “1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14 ve 15” değerleri denenmiştir. Algoritma başarı yüzdeleri Çizelge 3.33 ve Çizelge 3.34'te verilmiştir. En başarılı sonucun % 89,1137 ile k = 4 olduğu durumda elde edildiği görülmüştür. “k” değerine bağlı olarak algoritma başarı yüzdesinin değişimi Şekil 3.49'da verilmiştir.

Çizelge 3.33. Dinamik çalışma k-NN TF ağırlıklandırma sonucu (%10) (k=1...8)

Yüzde Değeri \ k	k=1	k=2	k=3	k=4	k=5	k=6	k=7	k=8
Doğru Sınıflandırma Yüzdesi (%)	88,8346	89,7418	88,2763	89,1137	87,7879	88,7648	87,3692	88,0670
Yanlış Sınıflandırma Yüzdesi (%)	11,1654	10,2582	11,7237	10,8863	12,2121	11,2352	12,6308	11,9330

Çizelge 3.34. Dinamik çalışma k-NN TF ağırlıklandırma sonucu (%10) (k=9...15)

Yüzde Değeri \ k	k=9	k=10	k=11	k=12	k=13	k=14	k=15
Doğru Sınıflandırma Yüzdesi (%)	87,4389	87,7181	86,7411	87,0202	86,6713	86,8807	86,8109
Yanlış Sınıflandırma Yüzdesi (%)	12,5611	12,2819	13,2589	12,9798	13,3287	13,1193	13,1891



Şekil 3.49. “k” değerine bağlı başarı yüzdesi değişimi (%10) (TF)

%15 oranı ile elde edilen veri kümesi

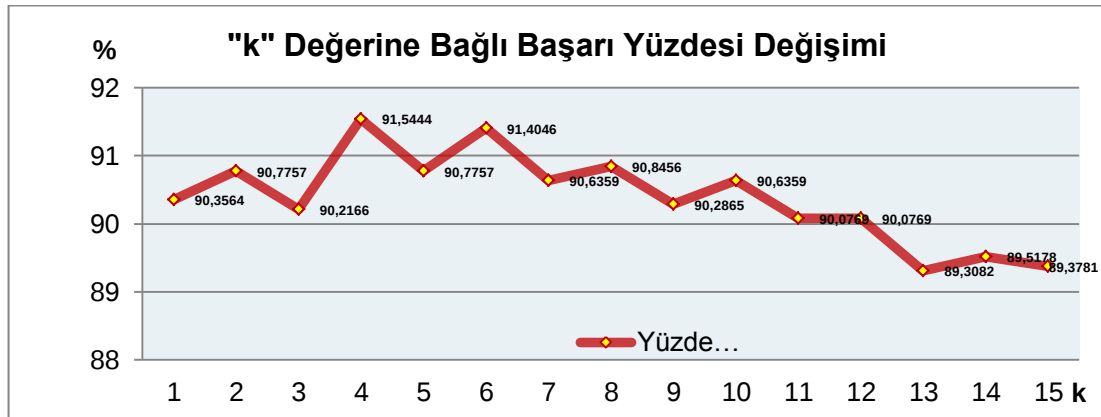
E-ticaret sitelerinin en az %15’inde ortak olarak geçen kelimeler dikkate alınarak elde edilen eğitim verisi üzerinde k-NN algoritması TF ağırlıklandırma yönteminin uygulanmasında “k” değeri için “1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14 ve 15” değerleri denenmiştir. Algoritma başarı yüzdeleri Çizelge 3.35 ve Çizelge 3.36’da verilmiştir. En başarılı sonucun % 91,5444 ile k = 4 olduğu durumda elde edildiği görülmüştür. “k” değerine bağlı olarak algoritma başarı yüzdesinin değişimi Şekil 50’de verilmiştir.

Çizelge 3.35. Dinamik çalışma k-NN TF ağırlıklandırma sonucu (%15) (k=1...8)

Yüzde Değeri \ k	k=1	k=2	k=3	k=4	k=5	k=6	k=7	k=8
Doğru Sınıflandırma Yüzdesi (%)	90,3564	90,7757	90,2166	91,5444	90,7757	91,4046	90,6359	90,8456
Yanlış Sınıflandırma Yüzdesi (%)	9,6436	9,2243	9,7834	8,4556	9,2243	8,5954	9,3641	9,1544

Çizelge 3.36. Dinamik çalışma k-NN TF ağırlıklandırma sonucu (%15) (k=9...15)

Yüzde Değeri \ k	k=9	k=10	k=11	k=12	k=13	k=14	k=15
Doğru Sınıflandırma Yüzdesi (%)	90,2865	90,6359	90,0769	90,0769	89,3082	89,5178	89,3781
Yanlış Sınıflandırma Yüzdesi (%)	9,7135	9,3641	9,9231	9,9231	10,6918	10,4822	10,6219



Şekil 3.50. “k” değerine bağlı başarı yüzdesi değişimi (%15) (TF)

%20 oranı ile elde edilen veri kümesi

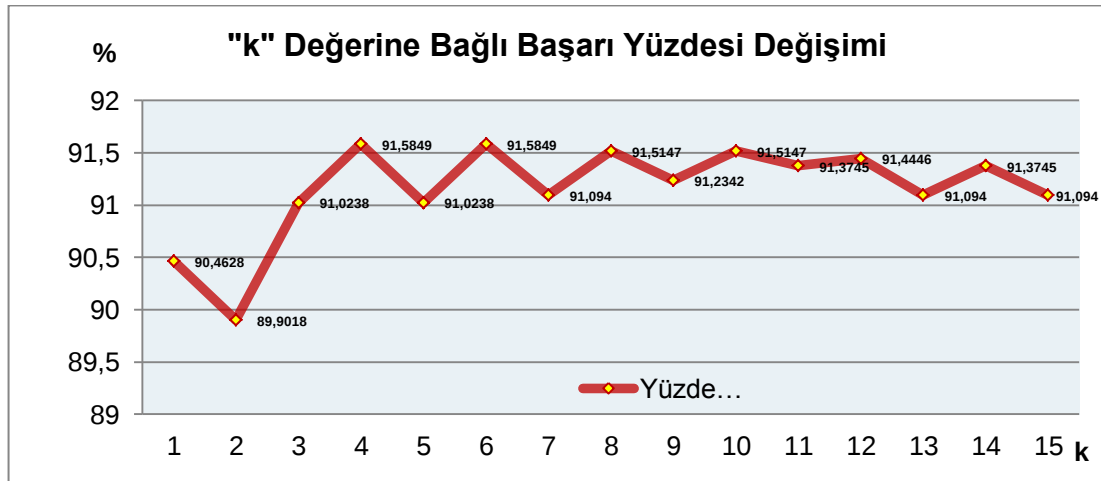
E-ticaret sitelerinin en az %20'sinde ortak olarak geçen kelimeler dikkate alınarak elde edilen eğitim verisi üzerinde k-NN algoritması TF ağırlıklandırma yönteminin uygulanmasında “k” değeri için “1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14 ve 15” değerleri denenmiştir. Algoritma başarı yüzdeleri Çizelge 3.37 ve Çizelge 3.38’dedir. En başarılı sonucun % 91,5849 ile k = 4 ve k = 6 olduğu durumda elde edildiği görülmüştür. “k” değerine bağlı algoritma başarı yüzdesi değişimi Şekil 3.51’dir.

Çizelge 3.37. Dinamik çalışma k-NN TF ağırlıklandırma sonucu (%20) (k=1...8)

k	k=1	k=2	k=3	k=4	k=5	k=6	k=7	k=8
Doğru Sınıflandırma Yüzdesi (%)	90,4628	89,9018	91,0238	91,5849	91,0238	91,5849	91,0940	91,5147
Yanlış Sınıflandırma Yüzdesi (%)	9,5372	10,0982	8,9762	8,4151	8,9762	8,4151	8,9060	8,4853

Çizelge 3.38. Dinamik çalışma k-NN TF ağırlıklandırma sonucu (%20) (k=9...15)

k	k=9	k=10	k=11	k=12	k=13	k=14	k=15
Doğru Sınıflandırma Yüzdesi (%)	91,2342	91,5147	91,3745	91,4446	91,0940	91,3745	91,0940
Yanlış Sınıflandırma Yüzdesi (%)	8,7658	8,4853	8,6255	8,5554	8,9060	8,6255	8,9060



Şekil 3.51. “k” değerine bağlı başarı yüzdesi değişimi (%20) (TF)

%25 oranı ile elde edilen veri kümesi

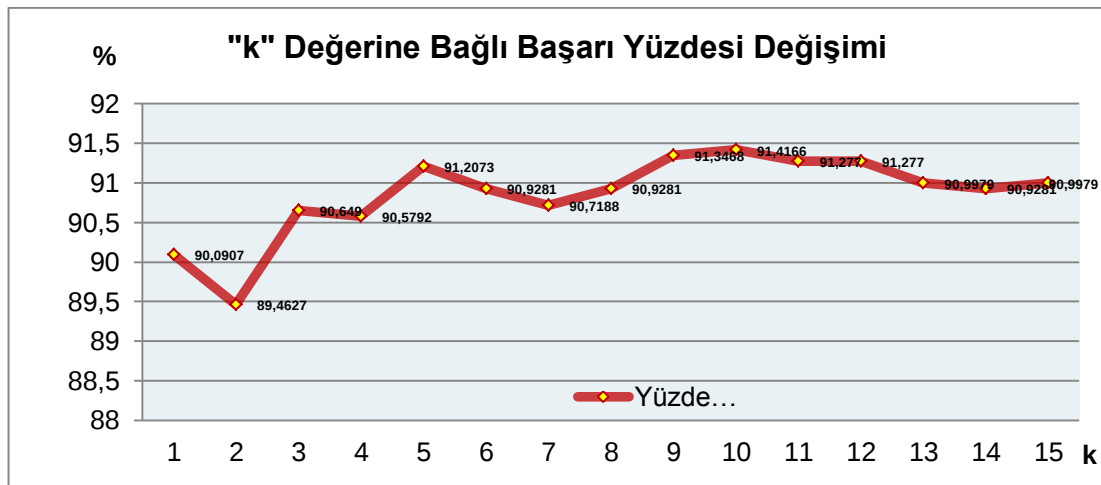
E-ticaret sitelerinin en az %25’inde ortak olarak geçen kelimeler dikkate alınarak elde edilen eğitim verisi üzerinde k-NN algoritması TF ağırlıklandırma yönteminin uygulanmasında “k” değeri için “1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14 ve 15” değerleri denenmiştir. Algoritma başarı yüzdeleri Çizelge 3.39 ve Çizelge 3.40’tadır. En başarılı sonucun % 91,4166 ile k = 10 olduğu durumda elde edildiği görülmüştür. “k” değerine bağlı algoritma başarı yüzdesinin değişimi Şekil 3.52’dedir..

Çizelge 3.39. Dinamik çalışma k-NN TF ağırlıklandırma sonucu (%25) (k=1...8)

Yüzde Değeri \ k	k=1	k=2	k=3	k=4	k=5	k=6	k=7	k=8
Doğru Sınıflandırma Yüzdesi (%)	90,0907	89,4627	90,6490	90,5792	91,2073	90,9281	90,7188	90,9281
Yanlış Sınıflandırma Yüzdesi (%)	9,9093	10,5373	9,3510	9,4208	8,7927	9,0719	9,2812	9,0719

Çizelge 3.40. Dinamik çalışma k-NN TF ağırlıklandırma sonucu (%25) (k=9...15)

Yüzde Değeri \ k	k=9	k=10	k=11	k=12	k=13	k=14	k=15
Doğru Sınıflandırma Yüzdesi (%)	91,3468	91,4166	91,2770	91,2770	90,9979	90,9281	90,9979
Yanlış Sınıflandırma Yüzdesi (%)	8,6532	8,5834	8,7230	8,7230	9,0021	9,0719	9,0021



Şekil 3.52. “k” değerine bağlı başarı yüzdesi değişimi (%25) (TF)

%30 oranı ile elde edilen veri kümesi

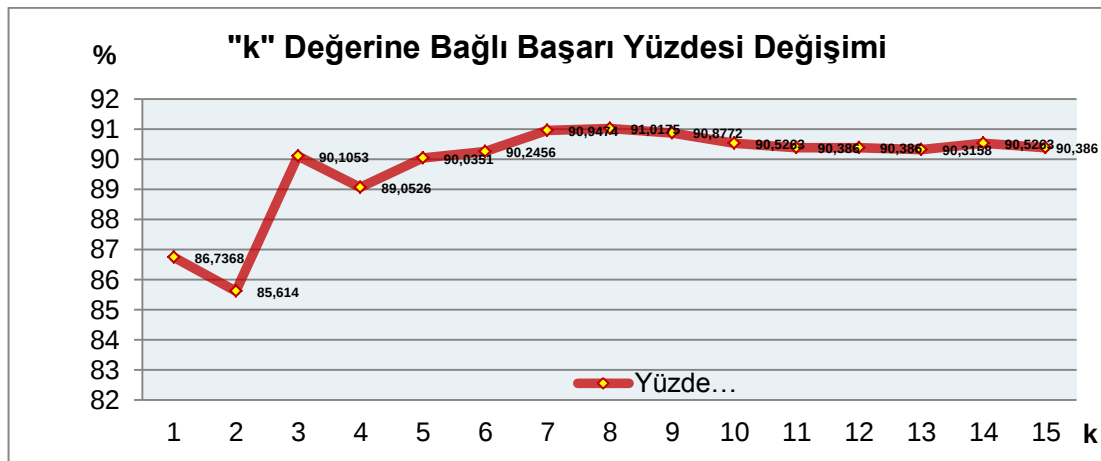
E-ticaret sitelerinin en az %30'unda ortak olarak geçen kelimeler dikkate alınarak elde edilen eğitim verisi üzerinde k-NN algoritması TF ağırlıklandırma yönteminin uygulanmasında “k” değeri için “1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14 ve 15” değerleri denenmiştir. Algoritma başarı yüzdeleri Çizelge 3.41 ve Çizelge 3.42’dedir. En başarılı sonucun % 91,0175 ile k = 8 olduğu durumda elde edildiği görülmüştür. “k” değerine bağlı olarak algoritma başarı yüzdesinin değişimi Şekil 3.53’tedir.

Çizelge 3.41. Dinamik çalışma k-NN TF ağırlıklandırma sonucu (%30) (k=1...8)

Yüzde Değeri \ k	k=1	k=2	k=3	k=4	k=5	k=6	k=7	k=8
Doğru Sınıflandırma Yüzdesi (%)	86,7368	85,6140	90,1053	89,0526	90,0351	90,2456	90,9474	91,0175
Yanlış Sınıflandırma Yüzdesi (%)	13,2632	14,3860	9,8947	10,9474	9,9649	9,7544	9,0526	8,9825

Çizelge 3.42. Dinamik çalışma k-NN TF ağırlıklandırma sonucu (%30) (k=9...15)

Yüzde Değeri \ k	k=9	k=10	k=11	k=12	k=13	k=14	k=15
Doğru Sınıflandırma Yüzdesi (%)	90,8772	90,5263	90,3860	90,3860	90,3158	90,5263	90,3860
Yanlış Sınıflandırma Yüzdesi (%)	9,1228	9,4737	9,6140	9,6140	9,6842	9,4737	9,6140



Şekil 3.53. “k” değerine bağlı başarı yüzdesi değişimi (%30) (TF)

%35 oranı ile elde edilen veri kümesi

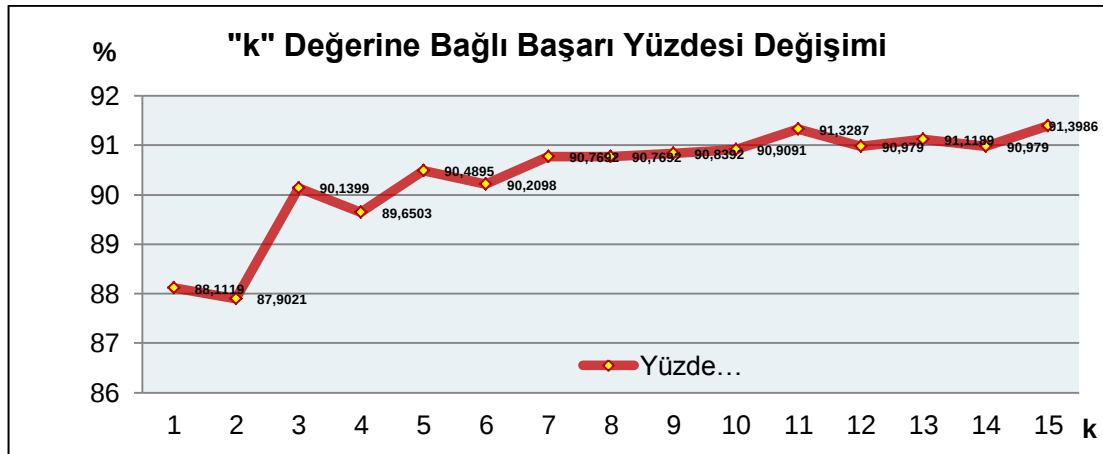
E-ticaret sitelerinin en az %35’inde ortak olarak geçen kelimeler dikkate alınarak elde edilen eğitim verisi üzerinde k-NN algoritması TF ağırlıklandırma yönteminin uygulanmasında “k” değeri için “1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14 ve 15” değerleri denenmiştir. Algoritma başarı yüzdeleri Çizelge 3.43 ve Çizelge 3.44’tedir. En başarılı sonucun % 91,3986 ile k = 15 olduğu durumda elde edildiği görülmüştür. “k” değerine bağlı olarak algoritma başarı yüzdesinin değişimi Şekil 3.54’tedir.

Çizelge 3.43. Dinamik çalışma k-NN TF ağırlıklandırma sonucu (%35) (k=1...8)

Yüzde Değeri \ k	k=1	k=2	k=3	k=4	k=5	k=6	k=7	k=8
Doğru Sınıflandırma Yüzdesi (%)	88,1119	87,9021	90,1399	89,6503	90,4895	90,2098	90,7692	90,7692
Yanlış Sınıflandırma Yüzdesi (%)	11,8881	12,0979	9,8601	10,3497	9,5105	9,7902	9,2308	9,2308

Çizelge 3.44. Dinamik çalışma k-NN TF ağırlıklandırma sonucu (%35) (k=9...15)

Yüzde Değeri \ k	k=9	k=10	k=11	k=12	k=13	k=14	k=15
Doğru Sınıflandırma Yüzdesi (%)	90,8392	90,9091	91,3287	90,9790	91,1189	90,9790	91,3986
Yanlış Sınıflandırma Yüzdesi (%)	9,1608	9,0909	8,6713	9,0210	8,8811	9,0210	8,6014



Şekil 3.54. “k” değerine bağlı başarı yüzdesi değişimi (%35) (TF)

%40 oranı ile elde edilen veri kümesi

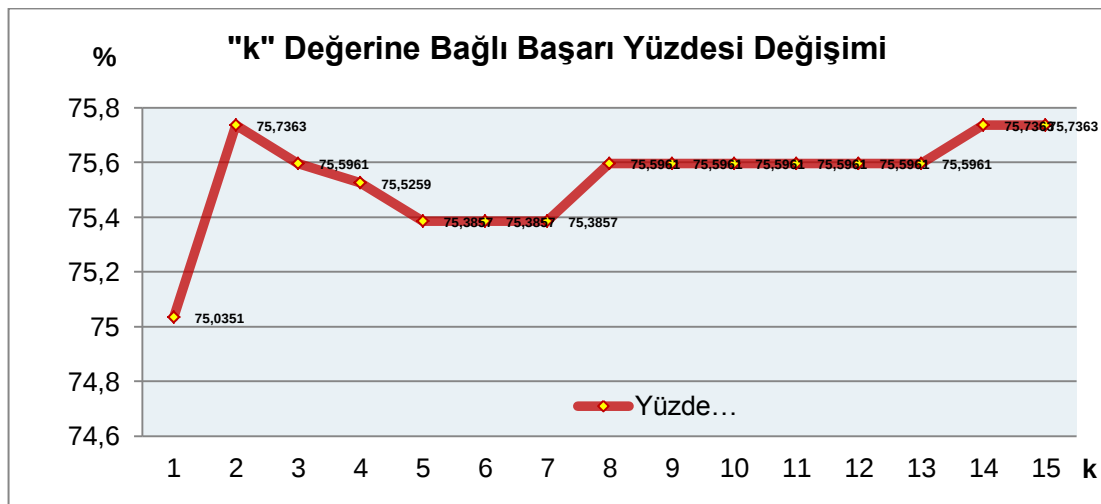
E-ticaret sitelerinin en az %40'ında ortak olarak geçen kelimeler dikkate alınarak elde edilen eğitim verisi üzerinde k-NN algoritması TF ağırlıklandırma yönteminin uygulanmasında “k” değeri için “1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14 ve 15” değerleri denenmiştir. Algoritma başarı yüzdeleri Çizelge 3.45 ve Çizelge 3.46'dadır. En başarılı sonucun % 75,7363 ile k = [2, 14 ve 15] olduğu durumlarda elde edildiği görülmüştür. “k” değerine bağlı algoritma başarı yüzdesi değişimi Şekil 3.55'tedir..

Çizelge 3.45. Dinamik çalışma k-NN TF ağırlıklandırma sonucu (%40) (k=1...8)

Yüzde Değeri \ k	k=1	k=2	k=3	k=4	k=5	k=6	k=7	k=8
Doğru Sınıflandırma Yüzdesi (%)	75,0351	75,7363	75,5961	75,5259	75,3857	75,3857	75,3857	75,5961
Yanlış Sınıflandırma Yüzdesi (%)	24,9649	24,2637	24,4039	24,4741	24,6143	24,6143	24,6143	24,4039

Çizelge 3.46. Dinamik çalışma k-NN TF ağırlıklandırma sonucu (%40) (k=9...15)

Yüzde Değeri \ k	k=9	k=10	k=11	k=12	k=13	k=14	k=15
Doğru Sınıflandırma Yüzdesi (%)	75,5961	75,5961	75,5961	75,5961	75,5961	75,7363	75,7363
Yanlış Sınıflandırma Yüzdesi (%)	24,4039	24,4039	24,4039	24,4039	24,4039	24,2637	24,2637



Şekil 3.55. “k” değerine bağlı başarı yüzdesi değişimi (%40) (TF)

k-NN algoritması TF ağırlıklandırma için sonuçların karşılaştırılması

Eğitim verisi üzerinde k-NN algoritması TF ağırlıklandırma yöntemi için farklı oranların kullanılmasıyla elde edilen sonuçların karşılaştırılması Çizelge 3.47’de verilmiştir. En başarılı sonucun %91,5849 başarıyla %20 ve k = 4 olduğu durumda elde edildiği gözlemlenmiştir.

Çizelge 3.47. k-NN TF ağırlıklandırma için sonuçların karşılaştırılması

Ölçüt \ Yüzde Değeri	0%	5%	10%	15%	20%	25%	30%	35%	40%
Başarı Yüzdesi (%)	85,2752	87,8049	89,1137	91,5444	91,5849	91,4166	91,0175	91,3986	75,7363
k Değeri	2	2	4	4	4	10	8	15	2,14,15

k-NN algoritması TF-IDF ağırlıklandırma yöntemi

Eğitim verisi üzerinde k-NN algoritması TF-IDF ağırlıklandırma yöntemi %0, %5, %10, %15, %20, %25, %30, %35 ve %40 değerleri ile uygulanmıştır.

%0 oranı ile elde edilen veri kümesi

E-ticaret sitelerinin en az %0’ında ortak olarak geçen kelimeler yani tüm kelimeler dikkate alınarak elde edilen eğitim verisi üzerinde k-NN algoritması TF-IDF ağırlıklandırma yönteminin uygulanmasında “k” değeri için “1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14 ve 15” değerleri denenmiştir. Algoritma başarı yüzdeleri Çizelge 3.48 ve Çizelge 3.49’da verilmiştir.

