



**T.C.
İSTANBUL ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ**



DOKTORA TEZİ

**MAKİNE ÖĞRENMESİ TEKNİKLERİNİ KULLANARAK TÜRKÇE
METİNLERDEN DÜŞÜNCE ÇIKARIMI VE DUYGU BELİRLEME**

Mine MERCAN

Bilgisayar Mühendisliği Anabilim Dalı

Bilgisayar Mühendisliği Programı

**DANIŞMAN
Prof. Dr. Ahmet SERTBAŞ**

Temmuz, 2018

İSTANBUL

Uygundur (17.09.2018)

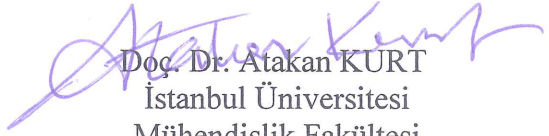
**Prof. Dr. Ahmet SERTBAŞ
Bilgisayar Mühendisliği
Bölüm Başkanı**

Bu çalışma, 12.07.2018 tarihinde aşağıdaki jüri tarafından Bilgisayar Mühendisliği Anabilim Dalı, Bilgisayar Mühendisliği Programında Doktora tezi olarak kabul edilmiştir.

Tez Jürisi


Prof. Dr. Ahmet SERTBAŞ (Danışman)
İstanbul Üniversitesi
Mühendislik Fakültesi


Prof. Dr. Selim AKYOKUŞ
Doğuş Üniversitesi
Mühendislik Fakültesi


Doç. Dr. Atakan KURT
İstanbul Üniversitesi
Mühendislik Fakültesi


Doç. Dr. Rüya ŞAMLI
İstanbul Üniversitesi
Mühendislik Fakültesi


Doç. Dr. Veli HAKKOYMAZ
Yıldız Üniversitesi
Elektrik-Elektronik Fakültesi



20.04.2016 tarihli Resmi Gazete’de yayımlanan Lisansüstü Eğitim ve Öğretim Yönetmeliğinin 9/2 ve 22/2 maddeleri gereğince; Bu Lisansüstü teze, İstanbul Üniversitesi’nin aboneli olduğu intihal yazılım programı kullanılarak Fen Bilimleri Enstitüsü’nün belirlemiş olduğu ölçütlere uygun rapor alınmıştır.

ÖNSÖZ

Doktora eğitimim boyunca bilgi ve tecrübesiyle yardımlarını eksik etmeyen her görüşmemizde yeni bir ufuk açan danışman hocam Prof. Dr. Ahmet SERTBAŞ'a,

Yüksek Lisans ve Doktora çalışmalarımda tecrübeleri ve bilgisiyle yardımlarını esirgemeyen hocam Doç. Dr. Atakan KURT'a,

Tez izleme jürimde bulunan ve bilgi ve tecrübesiyle tezime olan katkılarından dolayı hocam Prof. Dr. Selim AKYOKUŞ'a,

Veri bulmakta desteklerini eksik etmeyen bilgi ve tecrübelerinden yararlandığım Prof. Dr. Nevzat TARHAN'a, Doç.Dr. Levent METE'ye ve asistanlarına,

Birlikte çalışmaktan zevk aldığım bir çalışma arkadaşı olarak desteğini hep hissettiğim, bir çok alandaki bilgi ve tecrübesini benimle paylaşan, güzel sohbetler yaptığım Migena CEYHAN'a,

Hayatımı güzelleştiren ve hedeflerime giden yolda beni hep destekleyen kardeşim Fuat Kıygın ve eşi Filiz Kıygın'a,

Her zaman yanımda olan, bugünlerime çok katkıları olan annem Raziye Kıygın ve babam Nail Kıygın'a,

Son olarak yoğun çalışma dönemlerimde her zaman yanımda olan çocuklarım Fatih Mercan ve Alperen Mercan'a,

Teşekkür ederim.

Temmuz 2018

Mine MERCAN

İÇİNDEKİLER

Sayfa No

ÖNSÖZ	iv
İÇİNDEKİLER.....	v
ŞEKİL LİSTESİ	ix
TABLO LİSTESİ.....	x
SİMGE VE KISALTMA LİSTESİ	xvi
ÖZET	xvii
SUMMARY	xviii
1. GİRİŞ	1
2. GENEL KISIMLAR.....	5
2.1. PSİKİYATRİK RAHATSIZLIKLAR	5
2.1.1. Depresyon.....	6
2.1.2. Mani.....	6
2.1.3. Dikkat Eksikliği ve Hiperaktivite Bozukluğu (DEHB).....	7
2.2. DOĞAL DİL İŞLEME.....	7
Önceki Çalışmalar	9
2.2.1.1. Duygu Analizinde Önceki çalışmalar	11
2.2.1.2. Psikolojik rahatsızlıklar üzerine diğer dillerdeki çalışmalar	12
2.2.1.3. Türkçede Yapılan DDİ çalışmaları.....	16
2.2.1.4. Türkçe’de Yapılan Psikoloji Çalışmaları	18
3. MALZEME VE YÖNTEM.....	20
3.1. PSİKİYATRİK ÖLÇEKLER.....	20
3.1.1. Çocukluk Çağı Depresyon Ölçeği	20
3.1.2. Tematik Algı Testi.....	20
3.1.3. Beier Cümle Tamamlama Testi.....	21
3.2. PSİKOLOJİK SÖZLÜKLER.....	22
3.2.1. Harvard-III Sözlüğü.....	23
3.2.2. Harvard-IV Sözlüğü ve Diğer Kategoriler	29
3.3. MAKİNE ÖĞRENMESİ	35
3.3.1. Sınıflandırma Yöntemleri.....	35
3.3.1.1. Naive Bayes Sınıflandırıcı	36

3.3.1.2. SMO Sınıflandırıcı	38
3.3.1.3. J48 Sınıflandırıcı	38
3.3.2. Weka Programı	39
3.4. MORFOLOJİK ANALİZ.....	40
3.5. VERİ KÜMESİ	44
3.5.1. Çocuk Verisi (Testler)	44
3.5.2. Yetişkin Verisi.....	46
3.6. ÖZELLİK ÇIKARIMI	47
3.6.1. n-gram.....	48
3.6.2. Ters Doküman Frekansı	49
3.6.3. Ağırlıklı Logaritmik Benzerlik Oranı.....	49
3.7. ÖZELLİK VEKTÖRÜ OLUŞTURMA	50
3.7.1. Tekli Yöntem.....	50
3.7.1.1. Kelime/Kategori Frekansı	50
3.7.1.2. Kelime/Kategori varlığı	51
3.7.2. Liste Yöntemi	51
3.7.3. Karma yöntem	54
4. BULGULAR.....	57
4.1. ÇOCUK VERİSİ İKİLİ SINIFLANDIRMA BULGULARI.....	58
4.1.1. HV-III ile Çocuk Verisi İçin Bulgular.....	58
4.1.1.1. HV-III ile Depresyon x Sağlıklı Sınıflandırması Sonuçları	58
4.1.1.2. HV-III ile ADHD x Sağlıklı sınıflandırması sonuçları	60
4.1.1.3. HV-III ile Çocuk Verisinde Cinsiyet Sınıflandırması İçin Bulgular	62
4.1.2. HV-IV ile Çocuk Verisi İçin Bulgular	65
4.1.2.1. HV-IV ile Depresyon x Sağlıklı Sınıflandırması Sonuçları	65
4.1.2.2. HV-IV ile ADHD x Sağlıklı sınıflandırması sonuçları	67
4.1.2.3. HV-IV ile Çocuk Verisinde Cinsiyet Sınıflandırması İçin Bulgular	69
4.2. YETİŞKİN VERİSİ İKİLİ SINIFLANDIRMA BULGULARI	72
4.2.1. HV-III ile Yetişkin Verisi İçin Bulgular	72
4.2.1.1. HV-III ile Yetişkin Depresyon x Sağlıklı Sınıflandırması Sonuçları	73
4.2.1.2. HV-III ile Yetişkin Mani x Sağlıklı sınıflandırması sonuçları	74
4.2.1.3. HV-III ile Yetişkin cinsiyet sınıflandırması sonuçları	76
4.2.2. HV-IV ile Yetişkin Verisi İçin Bulgular	79
4.2.2.1. HV-IV ile Yetişkin Depresyon x Sağlıklı Sınıflandırması Sonuçları.....	79

4.2.2.2. HV-IV ile Yetişkin Mani x Sağlıklı Sınıflandırması Sonuçları.....	81
4.2.2.3. HV-IV ile Yetişkin Cinsiyet Sınıflandırması Sonuçları.....	83
4. 3. ÜÇLÜ SINIFLANDIRMA BULGULARI	90
4.3.1. HV-III ile Yetişkin Üçlü Sınıflandırması Sonuçları.....	91
4.3.2. HV-IV ile Yetişkin Üçlü Sınıflandırması Sonuçları	92
4.3.3. HV-III ile Çocuk Üçlü Sınıflandırması Sonuçları.....	93
4.3.4. HV-IV ile Çocuk Üçlü Sınıflandırması Sonuçları.....	95
5. TARTIŞMA VE SONUÇ	98
KAYNAKLAR.....	101
EKLER	108
EK 1. Psikiyatrik hastalıkların DSM-IV'e göre sınıflandırması (APA, 1994).....	108
EK 2. Harward-III sözlüğündeki kategoriler.....	110
EK 3. Harward-IV sözlüğündeki kategoriler.....	117
EK 4. Metot Kısaltmaları ve Açıklamaları.....	131
EK 5. Özellik seçimi yöntemlerine ve eşik değerlerine göre HV-III ile yapılan çocuk DEP x SAĞ sınıflandırmasının NB, SMO, J48 teknikleri uygulanarak elde edilen Liste ve Karma yöntem doğruluk sonuçları.....	132
EK 6. Özellik seçimi yöntemlerine ve eşik değerlerine göre HV-III ile yapılan çocuk ADHD x SAĞ sınıflandırmasının NB, SMO, J48 yöntemleri uygulanarak elde edilen doğruluk sonuçları.....	133
EK 7. Özellik seçimi yöntemlerine ve eşik değerlerine göre HV-III ile yapılan çocuk K x E sınıflandırmasının NB, SMO, J48 yöntemleri uygulanarak elde edilen doğruluk sonuçları.	134
EK 8. Özellik seçimi yöntemlerine ve eşik değerlerine göre yapılan çocuk DEP x SAĞ sınıflandırmasının HV-IV sözlüğü kullanılarak ve NB, SMO, J48 teknikleri uygulanarak elde edilen Liste ve Karma yöntem doğruluk sonuçları.	135
EK 9. Özellik seçimi yöntemlerine ve eşik değerlerine göre yapılan çocuk ADHD x SAĞ sınıflandırmasının NB, SMO, J48 yöntemleri uygulanarak ve HV-IV kullanılarak elde edilen doğruluk sonuçları.....	136
EK 10. Özellik seçimi yöntemlerine ve eşik değerlerine göre HV-IV ile yapılan çocuk K x E sınıflandırmasının MÖ yöntemleri uygulanarak elde edilen doğruluk sonuçları.	137
EK 11. Özellik seçimi yöntemlerine ve eşik değerlerine göre HV-III ile yapılan yetişkin DEP x SAĞ sınıflandırmasının NB, SMO, J48 teknikleri uygulanarak elde edilen Liste ve Karma yöntem doğruluk sonuçları.	138
EK 12. Özellik seçimi yöntemlerine ve eşik değerlerine göre HV-III ile yapılan yetişkin MAN x SAĞ sınıflandırmasının NB, SMO, J48 teknikleri uygulanarak elde edilen Liste ve Karma yöntem doğruluk sonuçları.	139