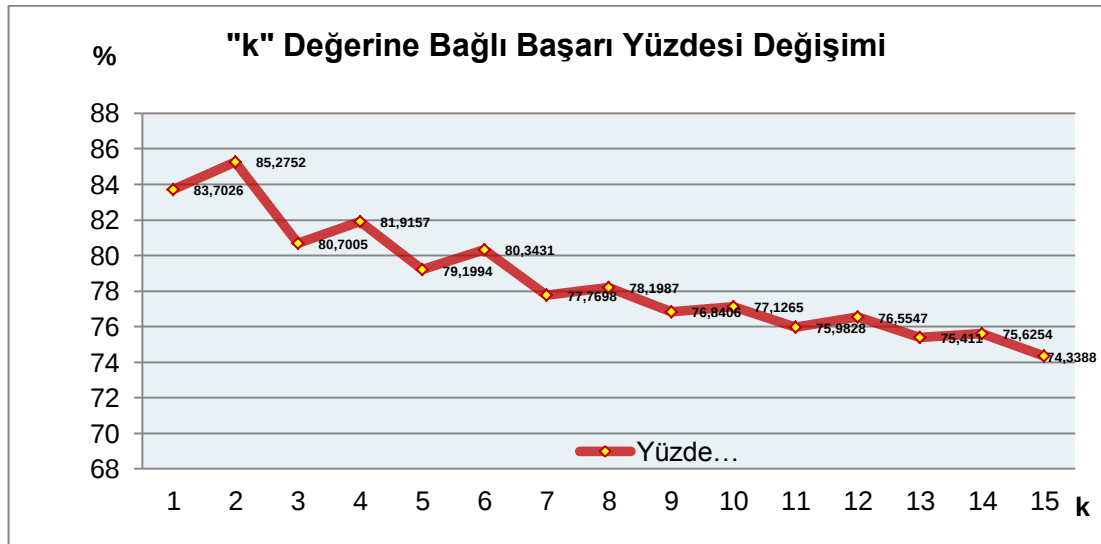
Çizelge 3.48. Dinamik çalışma k-NN TF-IDF ağırlıklandırma (%0) (k=1...8)

Yüzde Değeri \ k	k=1	k=2	k=3	k=4	k=5	k=6	k=7	k=8
Doğru Sınıflandırma Yüzdesi (%)	83,7026	85,2752	80,7005	81,9157	79,1994	80,3431	77,7698	78,1987
Yanlış Sınıflandırma Yüzdesi (%)	16,2974	14,7248	19,2995	18,0843	20,8006	19,6569	22,2302	21,8013

Çizelge 3.49. Dinamik çalışma k-NN TF-IDF ağırlıklandırma (%0) (k=9...15)

Yüzde Değeri \ k	k=9	k=10	k=11	k=12	k=13	k=14	k=15
Doğru Sınıflandırma Yüzdesi (%)	76,8406	77,1265	75,9828	76,5547	75,4110	75,6254	74,3388
Yanlış Sınıflandırma Yüzdesi (%)	23,1594	22,8735	24,0172	23,4453	24,5890	24,3746	25,6612

En başarılı sonucun % 85,2752 ile k = 2 olduğu durumda elde edildiği görülmüştür. “k” değerine bağlı olarak algoritma başarı yüzdesinin değişimi Şekil 3.56’da verilmiştir.



Şekil 3.56. “k” değerine bağlı başarı yüzdesi değişimi (%0) (TF-IDF)

%5 oranı ile elde edilen veri kümesi

E-ticaret sitelerinin en az %5’inde ortak olarak geçen kelimeler dikkate alınarak elde edilen eğitim verisi üzerinde k-NN algoritması TF-IDF ağırlıklandırma yönteminin uygulanmasında “k” değeri için “1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14 ve 15” değerleri denenmiştir. Algoritma başarı yüzdesi Çizelge 3.50 ve Çizelge 3.51’de verilmiştir.

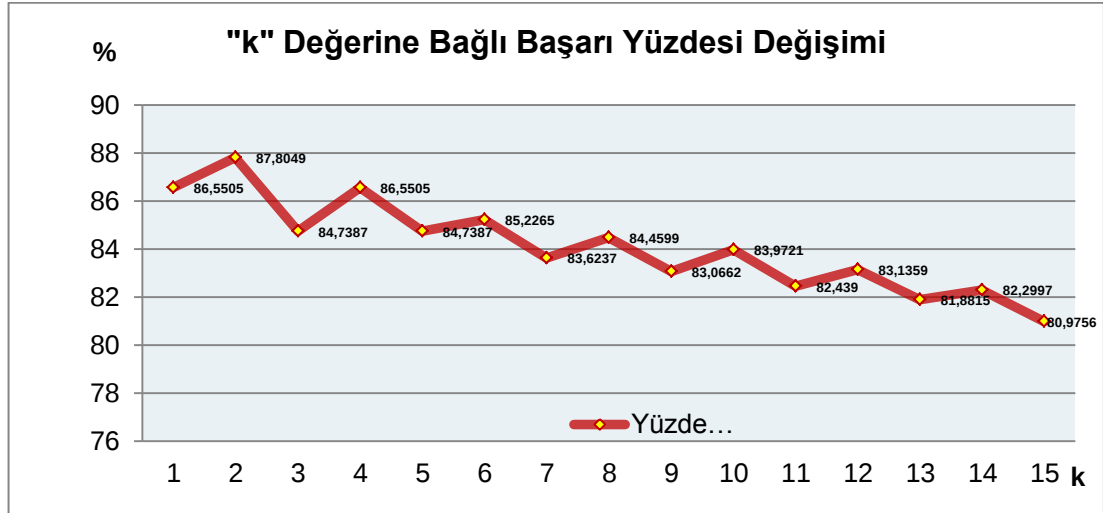
Çizelge 3.50. Dinamik çalışma k-NN TF-IDF ağırlıklandırma (%5) (k=1...8)

Yüzde Değeri \ k	k=1	k=2	k=3	k=4	k=5	k=6	k=7	k=8
Doğru Sınıflandırma Yüzdesi (%)	86,5505	87,8049	84,7387	86,5505	84,7387	85,2265	83,6237	84,4599
Yanlış Sınıflandırma Yüzdesi (%)	13, 495	12,1951	15,2613	13,4495	15,2613	14,7735	16,3763	15,5401

Çizelge 3.51. Dinamik çalışma k-NN TF-IDF ağırlıklandırma (%5) (k=9...15)

Yüzde Değeri \ k	k=9	k=10	k=11	k=12	k=13	k=14	k=15
Doğru Sınıflandırma Yüzdesi (%)	83,0662	83,9721	82,439	83,1359	81,8815	82,2997	80,9756
Yanlış Sınıflandırma Yüzdesi (%)	16,9338	16,0279	17,561	16,8641	18,1185	17,7003	19,0244

En başarılı sonucun % 87,8049 ile k = 2 olduğu durumda elde edildiği görülmüştür. “k” değerine bağlı olarak algoritma başarı yüzdesinin değişimi Şekil 3.57’de verilmiştir.



Şekil 3.57. “k” değerine bağlı başarı yüzdesi değişimi (%5) (TF-IDF)

%10 oranı ile elde edilen veri kümesi

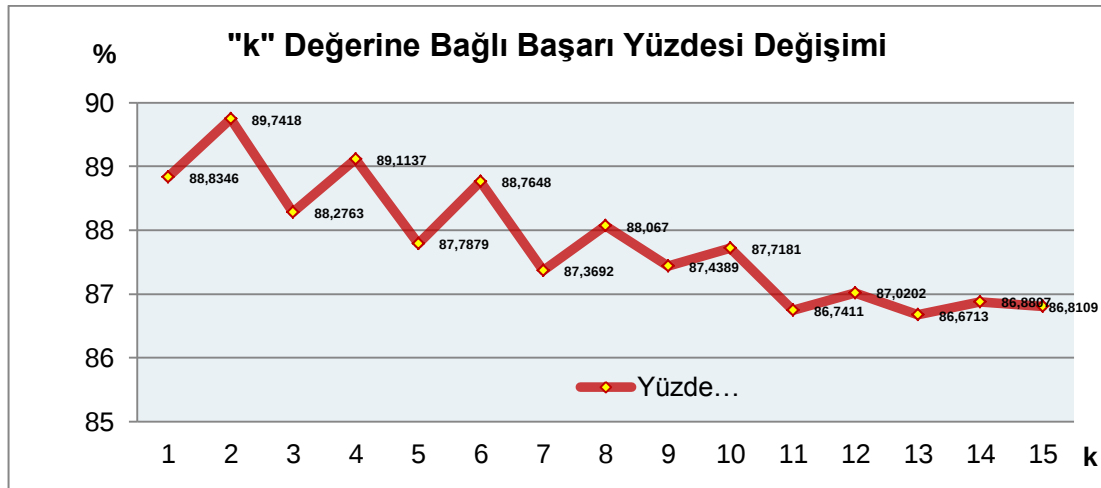
E-ticaret sitelerinin en az %10'unda ortak olarak geçen kelimeler dikkate alınarak elde edilen eğitim verisi üzerinde k-NN algoritması TF-IDF ağırlıklandırma yönteminin uygulanmasında “k” değeri için “1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14 ve 15” değerleri denenmiştir. Algoritma başarı yüzdeleri Çizelge 3.52 ve Çizelge 3.53'dedir. En başarılı sonucun % 89,1137 ile k = 4 olduğu durumda elde edildiği görülmüştür. “k” değerine bağlı algoritma başarı yüzdesi değişimi Şekil 3.58'dedir.

Çizelge 3.52. Dinamik çalışma k-NN TF-IDF ağırlıklandırma (%10) (k=1...8)

k	k=1	k=2	k=3	k=4	k=5	k=6	k=7	k=8
Doğru Sınıflandırma Yüzdesi (%)	88,8346	89,7418	88,2763	89,1137	87,7879	88,7648	87,3692	88,0670
Yanlış Sınıflandırma Yüzdesi (%)	11,1654	10,2582	11,7237	10,8863	12,2121	11,2352	12,6308	11,9330

Çizelge 3.53. Dinamik çalışma k-NN TF-IDF ağırlıklandırma (%10) (k=9...15)

k	k=9	k=10	k=11	k=12	k=13	k=14	k=15
Doğru Sınıflandırma Yüzdesi (%)	87,4389	87,7181	86,7411	87,0202	86,6713	86,8807	86,8109
Yanlış Sınıflandırma Yüzdesi (%)	12,5611	12,2819	13,2589	12,9798	13,3287	13,1193	13,1891



Şekil 3.58. “k” değerine bağlı başarı yüzdesi değişimi (%10) (TF-IDF)

%15 oranı ile elde edilen veri kümesi

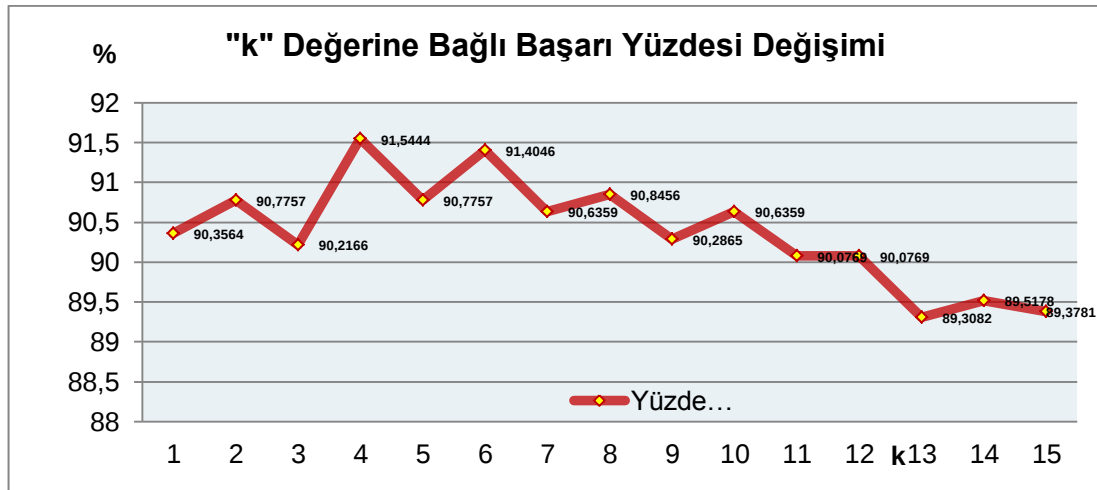
E-ticaret sitelerinin en az %15’inde ortak olarak geçen kelimeler dikkate alınarak elde edilen eğitim verisi üzerinde k-NN algoritması TF-IDF ağırlıklandırma yönteminin uygulanmasında “k” değeri için “1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14 ve 15” değerleri denenmiştir. Algoritma başarı yüzdeleri Çizelge 3.54 ve Çizelge 3.55’tedir. En başarılı sonucun % 91,5444 ile k = 4 olduğu durumda elde edildiği görülmüştür. “k” değerine bağlı algoritma başarı yüzdesi değişimi Şekil 3.59’dadır.

Çizelge 3.54. Dinamik çalışma k-NN TF-IDF ağırlıklandırma (%15) (k=1...8)

Yüzde Değeri \ k	k=1	k=2	k=3	k=4	k=5	k=6	k=7	k=8
Doğru Sınıflandırma Yüzdesi (%)	90,3564	90,7757	90,2166	91,5444	90,7757	91,4046	90,6359	90,8456
Yanlış Sınıflandırma Yüzdesi (%)	9,6436	9,2243	9,7834	8,4556	9,2243	8,5954	9,3641	9,1544

Çizelge 3.55. Dinamik çalışma k-NN TF-IDF ağırlıklandırma (%15) (k=9...15)

Yüzde Değeri \ k	k=9	k=10	k=11	k=12	k=13	k=14	k=15
Doğru Sınıflandırma Yüzdesi (%)	90,2865	90,6359	90,0769	90,0769	89,3082	89,5178	89,3781
Yanlış Sınıflandırma Yüzdesi (%)	9,7135	9,3641	9,9231	9,9231	10,6918	10,4822	10,6219



Şekil 3.59. “k” değerine bağlı başarı yüzdesi değişimi (%15) (TF-IDF)

%20 oranı ile elde edilen veri kümesi

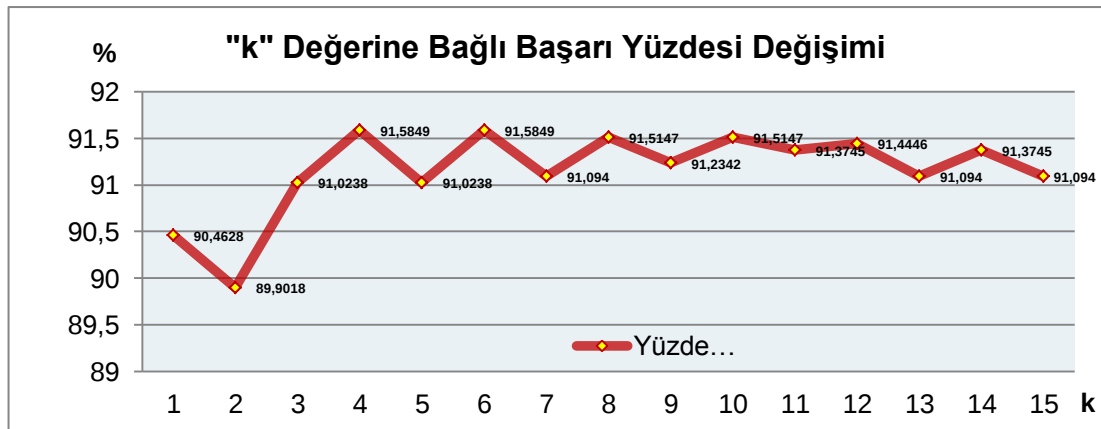
E-ticaret sitelerinin en az %20'sinde ortak olarak geçen kelimeler dikkate alınarak elde edilen eğitim verisi üzerinde k-NN algoritması TF-IDF ağırlıklandırma yönteminin uygulanmasında “k” değeri için “1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14 ve 15” değerleri denenmiştir. Algoritma başarı yüzdeleri Çizelge 3.56 ve Çizelge 3.57’de verilmiştir. En başarılı sonucun % 91,5849 ile k = 4 ve k = 6 olduğu durumda elde edildiği görülmüştür. “k” değerine bağlı olarak algoritma başarı yüzdesinin değişimi Şekil 3.60’ta verilmiştir.

Çizelge 3.56. Dinamik çalışma k-NN TF-IDF ağırlıklandırma (%20) (k=1...8)

Yüzde Değeri \ k	k=1	k=2	k=3	k=4	k=5	k=6	k=7	k=8
Doğru Sınıflandırma Yüzdesi (%)	90,4628	89,9018	91,0238	91,5849	91,0238	91,5849	91,0940	91,5147
Yanlış Sınıflandırma Yüzdesi (%)	9,5372	10,0982	8,9762	8,4151	8,9762	8,4151	8,9060	8,4853

Çizelge 3.57. Dinamik çalışma k-NN TF-IDF ağırlıklandırma (%20) (k=9...15)

Yüzde Değeri \ k	k=9	k=10	k=11	k=12	k=13	k=14	k=15
Doğru Sınıflandırma Yüzdesi (%)	91,2342	91,5147	91,3745	91,4446	91,0940	91,3745	91,0940
Yanlış Sınıflandırma Yüzdesi (%)	8,7658	8,4853	8,6255	8,5554	8,9060	8,6255	8,9060



Şekil 3.60. “k” değerine bağlı başarı yüzdesi değişimi (%20) (TF-IDF)

%25 oranı ile elde edilen veri kümesi

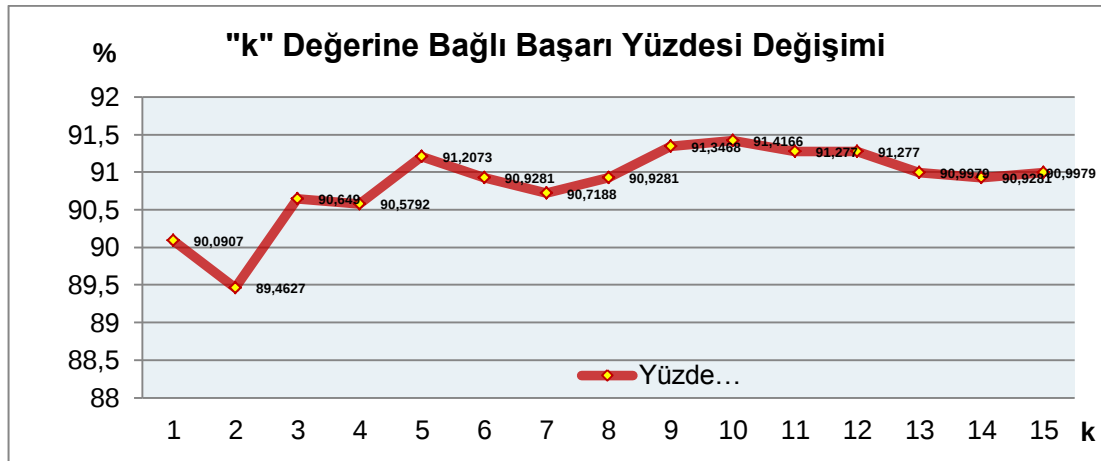
E-ticaret sitelerinin en az %25’inde ortak olarak geçen kelimeler dikkate alınarak elde edilen eğitim verisi üzerinde k-NN algoritması TF-IDF ağırlıklandırma yönteminin uygulanmasında “k” değeri için “1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14 ve 15” değerleri denenmiştir. Algoritma başarı yüzdeleri Çizelge 3.58 ve Çizelge 3.59’dadır. En başarılı sonucun % 91,4166 ile k = 10 olduğu durumda elde edildiği görülmüştür. “k” değerine bağlı algoritma başarı yüzdesi değişimi Şekil 3.61’dedir

Çizelge 3.58. Dinamik çalışma k-NN TF-IDF ağırlıklandırma (%25) (k=1...8)

k	k=1	k=2	k=3	k=4	k=5	k=6	k=7	k=8
Doğru Sınıflandırma Yüzdesi (%)	90,0907	89,4627	90,6490	90,5792	91,2073	90,9281	90,7188	90,9281
Yanlış Sınıflandırma Yüzdesi (%)	9,9093	10,5373	9,3510	9,4208	8,7927	9,0719	9,2812	9,0719

Çizelge 3.59. Dinamik çalışma k-NN TF-IDF ağırlıklandırma (%25) (k=9...15)

k	k=9	k=10	k=11	k=12	k=13	k=14	k=15
Doğru Sınıflandırma Yüzdesi (%)	91,3468	91,4166	91,2770	91,2770	90,9979	90,9281	90,9979
Yanlış Sınıflandırma Yüzdesi (%)	8,6532	8,5834	8,7230	8,7230	9,0021	9,0719	9,0021



Şekil 3.61. “k” değerine bağlı başarı yüzdesi değişimi (%25) (TF-IDF)

%30 oranı ile elde edilen veri kümesi

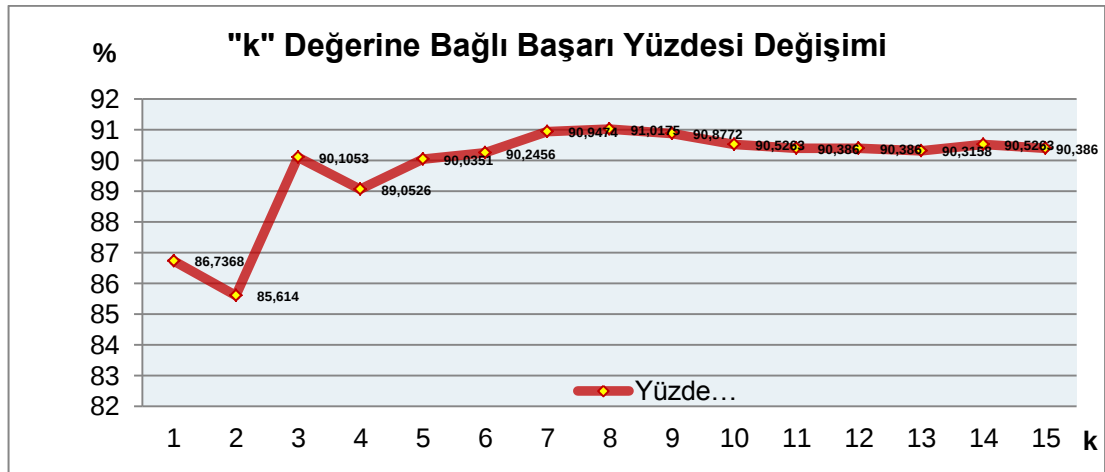
E-ticaret sitelerinin en az %30'unda ortak olarak geçen kelimeler dikkate alınarak elde edilen eğitim verisi üzerinde k-NN algoritması TF-IDF ağırlıklandırma yönteminin uygulanmasında “k” değeri için “1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14 ve 15” değerleri denenmiştir. Algoritma başarı yüzdeleri Çizelge 3.60 ve Çizelge 3.61’dedir. En başarılı sonucun % 91,0175 ile k = 8 olduğu durumda elde edildiği görülmüştür. “k” değerine bağlı algoritma başarı yüzdesi değişimi Şekil 3.62’dedir.

Çizelge 3.60. Dinamik çalışma k-NN TF-IDF ağırlıklandırma (%30) (k=1...8)

Yüzde Değeri \ k	k=1	k=2	k=3	k=4	k=5	k=6	k=7	k=8
Doğru Sınıflandırma Yüzdesi (%)	86,7368	85,6140	90,1053	89,0526	90,0351	90,2456	90,9474	91,0175
Yanlış Sınıflandırma Yüzdesi (%)	13,2632	14,3860	9,8947	10,9474	9,9649	9,7544	9,0526	8,9825

Çizelge 3.61. Dinamik çalışma k-NN TF-IDF ağırlıklandırma (%30) (k=9...15)

Yüzde Değeri \ k	k=9	k=10	k=11	k=12	k=13	k=14	k=15
Doğru Sınıflandırma Yüzdesi (%)	90,8772	90,5263	90,3860	90,3860	90,3158	90,5263	90,3860
Yanlış Sınıflandırma Yüzdesi (%)	9,1228	9,4737	9,6140	9,6140	9,6842	9,4737	9,6140



Şekil 3.62. “k” değerine bağlı başarı yüzdesi değişimi (%30) (TF-IDF)

%35 oranı ile elde edilen veri kümesi

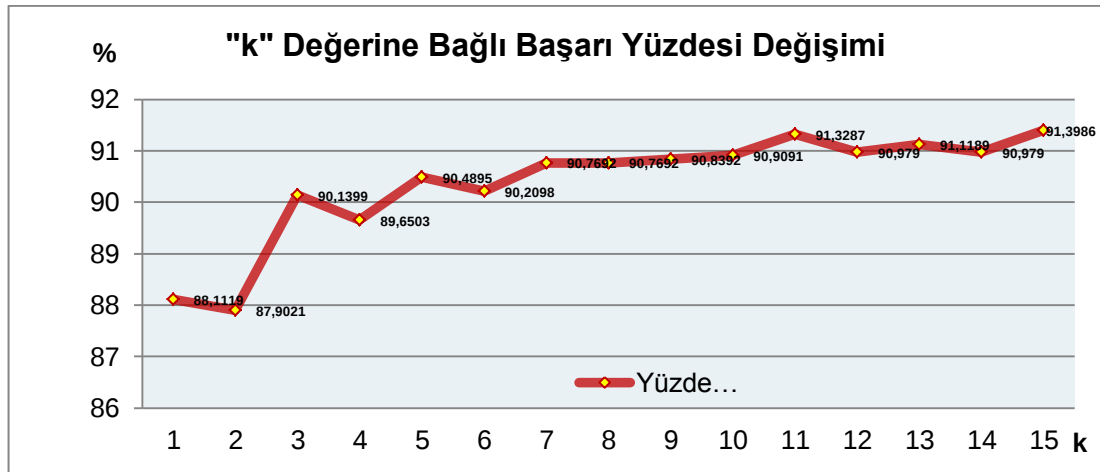
E-ticaret sitelerinin en az %35’inde ortak olarak geçen kelimeler dikkate alınarak elde edilen eğitim verisi üzerinde k-NN algoritması TF-IDF ağırlıklandırma yönteminin uygulanmasında “k” değeri için “1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14 ve 15” değerleri denenmiştir. Algoritma başarı yüzdeleri Çizelge 3.62 ve Çizelge 3.63’tedir. En başarılı sonucun % 91,3986 ile k = 15 olduğu durumda elde edildiği görülmüştür. “k” değerine bağlı algoritma başarı yüzdesi değişimi Şekil 3.63’tedir.

Çizelge 3.62. Dinamik çalışma k-NN TF-IDF ağırlıklandırma (%35) (k=1...8)

Yüzde Değeri \ k	k=1	k=2	k=3	k=4	k=5	k=6	k=7	k=8
Doğru Sınıflandırma Yüzdesi (%)	88,1119	87,9021	90,1399	89,6503	90,4895	90,2098	90,7692	90,7692
Yanlış Sınıflandırma Yüzdesi (%)	11,8881	12,0979	9,8601	10,3497	9,5105	9,7902	9,2308	9,2308

Çizelge 3.63. Dinamik çalışma k-NN TF-IDF ağırlıklandırma (%35) (k=9...15)

Yüzde Değeri \ k	k=9	k=10	k=11	k=12	k=13	k=14	k=15
Doğru Sınıflandırma Yüzdesi (%)	90,8392	90,9091	91,3287	90,9790	91,1189	90,9790	91,3986
Yanlış Sınıflandırma Yüzdesi (%)	9,1608	9,0909	8,6713	9,0210	8,8811	9,0210	8,6014



Şekil 3.63. “k” değerine bağlı başarı yüzdesi değişimi (%35) (TF-IDF)

%40 oranı ile elde edilen veri kümesi

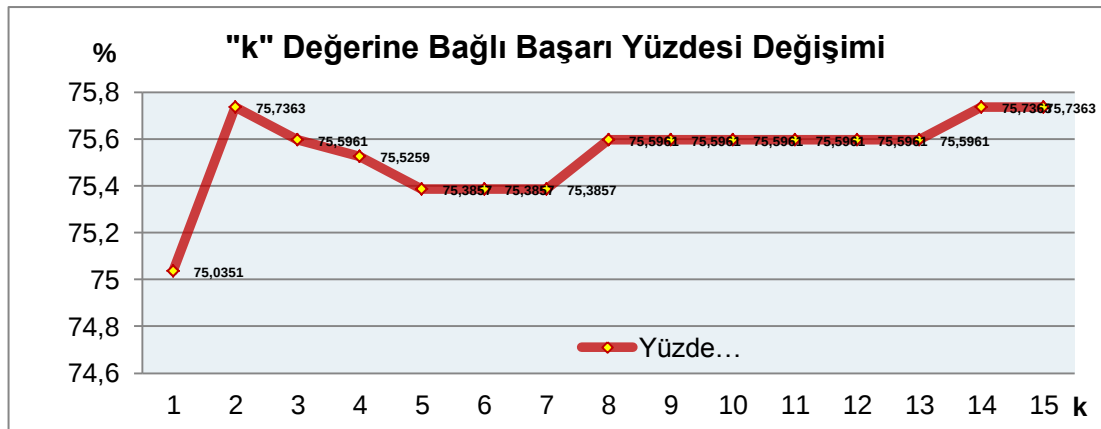
E-ticaret sitelerinin en az %40'ında ortak olarak geçen kelimeler dikkate alınarak elde edilen eğitim verisi üzerinde k-NN algoritması TF-IDF ağırlıklandırma yönteminin uygulanmasında “k” değeri için “1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14 ve 15” değerleri denenmiştir. Algoritma başarı yüzdeleri Çizelge 3.64 ve Çizelge 3.65'te verilmiştir. En başarılı sonucun % 75,7363 ile k = [2, 14 ve 15] olduğu durumlarda elde edildiği görülmüştür. “k” değerine bağlı olarak algoritma başarı yüzdesinin değişimi Şekil 3.64'te verilmiştir.

Çizelge 3.64. Dinamik çalışma k-NN TF-IDF ağırlıklandırma (%40) (k=1...8)

Yüzde Değeri \ k	k=1	k=2	k=3	k=4	k=5	k=6	k=7	k=8
Doğru Sınıflandırma Yüzdesi (%)	75,0351	75,7363	75,5961	75,5259	75,3857	75,3857	75,3857	75,5961
Yanlış Sınıflandırma Yüzdesi (%)	24,9649	24,2637	24,4039	24,4741	24,6143	24,6143	24,6143	24,4039

Çizelge 3.65. Dinamik çalışma k-NN TF-IDF ağırlıklandırma (%40) (k=9...15)

Yüzde Değeri \ k	k=9	k=10	k=11	k=12	k=13	k=14	k=15
Doğru Sınıflandırma Yüzdesi (%)	75,5961	75,5961	75,5961	75,5961	75,5961	75,7363	75,7363
Yanlış Sınıflandırma Yüzdesi (%)	24,4039	24,4039	24,4039	24,4039	24,4039	24,2637	24,2637



Şekil 3.64. “k” değerine bağlı başarı yüzdesi değişimi (%40) (TF-IDF)

k-NN algoritması TF-IDF ağırlıklandırma için sonuçların karşılaştırılması

Eğitim verisi üzerinde k-NN algoritması TF-IDF ağırlıklandırma yöntemi için farklı oranların kullanılmasıyla elde edilen sonuçların karşılaştırılması Çizelge 3.66'ta verilmiştir. En başarılı sonucun %91,5849 başarıyla %20 ve k = 4 olduğu durumda elde edildiği gözlemlenmiştir.

Çizelge 3.66. k-NN TF-IDF ağırlıklandırma için sonuçların karşılaştırılması

Ölçüt \ Yüzde Değeri	0%	5%	10%	15%	20%	25%	30%	35%	40%
Başarı Yüzdesi (%)	85,2752	87,8049	89,1137	91,5444	91,5849	91,4166	91,0175	91,3986	75,7363
k Değeri	2	2	4	4	4	10	8	15	2,14,15

Naïve Bayes algoritması ikili ağırlıklandırma (Multi-variate model)

E-ticaret sitelerinin en az %0, %5, %10, %15, %20, %25, %30, %35, %40'ında ortak olarak geçen kelimelerin anahtar sözlüğü oluşturmada kullanılması ile elde edilen eğitim kümesi üzerinde Naïve Bayes algoritması ikili ağırlıklandırma yöntemi kullanılmıştır. Eğitim kümesinin vektör uzay modeline ait kesitler verilmişti (Bkz. Şekil 3.8-3.16). Naïve Bayes başarı yüzdeleri Çizelge 3.67'de verilmiştir. En başarılı sonucun % 90,7188 ile e-ticaret sitelerinin en az %25'inde ortak olarak geçen kelimelerin dikkate alınmasıyla oluşturulan eğitim kümesi kullanıldığında elde edildiği görülmüştür.

Çizelge 3.67. Dinamik çalışma Naïve Bayes ikili ağırlıklandırma yöntemi sonucu

Kelime Yüzdesi \ Algoritma Yüzdesi	%0	%5	%10	%15	%20	%25	%30	%35	%40
Naïve Bayes (ikili) Doğru Sınıflandırma Yüzdesi (%)	85,2037	86,5505	88,2066	90,3564	89,9719	90,7188	88,8421	90,4196	72,8612
Naïve Bayes (ikili) Yanlış Sınıflandırma Yüzdesi (%)	14,7963	13,4495	11,7934	9,6436	10,0281	9,2812	11,1579	9,5804	27,1388

Naïve Bayes algoritması TF ağırlıklandırma (Multinomial model)

E-ticaret sitelerinin en az %0, %5, %10, %15, %20, %25, %30, %35, %40'ında ortak olarak geçen kelimelerin anahtar sözlüğü oluşturmada kullanılması ile elde edilen eğitim kümesi üzerinde Naïve Bayes algoritması TF ağırlıklandırma yöntemi kullanılmıştır. Eğitim kümesinin vektör uzay modeline ait kesitler verilmişti (Bkz. Şekil 3.17-3.25). Naïve Bayes başarı yüzdeleri Çizelge 3.68'da verilmiştir. En başarılı sonucun % 90,7433 ile e-ticaret sitelerinin en az %20'sinde ortak olarak geçen kelimelerin dikkate alınmasıyla oluşturulan eğitim kümesi kullanıldığında elde edildiği görülmüştür.

Çizelge 3.68. Dinamik çalışma Naïve Bayes TF ağırlıklandırma yöntemi sonucu

Algoritma Yüzdesi \ Kelime Yüzdesi	%0	%5	%10	%15	%20	%25	%30	%35	%40
Naïve Bayes (TF) Doğru Sınıflandırma Yüzdesi (%)	88,1344	88,7805	90,3001	90,3564	90,7433	89,4627	88,2807	88,7413	52,8050
Naïve Bayes (TF) Yanlış Sınıflandırma Yüzdesi (%)	11,8656	11,2195	9,6999	9,6436	9,2567	10,5373	11,7193	11,2587	47,1950

3.6. Statik ve Dinamik Çalışma Sonuçlarının Karşılaştırılması

Statik çalışmada uygulanan yöntemlerden k-NN algoritması için ikili (k=3), TF (k=4) ve TF-IDF (k=4) ile Naïve Bayes algoritması için ikili ve TF ağırlıklandırma yöntemi sonuçları Çizelge 3.69'da karşılaştırılmıştır. En başarılı sonuç k-NN algoritması TF ve TF-IDF ağırlıklandırma yöntemleri ile %92,2790 olarak elde edilmiştir. Dinamik çalışmada uygulanan yöntemlerden k-NN algoritması için ikili (k=6,7), TF (k=4) ve TF-IDF (k=4) ile Naïve Bayes algoritması için ikili ve TF ağırlıklandırma yöntemi sonuçları Çizelge 3.70'te karşılaştırılmıştır. En başarılı sonuca k-NN algoritması ikili ağırlıklandırma yönteminin %25 oranı ile elde edilen veri üzerinde uygulanması ile %91,8353 olarak k = 6 ya da k = 7 olduğu durumda ulaşılmıştır.

Çizelge 3.69. Statik çalışma algoritma sonuçlarının karşılaştırılması

Algoritma Ölçüt	k-NN İkili (k=3)	k-NN TF (k=4)	k-NN TF-IDF (k=4)	Naïve Bayes İkili	Naïve Bayes TF
Doğru Sınıflandırma Yüzdesi (%)	91,4643	92,2970	92,2970	90,9091	90,7009
Yanlış Sınıflandırma Yüzdesi (%)	8,5357	8,1888	7,7030	9,0909	9,2991
Doğru Sınıflandırma Sayısı	1318	1330	1330	1310	1307
Yanlış Sınıflandırma Sayısı	123	111	111	131	134
TP Oranı	0,819	0,923	0,923	0,909	0,907
FP Oranı	0,160	0,079	0,079	0,089	0,095
Kesinlik	0,848	0,923	0,923	0,910	0,907
Hassasiyet	0,819	0,923	0,923	0,909	0,907
F-Ölçütü	0,818	0,923	0,923	0,909	0,907

Çizelge 3.70. Dinamik çalışma algoritma sonuçlarının karşılaştırılması

Algoritma Ölçüt	k-NN İkili (k=6,7) (%25)	k-NN TF (k=4) (%20)	k-NN TF-IDF (k=4) (%20)	Naïve Bayes İkili (%25)	Naïve Bayes TF (%20)
Doğru Sınıflandırma Yüzdesi (%)	91,8353	91,5849	91,5849	90,7188	90,7433
Yanlış Sınıflandırma Yüzdesi (%)	8,1647	8,4151	8,4151	9,2812	9,2567
Doğru Sınıflandırma Sayısı	1316	1306	1306	1181	1212
Yanlış Sınıflandırma Sayısı	117	120	120	199	168
TP Oranı	0,918	0,916	0,916	0,856	0,878
FP Oranı	0,086	0,087	0,087	0,140	0,120
Kesinlik	0,919	0,917	0,917	0,859	0,879
Hassasiyet	0,918	0,916	0,916	0,856	0,878
F-Ölçütü	0,918	0,916	0,916	0,856	0,878

Statik ve dinamik çalışmada kullanılan algoritmaların kıyaslandığında iki çalışma arasındaki farkın çok az ((% 0,4437) olduğu görülmüştür. En başarılı sonuç statik anahtar kelimelerin seçilmesiyle yapılan statik çalışmada eğitim verisi üzerinde k-NN algoritması TF ve TF-IDF (k=4) ağırlıklandırma yöntemleri uygulanarak %92,2790 olarak elde edilmiştir.

Statik çalışma ve dinamik çalışma sonuçları arasındaki farkın çok az (% 0,4437) olmasının sebebi anahtar kelime sözlüğünün oluşturulması aşamasında statik olan statik çalışmada e-ticaret sitesi sınıfını yansıtan doğru anahtar kelimelerin seçilmiş olmasıdır. Statik çalışmadaki kelimelerin büyük bir bölümünün dinamik çalışmadaki kelimeler tarafından kapsandığı görülmüştür. Eğer statik çalışmanın anahtar sözlüğünü oluştururken e-ticaret sitelerinde ortak olarak sıklıkla geçtiği düşünülen kelimeler yerine ilgisiz kelimeler seçilmiş olsaydı modelin başarısının daha düşük olduğu görülecekti.

Her ne kadar kullanılan veri kümesine bağlı olarak statik çalışma dinamik çalışmadan daha başarılı bir sonuç üretmiş olsa da, birbirine çok yakın doğruluk derecesinde sonuçlar elde edilmiştir. Statik çalışma az bir farkla (% 0,4437) daha iyi olmasına rağmen sadece e-ticaret sınıfı üzerinde değil, farklı sınıflar için de uygulanabilirliği olduğu ve yenilenmeye açık olduğu için geliştirilen uygulama için dinamik çalışmanın kullanılması tercih edilmiştir. Dinamik çalışmada k-NN algoritması ikili ağırlıklandırma yönteminin %25 oranıyla elde edilen veri üzerinde uygulanması ile %91,8353 olarak $k = 7$ olduğu durumda en başarılı sonuca ulaşılmıştır.

Ardından bu bilgiler doğrultusunda oluşturulacak olan eğitim modelinin girdi olarak kullanılacağı bir uygulama ara yüzü geliştirilmiştir. Söz konusu model kullanılarak ara yüz ekranına girilecek sitelerin hangi sınıfa dâhil olduğu sonucuna ulaşılacaktır.

3.7. Uygulama Ara Yüzü

Uygulama ara yüzü Java programlama dili ve SWING¹ Kütüphanesi kullanılarak yapılmıştır. Kullanıcının uygulamayı kolay bir şekilde kullanabilmesi için sade bir ara yüz tasarlanmıştır. Uygulama ara yüzü iki ana ekrandan oluşmaktadır:

1. Sorgulama ekranı
2. Eğitim ekranı

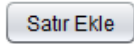
3.7.1. Eğitim ekranı

Eğitim ekranının amacı kullanıcının var olan eğitim kümesine yeni site isimleri ve bunlara ait sınıf bilgilerini girerek ekleme yapması ve daha sonra “Sorgulama Ekranı”nda kullanılmak üzere yeni bir eğitim modeli oluşturmasını sağlamaktır. Yeni eğitim modeli, Bölüm 3.5’te oluşturulan eğitim modelini oluştururken kullanılan internet sitelerinin listesine yeni site isimlerinin eklenmesiyle ve modelin bu yeni listeye göre tekrar eğitilmesiyle oluşturulmaktadır. Eğitim Ekranının görüntüsü Şekil 3.65’te verilmiştir.

Şekil 3.65. Uygulama eğitim ekranı görüntüsü

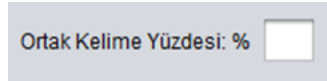
¹ SWING kütüphanesi Java dili ile görsel kullanıcı ara yüzü geliştirmekte kullanılmaktadır [51].

Eğitim ekranına veri kümesine eklenmesi istenen internet sitesinin adı yazılır, site e-ticaret sitesi ise “E-TİCARET SİTESİ” kutucuğu, e-ticaret sitesi değilse “E-TİCARET SİTESİ DEĞİL” kutucuğu işaretlenir. Birden fazla veri girilmek isteniyorsa Şekil 3.66’daki “Satır Ekle” düğmesine basılarak satır sayısı çoğaltılarak bilgilerin girişi yapılır.



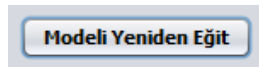
Şekil 3.66. Satır ekle düğmesi

Bölüm 3.6’da en iyi sonucu veren yüzde oranı %25 olarak bulunmuş ve mevcut eğitim modelini oluşturmada kullanılmıştı. Bununla birlikte kullanıcı eğer isterse Şekil 3.67’de “Ortak Kelime Yüzdesi” olarak adlandırılan, ortak kelimeleri oluştururken hesaba hatılacak en az e-ticaret sitesi sayısı alanını değiştirebilmektedir.



Şekil 3.67. Ortak kelime yüzdesi girişi

Şekil 3.68’deki “Modeli Yeniden Eğit” düğmesine basıldığı zaman girilen bilgiler hesaba katılarak yeni bir eğitim modeli oluşturulur.



Şekil 3.68. Modeli yeniden eğit düğmesi

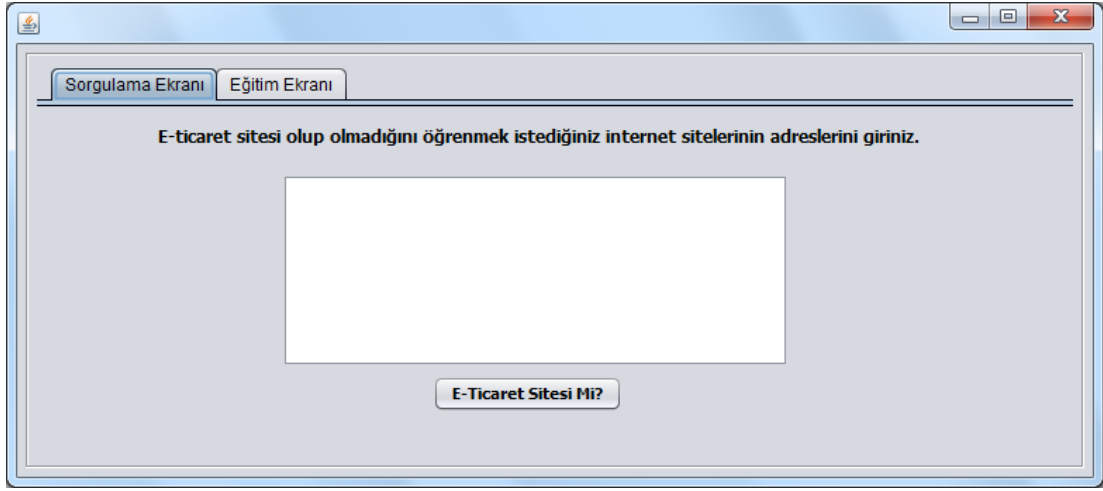
Yeni eğitim modeline ait sonuçlar Şekil 3.69’daki “Eğitim Modeli Sonuçları” kısmında ekrana gelir.

The screenshot shows a web application interface with two tabs at the top: 'Sorgulama Ekranı' and 'Eğitim Ekranı'. The 'Eğitim Ekranı' is active. Below the tabs, there is a header text: 'Eğitim verisine eklemek için site ismini ve e-ticaret sitesi olup olmadığı bilgisini giriniz.' (Enter the site name and whether it is an e-commerce site to add to the training data). Below this, there is a form with three input fields: 'SITE ADI' (Site Name), 'E-TİCARET SİTESİ' (E-commerce Site) with a checked checkbox, and 'E-TİCARET SİTESİ DEĞİL' (Not E-commerce Site) with an unchecked checkbox. There is a 'Satır Ekle' (Add Row) button next to the checkboxes. Below the form, there is a label 'Ortak Kelime Yüzdesi: %' followed by a small input field and a 'Modeli Yeniden Eğit' (Retrain Model) button. The bottom section is titled 'EĞİTİM MODELİ SONUÇLARI' (Training Model Results) and contains four sub-sections: 'Summary', 'Confusion Matrix', 'Detailed Accuracy By Class', and 'Predictions on Training Set'. Each sub-section has a large empty box for displaying results.