EK 13. Özellik seçimi yöntemlerine ve eşik değerlerine göre HV-III ile yapılan yetişkin K x E sınıflandırmasının NB, SMO, J48 teknikleri uygulanarak elde edilen Liste ve Karma yöntem doğruluk sonuçları.	140
EK 14. Özellik seçimi yöntemlerine ve eşik değerlerine göre yapılan yetişkin DEP x SAĞ sınıflandırmasının HV-IV kullanılarak elde edilen Liste ve Karma yöntem doğruluk sonuçları.	141
EK 15. Özellik seçimi yöntemlerine ve eşik değerlerine göre yapılan yetişkin MAN x SAĞ sınıflandırmasının HV-IV kullanılarak elde edilen Liste ve Karma yöntem doğruluk sonuçları.	142
EK 16. Özellik seçimi yöntemlerine ve eşik değerlerine göre yapılan yetişkin K x E sınıflandırmasının HV-IV sözlüğü kullanılarak elde edilen Liste ve Karma yöntem doğruluk sonuçları.	143
ÖZGEÇMİŞ	144

ŞEKİL LİSTESİ

Sayfa No

Şekil 3.1: Tematik algı testindeki fotoğraflardan bazıları.....	21
Şekil 3.2: Beier Cümle Tamamlama Testi'nden bir bölüm.	22
Şekil 3.3: Harvard-III sözlüğündeki bazı kategoriler ve içerdiği kelimeler.....	29
Şekil 3.4: Lineer bir SVM örneği (Platt J. C., 1998)	38
Şekil 3.5: Weka programının ekran görüntüsü.	40
Şekil 3.6: Morfolojik analiz programının aldığı veri yapısı.....	41
Şekil 3.7: Doküman önişlem aşamaları.....	43
Şekil 3.8: Doküman sınıflandırma aşamaları.	43
Şekil 3.9: Sistem şeması.....	44
Şekil 3.10: Bir deneğin verisi.....	46
Şekil 3.11: Örnek bir denek verisi. Sağlıklı ve Depresyon alt çizgilerine göre.	54
Şekil 4.1: MÖ yöntemlerine göre doğruluk sonuçları.....	88
Şekil 4.2: Eşik değerlere göre doğruluk sonuçları.	88
Şekil 4.3: Veri tiplerinde özelliklere göre doğruluk sonuçları.	89
Şekil 4.4: Kelime ve Kategori karşılaştırması.	90
Şekil 4.5: Özellik seçim metotlarının doğruluk sonucu karşılaştırmaları	90
Şekil 4.6: MÖ yöntemlerine göre üçlü sınıflandırma doğruluk sonuçları.	96
Şekil 4.7: Veri tiplerinde özelliklere göre üçlü sınıflandırma doğruluk sonuçları.....	97
Şekil 4.8: Özellik seçim metotlarının üçlü sınıflandırma sonucu karşılaştırmaları	97

TABLO LİSTESİ

Sayfa No

Tablo 3.1: Harvard-III Türkçe Sözlüğü'nün düzenlenmeden önceki yapısı.	27
Tablo 3.2: Harvard-III Sözlüğünün düzenlendikten sonraki yapısı.....	27
Tablo 3.3: Harvard-III sözlüğündeki bazı kategoriler ve numaraları.	28
Tablo 3.4: Harvard-IV İngilizce orijinal sözlüğü	30
Tablo 3.5: Harvard-IV sözlüğünün Türkçeye çevrilmesinin ilk aşaması.	30
Tablo 3.6: Türkçe Harvard IV sözlüğünün Türkçeye çevrilmesinin ikinci aşaması.	31
Tablo 3.7: Türkçe Harvard IV sözlüğünün Türkçeye çevrilmesinin üçüncü aşaması.....	31
Tablo 3.8: Morfolojik analiz programının kelime türleri.	41
Tablo 3.9: Morfolojik analiz programının örnek anlam kelimeleri.	41
Tablo 3.10: Verinin morfolojik analizi ve belirsizlik kaldırılması.	42
Tablo 3.11: Çocuk deneklerin bilgileri.	45
Tablo 3.12: Serbest konuşma deneklerin istatistik bilgileri.	47
Tablo 3.13: Cinsiyet sınıflandırması için yetişkin verisi.	47
Tablo 3.14: Cinsiyet sınıflandırması için çocuk verisi.	47
Tablo 3.15: Örnek bir cümlede unigram, bigram ve trigram öğeler.....	48
Tablo 3.16: Örnek bir denek için frekans tablosu	50
Tablo 3.17: Sınıflandırma için eğitim veri tablosundan bir bölüm.	51
Tablo 3.18: Sınıflandırma için verilerin kelime ve kategori varlık/yokluk tabloları.	51
Tablo 3.19: Sık kullanılan kelime listelerinden örnek bir bölüm.	53
Tablo 3.20: Örnek deneğin KelimeListe yöntemi için metin özelliklerinin eğitim verisine aktarılması.....	54
Tablo 3.21: Sistem eğitim veri tablosundan örnek bir bölüm.	54
Tablo 3.22: İlk aşamada çıkarılan frekans listeleri örneği.....	55

Tablo 3.23: Birleştirilmiş frekans listesi örneği.	55
Tablo 4.1: Çocuk DEP x SAĞ sınıflandırmasının NB, SMO, J48 teknikleri uygulanarak elde edilen Tekli yöntem doğruluk sonuçları.	59
Tablo 4.2: Çocuk DEP x SAĞ Sınıflandırmasında Karma, Liste ve Tekli yöntemlerinin kategori ve kelime olarak doğruluk sonuçlarının ortalamaları.	59
Tablo 4.3: DEP x SAĞ Sınıflandırmasında Karma ve Liste yöntemlerinin doğruluk sonuçlarının Eşik değerlere göre genel ortalamaları.	59
Tablo 4.4: Çocuk DEP x SAĞ Sınıflandırmasında Karma ve Liste yöntemlerinin kategori ve kelime olarak doğruluk sonuçlarının HS ve KS yöntemlerine göre karşılaştırması.	60
Tablo 4.5: ADHD x SAĞ sınıflandırmasında Tekli yöntem sonuçları.	60
Tablo 4.6: ADHD x SAĞ Sınıflandırmasında Karma, Liste ve Tekli yöntemlerinin kategori ve kelime olarak doğruluk sonuçlarının ortalamaları.	61
Tablo 4.7: Çocuk ADHD x SAĞ Sınıflandırmasında Karma ve Liste yöntemlerinin doğruluk sonuçları.	61
Tablo 4.8: ADHD x SAĞ Sınıflandırmasında Karma ve Liste yöntemlerinin kategori ve kelime olarak doğruluk sonuçlarının HS ve KS yöntemlerine göre karşılaştırması.	62
Tablo 4.9: Çocuk K x E sınıflandırmasında Tekli yöntem sonuçları.	62
Tablo 4.10: Çocuk K x E Sınıflandırmasında Karma, Liste ve Tekli yöntemlerinin kategori ve kelime olarak doğruluk sonuçlarının ortalamaları.	63
Tablo 4.11: Çocuk K x E Sınıflandırmasında Karma ve Liste yöntemlerinin doğruluk sonuçlarının Eşik değerlere göre genel ortalamaları.	63
Tablo 4.12: Çocuk E x K sınıflandırmasında Karma ve Liste yöntemlerinin kategori ve kelime olarak doğruluk sonuçlarının HS ve KS yöntemlerine göre karşılaştırması.	64
Tablo 4.13: HV-IV sözlüğü kullanılarak çocuk DEP x SAĞ sınıflandırmasının NB, SMO, J48 teknikleri uygulanarak elde edilen Tekli yöntem doğruluk sonuçları.	66
Tablo 4.14: DEP x SAĞ Sınıflandırmasında Karma, Liste ve Tekli yöntemlerinin kategori ve kelime olarak HV-IV ile doğruluk sonuçlarının ortalamaları.	66
Tablo 4.15: Çocuk DEP x SAĞ Sınıflandırmasında Karma ve Liste yöntemlerinin HV-IV ile doğruluk sonuçlarının Eşik değerlere göre en genel ortalamaları.	66

Tablo 4.16: HV-IV sözlüğü kullanılarak DEP x SAĞ Sınıflandırmasında Karma ve Liste yöntemlerinin kategori ve kelime olarak doğruluk sonuçlarının HS ve KS yöntemlerine göre karşılaştırması.	67
Tablo 4.17: ADHD x SAĞ sınıflandırmasında Tekli yöntem sonuçları.	68
Tablo 4.18: HV-IV kullanılarak çocuk ADHD x SAĞ sınıflandırmasında Karma, Liste ve Tekli yöntemlerinin kategori ve kelime olarak doğruluk sonuçlarının ortalamaları.....	68
Tablo 4.19: ADHD x SAĞ Sınıflandırmasında Karma ve Liste yöntemlerinin doğruluk sonuçlarının Eşik değerlere göre en genel ortalamaları.	68
Tablo 4.20: HV-IV kullanılarak ADHD x SAĞ sınıflandırmasında Karma ve Liste yöntemlerinin kategori ve kelime olarak doğruluk sonuçlarının HS ve KS yöntemlerine göre karşılaştırması.	69
Tablo 4.21: Çocuk K x E sınıflandırmasında Tekli yöntem sonuçları.	70
Tablo 4.22: Çocuk K x E Sınıflandırmasında Karma, Liste ve Tekli yöntemlerinin kategori ve kelime olarak doğruluk sonuçlarının ortalamaları.....	70
Tablo 4.23: Çocuk K x E Sınıflandırmasında Karma ve Liste yöntemlerinin doğruluk sonuçlarının Eşik değerlere göre genel ortalamaları.	70
Tablo 4.24: HV-IV sözlüğü kullanılarak yapılan E x K sınıflandırmasında Karma ve Liste yöntemlerinin kategori ve kelime olarak doğruluk sonuçlarının HS ve KS yöntemlerine göre karşılaştırması.	71
Tablo 4.25: Çocuk verisinde HV-III, HV-IV ve kelime karşılaştırması.	72
Tablo 4.26: Yetişkin DEP x SAĞ sınıflandırmasında HV-III ve NB, SMO, J48 yöntemleri uygulanarak elde edilen Tekli yöntem doğruluk sonuçları.....	73
Tablo 4.27: Yetişkin DEP x SAĞ sınıflandırmasında Karma ve Liste yöntemlerinin doğruluk sonuçlarının Eşik değerlere göre genel ortalamaları.....	73
Tablo 4.28: Yetişkin DEP x SAĞ Sınıflandırmasında Karma, Liste ve Tekli yöntemlerinin kategori ve kelime olarak doğruluk sonuçlarının ortalamaları.	74
Tablo 4.29: Yetişkin DEP x SAĞ Sınıflandırmasında Karma ve Liste yöntemlerinin kategori ve kelime olarak doğruluk sonuçlarının HS ve KS yöntemlerine göre karşılaştırması.....	74
Tablo 4.30: Yetişkin MAN x SAĞ sınıflandırmasında NB, SMO, J48 yöntemleri uygulanarak elde edilen Tekli yöntem doğruluk sonuçları.	75
Tablo 4.31: Yetişkin MAN x SAĞ sınıflandırmasında Karma ve Liste yöntemlerinin doğruluk sonuçlarının Eşik değerlere göre genel ortalamaları.....	75