Şekil 3.69. Eğitim modeli sonuçları ekranı

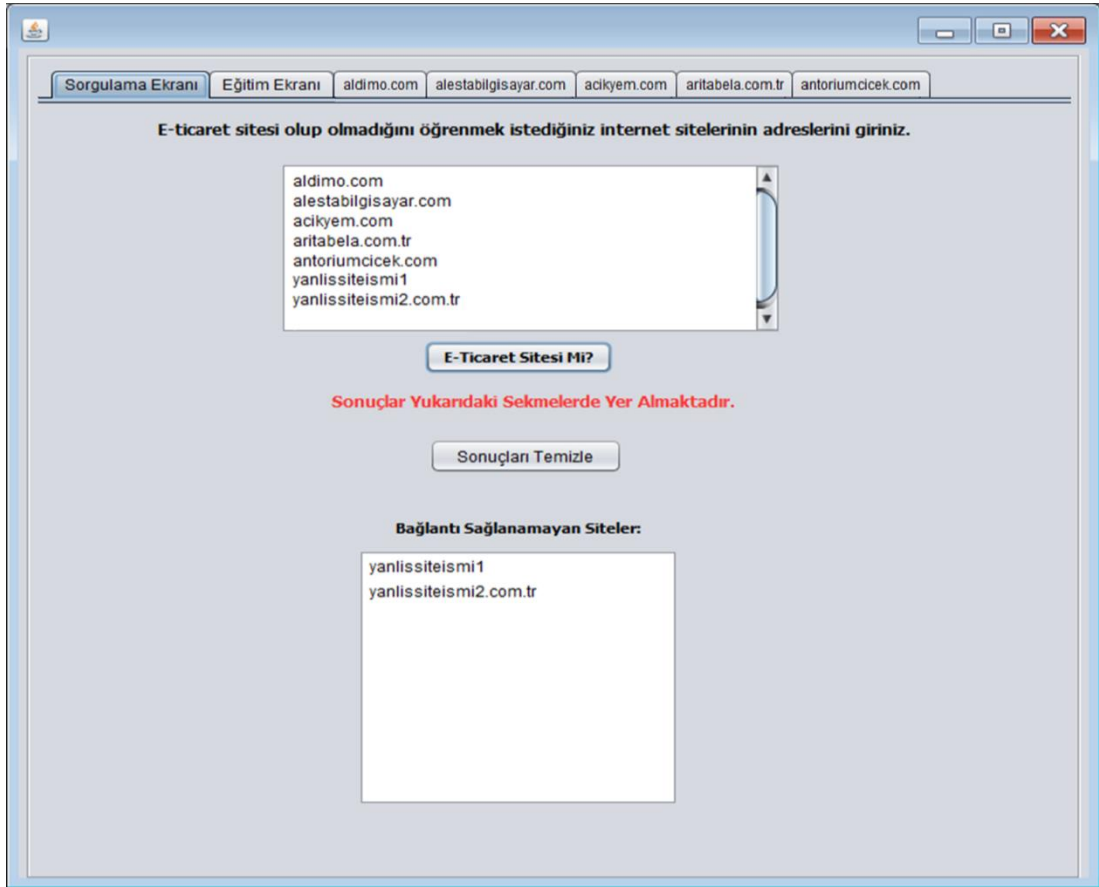
3.7.2. Sorgulama ekranı

Sorgulama ekranının amacı kullanıcının yazdığı internet sitesinin e-ticaret sitesi olup olmadığını sınanmaktadır. Sorgulanmak istenilen sitelerin isimleri ekrandan alındıktan sonra arka planda daha önce anlatılmış olan metin madenciliği işlemlerine ve “Eğitim Ekranı”nda oluşturulan modele göre sorgulama yapılmakta ve ardından sorgulama sonuçları ekrana gelmektedir. Eğer “Eğitim Ekranı” kullanılmadan “Sorgulama Ekranı” kullanılırsa Bölüm 3.6’da seçilmiş olan mevcut eğitim modeli kullanılmaktadır. Sorgulama ekranı görüntüsü Şekil 3.70’te verilmiştir.



Şekil 3.70. Sorgulama ekranı

Programın kullanımı Şekil 3.71 üzerinde örneklenerek anlatılmıştır.



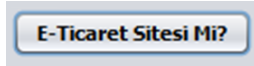
Şekil 3.71. Sorgulama ekranı ve sonuçları

Sorgu ekranındaki alana e-ticaret sitesi olup olmadığı öğrenilmek istenen internet sitelerinin adresleri Şekil 3.72'deki kısma yazılır.



Şekil 3.72. Site ismi giriş kısmı

Şekil 3.73'teki "E-Ticaret Sitesi Mi?" düğmesine basıldığı zaman girilen siteler önceden kurulan modele göre değerlendirilir.



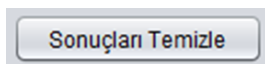
Şekil 3.73. E-ticaret sitesi mi? düğmesi

Listede yanlış yazım ya da sitenin aktif olmaması gibi sebeplerden dolayı bağlanılamayan siteler varsa bunlar Şekil 3.74'teki "Bağlantı Sağlanamayan Siteler" adı altında oluşturulacak listede görüntülenir.



Şekil 3.74. Bağlantı sağlanamayan siteler listesi

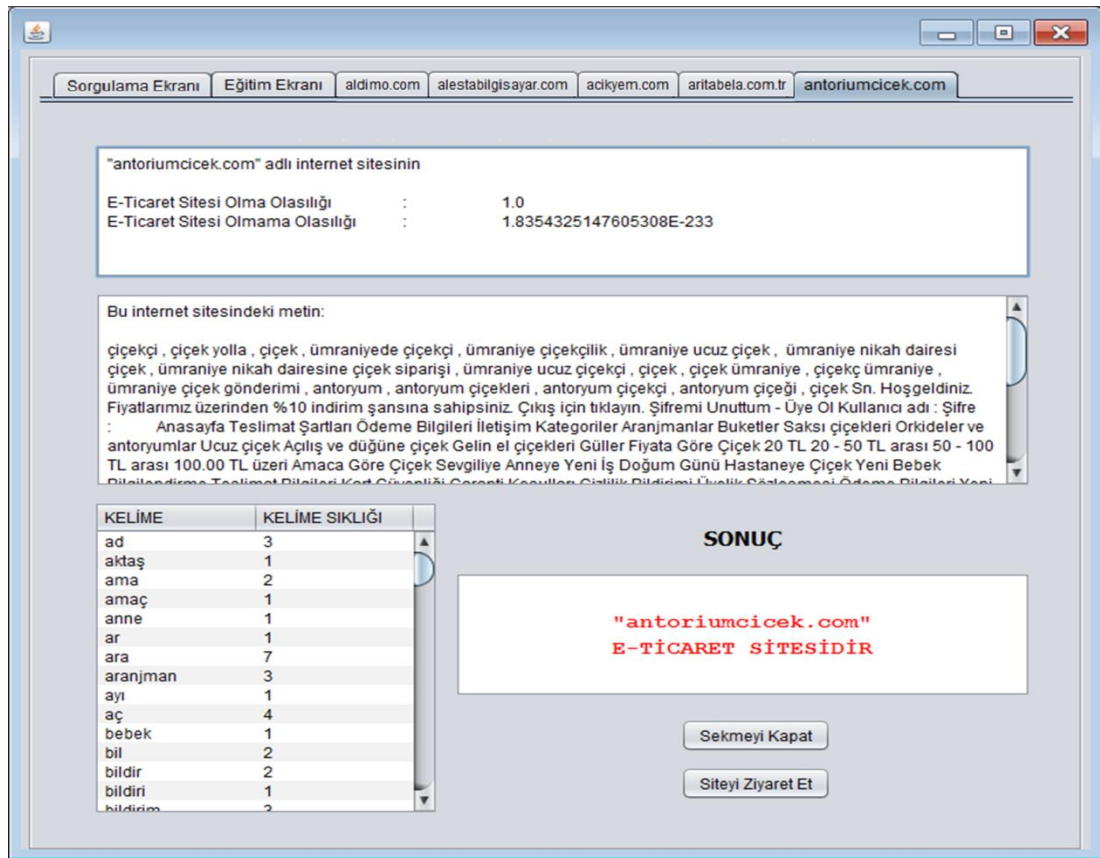
Şekil 3.75'teki "Sonuçları Temizle" düğmesine tıklandığı zaman şu işlemler gerçekleştirilir; açık olan sekmeler kapanır, "Bağlantı Sağlanamayan Siteler" listesi temizlenir ve "Bağlantı Sağlanamayan Siteler" listesi ekrandan kalkar.



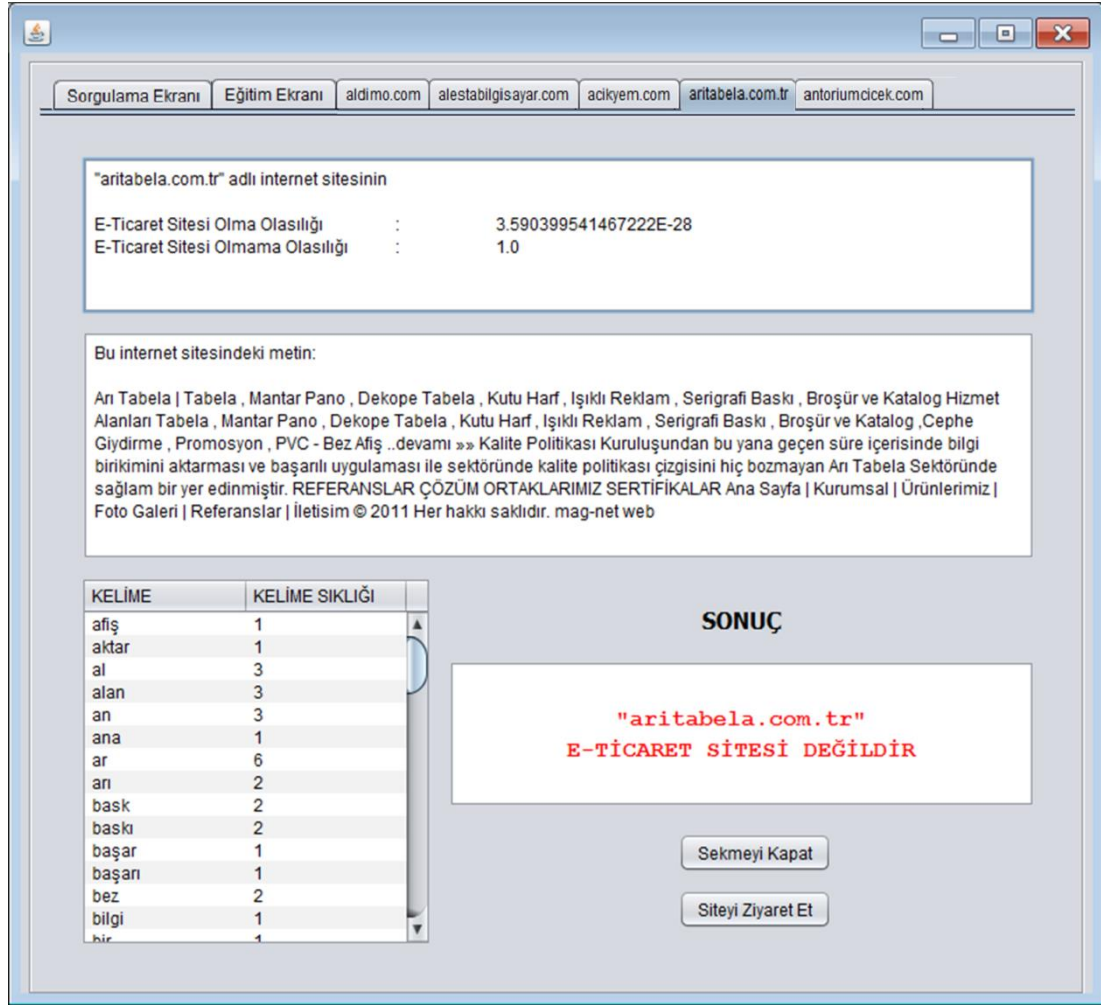
Şekil 3.75. Sonuçları temizle düğmesi

Uygulama bağlantı sağlanan ve işlemleri yapılan her bir site için sonuç verilerini içeren Şekil 3.76 ve 3.77'de gösterildiği gibi yeni bir sekme oluşturur. Bu sekmelerde girilen adrese ait şu bilgiler yer almaktadır:

- Siteye yönlendiren bir düğme,
- Sitede yer alan metin,
- Sitenin e-ticaret sitesi olma olasılığı,
- Sitenin e-ticaret sitesi olmama olasılığı ve
- Sonuç (sitenin e-ticaret sitesi olup olmadığı bilgisi). Sonuca ulaşmak için daha önce anlatılan işlemleri yapan Java kodu çalışmaktadır.



Şekil 3.76. Uygulama sonuç ekranı (e-ticaret sitesi)



Şekil 3.77. Uygulama sonuç ekranı (e-ticaret sitesi değil)

4. SONUÇLAR VE ÖNERİLER

Bu çalışmada kullanıcı tarafından belirtilen internet sitelerine bağlı olarak bu sitelerin içeriğini analiz eden, ardından metin madenciliği yöntemleri ile bu sitelerin e-ticaret sitesi olup olmadığına karar veren bir uygulama ve uygulama ara yüzü geliştirilmiştir.

Çalışmada öncelikle metin sınıflandırma kavramı ve önemi üzerinde durulmuştur. Ardından metin sınıflandırma süreci ve yöntemlerinden k-NN ve Naïve Bayes algoritmalar ele alınmıştır. Teorik bilgi küçük boyutlu bir örnek ile birlikte anlatılmıştır. Ardından bir uygulama ile bu algoritmalar gerçek hayata uyarlanmıştır. Bir kısmı elektronik ticaret sitelerinden oluşan bir eğitim kümesi ele alınmıştır. Eğitim kümesi ve sorgu dokümanı için veri elde etme aşamasında Java programlama dilinden yararlanılmıştır. Metin madenciliği aşamalarından doküman ön işleme, anahtar kelime oluşturma ve vektör uzay modelinin oluşturulması aşamalarının ardından kullanılmaya hazır hale gelen bu veriler k-NN ve Naïve Bayes algoritmalarıyla kullanılmıştır.

Uygulama iki çalışma ile yapılmıştır. Statik çalışmada anahtar kelimeler önceden belirlenmiş, dinamik çalışmada ise anahtar kelimelerin veri toplama işlemi sırasında dinamik olarak oluşturulması sağlanmıştır. Statik çalışma e-ticaret sitelerinde ortak olarak geçtiği düşünülen 30 anahtar kelimenin belirlenmesi ile k-NN (İkili, TF ve TF-IDF) ve Naïve Bayes (İkili ve TF-IDF) algoritmaları kullanılarak uygulanmıştır.

Statik çalışma için en başarılı sonuca k-NN algoritması TF ve TF-IDF ağırlıklandırma yöntemleri ile %92,2790 doğruluk yüzdesi ile ulaşılmıştır. Dinamik çalışmada veri elde etme aşamasında e-ticaret sitelerinde geçen tüm kelimeler ve ortak olarak geçen kelimelerin hesaba katılmasıyla anahtar kelimeler oluşturulmuş ve yine k-NN (İkili, TF ve TF-IDF) ve Naïve Bayes (İkili ve TF-IDF) algoritmaları kullanılarak yapılmıştır. E-ticaret sitelerinin %0, %10, %15, %20, %025, %30, %35 ve %40'ında ortak olarak geçen kelimeler ele alınmıştır. Veri kümesindeki e-ticaret sitelerinin %40'ından daha fazlasında aynı anda ortak olarak geçen hiç bir kelimenin (durak ve gereksiz kelimeler hariç) olmadığı gözlemlenmiştir. Bu 9 orana bağlı

olarak 9 ayrı veri kümesi oluşturulmuştur. k-NN algoritmasının her bir ağırlıklandırma yöntemi için 9 farklı yüzde değeri ve 15 farklı “k” değeri ile hesaplamalar yapılmış, program 27 (3 ağırlıklandırma * 9 yüzde) farklı veri kümesi ile toplam 153 ($9 \times 15(k\text{-}NN) + 9 \times 2(\text{Naïve Bayes})$) kez çalıştırılmıştır.

Statik ve dinamik çalışma sonuçları karşılaştırılmış ve çok az bir farkla statik çalışmanın daha iyi sonuç verdiği gözlemlenmiştir. Bunun sebebi statik çalışmadaki kelimelerin seçiminin doğru yapılmış olmasıdır. Geliştirilen uygulamada kendini uyarlayabildiği, sadece e-ticaret sınıfı üzerinde değil, farklı sınıflar için de uygulanabilir olduğu ve yenilenmeye açık olduğu için dinamik olan dinamik çalışmanın kullanılması tercih edilmiştir. Uygulama ara yüzünün k-NN algoritması ikili ağırlıklandırma yönteminin %25 oranı ve $k = 7$ değeri ile (%91,8353) elde edilen veri üzerinde uygulanmasına karar verilmiştir. Ardından bu sonuçlara bağlı olarak oluşturulacak olan eğitim modelinin girdi olarak kullanılacağı bir uygulama ara yüzü geliştirilmiştir. Uygulama ara yüzü eğitim ve sorgulama ekranlarından oluşmaktadır. Eğitim ekranında kullanıcı var olan eğitim kümesine yeni site isimleri ve bunlara ait sınıf bilgilerini girerek ekleme yapabilmekte ve daha sonra sorgulama ekranında kullanılmak üzere yeni bir eğitim modeli oluşturabilmektedir. Sorgulama ekranında ise kullanıcı e-ticaret sitesi olup olmadığını öğrenmek istediği site isimlerini ekrana girebilmekte ve varolan eğitim modeline göre sorgulama sonucunda girdiği sitelerin e-ticaret sitesi olup olmadığını öğrenebilmektedir.

Bu tez çalışması sonucunda kullanıma sunulan uygulama mali denetim amaçlı olarak kullanılabilir. Vergilendirmeden sorumlu birimler bu uygulamayı kullanarak e-ticaret sitelerini tespit edebilir, ardından bu site sahiplerinin vergisel yükümlülüklerini yerine getirip getirmediğini araştırabilir.

Pazarlamacılar tarafından pazarlamayı kolaylaştırmanın yöntemlerinden birisi pazar bölümlendirmesi yapmaktır. Pazar bölümlendirmesi değişik kriterlere göre yapılabilmektedir [52]. Ticaret yapılabilecek kişi ya da işletmeleri sınıflandırmaya tâbi tutarak, onları daha iyi anlayıp nasıl bir pazarlama yöntemi ile onlara ulaşmak gerektiği belirlenebilir.

Uygulamanın kullanım alanları şunlar olabilir:

- Ülkelerin vergi denetiminden sorumlu birimlerinin e-ticaret sitelerinin vergi mükellefiyeti tespiti için ön hazırlık,
- Dizin sitelerinde ilgili kategori altına yapılacak ekleme sırasında kategori önerme (mevcut durumda e-ticaret kategorisi),
- Endüstriyel pazarlamada pazar bölümlendirmesi yapmak için hedef kitlesi olarak kullanılacak e-ticaret sitelerinin tespiti:
 - Bankalar tarafından düşük faizli kredi verilmek üzere e-ticaret sitelerinin tespiti
 - GSM operatörleri tarafından uygun tarife kampanyası yapılmak üzere e-ticaret sitelerinin tespiti.

Bu uygulama e-ticaret sitelerinin sınıflandırılması için yapılmış olmasına rağmen, veri setleri farklı kategorilerden olacak şekilde seçildiği takdirde farklı konulardaki sınıflandırma işlemleri de yapılabilmektedir.

Bu tez çalışmasına dair öneriler şu şekilde sıralanmıştır:

- Uygulamada kullanılan %91,8353 oranına bakıldığında programın 100 siteden 8,1647 tanesinin yanlış sınıfa atandığı bilgisi görülmektedir. Bu noktada gerçekten e-ticaret sitesi olduğu halde e-ticaret sitesi değil ya da gerçekten e-ticaret sitesi olmadığı halde e-ticaret sitesi olarak sınıflandırılan sitelerin ortak özellikleri incelenmelidir ve neden bu şekilde değerlendirildiklerinin araştırılması gerekmektedir. Bu amaçla bağıntı analizi çalışması yapılabilir.
- Modellerin başarısında kampanya, fırsat, ilan, emlak, otomobil, ürün tanıtım ve fiyat karşılaştırma sitelerinin olumsuz yönde etkisinin olduğu düşünülmektedir. Çünkü bu siteler e-ticaret sitesi olmamasına rağmen içerdikleri kelimeler açısından algoritmalar tarafından e-ticaret sitesi olarak algılanabilmektedir. Bu konunun daha detaylı bir şekilde incelenerek sonuçlarının ortaya konması farklı bir çalışma olabilir.
- Bu tez çalışması e-ticaret sitelerinin belirlenmesi için yapıldı ancak ileride bu tez çalışması kullanılarak yapılacak çalışmalardan birisi farklı e-ticaret sektörlerine

yönelik anahtar kelimelerin ya da veri kümesinin belirlenmesiyle hangi e-ticaret sitelerinin hangi sektörde iş yaptığını analiz etmek olabilir. Sadece tekstil ya da sadece elektronik sektöründe iş yapanlar bulunup bu sektörlerle yönelik pazarlama stratejileri belirlenebilir. Bu çalışma sonuca göre ilgili sektörlerle uygun olarak bankalar tarafından kredi verme, satış yapma gibi işlemler yapılabilir. Ayrıca bu e-ticaret sitelerinin hangi bankalarla çalıştığını bulma ayrı bir çalışma konusu olabilir. Bu e-ticaret sektörlerine erişmek isteyen, onlara yönelik iş yapmak isteyen işletmeler pazar bölümlendirme yapmada bu çalışmayı bir araç olarak kullanabilir.

- Uygulamanın etkinliğinin artırılması için veri seti çoğaltılmalıdır.
- Eğitim modelinin oluşturulması sırasında daha çok sayıda algoritma kullanılarak farklı algoritmaların modelin başarısını nasıl etkilediği gözlemlenebilir.
- Eğitim modelinin oluşturulmasında her algoritmaya ait farklı parametreler denenerek eğitimde algoritma başarısını artıracak parametreler kullanılabilir.
- Kullanılan algoritmaların karşılaştırılması sırasında algoritma hızları ve hafıza kullanımı gibi farklı ölçütler de hesaba katılabilir.
- Eğitim modelinin oluşturulması sırasında Özellik Çıkarımı yapılarak anahtar kelime sözlüğünde yer alan kelimeler içerisinde daha az öneme sahip olanlar çıkartılarak modelin iyileştirilip iyileştirilemeyeceğinin araştırılması gerekmektedir.
- Uygulamanın başlangıç kısmına bir arama motoru eklenebilir. Arama motorunun bulunduğu siteler kullanıcıya sunulurken kullanıcıdan sınıf bilgisinin girilmesini isteyebilir. Bu veriler de eğitim verisine eklenir ve model tekrar eğitilebilir. Bu şekilde programın daha dinamik olması sağlanabilir.

KAYNAKLAR

1. İnternet : Bankalar Arası Kart Merkezi “İnternette Yapılan Kartlı Ödeme İşlemleri” http://www.bkm.com.tr/istatistik/sanal_pos_ile_yapilan_eticaret_islemleri.asp (2013).
2. İnternet : Türkiye Büyük Millet Meclisi “Elektronik Ticaretin Düzenlenmesi Hakkında Kanun Tasarısı” <http://www.tbmm.gov.tr/sirasayi/donem24/yil01/ss240.pdf> (2014).
3. İnternet : Resmi Gazete “İnternet Ortamında Yapılan Yayınların Düzenlenmesine Dair Usul ve Esaslar Hakkında Yönetmelik” <http://www.resmigazete.gov.tr/eskiler/2007/11/20071130-6.htm> (2014).
4. İnternet : Türkiye Büyük Millet Meclisi “Kişisel Verilerin Korunması Kanunu Tasarısı” <http://www2.tbmm.gov.tr/d23/1/1-0576.pdf> (2014).
5. İnternet : Resmi Gazete “Elektronik Haberleşme Sektöründe Kişisel Verilerin İşlenmesi ve Gizliliğinin Korunması Hakkında Yönetmelik” <http://www.resmigazete.gov.tr/eskiler/2012/07/20120724-2.htm> (2014).
6. İnternet : Resmi Gazete “Elektronik İmza Kanunu” <http://www.resmigazete.gov.tr/eskiler/2004/01/20040123.htm#1> (2014).
7. İnternet : Resmi Gazete “Tüketicinin Korunması Hakkında Kanun” <http://www.resmigazete.gov.tr/eskiler/2013/11/20131128-1.htm> (2014).
8. Taha, S. M., “Metin Madenciliği ile Doküman Demetleme”, Yüksek Lisans Tezi, *Gazi Üniversitesi Bilişim Enstitüsü*, Ankara (2011).
9. Vahaplar, A., İnceoğlu, M. M., “Veri Madenciliği ve Elektronik Ticaret”, *Türkiye’de İnternet Konferansı*, İstanbul, 1-3 (2001).
10. Guo, G. ve ark, “Using kNN Model for Automatic Text Categorization”, *Soft Computing Journal*, 10 (5): 423-430 (2006).
11. Karamcheti, A. C., “A Comparative Study on Text Categorization”, Yüksek Lisans Tezi, *Nevada Üniversitesi*, Las Vegas, 4, 29, 30 (2010).
12. Pilavcılar, İ. F., “Metin Madenciliği ile Metin Sınıflandırma”, Yüksek Lisans Tezi, *Yıldız Teknik Üniversitesi FBE Matematik Mühendisliği Anabilim Dalı*, 2, 4, 8, 21, 22 (2007).
13. Deng, Z., Zhang, M., “Improving Text Categorization Using the Importance of Words in Different Categories”, *Computational Intelligence and Security*, 3801: 458-463 (2005).