Tablo 4.32: Yetişkin MAN x SAĞ Sınıflandırmasında Karma, Liste ve Tekli yöntemlerinin kategori ve kelime olarak doğruluk sonuçlarının ortalamaları.	76
Tablo 4.33: Yetişkin MAN x SAĞ Sınıflandırmasında Karma ve Liste yöntemlerinin kategori ve kelime olarak doğruluk sonuçlarının HS ve KS yöntemlerine göre karşılaştırması.....	76
Tablo 4.34: Yetişkin K x E sınıflandırmasında NB, SMO, J48 yöntemleri uygulanarak elde edilen tekli yöntem doğruluk sonuçları.	77
Tablo 4.35: Yetişkin K x E sınıflandırmasında Karma ve Liste yöntemlerinin doğruluk sonuçlarının Eşik değerlere göre genel ortalamaları.	77
Tablo 4.36: Yetişkin K x E sınıflandırmasında Karma, Liste ve Tekli yöntemlerinin kategori ve kelime olarak doğruluk sonuçlarının ortalamaları.....	78
Tablo 4.37: Yetişkin K x E sınıflandırmasında Karma ve Liste yöntemlerinin kategori ve kelime olarak doğruluk sonuçlarının HS ve KS yöntemlerine göre karşılaştırması.....	78
Tablo 4.38: Yetişkin DEP x SAĞ sınıflandırmasında HV-IV sözlüğü kullanılarak elde edilen Tekli yöntem doğruluk sonuçları.....	80
Tablo 4.39: Yetişkin DEP x SAĞ sınıflandırmasında HV-IV kullanılarak Karma ve Liste yöntemlerinin doğruluk sonuçlarının Eşik değerlere göre genel ortalamaları.....	80
Tablo 4.40: Yetişkin DEP x SAĞ Sınıflandırmasında HV-IV kullanılarak Karma, Liste ve Tekli yöntemlerinin kategori ve kelime olarak doğruluk sonuçlarının ortalamaları.....	81
Tablo 4.41: Yetişkin DEP x SAĞ Sınıflandırmasında HV-IV kullanılarak elde edilen doğruluk sonuçlarının HS ve KS yöntemlerine göre karşılaştırılması.	81
Tablo 4.42: Yetişkin MAN x SAĞ sınıflandırmasında HV-IV sözlüğü kullanılarak elde edilen Tekli yöntem doğruluk sonuçları.....	82
Tablo 4.43: Yetişkin MAN x SAĞ sınıflandırmasında HV-IV kullanarak Karma ve Liste yöntemlerinin doğruluk sonuçlarının Eşik değerlere göre genel ortalamaları.....	82
Tablo 4.44: Yetişkin MAN x SAĞ Sınıflandırmasında HV-IV kullanarak Karma, Liste ve Tekli yöntemlerinin kategori ve kelime olarak doğruluk sonuçlarının ortalamaları.....	83
Tablo 4.45: Yetişkin MAN x SAĞ Sınıflandırmasında HV-IV kullanarak yapılan Karma ve Liste yöntemlerinin kategori ve kelime olarak doğruluk sonuçlarının HS ve KS yöntemlerine göre karşılaştırması.	83
Tablo 4.46: Yetişkin K x E sınıflandırmasında HV-IV sözlüğü kullanılarak elde edilen tekli yöntem doğruluk sonuçları.....	84

Tablo 4.47: Yetişkin K x E sınıflandırmasında HV-IV kullanılarak Karma ve Liste yöntemlerinin doğruluk sonuçlarının Eşik Değerlere göre genel ortalamaları.	84
Tablo 4.48: Yetişkin K x E sınıflandırmasında HV-IV sözlüğü kullanılarak Karma, Liste ve Tekli yöntemlerinin kategori ve kelime olarak doğruluk sonuçlarının ortalamaları.....	85
Tablo 4.49: Yetişkin K x E sınıflandırmasında HV-IV kullanılarak Karma ve Liste yöntemlerinin kategori ve kelime olarak doğruluk sonuçlarının HS ve KS yöntemlerine göre karşılaştırması.	85
Tablo 4.50: Yetişkin verisinde HV-III ve HV-IV karşılaştırması.	86
Tablo 4.51: HV-III ve HV-IV ile yapılan sınıflandırmalarda tüm veriler için genel Eşik Değer ortalamaları.....	87
Tablo 4.52: HV-III ve HV-IV ile yapılan sınıflandırmalarda tüm veriler için genel HS ve KS ortalamaları.....	87
Tablo 4.53: Yetişkin DEP x MAN x SAĞ sınıflandırmasında HV-III sözlüğü kullanılarak elde edilen Tekli yöntem doğruluk sonuçları.....	91
Tablo 4.54: Yetişkin DEP x MAN x SAĞ Sınıflandırmasında HV-III kullanılarak Karma, Liste ve Tekli yöntemlerinin kategori ve kelime olarak doğruluk sonuçlarının ortalamaları.....	92
Tablo 4.55: Yetişkin DEP x MAN x SAĞ Sınıflandırmasında HV-III kullanılarak elde edilen doğruluk sonuçları.	92
Tablo 4.56: Yetişkin DEP x MAN x SAĞ sınıflandırmasında HV-IV sözlüğü kullanılarak elde edilen Tekli yöntem doğruluk sonuçları.....	92
Tablo 4.57: Yetişkin DEP x MAN x SAĞ Sınıflandırmasında HV-IV kullanılarak Karma, Liste ve Tekli yöntemlerinin kategori ve kelime olarak doğruluk sonuçlarının ortalamaları.....	93
Tablo 4.58: Yetişkin DEP x MAN x SAĞ Sınıflandırmasında HV-IV kullanılarak elde edilen doğruluk sonuçları.	93
Tablo 4.59: Yetişkin DEP x ADHD x SAĞ sınıflandırmasında HV-III sözlüğü kullanılarak elde edilen Tekli yöntem doğruluk sonuçları.....	94
Tablo 4.60: Çocuk DEP x ADHD x SAĞ sınıflandırmasında HV-III kullanılarak Karma, Liste ve Tekli yöntemlerinin kategori ve kelime olarak doğruluk sonuçlarının ortalamaları.....	94
Tablo 4.61: Çocuk DEP x ADHD x SAĞ Sınıflandırmasında HV-III kullanılarak elde edilen doğruluk sonuçları.	94
Tablo 4.62: Çocuk DEP x ADHD x SAĞ sınıflandırmasında HV-IV sözlüğü kullanılarak elde edilen Tekli yöntem doğruluk sonuçları.....	95

Tablo 4.63: Çocuk DEP x ADHD x SAĞ Sınıflandırmasında HV-IV kullanılarak Karma, Liste ve Tekli yöntemlerinin kategori ve kelime olarak doğruluk sonuçlarının ortalamaları.....	95
Tablo 4.64: Çocuk DEP x ADHD x SAĞ Sınıflandırmasında HV-IV kullanılarak elde edilen doğruluk sonuçları.....	96



SİMGE VE KISALTMA LİSTESİ

Kısaltmalar	Açıklama
ADHD	: Attention Deficit Hyperactivity Disorder(DEHB)
ANX	: Anksiyete
APB	: Amerikan Psikiyatri Birliği
APA	: American Psychiatric Association
BEIER	: Beier Cümle Tamamlama Testi
BOW	: Bag of Words
CDI	: Children's Depression Inventory
ÇDÖ	: Çocuklar için Depresyon Ölçeği
DA	: Duygu Analizi
DDİ	: Doğal Dil İşleme
DEHB	: Dikkat Eksikliği Hiperaktivite Bozukluğu (ADHD)
DEP	: Depresyon
DM	: Düşünce Madenciliği
HV-III	: Harvard-III psikolojik sözlüğü
HV-IV	: Harvard-IV psikolojik sözlüğü
Idf	: Ters Doküman Frekansı (Inverse Document Frequency)
MÖ	: Makine Öğrenmesi
MAN	: Mani
NB	: Naive Bayes
SAĞ	: Sağlıklı
SMO	: Sequential Minimal Optimization
SVM	: Destek Vektör Makinesi (Support Vector Machine)
tf	: Terim Frekansı
TAT	: Tematik Algı Testi
WHO	: Dünya Sağlık Örgütü (World Health Organization)
WLLR	: Weighted Logaritmik Likelihood Ratio

ÖZET

DOKTORA TEZİ

MAKİNE ÖĞRENMESİ TEKNİKLERİNİ KULLANARAK TÜRKÇE METİNLERDEN DÜŞÜNCE ÇIKARIMI VE DUYGU BELİRLEME

Mine MERCAN

İstanbul Üniversitesi

Fen Bilimleri Enstitüsü

Bilgisayar Mühendisliği Anabilim Dalı

Danışman : Prof. Dr. Ahmet SERTBAŞ

Bu çalışma, Türkçe kullanımı ile psikolojik durumlar arasındaki ilişkiyi incelemektedir. Kullanılan Türkçe metinler yetişkinlerde depresyon, mani ve sağlıklı kişilerden, çocuklarda depresyon ve Dikkat Eksikliği Hiperaktivite Bozukluğu (DEHB) olan ve sağlıklı kişilerden toplanmıştır. Metinlerin analizinde çeşitli özellik çıkarım metotları denenerek en iyi sonuca ulaşılmaya çalışılmıştır. Özellik çıkarımı yöntemleri en genelde iki türden oluşmaktadır. *Liste* yöntemi ve *Tekli* yöntemdir. Liste yönteminde ayrıştırıcı etkisi olan kelimeler/kategoriler listelerde toplanmıştır. Tekli yöntemde ise kelimeler/kategoriler tek tek özellik olarak alınmıştır. Alt yöntemler ise 8 liste, 4 tekli ve 8 karma yöntemden oluşmaktadır. Özellikler morfolojik analizlerin sonuçları kullanılarak elde edilir. Metinlerin morfolojik analizleri, morfolojik analiz programı ile yapılmaktadır. Weka adlı başka bir programda Makine Öğrenmesi teknikleri kullanılarak sınıflandırma yöntemleri ile analiz yapılır. Tanı konulan kişilerin sözcük kullanımındaki farklılıkları belirlemek için bir uygulama geliştirilmiştir. Bu sistemde metin girilmesi sağlanarak kişinin ruhsal durumunu belirlemeye ilişkin ilk bulgular elde edilmiştir. Bu bulgular, Türkçede kelime kullanımının, insanların psikolojik durumları hakkında ipucu verdiğini göstermiştir. Bu çalışma psikiyatristler için teşhis koyma aracı bir uygulama programının ilk adımlarıdır.

Temmuz 2018, 163 sayfa.

Anahtar kelimeler: Metin Sınıflandırma, Özellik Seçimi, Makine Öğrenmesi

SUMMARY

Ph.D. THESIS

OPINION EXTRACTION AND SENTIMENT DETECTION FOR TURKISH DOCUMENTS USING MACHINE LEARNING TECHNIQUES

Mine MERCAN

İstanbul University

Institute of Graduate Studies in Science and Engineering

Department of Computer Engineering

Supervisor : Prof. Dr. Ahmet SERTBAŞ

This study examines the relationship between the use of Turkish and psychological states. The Turkish texts used were collected from depression, mania and healthy people in adults, depression and Attention Deficit Hyperactivity Disorder (ADHD) in children and healthy people. In analyzing the texts, various feature extraction methods have been tried to reach the best result. Feature extraction methods most commonly consist of two types. *List* method and *Single* method. In the List method, distinctive words/categories are collected in the lists. In the Single method, words/categories are taken as individual features. The sub-methods consist of 8 lists, 4 singles and 8 mixed methods. Features are obtained using the results of morphological analyzes. Morphological analysis of texts is done by morphological analysis program. In another program called Weka, analysis is done by classification methods using Machine Learning techniques. An application has been developed to identify differences in the word usage of diagnosed persons. The results of this application show that the use of vocabulary in Turkish gives clues about the psychological state of people. This study is the first step of an application program for the psychiatrist

July 2018, 163 pages.

Keywords: Text Classification, Feature Selection, Machine Learning

1. GİRİŞ

Karar verme sürecinde başka insanların ne düşündüğü merak edilir. Şirketler ürün pazarlama stratejilerini belirlerken, partiler seçim propagandalarını yaparken, üreticiler ürünlerinde yapacakları değişiklikleri belirlerken ve bunun gibi birçok durumda yöneticilerin geniş kitlelerin duygu ve düşünce çıkarımına ihtiyaçları vardır.

Kişilerin duygu ve düşüncelerini ifade etme yolları analiz edilerek çağrı merkezlerinin yükü hafifletilebilir, ürün pazarlama stratejileri geliştirilebilir, seçim sonuçları tahmin edilebilir veya psikolojik rahatsızlıklar belirlenebilir. Düşünce Madenciliği ve Duygu Analizi özellikle son yıllarda üzerinde çalışılmaya başlanan ve ortaya çıkardığı sonuçlar ile oldukça faydalı uygulamaların çıkmasına zemin hazırlayan önemli bir araştırma konusudur. Buradan yola çıkarak “Çok karmaşık bir canlı olan insanın söyleyemediklerini anlayabilir miyiz? Psikiyatristler danışanlarına yardımcı olurken duygu ve düşünce çıkarımı teknolojilerinden faydalanabilirler mi?” sorularına cevap aramaktadır.

Günümüzde, teknolojik gelişmelerin sonucu olarak, çok büyük miktarda veri birçok farklı gruptan insan tarafından erişilebilir duruma gelmiştir. Bu farklı gruplara örnek olarak iş dünyası, bilim insanları ya da ev hanımları verilebilir. Özellikle internetin geniş kullanımı insanlığa birçok imkân sağlamaktadır. Bugünün dünyasında:

- tüketiciler alışverişlerini forumlardaki ürün yorumlarını inceleyerek yapabilmektedir.
- firmalar müşteri memnuniyetlerini alışveriş sitelerindeki veya diğer yorum sitelerindeki müşteri yorumlarına göre ölçebilmektedir.
- internet sayfaları aracılığı ile birçok insan günlük yaşantısını dünyanın dört bir yanındaki insanlarla paylaşabilmektedir.

İstatistiksel araştırmalar, özellikle internet üzerinden bilgi paylaşımında hızlı bir artış olduğunu göstermektedir. Amerika’da yapılan bir çalışma (Purcell ve diğ., 2010), Amerikalıların %46’sının haber edinmek için günde 4 ile 6 arası kaynak kullanırken sadece %7’sinin tek kaynaktan haber aldığını söylemektedir. Diğer bir araştırma (Fox ve diğ., 2009) ise Aralık

2008’de Amerika’daki internet kullanıcılarının %20’sinin Twitter veya diğer bir paylaşım servisinde hesapları varken Eylül 2009’da bunun %33’e çıktığını belirtmektedir.

Yukarıda bahsedilen bilgi paylaşımındaki artma, bilimsel çalışmalarda da birçok alanda ilerleme ile sonuçlanmaktadır. Bu gelişmeler sayesinde bilim insanları çalışmaları için gereksinim duydukları kaynaklara dünyanın neresinde olursa olsun kolaylıkla erişebilmektedir. Dünyanın herhangi bir yerindeki bilim insanları ile fikir alışverişinde bulunabilmekte, birlikte çalışmalar yürütebilmektedir.

Teknolojik gelişmelerin bilimsel çalışmalara yaptığı diğer bir katkı da çalışmalarda kullanılmak üzere veri bulma işlemini kolaylaştırmış olmasıdır. İnternet sayesinde hazır veri bulunabilmekte, veri toplanabilmekte veya diğer bilimsel çalışmalardan temin edilebilmektedir. İnternet, özellikle sosyal bilimler alanındaki çalışmalar için büyük bir kaynak teşkil etmektedir. Bahsedilen gelişmeler disiplinler arası çalışmalara da olanak sağlamaktadır. Çalışmamızın da bunlardan biri olması amaçlanmıştır.

Bilimsel çalışmalarda kullanılan verilerin birçok kaynağı bulunmaktadır. Gazeteler, televizyonlar, radyolar, dergiler ve daha birçok medya bu kaynaklar arasında sayılabilir. Bu kaynakların en genişisi ise internettir. Blog yazıları, ürün yorumları, forumlar, paylaşım siteleri internetin sağladığı kaynaklardan bazılarıdır.

Yukarıda bahsedilen verilerin büyük bölümü yazı ya da konuşma formatındadır. Bu formatlardaki veri birçok bilim dalındaki çalışmalar ve iş dünyasında yapılabilmiş analizler açısından çok değerlidir. Bu verilerle yapılan anlamsal ya da yapısal analizler yardımıyla birçok faydalı sonuç ve bilgiye ulaşılabilir. Bu analizlerin iş dünyasında doğru yorumlanması ve dikkate alınması ile daha fazla müşteri memnuniyeti ve daha verimli çalışma düzeni elde edilebilir. Aynı zamanda firmalar bu gibi analizler sayesinde mevcut durumlarını daha açık ve ekonomik olarak görebilirler. Diğer taraftan sözü edilen formattaki verilerin analizi ile birçok bilim dalında çok sayıda buluş yapılabilir. Özellikle sosyal bilimlerde gelişme sağlanması ve yeni bakış açıları kazanılması açısından bu analizler büyük önem taşımaktadır.

Yazıların ya da konuşmaların analiz edilmesi ve elde edilen sonuçların kullanımı ile ilerleme sağlayabilen alanlardan biri de psikolojidir. Yukarıda bahsedilen kaynaklardan elde edilen verilerle yapılan analizler geniş kitlelerin psikolojik durumu hakkında birçok bulgu elde edilmesini sağlayabilir. Belirlenen herhangi bir grup insan psikolojik durumları üzerinde fikir

edinilerek sınıflandırılabilir ve bu sınıflara göre yeni bulgular elde edilebilir. Psikolojideki ana hedeflerden biri de kişilere tanı konmasıdır. Tanı koyma işlemi, yazı ve konuşma analizi aracılığı ile çok daha kolay ve verimli hale getirilebilir. Tüm bu amaçlar için Doğal Dil İşleme (DDİ) yöntemlerinin kullanımına ihtiyaç vardır.

Günümüzde yazı ya da konuşmanın elle analizi son derece zaman alıcı bir işlemdir. Bu nedenle büyük miktardaki verinin elle analizi için çok sayıda kişi ve maddi kaynak gerekmektedir. Burada yazı ve konuşmanın analizinde bilgisayar destekli yöntemlerin önemi ortaya çıkmaktadır. Bilgisayar destekli yöntemler sayesinde yazı ve konuşmaların analizi çok daha verimli, güvenilir ve ucuz olmaktadır. Yazı ve konuşma analizinde diğer bir önemli görev de verilerin gereksiz kısımlardan ayıklanması işlemidir ve bu işlem son derece zaman alıcıdır. Tüm bu sebeplerden dolayı bilgisayarlı analiz yöntemleri büyük öneme sahiptir.

Türkçede kullanılan dilin içerik analizinin gerçekleştirilmesi oldukça eski bir düşüncedir (Berelson, 1952) (Abadan, 1961). Ancak bilgisayarlı analiz uygulamaları daha yakın zamana dayanmaktadır. İngilizcede son yıllarda yoğun olarak çalışılmasına rağmen, Türkçede az sayıda çalışma bulunmaktadır. Başka dillerde yapılan çalışmalarda da elde edilen sonuçlar göz önüne alınınca, kişilerin kendilerini ifade için kullandıkları dil ile rahatsızlıkları arasında ilişki bulunduğu düşünülmektedir (Gottschalk ve Gleser, 1969) ki bu çalışmanın çıkış noktası bunu göstermeye yönelik olmuştur.