14. Yıldız, H. K. ve ark., “Metin Sınıflandırmada Yeni Özellik Çıkarımı”, **15. IEEE Sinyal İşleme ve İletişim Uygulamaları Kurultayı**, Eskişehir (2007).
15. Gandhi, V. C., Prajapati, J. A., “Review on Comparison between Text Classification Algorithms”, **International Journal of Emerging Trends & Technology in Computer Science**, 1 (3): 75-78 (2012).
16. Weiss, S. M. ve ark., “Maximizing Text-Mining Performance”, **Intelligent Systems and their Applications**, 14 (4): 63-69 (1999).
17. İlhan, U., “Application of KNN and FPTC Based Text Categorization Algorithms to Turkish News Reports”, Doktora Tezi, **Bilkent Üniversitesi Mühendislik ve Fen Fakültesi**, Ankara, 35 (2001).
18. Çalış, K., Gazdağı, O., Yıldız, O., “Reklam İçerikli Epostaların Metin Madenciliği Yöntemleri ile Otomatik Tespiti”, **Bilişim Teknolojileri Dergisi**, 6 (1), 1-7 (2013).
19. Wang, W., “An Empirical Study on Hierarchical Text Categorization”, Yüksek Lisans Tezi, **Guelph Üniversitesi**, Ontario, 6-7 (2007).
20. Hotho, A., Nürnberger, A., Paaß, G., “A Brief Survey of Text Mining”, **Ldv Forum**, 19-62 (2005).
21. Soucy, P.e Mineau, W., “A Simple KNN Algorithm for Text Categorization”, **IEEE International Conference**, 647-648 (2001).
22. Vidhya, K. A., Aghila, G., “A Survey of Naïve Bayes Machine Learning approach in Text Document Classification”, **International Journal of Computer Science and Information Security**, 7 (2): 206-211 (2010).
23. Schneider, K., “On Word Frequency Information and Negative Evidence in Naive Bayes Text Classification”, **Natural Language Processing – TAL**, Sydney, 474-486 (2004).
24. Akyokuş, S., “Anlamsal Web, Anlamsal Web Dilleri ve Araçları”, **Bilişim Kurultayı 2003**, İstanbul (2003).
25. İnternet : BICSI Information Technology Systems “Social Media to Social Business” https://www.bicsi.org/pdf/presentations/fall_13/Contractors%20and%20Social%20Media%20-%20Selso%20Huerta%20-%20NetPricer.net.pdf (2013).
26. Takçı, H., “Karakter Tabanlı Doküman Dili Tanıma Sistemi Tasarımı”, Yüksek Lisans Tezi, **Gebze Üniversitesi Gebze Yüksek Teknoloji Enstitüsü Mühendislik ve Fen Bilimleri Enstitüsü**, 16 (2005).

27. Chowdhury, G. G., “Natural Language Processing”, *Information Today*, Mazelya, 51-89 (2003).
28. Coşkun, C., Baykal A., “Veri Madenciliğinde Sınıflandırma Algoritmalarının Bir Örnek Üzerinde Karşılaştırılması”, *Akademik Bilişim 2011*, Malatya, (2011).
29. Silahtaroğlu, G., Demircan, F., “Çeviri Yazılımlarında Sözcüklerin Bağlam İçindeki Anlamını Algılamaya Yönelik Bir Uygulama”, *Akademik Bilişim 2013*, Antalya, (2013).
30. Akın, Y. K., “Veri Madenciliğinde Kümeleme Algoritmaları ve Kümeleme Analizi”, Doktora Tezi, *Marmara Üniversitesi Sosyal Bilimler Enstitüsü*, İstanbul, 55-56 (2008).
31. Koyuncugil, A. S., Özgülbaş, N., “Veri Madenciliği: Tıp ve Sağlık Hizmetlerinde Kullanımı ve Uygulamaları”, *Bilişim Teknolojileri Dergisi*, 2 (2): 21-32 (2010).
32. Manne, S., Kotha, S. K., Fatima, S. S., “Text Categorization with K-Nearest Neighbor Approach”, *Information Systems Design and Intelligent Applications 2012*, Hindistan, 413-420 (2012).
33. Wu, K. ve ark., “An Empirical Comparison of Min–Max-Modular K-NN with Different Voting Methods to Large-Scale Text Categorization”, *Soft Computing Journal*, 12: 647-655 (2008).
34. Kehagias, A. ve ark., “A Comparison of Word and Sense-based Text Categorization Using Several Classification Algorithms”, *Journal of Intelligent Information Systems*, 21 (3): 227-247 (2003).
35. Cingiz, O. M., Diri, B., “Classification of Microblogging Users”, *Signal Processing and Communications Applications Conference*, 20: 1-4 (2012).
36. Manning, C. D., Schütze C. D., “Foundations of Statistical Natural Language Processing 2nd ed.”, *MIT*, 299-301 (1999).
37. Altman, N. S., “An Introduction to Kernel and Nearest-Neighbor Nonparametric Regression”, *The American Statistician*, 46 (3), 175–185 (1992).
38. Eyheralendy, S., Lewis, D., Madigan D., “On the Naive Bayes Model for Text Categorization”, (2003).
39. Uzun, E., Agun, H. V., Yerlikaya, T., “Web Content Extraction by Using Decision Tree Learning”, *Signal Processing and Communications Applications Conference*, 20: 1-4 (2012).
40. Frank, E. ve ark., “WEKA Manual for Version 3-7-8.”, (2013).

41. Şeker, Ş. E., “İş Zekası ve Veri Madenciliği”, **Cinius**, İstanbul, 54 (2013).
42. Çobanoğlu, B., “Java ile Programlama ve Veri Yapıları 2. baskı”, **Pusula**, İstanbul, 11 (2010).
43. Shibu, S., Vishwakarma A., Bhargava, N., “A Combination Approach for Web Page Classification Using Page Rank and Feature Selection Technique”, **International Journal of Computer Theory and Engineering**, 2 (6): 897 (2010).
44. Dariush, A., “Meta-tag: A Means to Control the Process of Web Indexing”, **Online Information Review**, 27 (4): 238-242 (2003).
45. Gunter, B., “A Survey of Users in the UK”, **Aslib Proceedings** 57 (6): 513-522 (2005).
46. Menascé, D. A. ve ark., “A Methodology for Workload Characterization of E-Commerce Sites”, **Proceedings of the 1st ACM conference on Electronic commerce**, 119-128, (1999).
47. Tüzüntürk, S., “Veri Madenciliği ve İstatistik”, **Uludağ Üniversitesi İİBF Dergisi**, 29 (1): 65-90 (2010).
48. İnternet : Google Code “Zemberek” <http://code.google.com/p/zemberek/downloads/list> (2013).
49. Cengiz, C., Diri, B., “Türkçe Tweetlerden Soru İfadelerini Bulmak”, **Signal Processing and Communications Applications Conference**, 21 (2013).
50. Tee, J. Y. J., Soon, L. K., Ting, C. Y., “WebSum: Enhanced SumBasic Algorithm for Web Site Summarization”, **Data Mining and Optimization**, 137-142 (2012).
51. Alp, U. ve ark., “İnternet Üzerinden Uzaktan Robot Erişimi”, **11. Sinyal İşleme ve İletişim Uygulamaları Kurultayı**, 3 (2003).
52. Alabay, M. N., “Sosyal Medyada Tüketiciler ve Pazar Bölümleme Uygulamaları” (2011).

EKLER

EK-1. WEKA işlemlerini yapan Java kodu

```

private void doWekaIslemleri() {
    BufferedReader breaderTrain = null;
    BufferedReader breaderTest = null;
    try {
        breaderTrain = new BufferedReader(new FileReader("C://wekayaEgitim.arff"));
        breaderTest = new BufferedReader(new FileReader("C://wekayaTest.arff"));
        //Eğitim ve model oluşturma için
        Instances train = new Instances(breaderTrain);
        train.setClassIndex(train.numAttributes() - 1);
        breaderTrain.close();
        //Test için
        Instances test = new Instances(breaderTest);
        test.setClassIndex(test.numAttributes() - 1);
        breaderTest.close();
        //Algoritma ve model oluşturma
        IBk knn = new IBk(1);
        knn.buildClassifier(train);
        // Eğitim sonucu
        Evaluation evalTraining = new Evaluation(train);
        evalTraining.evaluateModel(knn, train);
        System.out.println("\nEĞİTİM");
        System.err.println(evalTraining.toSummaryString());
        System.out.println(evalTraining.toClassDetailsString());
        System.out.println(evalTraining.toMatrixString());
        // Eğitim sonucunu test üzerinde kullan
        Evaluation evalTestOnTraining = new Evaluation(test);
        evalTestOnTraining.evaluateModel(knn, test);
        System.err.println("SINAMA");
        Classifier classifier;
        try {
            classifier = (Classifier) weka.core.SerializationHelper.read("C:\\model.model");
            for (int i = 0; i < test.numInstances(); i++) {
                //Sınıf tahmini
                double pred = classifier.classifyInstance(test.instance(i));
                String classLabel = test.classAttribute().value((int) pred);
                System.out.print("\n\n" + (test.numInstances() - i) + "\t SONUÇ\t\t\t: ");
                System.err.println(classLabel);
                //Sınıf tahmin yüzdesi
                double[] preds = classifier.distributionForInstance(test.instance(i));
                System.out.println("E-Ticaret Sitesi Olma Olaslılığı\t: " + preds[0]);
                System.out.println("E-Ticaret Sitesi Olmama Olaslılığı: " + preds[1]);
                if (preds[0] > preds[1]) {
                    System.out.print("E-Ticaret Sitesi Olma Olaslılığı \t: ");
                    System.err.println(preds[0]);
                    System.out.println("E-Ticaret Sitesi Olmama Olaslılığı: " + preds[1]);
                } else {
                    System.out.println("E-Ticaret Sitesi Olma Olaslılığı \t: " + preds[0]);
                    System.out.print("E-Ticaret Sitesi Olmama Olaslılığı: ");
                    System.err.println(preds[1]);
                }
            }
        } catch (Exception e) {
            e.printStackTrace();
        }
    } catch (FileNotFoundException e1) {
        System.out.println("Weka işlemleri sırasında Hata Oluştı. 1");
        e1.printStackTrace();
    } catch (IOException e1) {
        System.out.println("Weka işlemleri sırasında Hata Oluştı. 2");
        e1.printStackTrace();
    } catch (Exception e1) {
        System.out.println("Weka işlemleri sırasında Hata Oluştı. 3");
        e1.printStackTrace();
    }
}

```

EK-2. Dizin sitelerinden bağlantı toplama Java kodu

```

package dizinsitesindenurltopla;

import java.io.BufferedReader;
import java.io.BufferedWriter;
import java.io.File;
import java.io.FileNotFoundException;
import java.io.FileReader;
import java.io.FileWriter;
import java.io.IOException;
import java.util.ArrayList;
import java.util.List;
import org.jsoup.Jsoup;
import org.jsoup.nodes.Document;
import org.jsoup.nodes.Element;
import org.jsoup.select.Elements;

public class DizinSitesindenLinkToplaAlisverisRehberiCom {

    public static List<String> extractLinks(String url) throws IOException {
        final ArrayList<String> result = new ArrayList<String>();
        Document doc = Jsoup.connect(url).get(); //Siteye bağlan
        //Site bilgisini içeren HTML etiketlerini seç
        Elements dizinsitesindenlinks = doc.select("span.urlimg");
        for (Element link : dizinsitesindenlinks) { //Her etiketin içerisindeki url'leri al
            result.add(link.text().replace("www.", "")); //başındaki "www."yi sil
        }
        return result;
    }

    public final static void main(String[] args) throws Exception {
        String readfile = "C:\\urltopla.txt"; //dosyadan dizin sitelerinin adlarını al
        BufferedReader br = new BufferedReader(new FileReader(readfile));
        BufferedWriter bw = null;
        try {
            //url'leri dosyaya yaz
            bw = new BufferedWriter(new FileWriter(new File("C:\\url toplanan.txt"), true));
            String line;
            //dosyada dizin sitesi adı oldukça her bir dizin sitesi için
            while ((line = br.readLine()) != null) {
                String site = line;
                //extractLinks(String url) metodunu çalıştır
                List<String> links = extractLinks(site);
                for (String link : links) {
                    bw.write(link); //url'leri dosyaya kaydet
                    bw.newLine();
                }
            }
        } catch (FileNotFoundException ex) {
            ex.printStackTrace();
        } catch (IOException ex) {
            ex.printStackTrace();
        } finally {
            try {
                try {
                    if (bw != null) {
                        bw.flush();
                        bw.close();
                    }
                } catch (IOException ex) {
                    ex.printStackTrace();
                }
            }
        }
    }
}

```

EK-3. Çizelge 3.1. Dizin sitelerinden bağlantı toplama Java kodundaki değişiklikler

SİTE ADI	KOD DEĞİŞİKLİĞİ
alisverisrehberi.com dizin.sitesi.web.tr netbul.gen.tr sitetanitimi.com	<pre> Elements dizinsitesindenlinks = doc.select("span.urlimg"); for (Element link : dizinsitesindenlinks) { result.add(link.text().replace("www.", "")); } </pre>
benimsayfam.org	<pre> Elements dizinsitesindenlinks = doc.select("a[title^=http://www.]"); for (Element link : dizinsitesindenlinks) { result.add(link.attr("title")); } </pre>
bilgirehberi.com	<pre> Elements dizinsitesindenlinks = doc.select("div.url"); for (Element link : dizinsitesindenlinks) { result.add(link.text().replace("http://.", "").replace("www.", "")); } </pre>
buneki.org	<pre> Elements dizinsitesindenlinks = doc.select("b") .not("a[href^=http://www.facebook]") .not("a[href^=http://www.live]") .not("a[href^=http://www.buneki]"); for (Element link : dizinsitesindenlinks) { result.add(link.attr("abs:href")); } </pre>
dmoz.org	<pre> Elements dizinsitesindenlinks = doc.select("a[href^=http://www.]") .not("a[href^=http://www.google]") .not("a[href^=http://www.mozilla]") .not("a[href^=http://www.bing]"); for (Element link : dizinsitesindenlinks) { result.add(link.attr("abs:href")); } </pre>
linkcenneti.net	<pre> Elements links = doc.select("a[href^=http://www.]") .not("a[href^=http://www.phplinkdirectory.com]") .not("a[href^=http://www.mozilla]") .not("a[href^=http://www.google]") .not("a[href^=http://www.bing]") .not("a[href^=http://www.phplinkdirectory]"); for (Element link : links) { result.add(link.attr("abs:href")); } </pre>
mavinokta.com	<pre> Elements dizinsitesindenlinks = doc.select("span.urlimg"); for (Element link : dizinsitesindenlinks) { result.add(link.text().replace("www.", "")); } </pre>
sektoreldizin.com	<pre> Elements dizinsitesindenlinks = doc.select("A[HREF^=http://www.]") .not("A[HREF^=http://www.google]") .not("A[HREF^=http://www.mavinokta]"); for (Element link : dizinsitesindenlinks) { result.add(link.attr("HREF") .replace("www.", "").replace("http://", "")); } </pre>
trdizin.org	<pre> Elements dizinsitesindenlinks = doc.select("a[href^=http://www.]") .not("a[href^=http://www.google]") .not("a[href^=http://www.mozilla]") .not("a[href^=http://www.bing]"); for (Element link : dizinsitesindenlinks) { result.add(link.attr("abs:href")); } </pre>
turkarama.com	<pre> Elements dizinsitesindenlinks = doc.select("a[HREF^=http://www.]") .not("a[HREF^=http://www.turkarama]"); for (Element link : dizinsitesindenlinks) { result.add(link.attr("HREF").replace("www.", "") .replace("http://", "")); } </pre>

EK-4. Çizelge 4.1. E-ticaret sınıfını oluşturan internet siteleri

07cicek.com	biberon.com.tr	derigiyimci.com	faturasoft.com
1001esya.com	bicakcim.com	deriline.com	fcmaps.com
1marka.com	bicaklar.com.tr	derimarket.net	feyizoglu.com
365alisveris.com	bicok.com	derindalis.com	feyzabilgisayar.com.tr
444cicek.com	bidunyaoyuncak.com	derinmor.com	fikus.com.tr
444dergi.com	bigbigstyle.com	derinshop.com	firebal.net
4gen.com.tr	bilbors.com	desa.com.tr	firsatbizde.com
4m4bilisim.com	bilenahsap.com	destinacicek.com.tr	firsatbufirsat.com
54cicek.com	bilgikare.com	devamerkezi.com	firsatlaritakipet.com
724hirdavatci.com	bilgisayardeposu.com	deveyuku.com	firsatvar.com
724magaza.com	bilgisayarmarket.com	deybi.com	fitnessdukkani.com
acarfoto.com.tr	bimeks.com.tr	dieselsaat.com	fiyatport.com
acelya.com.tr	binbirkitap.com	digimarketim.com	flamabayrak.net
acikyem.com	binbirmarka.com	digishoping.com	flo.com.tr
adainci.com	binbirsaat.com	dijital.com	floracay.com
adaletyayinevi.com	binicilikmalzemesi.com	dikeylimit.net	floracicek.com
adanacicekci.name.tr	birbuketmeyve.com	diktasld.com.tr	floraturk.com
adanacicekcilik.com	birdunya.com	dilekfener.com	folixir.com
adanacicekcilik.com.tr	birmutfak.com	dinamikbilisim.net	fonexkozmetik.com
adanacicekcilik.net	birnumaram.com	dincerkirtasiye.com	fotoakksesuar.com
adanaciceknet.com	bisikletcim.com	disbeyazlatıcı.com.tr	fotografium.com
adede.com.tr	bisskozmetik.com	disburada.com	fotografmerkezim.com
adepo.com	bitenekadar.com	dishetopuklar.com	foto-hali.com
adresecikofte.com	bitiklagelsin.com	disportaci.com	fotokopim.net
adrese mama.com	bitkici.com	divacicek.com	fotosende.com
agsmarket.com	bitkiselbakkal.com	divin.com.tr	frutation.com.tr
akademiegitim.com.tr	bitkiselpazar.com	diyetmax.com	fuficookie.com
akavm.com	bitkiselurunpazari.com	dkogold.com	funnababy.com
akercom.com.tr	bizde.com	dockersturkey.com	galacicek.com.tr
akillitasarruf.com	bizdekiler.com	dod.com.tr	gamateknoloji.com
akparca.com	bizegeline.com	dogaevinizde.com	gamer-market.com
aksesuaraski.com	bizimev.com.tr	dogalecza.com	garantialisveris.com
aksuvital-pazar.com	bizimnotebook.com	dogaltedavi.net	garantiofis.com
aktarland.com.tr	bjkstore.com.tr	dogancicek.com	gardeniacicek.com
aktifbebek.com	bkscombalkon.com	doganoyuncak.com	gardenyacicek.com
aktifmarketim.com	bkscombalkon.com	dogaspor.net	gaziantepecicek.com
aktifpedal.com	bluemavi.com	dogruhediyecicek.com	gercekbal.com
akuluaraba.gen.tr	bodrumcicek.com	dogumgunudavetiyesi.com	gercekucuzluk.com
akulumotor.com	bodrumdasatilik.com	dolichetekstil.com	gidercicek.com
al37.com	bolucicek.com	dolumshop.com	girnecicekcilik.com
alalimbakalim.com	bonnyfood.com	dolumturk.com	gittigidiyor.com

EK-4. Çizelge 4.1. E-ticaret sınıfını oluşturan internet siteleri (Devam)

alanadisitesi.com	bornovacicekcilik.com	dormeo.com.tr	giyimsepeti.com
alanyacicek.net	borsapazar.com	dostcicekcilik.com.tr	giyimzamani.com
alanyacicekcisi.com	boyner.com.tr	dostmedikal.com	gizilinci.com
albashop.com.tr	boyutstore.com	dovmemalzemesi.com	gizzaro.com
alextoysturkey.com	brightmobil.com	dovmemalzemesi.gen.tr	goblenci.com
alibabadogaltas.com.tr	btdepo.com	dpazar.com	goksallarcicekcilik.com
alinderi.com	btsony.com	dr.com.tr	gold.com.tr
alisveris.com	buderuskombifiyatlari.com	dreamersfigure.com	goldstore.com.tr
alisverisci.com	buduro.com	drhealthy.com.tr	goncacanta.com
alisverisdukkani.com	buketcicek.com	dshopping.com.tr	goprint.com.tr
alisverisim.net	buldumbuldum.com	dugundavetiyeleri.gen.tr	gorkemcicekcilik.com
alisveriskeyfim.com	bunlardanistiyorum.com	dukkkan34.net	gozdeavize.com.tr
alisverismagazasi.com	buradauygun.com	dukkkan360.com	gozdeciceksarayi.com
alisverisposeti.com	buramoda.com	dunyagrup.com	gozdeoyuncak.com
alpagut.com	burcucicekcilik.com.tr	dunyakitap.com	gozlemkitap.com
alpinist.com.tr	buroteknik.com	dunyapazarlama.com	gozlemyayincilik.com.tr
alsasaatcilik.com.tr	bursailan.com	durallar.com	greenpet.com.tr
alsatpazar.com	bursanur.com	durcleantemizlikmakinalari.com	grisepet.com
alternatifkitap.com	busrapirlanta.com.tr	dusyerisanalmagaza.com	grossdepot.com
altinsarrafi.com	butikbebe.com	e-arifoglu.com	grupfoni.com
aluser.com.tr	buyukmagaza.com	e-avize.com	gsmparcakontor.net
alveal.com	cabani.com.tr	e-bebek.com	guess-saat.com
alyanschi.com	caddegumus.com	ebilardo.com	gulactistore.com
anaokuluoyuncak.com	caddegumus.com	e-cadde.com	gulaylar.com
anatoliagumus.com	caliskanofis.com	ecicekci.com	gulciceksarayi.com
anindakapida.com	camasirkurutma.com	eczanegulden.com	gulistancicek.com
ankaracicek.com	camhediye.com	edipsaat.com	gulluoglushop.com
ankaracicekcilik.com	camurunleri.com	edo.com.tr	gulseven.com.tr
ankaraciceknet.com	canbilteknoloji.com	e-ekmek.com	gumuscu.net
ankaracicekpazari.com	candaoyuncak.com	e-fidancim.com	gumusdepo.com
ankaracicekyolla.name.tr	cankayacicekcilik.com	e-finspor.com	gumusdiyari.com
ankastreal.com	cantabak.com	efirsatci.com	gumusdukkani.com
ankastreunler.com	cantacenter.com.tr	eftelyacicek.com	gumustekin.org
annelutfen.com	capamakinasi.com.tr	e-gebe.com	gunboyu.com
antalyabegumcicek.com	careshop.com.tr	egitimseti.com	guneskitabevi.com
antalyacicek.com.tr	carrefour.com.tr	e-glorria.com	gunlukurun.com
antalyacicek.net	casperstore.net	e-hirdavat.com	gureryayinlari.com
antalyacicekcilik.com	casperturk.com	eiksir.com	guvensanat.com
antalyacicekcilik.net	casuskamera.com	ejoya.com	gym-market.com
antenci.net	casusununleri.com	e-kigili.com	hacdunyasi.com
antepecicek.com	cayomat.com	ekincioglutoys.com	hacimustafa.com

EK-4. Çizelge 4.1. E-ticaret sınıfını oluşturan internet siteleri (Devam)

antoriumcicek.com	cazipalisveris.com	ekiralikoto.com	hakanspot.com
appledunyasi.com	cazipsepet.com	eklite.com	hamileyastigim.com
aralgame.com	cdmarket.com.tr	ekocakgold.com	hanedanavm.com
arasyayincilik.com	cepdepo.com	ekolye.com	hanfendi.com
arceliksatici.com	cepmar.com	ekonomikayakkabi.com	hatemoglu.com
arcelikten.com	ceppaket.com	ekonomikmama.com	havalimanishop.com
arcpazar.com	ceptelesoncu.com	ekonomiksatis.com	hayallerdukkani.com
arekacicek.com	ceyizsarayi.com	ekopc.com	hayalpanosu.com
aresis.com.tr	ceyto.com	ekozmetik.com	haytapshop.com
arikaltin.com	ceyyiz.com	eksenbilgisayar.com	hazerspor.com
arivitalonlineshop.com	cicekacil.com	ekstremakvaryum.com.tr	hazinex.com
arkadas.com.tr	cicekal.net	e-kundura.com	hediye.com.tr
arkeolojisanat.com	cicekalsam.com	elektrikdeposu.com	hediyebahcesi.com
arselcicek.com	cicekbahcem.com	elektrikdukkani.com	hediyebizden.com
arsiv.com	cicekbiz.com	elektriklialet.com	hediyecek.com
arslanzademagaza.com	cicekbul.com	elektriklisupurgetorbasi.com	hediyecek.com.tr
arzumurunleri.com	cicekciarzum.com	elektrikmagazam.com	hediyyedenizi.com
arzustore.com	cicekcilersitesi.com	elektrik-malzemeleri.com	hediyyede.com
asdedektor.com	cicekclub.com	elektrikpazar.com	hediyeefabrikasi.com
aseriyorum.net	cicekclub.com	elektriksepetim.com	hediyyelerdunyasi.com.tr
askermekani.com	cicekdukkani.net	elektrodepo.com	hediyeekartim.com
asteknikhirdavat.com.tr	cicekgonder.com	e-literayayin.com	hediyyerim.com
asyabal.com	cicekgonderin.com	elmacipazari.com	hediyyesaati.com
asyatip.com	cicekgondersene.com	elmalibilgisayar.com	hediyyevi.com
ateliercreart.com	cicekhatti.net	elmasepeti.com	hemenbenimolmali.com
atelsan.com.tr	cicekizmir.com	elnuru.com	hemenevinde.com
ataskamp.com	ciceklerdeyasi.com	elyapimlari.com	hemerapirlanta.com
avantajim.com	cicekmalzemeleri.com	emecmua.com	hepindirimde.com
avantajliurunler.com	cicekmarket.com	em-er.com	hepindirimde.com
avcimekani.com	cicekmarket.com	e-minico.com	hepsiantep.com
avizemarketim.com	cicekmax.com	emrehome.com	hepsiburada.com
avmarketi.com	ciceknet.com	emskirtasiye.com	hepsidijital.com
avpazarim.net	cicekpasaji.com	enginyayinevi.com.tr	hepsidogal.com
avrupatr.com	cicekpazari.com.tr	enilginc.com	hepsielgde.com
ayakkabigar.com	cicekpostasi.com	eniyial.com	hepsikitap.com
ayazshop.com	ciceksecin.com	eniyisinden.com	hepsimarkette.com
aydisplay.com	ciceksepeti.com	eniyiteknoloji.com	hepsimutlak.com
ayhancocuk.com	ciceksepeti.net	enkolay.com.tr	hepsindenal.com
ayyildiziletisim.com	cicekservisi.net	ennucuz.com	hepsinialalim.com
babybanzturkey.com	ciceksiparis.com	enpa.com.tr	hepsisony.com
bagcilarcicek.com	ciceksiparisi.name.tr	ensarshop.com	hepsivar.com

EK-4. Çizelge 4.1. E-ticaret sınıfını oluşturan internet siteleri (Devam)

balerin.com	ciceksiparisi.net	enucuzreyon.com	herkesemobilya.com
balikmarketim.com	cicekvakti.com	enuygunpirlanta.com	hermeskitap.com
balkotrade.com	cicekyolla.com.tr	epazaar.com	herseyahsap.com
balonbuketi.com	cikolatayolla.com	epiteks.com	herteknoloji.com
balorama.com.tr	cilekcicek.com	erdoganlarbisiklet.com	hikmetkitap.com
bamboudavetiye.com	cimri.com	eren.com.tr	hindirim.com
baneva.com	cinaryayincilik.com.tr	erengul.com	hipernex.com
banyobanyo.com	civcivbebe.com	ereyon.com.tr	hipir.com
baranbilgisayar.com.tr	clubsafari.com.tr	erkamalisveris.com	hirdavat.com
barcin.com	cocukvebebek.com	erkinyapi.com.tr	hirdavatfirsati.com
barisguvenlik.com	colonyo.com	erlastik.com	hirdavatmarketim.com
barkodburada.com	comfydesk.com.tr	erolmobilya.com	hizlial.com
barkodpazari.com	cs-cartr.com	erolmobilya.net	hizlialisveris.com
basakcicekcilik.com	cup.com.tr	erospirlanta.com	hizlipc.com.tr
basarionline.com	dagevi.com	eroticshop.gen.tr	hobbygaleri.com
basmarket.com	daghdash.com	erotikshop.web.tr	hobbytime.com.tr
basmatik.com	dalkilicspor.com	esabilgisayar.com.tr	hobievi.com
batangemi.com	damlaltd.com	escombilisim.net	hobim.net
batipetmarket.com	dantelci.com	eskiesyaalimsatim.com	hobisitesi.com
bayraksatinal.com	dataservis.net	eskikitap.org	horizonav.com.tr
bayspor.com	davetiye.gen.tr	eskisehircicek.net	hotic.com.tr
bebekchik.com	davetiyeevi.com	eskisehircicekcilik.com	hukukkitabi.com
bebekcocukdunyasi.com	davetiyeseker.com	eskisehircicekcilik.net	hukukmarket.com
bebekhane.com	davetiyesiparisi.com	eskisehironlinecicek.com	icgyimshow.com
bebekmagazasi.com	daytonamotor.com	esmgim.com	idealcicekcilik.com
bebekom.com	db-dizayn.com	euroakvaryum.com	idealhediye.com
bebek.net	dedektorcu.com	eurocicek.com	ideefixe.com.tr
bebekurunleri.blogcu.com	dedektormarket.com	evbutik.com	idefix.com
bebemax.com	dedenden.com	evcilal.com	ifclock.com
bebeshop.com.tr	defnesabun.com	evcilbesinleri.com	igneiplikburada.com
beceriklialetler.com	degisimkitap.com	evcilimpet.com.tr	ihsan.com.tr
bellabutik.com	dekorpark.com	evcilmarket.com	ihtiyacinne.com
bellagloria.com	dekorpazari.com	evdealisveris.com	ikea.com.tr
bellona.com.tr	dek-tech.com	evdekimarketim.com	ikinciel.org
benceiyi.net	deltabisiklet.com	evegilir.com	ikincielegel.com
benimevim.com.tr	demirciogluoto.com	e-veptron.com	ikitikla.com
betashoes.com	denokids.com	evimdeko.com	moda12.com
bey.com.tr	dentalmarketim.com	evsever.com	novabazar.com
beyazadam.com	dentalticaret.com	evsuaritma.com	ofisegelsin.com
beyazdepo.com.tr	dentrealmarket.com	evzane.com	onlinekuyumcu.com
beyazsyacenter.com	dentstore.net	exa.com.tr	platformkitap.com

EK-4. Çizelge 4.1. E-ticaret sınıfını oluşturan internet siteleri (Devam)

beyazesyam.com	depodent.net	farkliyayincilik.com	saglikdemeti.com
beyazfirin.com	depokitap.com	farmasaglik.com	truvakitap.com
beyazsilver.com	deppo.com	farmasante.com	
beyazvitrin.com	dergikapinda.com	farukgulluoglu.com.tr	

EK-5. Çizelge 5.1. E-ticaret olmayanlar sınıfını oluşturan internet siteleri

1001araba.com	baloncu.gen.tr	damlagumus.com	fatihmetal.net
1e1indirim.com	baloncu.name.tr	damlakaya.com.tr	favoriboya.com
2elmotosiklet.com	balonlar.com	darmada.com	faziletmobilya.com
3anatural.com	balonpark.net	datehediyelikesya.com	fembilgisayar.com
3bdizayn.com.tr	balon-tr.com	daxguvenlik.net	fensiguzellik.com
724depo.com	balparmak.com.tr	decofire.com	feravize.com
abi.com.tr	banyodolaplari.info	defnemoda.com	fersanasansor.com.tr
abiye.tv.tr	basariburo.com	dekamuhendislik.com.tr	fezayapi.com
abiyeelbise.com	baskentmuhendislik.com	dekas.com	fideland.com
abiyegiyim.net	bayankemeri.com	dekorasyondizini.org	figes.com.tr
abiyeler.org	bayansporgiyim.com	dekorcini.com	figurin.us
acarka.com	bayilan.com	dekorcini.com.tr	filizdogaltas.com
acayipshop.com	bayrakajans.com	dekormarket.com	finahsapevler.com
acdc.com.tr	bayrak-diregi.com	dekormobilyam.com	fineline.com.tr
acdcelektronik.com	baysanbayrak.com	dekorsanat.com	fins.com.tr
acelik.ws	baysanlarcambalkon.com	dekosetbanyo.com	firintas.com.tr
acente.net	bbskart.com	demirelektronik.com	firsatvitriini.com
acilensatiyorum.com	bebekmarket.com	demrealestate.com	fitmarket.net
acilplaket.com	bebekodasial.com	deniz-bayrak.com	fiyatuzmani.com
aciltercume.com.tr	bebekstore.com	denizlerkitabevi.com	flamabayrak.com
aclperde.com	bebekurun.com	denizpazari.com	flexavantaj.webnode.com.tr
adalya.de	bedavamelodi.vazgectim.com	denizstore.com	flexxi.net
adampos.com	begendik.com.tr	dersaadet.com.tr	foks.com.tr
adanakentrehberi.com	belsas.com.tr	destinacicekcilik.com	formapromosyon.com
adanzye.com	bembilgisayar.com	develibazalt.com	forummarmaraavm.com
adaptordunyasi.com	bemolmusic.com	devrenilan.com	foto7-24.com
adaptormarket.com	berkedekor.com	diamond.com.tr	fotografkagitlari.com
address34.com	berkemutfakdolabi.com	didimfatihehlak.com	fotokopiservis.biz
adela.com.tr	berkeyapi.com	dijimecmua.com	fotomagix.com
adengiyim.com	berrakmakarna.com.tr	dijitalmuzik.com	fotoozen.com
adrenalin.com.tr	berrakticaret.com	dilani.com	fotoparca.com
adresmutfak.com	besiciler.com	dincermobilya.com	fotoremzi.com
adtmsilcilik.com.tr	bestkartne.com	discovery-tm.com	franke.com
advesreklam.com	bestnet.com.tr	diyetonline.net	fuarstandi.com
afled.tr.gg	bestyazilim.com.tr	dizaynperde.net	fulminakesuar.com
agackesen.com.tr	betadental.com.tr	dkm.com.tr	furkanpimapen.com
agacsan.com.tr	beyazbonbon.com	doganerreklam.com.tr	furuztektstil.com
ag-forkliftparts.com	beyazesyabildi.com	dogtas.com.tr	furyaorg.com
agroma.8m.com	beyazesyaservisler.com	doguscatering.com	galaksirehberi.com
ahsap.galerisi.com	beyazfrezya.com	dogusca.com.tr	galaxydesign.net
ahsapoyuncak.com	beyazlale.com.tr	dogusyapiotomasyon.com	galeriebru.com

EK-5. Çizelge 5.1. E-ticaret olmayanlar sınıfını oluşturan internet siteleri (Devam)

ahsapsan.com	beypazari.biz	domainhane.net	galvanizlisac.net
ajansglobal.com	bez-bebek.com	domainsiteniz.com	garantialsat.com.tr
ajanstekno.com	bgstore.com.tr	domainstandi.com	gazibilisim.com.tr
akademicicek.com	biarada.com	domainstandi.com	ge.com
akademikitavevi.com	biarastir.com	dorsankapi.com	gebeduve.tr.gg
akasyacicekcilik.com	bicerkardesler.com.tr	dortmevsimcicek.com	gegi.com.tr
akcelikboru.com	bilalart.com	dorukticaret.com.tr	gelal.com
akerbilardo.com.tr	bilalart.com	doryaspor.com	gelindamat.com
akhanpromosyon.com	biletix.com	dostemlakofisi.com	gelinlik.com.tr
akman-tekstil.com	bilgicemberi.com	doviz.com	gelinlika.com
akmen.com.tr	bilgisayarci.com	doyum.com.tr	gelinlikbak.com
akmtele.com	bilgiser.com.tr	doyummakarna.com	gelinlikler.name
akordion.org	bilgiset.com	dpperfumum.com.tr	gelinlikmodelleri.name
akorganikgubre.com	bilgitek.com.tr	drycenter.com	gelinliknet.com
akpinarsu.com	bilgiyayinevi.com.tr	drysoft.net	gelinsac.com
aksamto.com	bilkosis.com.tr	dugunajandasi.com	gems.com.tr
akspotsogutma.com	billuraydinlatma.com	dugunorganizasyonu.cc	genclersahaf.com
aksusekerleme.com	bilmak.com.tr	dunyasoft.com	generaloto.com.tr
aksutost.com	bilsenyufka.com	durakraft.com.tr	gentra.com.tr
aktivizyon.com	bimomak.com	duruesans.com	gereko.com
akyapidizayn.com	bind.com.tr	dusakabin.org	gevezetisort.com
alamac.com.tr	bintasmobilya.com	duvarkagididunyasi.com	gimkemer.com
albayrakotomotiv.com	bircicekyolla.com	duvarkagiditurkiye.org	gingersbutik.com
albenimobilya.com	birebir.com	duvarstickeri.com	girenbuluyor.com
albiaraba.com	birebiriletisim.com	duygugelinlik.net	giyimrehberi.com
albilnet.com	birmarket.com	duzcekontormarket.com	gizlimoda.com
aldimo.com	birordanbirburdan.com	dworldshop.com	gizteknik.com.tr
alfaklima.com.tr	bisiklet.com.tr	ebobinaj.com	globalkontrol.com
alfar.com.tr	bisikletforum.com	ebruhan.com	gokkusagicicek.com
algiy.com	bismak.com.tr	eceskibeton.com	goklerbijuteri.com
algorizma.com	bitiklabul.com	ececicek.com	goldenbilgisayar.com
alismobilya.com	bitkidiyari.com	eciceksepeti.com	goldgenie.com.tr
alissenuera.com	bizim-avrupa.com	ecrinicgiyim.com	gorkemilaclama.com
alisveriscanavari.com	bks.com.tr	edipenerji.com	gosterikostum.com
alize.org	bksam.com	efalarm.com	gozde-emlak.com
alkabin.com	bktasarim.com	efeoptik.com.tr	gozdeisguvenligi.com
alomarangozum.net	bky.com.tr	efes-ltd.com.tr	grandyavuzhotel.com
alonet.net	bodekgroup.com	efilateli.com	greens.com.tr
alpena.com.tr	bodrumnazaremlak.com	eforcadir.com	grundig.com.tr
alpera.org	boer.com.tr	eforgroup.net	gulaydincatering.com
alperen.com.tr	bolivo.com	egecadir.com	gulenoto.com

EK-5. Çizelge 5.1. E-ticaret olmayanlar sınıfını oluşturan internet siteleri (Devam)

alsancakbayrak.com.tr	boyunaskiipi.com	egeotoborsa.com	gulgayrimenkul.com
altinaramadedektoru.com	boyut.com.tr	egeotomatikkapi.com	gulmum.net
altinbezeme.com.tr	boyut3d.com	e-hosver.com	gultemizlik.com
altin-doviz.com	bozcaadaelektrik.com.tr	e-kent.com	gun-ay.com
alurad radyator.com	bridgestone-lastikturk.com	ekipsaat.com	gunayikinciel.com
alyansci.net	brita.tc	ekolbil.com	gunduztrafo.com
ambalajsektoru.com	bsabilgi.com.tr	ekontor.net	gunesintamicinde.com
anadolutel.com.tr	btm.com.tr	ekotekklima.com	gunestekstil.com
anahtarcilar.com	bulgurium.com	ekotel.com.tr	gunhanemlak.com
anahtarcozum.com.tr	burakcelikkapi.net	ekscambalkon.com	guralporselen.com.tr
andac.com.tr	burakcelikkasa.com	eksiogluvedenevenakliyat.com	gurayfidan.com
andajans.com	buraktreyler.com.tr	ektassigorta.com	gurbuzevaletleri.com
anelsatr.com	burambal.com.tr	elektal.com.tr	gurel.com.tr
anfomak.com	burayperuk.com	elektrikcimiz.com	gurlesin.com.tr
aniyayincilik.com.tr	burcakkumas.com	elektriklikombi.info	gurmobilya.com.tr
ankapotent.com	burcu.com	elektro-mak.com.tr	gurmobilya.tr.gg
ankarapromosyon.com	burcum.com.tr	elektromega.com	gursel.com
annebebe.com	butterfly.com.tr	elektrosonik.com.tr	gurtaks.net
antalyamigros.com	buyrunefendim.com	elektrospot.com	gurzemin.com
antalyaorganizasyon.tr.gg	byahsap.com	elephant.web.tr	guvender.com.tr
antalya-seriilan.com	bycicili.com	eliabilisim.com	guvenmanti.com.tr
antebim.com	byoral.com	elifnurhali.com	gym-center.com
anteplilercelikkasa.com	cadirci.com	elitshop.com	hakanmetal.com.tr
antes.com.tr	cadirdunyasi.com	elitstarcambalkon.com	hakayrulman.com
antikas.com	cagdasdiscephe.com	elmasis.com	halilali.com
apsis.com.tr	caglayan.com	elmaskasa.com	hamarateller.com
aquasafer.com	cagriomotiv.com	elmayayinevi.com	hamilegiyim.deyiz.biz
arakibulaki.com	cakmaklimarket.tr.gg	elsaofis.com	hancerli.com.tr
arascanta.com	calimera.com.tr	eltem.com.tr	harman.com.tr
arasmutfak.com	caliskangroup.com	emarket.com	harmantasarim.com
arasta.com	caliskanpazarlama.com.tr	ematem.com.tr	harput.net
arelchristin.com.tr	cambalkonum.com	emektugla.com	hasanguler.com
aribilisim.net	camine.com.tr	eminmakine.net	hasdem.com.tr
ariciboya.com	camlamabalkon.com	emiraluminyum.com	hasek.com.tr
arinoks.com.tr	campingcadir.com	emirudlari.com	haserekontrol.com
aripen.com	cancancatering.com	emkoelektronik.com.tr	hasermonopomp.com
arisanat.com	cankayaperde.com	emlak-ilan.net	haskonteyner.com
arisdiamond.com	canlitv.net	emlakmugla.com	hassuaritma.com.tr
arisutu.org	canon-servisi.com	emlakozdemir.com	hayatyapi.com.tr
aritabela.com.tr	cansukozmetik.com	emlakrehberi.com.tr	hayirliis.8m.net
aritasinsaat.com	cansupanel.com	emlaksahibi.com	hazalkozmetik.com

EK-5. Çizelge 5.1. E-ticaret olmayanlar sınıfını oluşturan internet siteleri (Devam)

ariyonca.com	cantaimalat.org	emmi.com.tr	hazirkontor.com
arliva.com.tr	careemlak.com	empasemlak.com	hazirpanelkapi.com
armadatekstil.net	carmenorganizasyon.org	enboyperde.com	hdkofismobilya.com
armadayazilim.com	carsi.com.tr	enerjitasarrufu.net	healthygermany.com
armaperde.com	carwax.com.tr	enginhavaiфик.com	hedefelektrik.com
artem.com.tr	cascahavyuz.com	enkay.com.tr	hediye24.net
artevetta.com	casperci.com	entertainmentgrup.com	hediyeal.org
artista.com.tr	casuscell.com	enucuzarac.com	hediyececell.com
aryada.com	catikaplamaizolasyon.net	enucuzbul.org	hediyeelikesya.info
arzumsan.com.tr	catkayapvc.com	enuygunteklif.com	hediyeelsak.com
asansekerleme.com	caykur.gov.tr	e-piknik.com	hediyeinyap.com
asilyayin.com.tr	cayyoluhalilikama.net	e-piramit.com	hediyeoner.com
askepenk.com	cecenplastik.com	epsilonyayinevi.com	hediyeonerisi.com
askimozbek.com	celikkapi.eu	erdembambu.com.tr	hediyepaketim.net
aslanbijuteri.com.tr	celikkasaci.net	erden.com.tr	hekimogluvedeneve.com
aslanguvenlik.com	celikray.com	erenler.com.tr	hekimogluvedenevenakliye.com
aslans.com.tr	cembotanic.com.tr	ergazadana.com	helenal.com.tr
asprefabrik.com	cemrereklam.net	erginkemer.com	helicamcopter.com
astech.com.tr	cemsunger.com	erkailan.com.tr	heliniletisim.tr.gg
astool.com.tr	cemtarim.com.tr	erkaseriilan.com	hemalhemsat.com
asudeorganizasyon.com	cennetemlak.net	erkekolsaati.org	hem-al-hem-sat.com
asyabant.com	cep.gen.tr	erkekyuzukleri.com	hemensat.com
asyalale.com	cepkeyfi.com	ermas.com.tr	hemkazanhemeglen.tr.gg
atacicek.com	ceptelhasanesi.com	erolmobilya.tr.cx	hepsiyorenden.com
atakol.com	cerenraf.com	erolozlem.com	hercebeuygun.com
atalarsapka.com	cesaperde.com	eromguvenlik.com	hermarkakontor.telekontor.com
atalayseramik.com	cesitliilanlar.com	erpa.com.tr	herseykampanya.com
ateksis.com	cesurbranda.com	er-pa.com.tr	hes.com.tr
atesanit.com	cetinler.com.tr	ersahingold.com	hesapliavm.com
atkiatkisi.com	cetsanhidrolik.com	ersaraf.com	hidrolikmarket.net
atkiimalat.com	cevhercelik.net	ersoyalyans.com	hiperteknoloji.com
atlas-muhendislik.com	ceyiz.com.tr	ertugrultuzemenemlak.com	hipokrat.com.tr
atomika.com.tr	ceylancelikkapi.com	eryigit.com.tr	hitbox.com.tr
atvkontor.com	ceylanceyiz.com	erzincantadin.com	hitechtr.com
autosamolepky.cz	cezayirantik.com	esemspor.com.tr	hobi2001.com
autoscout24.de	chipbayi.com	eskiesya.gen.tr	hobikozmetik.com
avantajkontor.com	cicek.cc	eskisehircilingir.net	holidaybau.com
avantajyakala.com	cicekalalim.com	esproit.com	homelet.com.tr
avbicagi.com	cicekci-cicek.com	estatrafo.com	homesantalya.com
avcinindunyasi.com	cicekcilik.us	esyaalanlar.com	hometransturk.com
avmgazette.com	cicekcisizin.com	eticaret.org	homyy.com