Tezde, sıklıkla rastlanan hastalıklardan depresyon, mani, dikkat eksikliği-hiperaktivite gibi rahatsızlıklar incelenmiştir. Çalışmamızda, sözü edilen rahatsızlıklara sahip bulunan hastalardan ve sağlıklı kişilerden yazılı veya sözlü olarak kayıtlar alınmıştır. Hastalardan ve sağlıklı kişilerden elde edilen bu veriler, Makine Öğrenmesi (MÖ) ve DDİ teknikleri kullanılarak işlenmiş ve sonuçta belirli bir hastalıkta kullanılan dilin ayırt edici özellikleri belirlenmiştir. Belirlenen dil özelliklerini altyapı olarak kullanan bir bilgisayar destekli psikiyatrik tanı sistemi kurulmuştur. Bu sistemde metin girilmesi sağlanarak kişinin ruhsal durumunu belirlemeye ilişkin ilk bulgular elde edilmiştir. Çalışmanın sonunda, kişilerin rahatsızlıkları ile ilgili kesin bir teşhis elde etmek değil, yoğun analiz gerektiren ortamlarda uzmanların işini kolaylaştıran ön bir analiz yapmak hedeflenmiştir.

Bundan sonraki bölümler şu şekilde düzenlenmiştir:

İkinci bölümde, psikiyatrik rahatsızlıklar ve DDİ hakkında genel bilgiler ile önceki çalışmalar verilmiştir. Üçüncü bölümde, psikiyatrik araçlar, kullanılan MÖ teknikleri ile kullanılan araç programlar anlatılmaktadır. Dördüncü bölümde ise çalışılan veri kümeleri, özellik çıkarımı ve metotlar ile deneysel sonuçlar verilmiştir. Son bölümde, deneysel sonuçlar hakkında yorumlara yer verilmiş, özet ve sonuç ile birlikte gelecek çalışma önerilerinde bulunulmuştur.



2. GENEL KISIMLAR

Çalışmamız DDİ alanının bir alt alanı olan Duygu Analizi çalışmasıdır. Bu bölümde analizi yapılan metinlerin alındığı deneklerde bulunan psikiyatrik rahatsızlıklar genel olarak anlatılacaktır. Ayrıca DDİ ile ilgili daha önce yapılan çalışmalar incelenecektir.

2.1. PSİKİYATRİK RAHATSIZLIKLAR

Psikiyatri terimi, zihin, akıl ya da ruh anlamına gelen Yunanca "psyche" ve şifa anlamına gelen "iatreia" kelimelerinden türetilmiştir. 18. yüzyıla kadar zihinsel hastalıklar genellikle şeytanın insanı ele geçirmesi olarak görülmüştür, ancak zamanla tedaviyi gerektiren bir hastalık olarak düşünölmeye başlanmıştır. (Lomax, 2015)

Dünya Sağlık Örgütü (World Health Organization-WHO) ruhsal sağlık tanımını, “Bireyin kendi potansiyelini gerçekleştirebileceği, yaşamın normal stresleriyle baş edebileceği, verimli çalıştığı ve topluma katkıda bulunduğı bir huzur, iyi hissetme, mutluluk halidir.”, olarak tanımlar. Psikolojik rahatsızlıklar farklı belirtilerle geniş bir terimdir. Bununla birlikte, rahatsız edici düşüncelerin, duyguların, davranışların ve başkalarıyla olan ilişkilerin bir kombinasyonu şeklinde karakterize edilir. Bazıları depresyon, kaygı, davranış bozuklukları, bipolar bozukluklar ve şizofrenidir. Bu bozuklukların birçoğı başarıyla tedavi edilebilir. (WHO, 2014)

DSM (Psikolojik Rahatsızlıkların Teşhis ve İstatistik El Kitabı: The Diagnostic and Statistical Manual of Mental Disorders) Psikolojik Rahatsızlıkların sınıflandırma sistemidir. Amerikan Psikiyatri Birliğı - APB (American Psychiatric Association - APA) tarafından tanı ölçütü olarak yayınlanmıştır. (Helzer, 2009). Psikiyatrik tanı için Amerika ve Türkiye’de kullanılan bir başvuru kitabı görevini görür. Dördüncü Baskısı 1994 güncellemesidir ve DSM-IV olarak adlandırılır, DSM-V ise 2013 de güncellenen baskısıdır.

Psikolojik rahatsızlıkların tanımının kapsamlı olması ve tanı koymanın güç olması, dünya çapında yaygınlığının hesaplanmasını güçleştirmektedir. Bununla birlikte herhangi bir DSM hastalığının yaşam boyu yaygınlık oranı dünya çapında %18-%36 aralığındadır. (Kessler ve diğ, 2009)

Psikiyatrik hastalıkların DSM-IV’e göre sınıflandırması (APA, 1994) EK1’de bulunmaktadır.

2.1.1. Depresyon

Depresyon, hüznün, ilgi veya neşe kaybı, suçluluk duygusu veya değersizlik, uyku veya iştah bozukluğu, yorgunluk ve dikkat dağınıklığı ile kendini gösteren bir psikolojik rahatsızlıktır. Depresyon, insanların işyerinde veya okulda çalışabilmelerini ve günlük yaşantılarıyla baş edebilmelerini önemli derecede bozabilir, uzun süreli olabilir veya tekrarlayabilir. Şiddetli depresyon intihara yol açabilir (WHO Regional Office for Europe, 2016). Depresyon karmaşık bir sorundur ve onu yalnızca duygusal değişme bozukluğu olarak kabul etmek aşırı basitleştirir, çünkü duygusal etkileri, aynı zamanda fizyolojide değişiklikler ve çocukların bilişsel ve sosyal işlevlerinde değişiklikler içerir. Bu gerçek, bireylerde biyolojik, psikolojik ve sosyal karmaşıklığın çok seviyeli gelişimsel entegrasyon süreçlerinin kavranmasını zorlar. (McCauley ve diğ., 2001) (Avanci ve diğ.,2012)

Depresyon yaygın bir psikolojik rahatsızlıktır ve dünya genelinde günlük yaşamda yetersizliğin başlıca nedenlerinden biridir. Dünyada her yaştan yaklaşık 350 milyon kişi depresyondan etkilenmiştir. Kadınlarda erkeklerden daha fazla görülür. (WHO Regional Office for Europe, 2016)

2.1.2. Mani

Yaygın olarak manik depresyon adıyla bilinen iki kutuplu (bipolar) bozukluktur. Kişinin ruh hali, enerji ve işlev kabiliyetinde kaymalara neden olan bir beyin rahatsızlığıdır. Bipolar bozukluğun belirtileri, sosyal ilişkilerde bozukluk, çalışma ya da okula gitme zorluğu ve hatta intihar ile sonuçlanabilir. Genellikle normal ruh halleri dönemleri vardır. Ancak tedavi edilmediğinde, bipolar bozukluğu olan insanlar ruh halindeki değişimlerini yaşamaya devam eder. Bu hastalığın iyi tarafı tedavi edilebilir olması ve hastalığa sahip kişilerin tam ve üretken yaşamlarına devam edebilmesidir.

Bipolar bozukluk, aşırı duygusal salınımlara neden olabilir. Aşırı derecede sinirli, üzgün ve umutsuzluktan çok büyük mutlu dönemlere gidip gelmeler ve bu ikisi arasındaki dönemlerde de çoğunlukla normal ruh hallerinde olurlar. Yüksek ve alçak dönemlere mani ve depresyon atakları adı verilir (Parekh, 2015b)

2.1.3. Dikkat Eksikliği ve Hiperaktivite Bozukluğu (DEHB)

Çocuklarda en yaygın psikolojik rahatsızlık Dikkat Eksikliği Hiperaktivite Bozukluğu – DEHB rahatsızlığıdır (Attention Deficit Hyperactivity Disorder-ADHD). DEHB, birden fazla zeminde (ör. Okul ve ev) oluşan bir davranış kalıbı ile karakterizedir ve sosyal, eğitimsel veya mesleki konularda başarısızlıkla sonuçlanabilir. Belirtileri ayrıntılara dikkat etmeme, görev ve faaliyetlerin düzenlenmesindeki zorluk, aşırı konuşmak, yerinde duramamak gibi davranışlardır. (APA, 2013) (Powell ve Powell, 2016)

Depresif belirtileri inceleyen yurt dışı çalışmalarda depresif belirtilerin çocuklarda %10-%13, ergenlerde ise %21 ile %56 arasında değişen oranlarda görüldüğü bulunmuştur. Ülkemizde ilköğretim öğrencileri arasında yapılan bir çalışmada %30 oranında orta, %6 oranında ağır depresif belirti, lise öğrencilerinde de %27 oranında depresif belirti saptanmıştır. Çocuk ve ergenlerde depresif bozukluğun sıklığı incelendiği çalışmalar gözden geçirildiğinde; depresif bozukluğun yaşla artış gösterdiği, okul öncesi çocuklarda %1'den, okul çağına %3'ten, az olduğu, ergenlerde ise bu oranın %20'ye kadar çıktığı görülmektedir (Alşen, 2012).

2.2. DOĞAL DİL İŞLEME

DDİ, bilgisayarların doğal dil metinlerini veya konuşmalarını anlamak ve işlemek için nasıl kullanılabileceğini araştıran bir araştırma ve uygulama alanıdır (Chowdhury, 2005). Tarihsel olarak DDİ, bilgisayar bilimlerinde, elektrik mühendisliğinde dil bilimde ve psikolojide farklı farklı bölümler olarak ele alınmıştır. Bu bölümler bazı alanlarda örtüşür;

- Dilbilimde » Hesaplamalı dilbilim
- Bilgisayar biliminde » DDİ
- Elektrik mühendisliğinde » Ses tanıma, konuşma tanıma
- Psikolojide » Hesaplamalı psiko-dilbilim

Bu bölümde DDİ'nin gelişimine katkısı olan farklı tarihi çalışmalar özetlenecektir. Ses işleme DDİ'nin konusuna girmesine rağmen bizim konumuz dışında kalmaktadır.

Dil işlemeyi diğer veri işlemekten ayıran özellik, DDİ'nin dili insanın kullandığı yapıda ele almasıdır. Örneğin satır ve karakter sayma DDİ'nin konusu değildir. Ancak kelime sayma dil işlemenin konusuna girer. DDİ konuları dili yapısal veya anlamsal olarak işleyen alanlardır.

DDİ'nin konusu olan problemlerden bazıları aşağıdadır.

- Otomatik konuşma tanıma
- Doğal dili anlama
- Doğal dili üretme
- Makine çevirisi
- Soru cevaplama sistemleri
- Web özetleme
- Bilgi çıkarımı
- Kelime algı belirsizliği çözümlemeleri

Yukarıdaki problemlerin hepsi araştırma kapsamındadır. Bu teknolojilerle yapılmış ticari uygulamalar da bulunmaktadır.

Aşağıdaki örnekler şu anda yaygın olarak kullanılan ve bu eğilimi yansıtan sistemlerdir.

- Seyahat şirketleri müşterilerin rezervasyon, kalkış, varış saatleri için konuşan sistemler kullanmaktadır.
- Telefon yapım şirketleri, otomatik ses tanıma ve otomatik metin okuma sistemleri uyguluyorlar. Bu kullanıcıların ses ile komut verme, eğlence ve yön bulma sistemlerini çalıştırmalarını sağlıyor. Iphone uygulaması olan Siri de bunlardan biridir. Benzer şekilde, konuşan diyalog sistemi, çağrı merkezlerinde de kullanılmaktadır.
- Google bir dilde arama yaparken çeviri servislerini kullanıp bilgi çıkarımı sağlamaktadır. Kullanıcı kendi dilinde arama sorgusu yapıp başka dilde bir sayfada bulabiliyor. Google sorguyu diğer dile çevirip ilgili sayfaları buluyor ve sonra kullanıcının diline çeviriyor. Ayrıca Google Çeviri sistemi de ücretsiz olarak internette eskisinden çok daha iyi olarak hizmet vermektedir.
- Pearson gibi, ETS Tests and Products test servisleri gibi büyük eğitim yayın şirketleri, binlerce öğrencinin yazdığı metinleri analiz eden, notlandıran ve değerlendiren, otomatik sistemler kullanıyorlar.
- Metin analizi kullanan şirketlerin sayısı son zamanlarda oldukça artmıştır ve bu şirketler akıllı pazarlama servisleri sağlamaktadır. Bunlar kullanıcının düşüncesini, tercihlerini web

bloglarından, tartışma forumlarından ve kullanıcı gruplarından otomatik olarak ölçen sistemleri kullanıyorlar.

- Emniyet güvenlik birimleri de metin analizi sistemlerini suçluları yakalamak için kullanmaktadır.

Önceki çalışmalar;

DDİ çalışmaları 1940'lara kadar uzanır. DDİ'nin ilk uygulaması 2. Dünya Savaşı sırasında düşman kodlarını kırmak için geliştirilen bir Makine Çevirisi uygulamasıydı. 1950'de Alan Turing, daha sonra Turing Testi (Turing, 1950) olarak adlandırılan bir sistem önerdi. Bu sistem bir bilgisayarın bir insanı taklit edebilme yeteneğini gösteren bir oyundu. İnsanın, görmeden iletişime geçtiği şeyin makine mi yoksa insan mı olduğunu tahmin etmesi esasına dayanmaktadır. Eğer denek, karşısındakinin makine olduğunu anlayamazsa, makinenin bir nevi “düşünme” yetisine sahip olduğu söylenebilir. (Turing, 1950)

1957'de Noam Chomsky, evrensel bir dilbilgisi olduğunu ve tüm dillerin bu dilbilgisinin özelliklerine sahip olduğunu ileri süren bir fikir (Chomsky, 1957) öne sürdü.

1966'da ELIZA programı ile Turing'in öngörülerinin bir kısmı gerçekleşti (Weizenbaum, 1966). ELIZA, kullanıcıyla sınırlı bir konuşma şekli yürüten erken bir DDİ sistemi idi.

1960'larda DDİ alanında içerik analizi çalışmaları da başlamıştır. Zaman içinde üç temel olgu belirleyici olmuştur. Birincisi; bilgisayarların kullanımı, ikincisi; sözel olmayan iletişime ilginin ve semiyolojinin gelişimi ve üçüncüsü de dilbilim çalışmalarındaki gelişmelerdir. Bilgisayarların ilk olarak ortaya çıktığı 1944 yılından sonra hızla gelişen bilgisayar teknolojisinin içerik analizine sağladığı olanaklar "General Inquirer" (Stone ve diğ., 1966) adlı kitapta değerlendirilmiştir. İçerik analizi sosyal bilimlerin alanına giren konularda sosyal gerçeği araştırmada, deneysel ve gözleme dayalı yöntemlerin uygulanamadığı durumlarda kullanılabilir bir konuma gelmiştir.

1966'da Otomatik Dil İşleme Danışmanı Komitesi (ALPAC: Automatic Language Processing Advisor Committee), araştırmaların beklentileri karşılayamadığını gösteren bir rapor yayınlamıştır. Raporun ardından DDİ üzerine yapılan araştırmalar uluslararası düzeyde azalmıştır. 1970'de kurulan SHRDLU (Winograd, 1971), bu durgunluğun ardından başlanan

başarılı çalışmaların başında gelmektedir. Sistem, normal İngilizce dilinde sorulan soruları yanıtlar, komutları uygular. Talepleri bir grafiksel bloklar dünyası olarak gerçekleştiren bir yazılımdır. Daha sonra Roger Schank ve bazı meslektaşları insanın kavramsal bilgilerini inceledi. Bunlardan biri Robert P. Abelson ile olan bilgi sistemi teorisi üzerine yapılan çalışmadır (Abelson ve Schank, 1977).

Bo Pang ve Lillian Lee'ya göre inançlarla ilgili erken projeler duygu ve düşünce analizinin öncüleri olarak sayılabilir (Pang ve Lee, 2008). Bu projelerden iki örnek “İnanç Sistemlerinin Bilgisayar Modelleri” (Carbonell, 1979) ve “İnançlar, Düşünceler ve Çoklu Ortamlar” dır (Wilks ve Bien, 1983).

1982’de Stanley R. Rosenschein ve Stuart M. Shieber’in çalışması (Rosenschein ve Shieber, 1982), kompozisyon modeli-teorik semantiği yansıtan sözdizim odaklı çeviri sistemidir.

1980’lerin sonundan itibaren DDİ’ye MÖ yaklaşımları girmiştir. Bilgisayar bilimlerindeki gelişmelerle 1990’lar DDİ’nin yükseliş dönemi haline gelmiştir. Olasılık ve veri-güdümlü modeller standart oldu. Doğal dillerdeki belirsizlikler DDİ konusunda çalışanlar için en büyük zorluktur. Bunları aşabilmek için dil verilerinden oluşturulmuş olasılık modellemelerinden faydalanılır. Bu araştırmalardan biri, Adam L. Berger ve arkadaşlarının (Berger ve diğ., 1996) çalışmasıdır. Çalışma maksimum entropi modellerini otomatik olarak oluşturmak için maksimum olasılık yaklaşımı sunmaktadır.

1990’larda bilgisayarlarda hız ve bellek artışı, DDİ’nin alt alanlarında ticari olarak patlamaya sebep olmuştur. Yazım ve dilbilgisi yanlışlarını düzeltme alanlarında çalışmalar yapılmıştır. Daha sonra Web’in gelişimi, dil tabanlı bilgi çıkarımında ihtiyacı belirgin olarak ortaya çıkarmıştır.

MÖ metotlarından SVM (Cortes ve Vapnik, 1995) multinomial lojistik regresyon (MaxEnt) (Berger ve diğ., 1996) ve grafik Bayesian modelleri (Pearl, 1998) gibi teknikler, DDİ’de standart uygulama haline gelmiştir.

DDİ’nin araştırmalarından biri (Liu ve diğ., 2007), düşünce madenciliğinde düşük kaliteli ürün eleştirilerini belirleme sorununu ele almıştır. Çalışmada, bir sınıflandırma yöntemi kullanarak ürün eleştirilerinin bilgi içeriği, okunabilirliği ve tarafsızlığı araştırılmıştır. Bu bir düşünce

madenciliği uygulamasıdır ve iki aşamalı bir çerçeve oluşturulmuştur. İki kişi tarafından etiketlenmiş 4909 eleştiri incelenmiştir. Ortalama doğruluğu %78'dir.

Bir başka araştırmada (Jin ve diğ., 2007), web sayfalarının hassas içerik sınıflandırması yapılmıştır. Çalışma, reklam vermek için uygun web sayfası belirlemede önemlidir. Yaklaşık 200 web sayfası bazı kategorilere göre etiketlenmiştir. Bu sayfalar rastgele olarak eğitim ve test kümesi olarak ayrılmıştır. Toplanan başka sayfalar ise değerlendirme seti oluşturmak için etiketlenmiş ve etiketlenmemiş sayfalar olarak kullanılır. Ortalama doğruluk hiyerarşik sınıflandırıcılarda %57, ikili sınıflandırıcılarda %78,5'dir.

2.2.1.1. Duygu Analizinde Önceki çalışmalar

İnternetteki büyük veriye kolay erişim nedeniyle Düşünce Madenciliği (DM) ve DA araştırmaları artmaktadır. DA, kaynak materyaldeki öznel bilgileri saptamak ve çıkarmak için metin analizi ve hesaplamalı dilbilim kullanmaktadır. DM, insanların ifadelerinden belirli bir konu ya da problem hakkında otomatik olarak bilgi çıkarma sürecidir (Sindhu ve diğ., 2014). Alternatif olarak DA, bir metnin otomatik olarak sınıflandırılması olup, yazarın belirli bir konuyla ilgili tutumunu belirlemeye çalışmaktadır. Bu tekniklerin araştırmaları ve uygulama alanları geniş bir yelpazede artarak devam etmektedir.

A. Abbasi ve H. Chen'in çalışmasında karanlık web forumlarının (Abbasi ve Chen, 2007) duygusal yoğunluk analizleri yapılmıştır. Aşırılık yanlısı grupların forum ilanları analizine yönelik bir sistem sunarlar. Analizin kullandığı öznitelikler, mutluluk, öfke, hüzn, korku gibi hisleri temsil eden terimlerdir. Sistem iki kısımdan oluşur. Birinci bölüm, bir duygu sözlüğü oluşturmaktır. İkinci bölüm, yoğunluk analizi tekniğinin uygulanmasıdır. 16 forumdan gelen mesajlar toplanmış ve doğrusal regresyon analiziyle analiz edilmiştir. Bunun gibi araştırmalarda, forumlar veya bloglar gibi internet hizmetlerinin, eleştiri ve duygusal bilgilere sahip olmaları nedeniyle, duygu analizi için iyi metinler olduğu kabul edilir. İnsanların kendi deneyimleri, görüşleri ve duygularını diğer insanlar tarafından okunması ve yorumlanması için yazılan bloglar ve açık günlükler de bu metinlerdendir.

Fukuhara ve arkadaşları (2007), belirli zaman aralığında olan toplumsal olayları inceleyerek insanların olaylar karşısındaki duygu eğilimlerini belirli bir zaman aralığında olan olayların etkisini analiz eden bir çalışma sunmuşlardır.

Bir başka çalışmada ise Aman ve Szpakowicz bir metindeki cümleleri duygu sözlüğünden türetilen özellikleri kullanarak, Ekman'ın altı temel duygu kategorisine (mutluluk, üzüntü, öfke, tiksinti, sürpriz, korku) otomatik olarak sınıflandırılmıştır. Ek olarak, duygusal özellikler ile korpus esaslı unigram özelliklerinin birleşimi kullanılarak üstün sınıflandırma performansı elde edilmiştir (Aman ve Szpakowicz, 2008).