EK-5. Çizelge 5.1. E-ticaret olmayanlar sınıfını oluşturan internet siteleri (Devam)

avmindirim.com	cicekgold.com	etiketmobilya.com	horozyapi.com.tr
avon.com.tr	ciceklist.com	etiketonline.com	hugedomains.com
avonni.com	cicekstore.com	etimesguttasatilikdaireler.com.tr	huncular.com
avtateknoloji.com.tr	cicekzamani.com	etkilisozler.com	hurriyetoto.com
ayakkabidunyasi.com.tr	ciftcilerltd.com	etkinilanlar.com	huzuryayinevi.com.tr
ayanebru.com	cigdemcicek.com.tr	etm.com.tr	hys-avm.com
ayareletronik.com	cikas.com	etyumusatici.com	ica.com.tr
ayberticaret.com	cilamakinasi.net	eurochrom.com	iccamasirci.com
aydinaras.com.tr	cilek.com	eurofire-tr.com	ickiburada.com
aydinmobilya.com.tr	cinarboya.com	euromy.net	idas.com.tr
aygips.com.tr	cinargiyim.nice-forum.net	evbazar.com	ideaaksesuar.com
aykarmak.com	cinartesisat.com	evdeiskur.com	idealfirsat.com
aykarteknoloji.com	cinseluyaricilar.com	evdenekgelir.com	ihlastemizlikrobotu.net
aynayayinevi.com	cliffroad.com.tr	evdenisyap.com	ikiapc.com
ayreks.com.tr	cnm.com.tr	evduvarkagidi.com	ikincielaraba.net.tr
aytac.com.tr	cocacola.com.tr	evihan.com	ikincielesya.biz
aytasarim.com	cocuktancocuga.com	evimintaci.com	ikincielgazetesi.com
ayyildizdavetiye.com	coksatan.net	evlendikce.com	ikincielmobilya.com
ayyildizoto.com	colors2000.com	evlilikmarket.com	ikincielorg.com
bafayag.com	columbustemizlikmakinası.com	evlilikruyasi.com	ikincieloto.biz
bafayag.net	copmatik.com	evliliksiteleri.com	ikiyildiz.com
bahcecicek.com	coriano.com.tr	evsinemasistemleri.net	iklimsadegirmenonu.com
bakirkoyemlak.org	coskunet.com.tr	evtekspende.com	itsbeauty.com
balacouture.com	crowguvenlik.com	evturnakliyat.com	kayseritv.com
balciyakup.com	crystagel.com.mx	eyka.com.tr	mymachineparts.com
balgibialisveris.com	daigelinlik.com	ezcam.com.tr	seyrankoltuk.com.tr
balikcmarket.com	damla-etiket.com	fabervana.com.tr	tekinacar.com.tr
balimuzayede.com	damlaetiket.net	fancy-jewellery.com	webs.com
balon.com.tr	damlaetiketci.net	fastltd.net	

EK-6. Durak ve gereksiz kelimelerin listesi

“a, aa, acaba, altı, altmış, ama, ancak, arada, artık, asla, aslında, ayrıca, az, b, bana, bazen, bazı, bazıları, bazısı, bb, belki, ben, benden, beni, benim, beri, beş, bile, bin, bir, birçoğu, birçok, birçokları, biri, birisi, birkaç, birkaçı, birkez, birşey, birşeyi, biz, bizden, bize, bizi, bizim, böyle, böylece, bu, buna, bunda, bundan, bunlar, bunları, bunların, bunu, bunun, burada, bütün, c, cc, ç, çoğu, çoğuna, çoğunu, çok, çünkü, d, da, daha, dahi, dd, de, defa, değil, demek, diğer, diğeri, diğerleri, diye, doksan, dokuz, dolay, dolayısıyla, dört, e, edecek, eden, ederek, edilecek, ediliyor, edilmesi, ediyor, ee, eğer, elbette, elli, en, etmesi, etti, ettiği, ettiğini, f, fakat, falan, felan, ff, filan, g, gene, gg, gibi, göre, ğ, ğğ, h, hâlâ, halen, hangi, hangisi, hani, hatta, hem, henüz, hep, hepsi, hepsine, hepsini, her, her biri, herhangi, herkes, herkese, herkesi, herkesin, hh, hiç, hiç kimse, hiçbir, hiçbir, hiçbirine, hiçbirini, ı, u, i, için, içinde, ii, iki, ile, ilgili, ise, işte, itibaren, itibarıyla, j, jj, k, kaç, kadar, karşı, katrilyon, kendi, kendilerine, kendine, kendini, kendisi, kendisine, kendisini, kez, kırk, ki, kim, kimden, kime, kimi, kimin, kimisi, kimse, kk, l, ll, m, madem, mı, mi, milyar, milyon, mm, mu, mü, n, nasıl, ne, ne kadar, ne zaman, neden, nedenle, nedir, nerde, nerede, nereden, nereye, nesi, neyse, niçin, niye, nn, o, olan, olarak, oldu, olduğu, olduğunu, olduklarını, olmadı, olmadığı, olmak, olması, olmayan, olmaz, olsa, olsun, olup, olur, olursa, oluyor, on, ona, ondan, onlar, onlara, onlardan, onları, onların, onlari, onu, onun, oo, orada, otuz, oysa, oysaki, ö, öbürü, ön, önce, öö, ötürü, öyle, p, pek, pp, r, rağmen, rr, s, sadece, sana, sanki, sekiz, seksen, sen, senden, seni, senin, siz, sizden, size, sizi, sizin, son, sonra, ss, ş, şayet, şey, şeyden, şeye, şeyi, şeyler, şimdi, şöyle, şş, şu, şuna, şunda, şundan, şunlar, şunları, şunu, şunun, t, tabi, tamam, tarafından, trilyon, tt, tüm, tümü, u, uu, ü, üç, üü, üzere, v, var, vardı, ve, veya, veyahut, vv, y, ya, ya da, yani, yapacak, yapılan, yapılması, yapıyor, yapmak, yaptı, yaptığı, yaptığını, yaptıkları, yedi, yerine, yetmiş, yine, yirmi, yoksa, yüz, yy, z, zaten, zira, zz, acilen, açık, açıkça, açıktan, adamakıllı, adamca, adeta, ağırdan, ağızdan, ailece, ait, akabinde, akıllıca, aksine, akşamları, akşamüzeri, alabildiğine, alalecele, amiyane, amma, anca, ani, aniden, ansızın, antrparantez, aracılığıyla, aralık, aralıksız, aslen, atfen, aval, ayık, aynen, ayrı, azıcık, bade, bari, basbayağı, başkası, be, beden, bende, beraber, bilfiil, bilgece, bilhassa, birazdan, birazı, birbiri, birden, birdenbire, birileri, birlikte, bizatihi, bizimki, bizzat, bolca, boşuna, boyuna, böylelikle, böylesine, bugün, burası, caba, capcanlı, car, centilmence, cevaben, civil, cidden, ciddi, cömertçe, cumburlop, cümbür, çabucak, çabuk, çaktırmadan, çala, çan, çarçabuk, çaresiz, çatır, çıldır, çılgınca, çır, çtır, çoğun, çokluk, dahası, dahiyane, daim, daima, dandini, dank, dar, değin, dek, deminden, derhal, derinliğine, dışarı, dik, dimdik, dinen, direkt, diyerek, doğru, doğruca, doğrusu, doludizgin, dostane, dostça, durmadan, dün, dünden, düpedüz, dürüst, düşmanca, düşünce, düz, düzgün, ebediyen, eğik, eğreti, eğri, eh, eksik, elbet, eleştirel, ender, epey, epeyce, erkekçe, erken, erkenden, esefle, eskiden, etkili, evet, ewel, evvelce, evvel, ezberden, ezberi, ezberde, fahri, falanca, fazlaca, fena, fevkalade, fir, fii len, filanca, fütursuzca, galiba, gayri, gece, geceleri, geceli, geç, geçe, geçenlerde, gele, genellikle, genişçe, gerçek, gerçekten, gereğince, gerek, geri, gıcır, gırla, girt, gıyaben, gibisinden, giderayak, giderek, gine, gitgide, gittikçe, gizlice, görece, görünürde, görünüşte, götürü, gücün, güçlülük, günce, gündüz, gündüzleri, gü, güya, güzel, güzelce, güzellekle, habersiz, hafifçe, hafiften, hakikaten, hakkında, hala, halbuki, haliyle, hant, harıl, hariç, hart, hasbelkader, hatır, hayır, hayli, haza, hazır, he, helal, helalinden, hele, hemen, hepten, herhalde, heyetiyle, hı, hırıl, hızla, hızlı, homur, hop, hoş, hot, hüngür, ışıl, içten, içtenlikle, idarece, iğneleyici, ihtimal, iken, ila, ilaveten, ilelebet, ileride, ileride, ilişkili, ilk, ilkin, illa, ille, inadına, incecik, indinde, inim, insafsızca, insanca, isli, ister, istisnasız, itibarıyla, ivedilikle, iyi, iyice, iyilikle, izinsiz, kabaca, kabl, kaç, kadınca, kah, kala, kanalıyla, kanunen, kararınca, karşı, karşılıklı, karşılıksız, kasım, kasten, katı, katır, katıyın, kazaen, kazara, kazasız, kendiliğinden, kendinden, kerhen, kesinlikle, kestirmeden, keşke, keza, kışın, kırır, kıpır, kırım, kırt, kıs, kısa, kısaca, kısmen, kışın, kıtır, kıyasıyla, kıyasla, kolay, kolayca, kolaylıkla, kontra, korkakça, kös, kötü, küldür, küs, küt, lakın, langur, lap, lava, lop, lök, lütfen, maail, maalesef, maddeten, mademki, mahmur, mealen, mecburen, meğer, meğerse, memnuniyetle, merhaba, mesela, metazori, muhakkak, muhkem, muhtemelen, mukabil, mutlak, mutlaka, müft, münasebetiyle, müşterek, mütemadiyen, müthiş, naçizane, nadiren, nafiye, naklen, nasılsa, nedeniyle, nedense, nere, neredeyse, nereli, neresi, neye, neyi, nezdinde, nice, nispeten, nitekim, oldukça, onca, onda, ora, oradan, orası, ortaklaşa, ortalıkta, öğlen, öğleyin, önceden, önceleri, öncelikle, örneğin, öte, öylece, öylesine, özellikle, özgürce, pat, pattadak, pekala, peki, pervasızca, peşin, peşinen, peyderpey, pit, puf, pürdikkat, rahat, rahatça, rastgele, resmen, ruhen, sabah, sabahları, sabahleyin, sağlam, sahi, sahiden, sakarca, sakın, salimen, salt, samimiyetle, serbestçe, sessizce, seyrek, sıkça, sıkıca, sıradan, sıralı, sırasıyla, sırf, sonradan, sonraları, sorumsuzca, sözde, suçüstü, süratle, şahane, şahsen, şairane, şakkadak, şaşaalı, şık, şımarık, şıppadak, şimdiden, şurada, şurası, ta, tabiatıyla, takriben, tamamen, tedricen, tekrar, temiz, tepetakkak, tercihan, ters, tersine, tesadüfen, tez, tubben, tıpatıp, tıpkı, topluca, toptan, tuhaf, tutarak, tutkulu, tüm, tüm, uğrun, ustaca, ustalıkla, usulca, uyarınca, üst, üstüne, üzeri, üzerinde, üzerine, vasıtasıyla, vekaleten, vev, vira, yalandan, yalnız, yalnızca, yanlış, yanlışlıkla, yarı, yarin, yavaş, yavaşça, yayan, yayılı, yazık, yeniden, yerinde, yeter, yeterince, yığınla, yok, yoluyla, yudum, yukarda, yukarıda, yukarıdan, yürekten, yüzünden, yüzüstü, zahir, zamanında, zamanla, zararına, zıp, ziyade, zor, zoraki, zorla, zorlukla, adana, adıyaman, afyonkarahisar, ağı, aksaray, amasya, ankara, antalya, ardahan, artvin, aydın, balıkesir, bartın, batman, bayburt, bilecik, bingöl, bitlis, bolu, burdur, bursa, çanakkale, çankırı, çorum, denizli, diyarbakır, düzce, edirne, elazığ, erzincan, erzurum, eskişehir, gaziantep, giresun, gümüşhane, hakkari, hatay, iğdir, ısparta, istanbul, izmir, kahramanmaraş, karabük, karaman, kars, kastamonu, kayseri, kırıkale, kırklareli, kırşehir, kilis, kocaali, konya, kütahya, malatya, manisa, mardin, mersin, muğla, muş, neşehir, niğde, ordu, osmaniye, rize, sakarya, samsun, siirt, sinop, sivas, şanlıurfa, şırnak, tekiadağ, tokat, trabzon, tunceli, uşak, van, yalova, yozgat, zonguldak”

EK-7. %0 oranı ile elde edilen veri kümesi için anahtar kelimeler (Devam)

kırık, kırış, kırışık, kısıl, kısayol, kışkan, kısmi, kısık, kısım, kısır, kısıt, küt, kıta, kıvam, kıvanç, kıvıl, kıvılcım, kıvrı, kıvrıcık, kıy, kıyafet, kıyas, kıyma, kıymet, kıyı, kıyıla, kız, kızak, kızakla, kızar, kızart, kızartı, kızgın, kızmabirader, kızıl, kızılây, kızılcahamam, kızılderili, kızılçam, kızılötesi, kış, kışla, kışlak, la, labirent, laboratuvar, lacivert, laden, ladin, lagos, lahana, lahmacun, lain, lak, lake, laktoz, lal, lale, laleazar, lam, lama, lamba, lambala, lambri, lame, lane, lanetle, langirt, lanolin, largo, lastik, lata, lateks, latif, latife, latin, latince, lav, lavabo, lavanta, layık, layıkıyla, laz, lazer, lazım, lazımlık, laçin, lebib, lefköş, legal, leh, lehim, lehimle, lehçe, leke, lekele, lenf, lenger, lens, leopar, levazım, levent, levha, levye, ley, leyla, leyli, leziz, lezzet, libero, libre, lider, lif, lift, lig, lika, likit, likör, liman, limit, limited, limon, linet, link, linoilyum, linux, lira, lisan, lisans, lise, liste, listele, literatür, litografi, litre, lityum, liva, livar, loca, logaritma, logo, logos, lojistik, lok, lokal, lokma, lokman, lokomotif, lokum, lor, losyon, lotus, loğusa, lpg, lut, lösemi, lübnan, lüfer, lüks, lütfi, lütfullah, lütuf, mac, macar, macaristan, macera, macun, macunla, madalya, madalyon, madde, maddi, maden, madeni, maestro, mafya, magazin, magma, magnezyum, mahal, mahalle, mahalli, mahfuz, mahir, mahkeme, mahlep, mahmud, mahmudiye, mahmut, mahpus, mai, mail, main, majör, makale, makam, makara, makarna, makas, makasla, makedonya, maket, maki, makine, makro, maksat, maksimum, makul, makyaj, mal, mala, malakit, malezya, mali, malik, malikane, maliye, maliyet, malkara, malkoç, malt, malta, maltepe, maltız, malzeme, mama, mamak, mambo, mamul, mamut, manav, manavgat, mancınık, manda, mandal, mandalina, mandalla, mandar, mandolin, manevra, manga, mangal, mango, mani, manifesto, manikür, manken, manolya, mansur, mantar, mantarla, mantis, mantı, mantık, manyas, manyetik, manyetometre, manzara, manzur, manşet, manşon, mapa, maral, marangoz, margarin, marifet, marina, marjinal, mark, marka, markala, marke, market, markiz, marley, marmara, marmaris, marmelat, mars, mart, martin, martı, marul, marya, marş, mas, masa, masaj, masal, masat, masaüstü, masif, mask, maskara, maske, maskele, maskot, maslak, maslavi, mason, masraf, mastar, master, masum, masumiyet, masura, mat, matara, matbaa, matbu, matem, matematik, materyal, matiz, matkap, matris, maun, maval, mavi, mavna, maya, maydanoz, maymun, mayo, mayonez, mayıs, mazak, mazhar, mazi, mazlum, mazot, mazurka, maç, maça, mağara, mağaza, mağdur, maş, maşa, maşala, maşallah, maşrapa, meal, meb, meblağ, mebrure, mecbur, mecdi, meciydiye, meciydiyeköy, meclis, mecmua, mecrâ, medar, medeni, medeniyet, medikal, medine, meditasyon, medya, mega, megafon, mehdi, mehmet, mehtap, mekani, mekanik, mekanizma, mekik, meksika, mektup, mela, melahat, melami, melas, melda, melek, melezi, melih, meliha, melik, melike, melisa, melodi, meltem, meme, memnu, memnun, memnuniyet, memur, men, menderes, mendil, menekşe, menemen, menfez, mengen, mengene, meni, menopoz, menteşe, mentol, menus, menzil, menü, mera, merak, meral, meram, merasim, mercan, mercek, mercimek, merdan, merdane, merdanele, merdiven, meridyen, merinos, meriç, merkez, merkezi, mermer, mermi, mert, mertek, merter, merve, meryem, merzifon, mesafe, mesai, mesaj, mesane, mesel, mesele, mesh, meslek, mesleki, mest, mesut, met, meta, metafizik, metal, metalik, metalürji, metanet, mete, meteoroloji, metin, metis, metodoloji, metot, metre, metrekarre, metres, metrik, metro, metronom, metropol, metruk, metruke, mevcut, mevla, mevzana, mevlit, mevsim, mevzuat, meyan, meydan, meyhaneye, meyil, meyve, meyyal, mezar, mezat, meze, mezhep, mezopotamya, mezun, mezuniyet, mezura, meç, meşale, meşe, meşgul, meşhur, meşyon, mit, mitoloji, mizah, miço, mk, mobil, mobilya, moda, model, modelaj, modelle, modem, modern, modifikasyon, modül, modüler, mola, molekül, molla, moment, monitör, monografi, monopol, monoray, mont, montaj, monte, mor, morfoloji, motel, motif, motivasyon, motopomp, motor, motosiklet, mozaik, mozambik, muadil, muallim, muamele, muammer, muayene, mucid, mucit, mucize, mudanya, mudurnu, mufla, muhabbet, muhabir, muhafaza, muhafazakar, muhakeme, muhallesi, muhammed, muhammedi, muhammet, muharrem, muhasebat, muhasebe, muhayyile, muhiddin, muhit, muhlis, muhsin, muhtar, muhtar, muhtelif, muhterem, muhteşem, mukadder, mukaddes, mukafı, mukavva, mum, mumhane, mumla, mumya, murad, muradiye, murat, murathan, murt, murtaza, musa, musallat, musevi, muska, musluk, mustafa, mut, mutaf, mutfak, mutki, mutlu, muz, muzaffer, muzip, muştı, möble, mübarek, müberra, mücadele, mücahid, mücahit, mücevher, mücevherat, mücteba, müdahale, müdür, müfettiş, müfid, müfide, müfit, müge, mühendis, mühre, mührele, mühür, mühürdar, mühürle, müjde, müjgan, mükemmel, mülakat, mülayim, mülk, mülkiyet, mümessil, mümin, mümine, mümkün, münevver, münib, münir, münire, mürdüm, mürekkep, mürekkeple, mürsel, mürver, müsabaka, müşil, müslim, müslüman, müstakil, müstesna, müteahhit, müzeyde, müze, müzeyyen, müzik, müzikal, müzisyen, müşavir, müştehir, müşteri, mihla, miknatıs, mur, murra, murilda, mısır, mızıka, na, nabi, nabız, nacak, naci, nadide, nadir, nafaka, naftalin, nahit, nail, naim, nakib, nakil, nakit, nakkaş, nakli, nakliyat, nakliye, nakış, nakışla, nakışbendi, nal, nalan, nalbant, nalbur, naldöken, nallıhan, nalın, nam, namaz, name, namibya, namlu, namık, nane, nanoteknoloji, nar, nardin, narenciye, nargile, narin, narlıdere, narman, narçiçek, nas, nasr, nasrullah, nasuh, nasuhi, nasır, navigasyon, naylon, naz, nazan, nazar, nazariyat, nazif, nazife, nazik, nazlı, nazmi, nazım, nazır, nebi, nebil, necat, necati, neccar, necdet, necef, necip, necla, necmettin, nedim, nefes, nefis, nefise, nefret, negatif, nehir, nektar, nem, nemlen, nemrut, neon, nergis, nermin, nesefi, nesibe, nesil, nesim, nesir, neslihan, nesrin, net, netice, neva, nevi, nevin, nevres, nevresim, nevzat, ney, nezih, nezihe, nezir, nezt, neşe, neşet, neşir, neşriyat, nida, nigar, nihad, nihâl, nihale, nihan, nihât, nihavend, nihayet, nikah, nikan, nikaragua, nikel, nikelaj, nikotin, nil, nilay, nilgün, nilüfer, nimet, nine, ninni, nisan, nispi, nitele, nitelik, niyazi, niyet, niza, nizam, nizami, nişan, nişanla, nişasta, noel, nohut, nokta, noktala, norm, normal, norveç, nostalji, nostalgik, not, not, noter, nova, nuh, numan, numara, numarada, numune, nur, nuran, nuray, nurbaki, nurdan, nureddin, nurefşan, nurettin, nurgül, nurhan, nuri, nursel, nursen, nurten, nurullah, nurşen, nusret, nutki, nutuk, nöbet, nöroloji, nötr, nüfuz, nükleer, nüve, nüzhet, obje, objektif, obruk, ocak, od, oda, odak, odakla, odun, of, ofis, ofset, ogün, oh, ohm, oje, ok, okan, okay, okey, oksijen, oksit, oktan, oktay, oku, okul, okur, okutan, okutman, okuyan, okyanus, okyar, okyay, ol, olanak, olay, olcay, olgu, olgun, olguner, olimpiik, olimpiyat, olivin, olmadık, olta, oltu, oluk, olumlu, olumsuz, oluş, oluşum, om, oma, omler, omurga, omurilik, onar, onaran, onat, onay, onayla, onbaşı, oniks, onkoloji, onlarca, onur, onurhan, onursan, opal, operasyon, operatör, opsiyon, optik, optimist, optimum, orak, oral, oran, oranla, org, organ, organik, organizasyon, organize, organizma, orgun, orhan, orhangazi, orijinal, orkide, orkun, orman, orta, ortak, ortaköy, ortalı, ortalama, ortam, ortaokul, ortay, ortodonti, ortopedi, ortopedik, oruç, oryantal, oskay, osman, osmancık, osmangazi, osmanlı, osmanlıca, ot, otac, otacı, otantik, otel, oto, otobiyografi, otobiyografik, otobüs, otogar, otokar, otoklav, otomasyon, otomat, otomatik, otomobil, otomotiv, otopark, topsi,