Diğer bir çalışmada, ikili sınıflandırma ile Arapça metinlerde gözetimli (supervised) DA yapılmıştır (Abdul-Mageed ve diğ., 2014). İngilizce dili ile karşılaştırıldığında Arapça gibi morfolojik olarak zengin ve karmaşık dillerde duygusal analiz yapmak oldukça zordur. Bu bakımdan önemli bir çalışmadır.

Pang ve Liu 2008 yılında Duygu ve Düşünce Çıkarımı (DÇ) alanında genel bir derleme yapmışlar ve o güne kadar yapılan araştırmalarla ilgili etraflı incelemeye yer vermişlerdir (Pang ve Lee, 2008).

2.2.1.2. Psikolojik rahatsızlıklar üzerine diğer dillerdeki çalışmalar

Psikolojik rahatsızlıklar, modern dünyanın önemli sorunlarından biridir. Tüm toplumları yaygın olarak etkilemektedir. Günümüzde çoğu insan ruh hastalıkları ile başa çıkmaya çalışmaktadır. Bazen rahatsız olup olmadığını bile bilmeden sorunlarını çözmeye çabalamaktadır. Çünkü bir insanın psikolojik olarak sağlıklı olduğunu ancak psikiyatristler belirleyebilir. Çeşitli tiplerde ve belirtilerde görülebilen psikolojik rahatsızlıklar genellikle anormal düşüncelerin, algıların, duyguların, davranışların ve başkalarıyla olan ilişkilerin bir kombinasyonu ile karakterizedir (WHO, 2014).

Ruh sağlığı ve hastalıkları, bireyler ve toplumla ilgili birçok parametreyi kapsar. Bireysel seviye, kişinin düşüncelerini, duygularını, davranışlarını ve başkalarıyla olan etkileşimlerini yönetme becerisi gibi niteliklerden oluşur. Bireysel olmayan diğer faktörler, ulusal politika, sosyal bakım, koruma ve destek, yaşam standartları, çalışma koşulları, toplum desteği olarak listelenebilir (WHO, 2014).

Ruhsal bozuklukların yüksek riskleri ve etkileri vardır. Bu bozuklukların erken teşhisi, daha iyi ve hızlı sonuçlar elde etmek için yararlı olabilir. Sıklıkla karşılaşılan ruhsal rahatsızlıkların çoğunda teşhis ve tedavi yetersiz kalır (Kessler ve diğ., 1998).

Erken müdahale edilmemesi toplumu olumsuz etkiler ve ileriye dönük maliyetlerin artmasına neden olur. Psikiyatrik hastalık maliyeti, Dünya Sağlık Örgütü (WHO) tarafından sağlanan verilere göre önemli etkilerle büyümeye devam etmiştir. Zihinsel rahatsızlığı olan insanların ortalama yaşam süresi, genel nüfustan yaklaşık 20 yıl daha kısadır. Buradaki ana faktör, kronik hastalıkların bu insanlar arasında daha yaygın olması ve bu kişilere erişimin daha zor olması ve bazen tedavinin kalitesiz olmasıdır (WHO, 2013).

Düşük ve yüksek gelirli ülkelerde, tedavi görmeyen zihinsel engelli bireylerin yüzdesi, sırasıyla %76-85 ve %35-50 arasında değişmektedir. Tedavi görenlerden birçoğu için kalitesiz bakım bu durumdan çok daha kötüdür. Bilinen istatistikler, bu gerçekleri göstermekte ve intihar eden kişilerin yaklaşık %20'sinin önceki bir hafta içinde ve %40'ının bir ay içinde pratisyen hekim veya psikiyatri servisleriyle görüştikleri iddia edilmektedir (WHO, 2014).

Psikolojik sağlık sorunları için anahtar çözüm, teknolojik gelişmelerde bulunabilir. Teknoloji hayatımızın önemli bir parçası olmuştur. Faaliyetlerimizin neredeyse tamamı internet teknolojisine büyük ölçüde bağımlıdır. Günlük hayatın etkinliklerinde bilgisayarları, akıllı telefonları ve tabletleri kullanmak çok veya az sürelerle zorunlu olmuştur. Bugün insanlar, teknolojiyle verilere kolay erişime sahiptir ve bu durum birçok değişime neden olmaktadır. Günümüz dünyasında insanlar sanal ortamda yorumlarını, duygularını, düşüncelerini, fikirlerini ve deneyimlerini özgürce paylaşmak için sosyal bloglar kullanmayı tercih ederler. Bu gerçek, dijital veri okyanusu olan internet üzerinden iletişimi sağlar. İnternet üzerindeki bu büyük dijital veri diğerlerine kıyasla çok daha kolay ve etkili bir veri analiz kaynağıdır. İnternetteki veriler genel olarak insanların görüşlerini paylaştığı herhangi bir platformdan edinilebilecek metin biçimindedir. Ayrıca, bu veriler kültür, yer, etnisite, duygular vb. ile ilgili çeşitli farklılıkları yansıtır. Dahası, genel olarak insanlar, internet üzerinden bir şeyler paylaştıklarında, özellikle de sosyal bloglarda kendi görüşlerini paylaştıklarında başka bir kişinin veya grubun etkisi altına girmez. Bu tür veriler, diğer kaynaklara kıyasla daha güvenilir ve analiz edilmesi kolaydır. Kişilerin görüşleri, karakteristikleri, eğilimleri gibi konularla ilgili herhangi bir çalışma yapan bilim insanları için gereklidir ve bu alanda çalışmaların artmasına sebep olmuştur.

Bilgisayar uygulamaları bir kişinin psikolojik durumunu metin veya konuşma verileri analiziyle tanımlamak için de kullanılabilir, çünkü kişilerin ifadelerinin mevcut ve hatta geçmiş durumları hakkında ipucu ve izleri gizlediğine inanılmaktadır (Pennebaker, 2007).

Dil kullanımı, psikiyatrik bozuklukların yol açtığı duygusal iniş çıkışlar, değişken ruh durumu ve insan zihniyle ilgili önemli kanıtlar sağlar (Saleem ve diğ., 2012). Son yıllardaki araştırmalar, duyguların kişinin konuşmasını etkileyeceğini gösterir (Colace ve diğ., 2014). Duygular, genellikle istemdişi olan, pozitif veya negatif psikolojik tepkiler ve fiziksel ifadeleri uyandıran karmaşık psikofiziksel süreçlerdir. Duygular genellikle, gerçekte ya da hayal gücü içindeki öğeler, nesneler ya da ilişkiler hakkındaki hisler, algılar ya da inançlarla ilgilidir. Bunlar genellikle bilinçli çaba yerine, kendiliğinden ortaya çıkar. Bu his, bireyin konuşmasında olumlu ya da olumsuz etkiye sahiptir. Özellikle bir kişinin yüz ifadelerinin algılanabildiği yüz yüze konuşma ile karşılaştırıldığında metin içindeki duyguları tespit etmek zor olsa da imkânsız değildir. Duygu Analizi (DA) bu gerçeğe ilişkin umut verici bir yaklaşım ve araştırma alanıdır. DA bir metinde ifade edilen görüş, duygu ve duyguların hesaplamalı çalışmasıdır (Colace ve diğ., 2014).

Psikolojik rahatsızlıkların teşhisi konusundaki standart yaklaşım, klinik olarak uygulanan bir dizi tanı amaçlı görüşme ve testtir (Weathers ve diğ., 2001). Bununla birlikte, bu testleri kullanan hastaların değerlendirilmesi pahalı ve zaman alıcı olmaktadır. Dahası, zihinsel hastalıklarla ilişkili damgalanmadan etkilenen bireyler ve onların aile üyeleri tarafından yanlış bilgilendirmeye sebep olabilir. Bu yanlış bilgilendirme testleri güvenilirmez yapar (Saleem ve diğ., 2012).

Sözü edilen sebeplerden dolayı gün geçtikçe zihinsel bozuklukları analiz eden bilgisayar uygulamaları alanına giderek artan bir ilgi vardır. Psikiyatrik bozuklukların tanısı ve hasta takibi DA teknikleri kullanan bilgisayar uygulamaları ile desteklenebilir. Bu gerçekler bilim adamlarını, metin analiziyle, sayısal ortamda birçok kişi tarafından dinamik olarak üretilen mevcut verileri kullanmaya doğru yönlendirir. Bu nedenle, zihinsel rahatsızlıkların tanısı, bizim çalışmamızın konusu olarak potansiyel fayda alanı olarak seçilmiştir.

Bilim adamları yıllardır doğal dil ile psikoloji arasındaki ilişkiyi tartışırlar. McClelland, bunlardan birisi olarak gıda yoksunluğu yaşayan insanların 1, 4 veya 6 saat boyunca yazdıkları hikayeleri analiz etmiştir (Atkinson ve McClelland, 1948). Gottschalk ve Gleser bu düşüncüyü 1969'da çalışmalarıyla desteklemiştir (Gottschalk ve Gleser, 1969). Tucker ve Rosenberg (1975), 10 şizofreni hastasının rüya metinlerini ve serbest konuşmalarını, 10 şizofreni olmayan deneginkinden ayırmak için bilgisayar analizini kullanmıştır. Sonuç olarak şizofreni hastalarının aslında içlerinde var olan huzursuzluğa ve karmaşaya karşı telaşlı bir savunma

içinde olduklarını ve bu savunmayı kullandıkları dilde kendilerini zaman ve mekan içinde konumlandırarak yansıtıyorlardı.

Başka bir çalışmada dört tanı grubunda toplamda 71 (Somatizasyon, Paranoya, Kanser, Depresyon) hastadan alınan serbest konuşma örnekleri ile konuşma içeriği analizi yapılmıştır (Oxman ve diğ., 1988). Gotschalk-Glessner (G-G) ve General Inquirer (GI) olmak üzere iki nicel analiz yöntemi kullanılmıştır. G-G yönteminde kullanılan özellikler, kaygı (ölüm, terk edilme, yaralanma, aşağılanma kaygıları), düşmanlık (kendine, başkalarına veya karşılıklı düşmanlık), depresyon, umut, bilişsel-zihinsel bozulma ve sosyal yabancılaşma, kişisel bozulma, GI yönteminde ise özellikler Harvard-III psikolojik sözlüğünde birinci sınıftaki 55 kategoridir. Bunlar varlıklar, süreçler ve niteleyicileri içerir.