EK-7. %0 oranı ile elde edilen veri kümesi için anahtar kelimeler (Devam)

otorite, otur, oturacak, oturak, oturum, ov, ova, oval, ovala, oy, oya, oyal, oyla, oyna, oyun, oyuncak, ozalit, ozan, ozon, ozonla, oğlak, oğlan, oğul, oğuz, oğuzalp, oğuzhan, pabuç, padişah, pafta, paha, pahal, pahalı, paket, pakette, pakistan, paksan, paksoy, paksu, pal, pala, palan, palandöken, palas, palaska, palet, palmiye, palto, palyaço, pamir, pamuk, panama, panda, panel, panik, panjur, pano, panorama, pansiyon, pansuman, panter, pantograf, pantolon, panzehir, papa, papatya, papağan, papirüs, papyon, para, parabol, paradigma, paraf, parafin, paraguay, paralel, paravan, parazit, paraşüt, pardus, pardösü, parfüm, parfümeri, park, parka, parke, parla, parlak, parmak, parola, pars, parsa, parsel, parselle, parti, partner, parttime, parça, parçala, parşömen, pas, pasaj, pasaport, pasif, paslan, paspas, paspasla, pasta, pastal, pastel, pastil, pastörize, pastırma, pata, patates, paten, patent, pati, patik, patinaj, patla, patlak, patlıcan, patoloji, patolojik, pay, payam, paydos, payla, paylaş, pazar, pazarla, pazartesi, paça, paşa, pedagog, pedal, pediatri, pedikür, pedodonti, pedometre, pehlivan, peker, pekmez, pekşen, pelerin, pelesenk, pelikan, pelin, pelit, pelüş, pembe, pencere, pendik, penguen, penis, penisilin, pens, pense, penye, pençe, pepe, perakende, perdah, perde, perdele, performans, pergel, pergelde, pergola, peri, peridot, perihan, periyodik, perlon, perma, peroksit, peron, personel, perspektif, peru, peruk, pervane, pervari, pervaz, pervin, perviz, perçin, perşembe, pes, petek, petrol, petunya, peyami, peygamber, peyman, peynir, peyzaj, peçete, peş, peştamal, pi, pigment, pijama, pik, pikap, pike, piknik, piko, pil, pilav, piliç, pilot, pim, pinhan, pinpon, pipet, pipo, pir, piramit, pire, piring, pirit, pis, pislik, pist, piston, piyade, piyale, piyango, piyanist, piyano, piyasa, pizza, piş, pişik, pişir, pişman, pişmaniye, pişti, plaj, plak, plaka, plaket, plan, planet, planetaryum, planla, planya, planyala, plaster, plastik, platform, platin, plato, platonik, plazma, po, poker, polat, polatlı, polemik, polen, polietilen, poligon, poliklinik, polimer, polis, polisaj, polisiye, politika, polişe, poliüretan, polo, polonez, polonya, polyester, pomat, pompa, pompala, ponpon, pop, popüler, popülizm, porselen, porsiyon, porsuk, portakal, portatif, portbebe, portekiz, portföy, portmanto, porto, portre, pos, post, posta, postala, poster, postiş, pot, pota, potansiyel, potas, pound, poyraz, pozantı, pozisyon, pozitif, poğaç, poşet, poşetle, pratik, pratika, prematüre, prens, prenses, prensip, pres, presle, prestij, presto, prezervatif, prim, primat, printer, priz, prizma, problem, prodüksiyon, profesyonel, profesör, profil, program, programla, proje, projeksiyon, projektör, promosyon, prosedür, proses, prostat, protein, protesto, protez, protokol, proton, prova, prusya, pruva, prömiyer, psikiyatri, psikiyatrist, psikoloji, psikolojik, psikoterapi, ptt, puan, puanla, puding, pudra, pudrala, pufla, puhu, pul, pulla, puma, punto, pupa, puro, pus, puset, pusula, pörtlek, püf, pür, püre, püren, pürüz, püskül, püskür, pütür, pınar, pınarbaşı, pır, pırlanta, pırl, pırlıltı, quantum, rabbani, rabi, rabia, raci, radar, radife, radikal, radyan, radyasyon, radyatör, radyo, radyoaktif, radyografi, radyoloji, raf, rafet, rafine, rafting, rafya, ragıp, rahatla, rahim, rahime, rahle, rahman, rahmi, raid, raif, rakam, rakamla, racket, rakip, rakkase, rakor, raks, raki, rakim, ram, ramazan, rami, rampa, randevu, ranza, rap, rapor, raporla, raptiye, rasathane, rasih, rasim, raspa, raspala, raf, ray, rayba, razi, razı, rağbet, rc, re, reaktör, reasürans, recai, recep, redaksiyon, redresör, refakat, referans, refik, refika, refleks, reflektör, regülatör, rehabilitasyon, rehber, reis, rekabet, reklam, rekor, remiz, remzi, remziye, rende, rende, rengarenk, renk, renkle, reproduksiyon, resen, resif, resim, resimle, resital, resmi, ressam, rest, restoran, restorasyon, restore, resul, retina, revir, revolver, revü, rey, reyhan, reyon, rezan, reze, rezene, rezervasyon, rezil, rezistans, reçel, reçete, reçine, reşat, reşat, reşit, rica, rika, ring, risale, risk, ritm, riva, riya, robot, rodaj, rodeo, rodyum, roket, rol, rom, roma, roman, romantik, romantizm, romanya, romatizma, rot, rota, rotasyon, rotıl, rotor, roza, rozet, rtük, ruble, ruby, rugan, ruh, ruhi, ruhsal, ruhsat, ruj, rujla, rukiye, rulman, rulo, rum, rumeli, rumi, rus, rusya, rusça, röfle, röle, rölyef, römork, röntgen, röportaj, rötür, rötüş, rüknetin, rüstem, rütbe, rüya, rüzgar, rüşti, rüşvet, rıdvan, rıfat, rih, rihtim, rıza, sa, saadet, saat, saba, sabahattin, saban, sabi, sabiha, sabit, sabitle, sabo, sabri, sabun, sabunla, sabır, sac, sacayak, sacayağı, sacide, sada, sadak, sadakat, sadaret, sadberk, sade, sadet, sadeyağ, sadi, sadreddin, sadri, sadullah, sadık, saf, safa, safari, safer, saffet, safi, safir, safiye, sahra, safran, safranbolu, safvet, saha, sahabe, sahaf, sahan, sahih, sahil, sahip, sahne, sahnele, sahra, sahte, sahtekar, sahur, said, saim, saimbeyli, saime, sak, saka, sakal, sakar, sakat, sakata, saki, sakin, sakinle, sakla, saklan, sakı, sakı, sakınca, sakız, sala, sala, salacak, salah, salam, salamura, salat, salata, salatalık, saldırı, salep, salgıla, salih, saliha, salim, salime, salkım, salla, sallantı, salman, salon, saltanat, saltık, salyangoz, salça, salçala, salı, salık, salın, salıncak, salınım, saliver, sam, saman, samanyolu, samet, sami, samur, samuray, san, sanal, sanat, sanatkar, sanatoryum, sanayi, sancak, sancar, sancı, sandalet, sandalye, sandviç, sandık, sandıkla, sandıklı, saner, sani, saniye, santra, santral, santrifüj, sap, sapa, sapan, sapanca, sapla, saplantı, sapta, sapık, sar, sara, sarar, sararti, saray, saraç, sarban, sardunya, sarf, sargut, sargı, sari, sark, sarkaç, sarkıt, sarma, sarmal, sarmaşık, sarnıç, sarp, sarraf, sarrafiye, sars, sarsıl, sari, sarık, sarıkamış, sarıkaya, sarıkız, sarımsak, sariyer, sat, saten, satranç, satürn, satı, satıcı, satılık, satır, satış, sauna, sav, sava, savan, savana, savaş, savaştı, savcı, saver, savni, savun, savunma, say, saya, sayar, sayaç, saydam, saye, sayfa, saygı, saygın, sayı, sayıl, sayılı, sayılı, sayım, sayın, sayısız, sayısız, sazan, saç, saçak, saçma, saçı, sağ, sağla, sağlamer, sağlıklı, sağmal, sağ, sağın, sağır, seans, seba, seben, sebeb, sebil, sebze, seccade, secde, seci, seda, sedat, sedef, sedir, sedye, sefa, sefer, seferi, seferihisar, segman, seher, sehpa, sek, sekizgen, sekreter, seks, sektör, sel, selahaddin, selam, selami, selamla, sele, selen, selim, selin, selis, selma, selman, selva, selçuk, selçuklu, selülit, selülozik, sema, semaver, semavi, semazen, sembol, semer, semerle, semi, semih, semiha, semine, seminer, semir, semiz, sempatik, sempozyum, semra, semt, sene, senaryo, senatör, sendrom, sene, senem, senet, senih, seniha, senkronizasyon, sensör, sentetik, sentez, sepet, sepetle, ser, sera, seramik, serbest, sercan, serdar, seren, serenat, sergi, sergile, serhan, serhat, seri, serigrafi, serik, serim, serin, serkan, serkiz, sermaye, sermin, serp, serpil, serra, sert, sertel, sertifika, sertlik, serum, server, servet, servi, servis, serçe, serüven, ses, sesle, sesli, sessiz, set, sev, seval, sevcan, sevdä, seven, sevengül, sever, sevgi, sevgilim, sevi, sevil, sevim, sevimli, sevin, sevinç, seviye, sevk, sevkıyat, seyahat, seyahatname, seydişehir, seyhah, seyfettin, seyfi, seyfullah, seyhan, seyit, seyitgazi, seylan, seyrel, seyret, seyyar, seyyid, seyyit, sez, sezai, seza, sezek, sezen, sezer, sezi, seziğin, sezi, sezon, seç, seçenek, seçi, seçil, seçim, seçki, seçkin, seçmen, seçir, seçirt, si, sibel, sicil, sifon, sifonla, sigar, sigara, sigorta, sigortacı, sigortala, sihir, sihirbaz, sık, sika, sikke, sikkele, siklon, sil, silah, silahla, silecek, ilgi, silikon, silindir, silindirik, silis, silivri, silk, silme, silsile, sim, sima, simge, simgele, simit, simurg, simya, simülasyon, sin, sina, sinagog, sinameki, sinan, sincan, sincap, sindir, sindirim, sine, sinek, sineklik, sinema, sinerji, sini, sinir, sinirli, sintine, sinyal, sinüs, sinüzit, sipariş, siper, siren, sirk, sirke, sirkülasyon, siroz, sirrus, sirtaki, sis, sistem, sistematik, sistit, site, sitem, sitil, siva, siverek, sivil, sivilce, sivri, sivrihisar, sivrisinek, sty, siya, siyah, siyasal, siyaset, siyasi, siyonist, siğil, ski, slip, slogan, slovakya, slovenya, smokin, sn, soba, soda, sodyum, sof, sofa, sofı, sofistike, sofrä, sohbet, sok, sokak, soket, sol, solaryum, solfej, solgun, solla, solmaz, solo, solu, solunum, solüsyon, som, soma, somer, somon, somun, sonar, sonat, sonbahar, sonda, sondala, soner,

EK-7. %0 oranı ile elde edilen veri kümesi için anahtar kelimeler (Devam)

yavuzalp, yavuzcan, yavuzer, yay, yaya, yaygı, yaygın, yayla, yaylım, yayıl, yayım, yayımla, yayın, yayınla, yaz, yazar, yazgan, yazlık, yazma, yazman, yazı, yazıla, yazılım, yazın, yazır, yazıt, yağ, yağcı, yağdanlık, yağla, yağmur, yağmurla, yağın, yağız, yaş, yaşa, yaşam, yaşantı, yaşar, yaşlan, ye, yed, yedek, yedekle, yediveren, yegane, yek, yekpare, yel, yelda, yeldan, yelek, yelekle, yelken, yelpaze, yem, yemek, yemen, yemeni, yemin, yemle, yener, yeni, yenile, yenimahalle, yenisey, yeniçeri, yeniş, yenişehir, yepyeni, yer, yeraltı, yerel, yerin, yerköy, yerleş, yerli, yeryüzü, yet, yetenek, yeterli, yetersiz, yeti, yetik, yetim, yetiş, yetişkin, yetki, yetkin, yeğ, yeğen, yeğın, yeğle, yeşer, yeşil, yeşilova, yeşilyurt, yeşim, yitik, yiyecek, yiğit, yoga, yokuş, yol, yolcu, yoldaş, yolla, yolluk, yolsuz, yon, yonca, yonga, yor, yordam, yorga, yorgan, yorgun, yorum, yorumla, yosun, yoğ, yoğun, yoğur, yoğurt, yoğuş, ysk, yıl, yufka, yukarı, yulaf, yum, yuma, yumak, yumurcak, yumurta, yumurtalık, yumuşa, yumuşacık, yumuşak, yun, yunan, yunanca, yunanistan, yunus, yurdağül, yurt, yurttaş, yusuf, yusufoğlu, yuva, yuvalla, yuvarla, yuvarlak, yön, yönel, yönelik, yönet, yönetim, yönetmelik, yönetmen, yönel, yöntem, yöre, yöruk, yüce, yücel, yük, yükle, yüklem, yüklük, yüksek, yüksekova, yüksel, yükselti, yüküm, yünni, yün, yüreğ, yüreğir, yürü, yürüt, yürüteç, yüzbaşı, yüzde, yüzey, yüzle, yüzlerce, yüzyıl, yüzük, yık, yıka, yıl, yılan, yılbaşı, yıldır, yıldırım, yıldız, yıldızla, yilla, yılmaz, yıpran, yurt, yırtmac, zafer, zahide, zahit, zahmet, zaid, zaim, zakir, zaman, zambak, zambiya, zar, zara, zarar, zarf, zarfla, zargana, zarif, zatürre, zavallı, zayıf, zayıfla, zebra, zeheb, zehir, zehra, zeka, zekai, zekeriya, zeki, zekiye, zeliha, zemin, zemzem, zen, zencefil, zenci, zengin, zenginle, zenne, zeplin, zer, zerdeçal, zerrin, zevk, zeybek, zeynel, zeynelabidin, zeynep, zeytin, zeytinburnu, zeytinlik, zeytinyağ, zift, zihin, zihni, zil, zile, zimmet, zincir, zincirle, zinde, ziraat, zirkon, zirve, ziverbey, ziya, ziyaret, ziyaretçi, ziyet, zoka, zombi, zona, zootehni, zorlu, zorunlu, zorunluluk, zum, zübeyde, zübeyir, zübeyr, zülal, zülüf, zümrüt, zürafa, zıbın, zimba, zimbala, zımpara, zımparala, zipkin, zipkinla, zıpla, zipzip, zirh, zivir, çaba, çabala, çadır, çak, çakal, çakar, çakmak, çaki, çakıl, çakır, çakşır, çal, çalar, çalgı, çalı, çalık, çalım, çalış, çalışkan, çalıştay, çam, çamaş, çamaşır, çamlidere, çamur, çanak, çandır, çankaya, çanta, çap, çapa, çapak, çapala, çapan, çapanoğlu, çapari, çapraz, çardak, çare, çark, çarkifelek, çarp, çarpı, çarpış, çarık, çarşaf, çarşafı, çarşamba, çarşı, çat, çatak, çatal, çatalca, çatla, çatlak, çatı, çavuş, çay, çaydanlık, çayır, çayırılı, çağ, çağa, çağanoz, çağatay, çağdaş, çağla, çağlar, çağlayan, çağman, çağrı, çağrış, çağıl, çağın, çağır, çebi, çek, çekecek, çeker, çekerek, çeki, çekil, çekiliş, çekim, çekimle, çekin, çekirdek, çekirge, çekic, çekicile, çekiş, çekme, çekmece, çekyat, çekçek, çekül, çeledi, çelenk, çelik, çelikel, çelikle, çelişki, çember, çemen, çene, çengel, çer, çerez, çeri, çerkezköy, çerkeş, çerçeve, çerçevele, çete, çetin, çevik, çevir, çeviri, çevirme, çevirmen, çevre, çevren, çevri, çevrim, çeyiz, çeyrek, çece, çeçen, çeşit, çeşitle, çeşitli, çeşme, çeşni, çift, çifte, çiftle, çifteler, çiftlik, çiftçi, çiklet, çikolata, çile, çilek, çilimli, çilingir, çim, çimen, çin, çince, çini, çinko, çip, çipura, çiroz, çise, çit, çita, çiti, çitlembik, çivi, çivile, çiz, çizge, çizgen, çizgi, çizi, çizik, çizim, çizme, çiçek, çiçekle, çiğ, çiğde, çiğdem, çiğli, çiğne, çiğnem, çış, çoban, çocuk, çolak, çomak, çor, çorak, çorap, çorba, çorlu, çorman, çoruh, çotanak, çoğal, çoğul, çoğunluk, çubuk, çukurova, çul, çuval, çök, çöker, çöm, çömlek, çöp, çörek, çöz, çözüm, çözümle, çözün, çözünürlük, çürü, çürük, çıban, çık, çıkar, çıkarma, çıkart, çıkma, çıkmaz, çıkırık, çıktı, çiki, çikin, çıkış, çılgın, çınar, çınarcık, cıngırak, çınla, çıplak, çıra, çirak, çırp, çırpma, çırpı, çırpıcı, çırçır, çırılçıplak, çıt, çita, çitçit, çiğ, çiğir, öbek, öbür, öcal, öd, öde, ödem, ödemiş, ödül, ödüllendir, ödüün, öfke, öge, öke, öksürük, öksüz, ökten, ökten, öl, öldür, ölç, ölçek, ölçü, ölçümle, ölü, ölüm, ölümcül, ömer, ömerli, ömür, önal, öney, öncü, önder, önem, önemse, önen, öner, öneri, öngör, önle, önlem, önlük, öp, öpücük, ör, ördek, ören, örge, örgen, örgü, örgün, örgüt, örgütle, örnek, örmekle, örs, ört, örtü, örümcek, öss, öt, öteki, ötüken, öv, öveç, övünç, öykü, öyküle, öz, özak, özal, özalp, özbek, özbekistan, özbek, özdek, özdeş, öze, özel, özen, özet, özelle, özge, özgü, özgül, özgüle, özgün, özgür, öze, özlem, özne, özsever, özü, ögle, öğren, öğrenci, öğret, öğreti, öğretim, öğretmen, öğüt, ücret, ücretle, üfle, ülke, ülken, ülker, ülkü, ülser, ümit, ümmühan, ümran, ümraniye, ün, ünäl, ünalan, üniforma, ünite, üniversal, üniversite, ünlü, ünlüler, ünsal, ünver, ünverdi, ünzile, ünüvar, ürdün, üre, ürem, üret, üretim, ürk, üroloji, ürtiker, ürün, üs, üsküdar, üste, üstelik, üstlen, üstün, üstündağ, üt, ütopya, ütü, ütüle, üveyik, üye, üz, üzer, üzgü, üzgün, üzül, üzüm, üzüntü, üçle, ihlamur, ilga, ilgar, ilgaz, ilgın, ilıcak, ilık, ilım, ırk, ırmak, ırs, ıskarmoz, ısla, ıslah, ıslak, ıslan, ıslık, ısmarla, ıstampı, ısı, ısın, ısır, ısırgan, ısıt, ıtriya, ıtır, ıvır, ızgara, ışı, ışıık, ışııklan, ışılda, ışıldak, ışıltı, ışın, ışınsu, şaban, şablon, şadan, şadi, şadiye, şafak, şaft, şah, şahan, şahin, şahinbey, şahiner, şahit, şahsiyet, şahtere, şahıs, şaik, şair, şaka, şakir, şal, şalgam, şalter, şalvar, şamandıra, şamdan, şamil, şampiyon, şampuan, şampuanla, şan, şanlı, şans, şantiye, şapka, şarani, şarap, şarj, şarjör, şark, şarküteri, şarkı, şart, şartla, şase, şasi, şavk, şaziye, şaş, şaşkın, şaşır, şeb, şebboy, şebeke, şebnem, şef, şeffaf, şefik, şefika, şeftali, şehir, şehit, şehitkamil, şehmini, şehri, şehriban, şehriye, şehzade, sek, şeker, şekerle, şekerleme, şekerpare, şekil, şekillen, şekli, şeker, şelale, şema, şems, şemseddin, şemsettin, şemsi, şemsinur, şemsiye, şen, şencan, şener, şengil, şengül, şeniz, şenkal, şenlik, şentürk, şerafeddin, şerbet, şerbetle, şeref, şereflikoçhisar, şerh, şerif, şerife, şerit, şeritle, şevket, şeyban, şeyda, şeyh, şeytan, sezlong, şeşbeş, şiddet, şifa, şifon, şifre, şifrele, şür, şikayet, şile, şilte, şimal, şimşek, şimşir, şinasi, şipşak, şiraz, şirin, şirket, şirvan, şist, şizofren, şizofreni, şiş, şişe, şişele, şişhane, şişir, şişkin, şişli, şişman, şnorkel, şofben, şok, şort, şua, şuayip, şubat, şube, şule, şurup, şöbiyet, şölen, şömine, şömez, şövale, şövalye, şükran, şükri, şükür, şüphe, şımar, şınav, şıpsavdi, şırına

EK-8. %5 oranı ile elde edilen veri kümesi için anahtar kelimeler

“ad, ada, adet, adres, ahşap, aksesuar, al, ala, alan, alet, alt, altın, alın, alışveriş, amaç, an, ana, anahtar, anne, ar, ara, araba, aranjman, araç, arkadaş, at, ay, ayak, ayakkabı, ayn, aç, ağ, bahçe, bak, bakı, bakım, bank, banka, banyo, bas, baskı, bası, bay, bayan, bayi, baş, bebek, beyaz, bil, bildir, bildiri, bildirim, bilgi, bilgisayar, bitki, boy, boş, buket, bul, bulun, büyük, cam, can, cep, cihaz, cm, dahil, deniz, deri, destek, detay, deva, devam, değ, değiş, dijital, dil, doğu, doğum, duvar, dünya, düğ, düğü, düğün, dış, edin, ek, ekle, ekran, el, elektrik, elektronik, erkek, et, ev, eğit, eğitim, eşya, fark, firma, fiyat, form, fotoğraf, fırsat, garanti, gel, gelin, geliş, genel, geçmiş, gir, giriş, git, giy, giyim, giz, gizli, grup, gönder, gönderi, gör, görün, göster, göz, gül, gün, güneş, günü, güven, ha, haber, haberdar, hafta, hak, hal, harita, hatırla, hava, havale, hazırla, hediye, hesap, hizmet, hız, iade, il, ilet, in, indir, iste, içerik, iş, işle, işlem, kablo, kal, kalite, kalp, kamera, kampanya, kap, kapı, kar, kargo, kart, kas, kasa, kat, katalog, kategori, kay, kaynak, kayıt, kdv, kes, kitap, kişi, kol, koltuk, koru, koşul, kredi, kulak, kullan, kur, kuru, kurum, kut, kutu, küçük, kırmızı, liste, makine, malzeme, mark, marka, markala, masa, mavi, mağaza, merkez, merkezi, mesafe, mini, model, motor, mutfak, müzik, müşteri, numara, oda, ofis, oku, ol, oluş, orkide, orta, oto, oy, oyun, oyuncak, paket, parça, pembe, politika, renk, resim, saat, sade, sak, saklı, saksı, san, sat, satı, satış, sayfa, sayı, saç, sağ, sağlık, sepet, seri, servis, ses, set, sev, sevgi, seç, seçenek, seçi, sipariş, sistem, site, siyah, sor, soru, spor, stok, su, sun, söz, sözleş, sür, süre, sık, tablet, tak, takip, taksit, takı, takım, taraf, tarih, tasar, tasarımı, taş, taşı, tek, teknik, tekstil, tel, telefon, temizle, teslimat, tip, tipi, tl, top, toplam, tv, tüket, türk, türkiye, tıkla, ucuz, ula, ulaş, unut, uy, uygula, uygun, uç, vazo, ver, video, yap, yapı, yar, yardım, yas, yatak, yaz, yazı, yazılım, ye, yeni, yer, yol, yıl, zaman, ziyaret, ziyaretçi, çal, çalış, çamaşır, çanta, çek, çeşit, çift, çiçek, çiçekle, çocuk, çöz, çözüm, çık, öde, ödem, özel, ücret, üre, üret, ürün, üs, üye, ısıt, şart, şekil, şifre”

ÖZGEÇMİŞ

Kişisel Bilgiler

Soyadı, adı : KAŞIKÇI, Tuğba
Uyruğu : T.C.
Doğum tarihi ve yeri : 31.01.1988 Konya
Medeni hali : Evli
Telefon : 0 (545) 256 46 65
Faks : -
e-mail : ozdemir.tugba@gazi.edu.tr.

Eğitim

Derece	Eğitim Birimi	Mezuniyet tarihi
Lisans	Çankaya Üniversitesi / Endüstri Mühendisliği	2010
Lise	Mimar Sinan Lisesi	2005

Yabancı Dil

İngilizce (A)

Hobiler

Bilgisayar teknolojileri