Deneklerin teşhislerini bilmeyen iki psikiyatrist tarafından da metinler okunmuş ve sınıflandırma yapılmıştır. Sistematiik yapılan sınıflandırmalar ile psikiyatristlerin metinleri okuyarak yaptığı sınıflandırmalar karşılaştırılmıştır. Sonuçlar, dildeki kelime seçiminin sistematiik olarak hesaplanmasından elde edilen verilerin hastaları kendi tanı gruplarına sınıflandırmak için kullanılabileceğini ve bu sınıflamanın psikiyatrik değerlendiriciler tarafından yapılan etiketleme ile karşılaştırıldığında iyi olduğunu göstermektedir. Doğruluk ortalamaları psikiyatrik değerlendiriciler için %66, G-G yönteminde %62 ve GI yönteminde ise Harvard-III için %80 bulunmuştur (Oxman ve diğ., 1988). Sınıflandırmalar diskriminant analizi yöntemi ile yapılmıştır. Aynı verilerle yapılan başka bir çalışmada ise Harvard-IV için %85 olarak bulunmuştur (Rosenberg ve diğ., 1990). Bu çalışmada içeriğe duyarlı iki bilgisayarlı sınıflandırma (GI ve G-G) ve insan puanlama sınıflandırması sonuçları karşılaştırılmıştır. İnsan puanlaması cümlelerdeki duygusal ifadelerin kullanılarak cümle puanlamasına dayanmaktadır. Bu puanlamalar G-G sistemini geliştiren bir uzmana yaptırılmıştır. Bilgisayar sınıflandırmaları ise biri tek tek kelime frekanslarına diğeri morfolojik analizden sonraki tekniklerine dayanmaktadır. Dört tanısai gruptaki 71 denekten serbest konuşma örnekleri alınarak üç yöntemle diskriminant analizi yapılarak sınıflandırma yapılmıştır. Üç yöntemin sınıflandırma performansını karşılaştırırken, her iki bilgisayarlı yöntemin de insan puanlı yöntemden daha doğru olduğu tespit edilmiştir.

Bir başka çalışmada 95 gönüllünün serbest konuşma örnekleri ve Tematik Algı Testi (TAT) yanıtları, General Inquirer içerik analizi bilgisayar programı ve Harvard III Psiko-sosyolojik sözlük (Stone ve diğ., 1966) kullanılarak karşılaştırılmıştır. Bireysel farklılıkların tahmininde

TAT verilerinin serbest konuşma metinlerinden daha iyi sonuç verdiği bulunmuştur (Schnurr ve diğ., 1992).

İlk çalışmaların ardından psikolojideki metin analizi çalışmaları hız kazanmıştır. Günümüzde araştırmacı ve akademisyenler bilgisayarlı yöntemlerle çok miktarda veriyi işleme imkânı bulmuşlardır. Pennebaker (2007), sağlık psikolojisi araştırmacılarının kelimeleri, doğal dili ve ifade hakkında incelemeler yapmak için yeni yöntemler sunan yeni bir çağın eşiğinde olduğunu iddia etmiştir.

Aynı düşünceleri benimseyen araştırmacılar, bilgisayarın olanaklarını kullanarak çalışmaları geliştirdiler. Sözü edilen değerlerin psikolojik ve hesaplamalı dilbilim üzerine birçok disiplinler arası araştırma için bir araç olarak kullanılan Dilbilimsel Sorgulama ve Sözcük Sayımı (LIWC) olduğu önemli araştırmalardan biridir (Pennebaker ve diğ., 2001). LIWC, kelime kategorilerinden oluşan ve yaklaşık 4500 kelimeyi ve kelime gövdelerini içeren bir sözlüğü kullanır; insanların sözlü ve yazılı konuşmalarında sunulan duygusal, bilişsel ve yapısal bileşenleri incelemek için kullanılan bir metin analiz programıdır.

Diğer bir araştırmada (Cohen ve diğ., 2008) 35 erkek 33 kadının 3 dakikalık serbest konuşma metinlerinin dil analizini yapıldı ve doğal konuşma analizi ile kişiliği ölçmenin mümkün olabileceği kanıtlandı.

2.2.1.3. Türkçede Yapılan DDİ çalışmaları

DA veya DM uygulamaları bilgisayar destekli olarak çoğunlukla İngilizce’de gerçekleştirildi. Türkçe’de metin analizi çalışmaları 60’lı yıllarda başlamakla birlikte bilgisayar uygulamaları son yıllarda başlamış ve ihtiyaç artmasıyla hız kazanmış olsa da yeterli düzeyde değildir. Türkçe’de bilgisayar destekli olmayan en eski içerik analizi çalışmalarından biri "Cumhuriyet ve Ulus Gazeteleri Hakkında Muhteva Tahlili" dir (Abadan, 1961). Eski çalışmalarda veriler küçük çaptadır ve elle işlenmiştir. Yakın zamanlarda ise internet ortamında düşünce belirten yazılar hızla artmış ve elle incelenmesi imkânsız hale gelmiştir. Verilerin dijital ortamda otomatik olarak işlenmesi önem kazanmıştır.

Otomatik olarak yapılması için öncesinde otomatik olarak morfolojik analiz yapılması ve anlam belirsizliklerinin kaldırılması gerektirmektedir. Bunlardan biri 1994 yılında Kemal Oflazer’in morfolojik analiz çalışmasıdır. 23000 kelimedenden oluşan kelime kökü sözlüğüne dayanmaktadır.

İki aşamalı kurallara dayanan ve sonlu durum makineleri ile yapılan morfolojik analiz sistemidir (Ofłazer, 1994). Zemberek isimli morfolojik analiz çalışması ise başka bir çalışmadır. İstatistiksel bir modeldir (Akın ve Akın, 2007). DDİ'nin en büyük zorluğu belirsizlik giderilmesidir (Yüret ve Türe, 2006). Türkçe gibi eklemeli dillerde belirsizlik çok yüksektir. Metinlerde genellikle kelimelerin çoğu morfolojik olarak belirsizdir. Morfolojik anlam ayırımı için kural tabanlı bir model sunulmaktadır. Kurallar denetimli eğitim kullanarak yeni bir karar listesi öğrenme algoritması tarafından oluşturulur. Doğruluğu %96'dır ve istatistiksel modellerin biraz üzerindedir (Yüret ve Türe, 2006).

Türk dili üzerine yapılan diğer önemli çalışma DİLMAÇ'tır (Shylov, 2008). Türkmen Türkçe morfolojik analiz ve makine çevirisi sistemidir. DİLMAÇ kullanılarak Kazan Tatar dili (Gökgöz ve diğ., 2011) ve Kazak dilinde iki aşamalı morfolojik analiz yapılmıştır (Zafer ve diğ., 2011).

Metin sınıflandırma üzerine yapılan çalışmalar genellikle morfolojik analiz programlarını kullanmaktadırlar. Bu alanda yapılan çalışmalardan biri yazar belirlemedir (Amasyalı ve Diri, 2006). Bu çalışmada makine öğrenmesi teknikleri kullanılmış ve iyi sonuçlar elde edilmiştir. Gazetelerden toplanan makaleler üzerinde n-gram modeli uygulanmış ve yazarın cinsiyeti ve yazının türünün belirlenmesine çalışılmıştır. Metnin yazarını belirleme başarısı, metnin türü ve yazarın cinsiyeti sırasıyla %83, %93 ve %96 olarak elde edilmiştir. Yine metin sınıflandırmada yapılan diğer bir çalışmada Web'den alınan veri seti altı kategoriye (otomobil, politika, ilaç, dergi, ekonomi ve spor) ayrılmış 600 metin dokümanından oluşmaktadır. Her kategoride 100 belge bulunmaktadır. Unigram, bigram, trigram olarak kelime özellik seçiminde en iyi sonucu unigram vermiştir. Doğruluk sonuçları sırasıyla %95, %81 ve %41'dir (Güran ve diğ., 2009).

Türkçe dilinde yapılan diğer bir çalışma Türkçe film eleştirilerinden oluşan bir veri kümesindeki çalışmadır (Eroğul, 2009). Bu eleştiri metinleri pozitif, negatif veya nötr olarak sınıflandırılmıştır. Eleştirilerde kullanılan kelimelerin köklerini ve eklerini bulmak için Zemberek aracı kullanılmıştır. N-gram, sözcük türü ve olumsuzluk etiketlemeleri kullanılırken, Destek Vektör Makinesi (Support Vector Machine-SVM) algoritması, pozitif-negatif sınıflandırmada veriyi %85 doğrulukla sınıflandırmıştır. Türkçede yapılan diğer bir çalışmada ise (Kaya ve diğ., 2012) Türkçe siyasi haberler, etiketlenmiş veriler kullanılarak analiz edilmiştir. Farklı MÖ algoritmaları kullanılarak sonuçlar elde edilmiştir. Zemberek aracı kullanılarak metindeki kelimeler kök ve eklerine ayrıldıktan sonra, Maksimum entropy n-gram

dil modeli, Naïve Bayes (NB) Sınıflaması ve SVM algoritmaları uygulanarak doğruluk oranları %65'den %77'ye kadar değişen oranlarda sonuçlar elde edilmiştir. Türkçe için yapılan başka bir çalışmada (Orhan ve diğ., 2014) müşterilerin ürün incelemeleri, dil özelliklerini göz önüne alarak bunların polaritesini belirlemek üzere incelendi ve çok yüksek doğruluk sonuçları elde edildi. Duygu analizi çalışmaları ise daha az yapılmaktadır. Bunlardan biri olan yüksek lisans tezinde iki tür veri üzerinde duygu analizi yapılmıştır. Biri İngilizce gerçek duygusal metinlerden çeviri diğeri ise Türkçe çocuk hikayeleridir. Denetimli eğitim için etiketlemeler el ile yapılmıştır. Makine öğrenmesi teknikleri kullanılmıştır. Çeviri metinler için en yüksek %81,3 ve çocuk hikayeleri için %76,8 doğruluğa kadar ulaşılmıştır (Boynukalın, 2012) (Boynukalın ve Karagöz, 2013).

2.2.1.4. Türkçede Yapılan Psikoloji Çalışmaları

Kişilerin konuşma esnasında kullandıkları sözcüklerin ruhsal durumlarına ilişkin ipuçları verdiği bilinmektedir (Viney, 1983). Sözcük kullanım özelliklerinden yola çıkarak ruhsal durumların belirlenmesi çalışmaları Türkçe’de oldukça azdır.

İlk çalışma 1993 yılında tamamlanan Levent Mete ve arkadaşlarının ruhsal durumun belirlenmesindeki çalışmadır (Mete, L., Schnurr, P.P., Rosenberg, S.D., Oxman, T.E., Doğaner, İ., Sorias, S., 1993). Bu çalışmada 50 şizofreni, 50 manik, 50 depresyon rahatsızlığı olan ve 50 hasta olmayan kişinin 5’er dakikalık serbest konuşma örneklerinin metinlerini analiz yapılmıştır. Sözcük kategorilerini bulmak için Harvard-III Psikolojik Sözlüğünün (Montague, 1971) Türkçe versiyonu kullanılmış ve analiz etmek için GI (Stone ve diğ., 1966) bilgisayarlı içerik çözümleme programı kullanılmıştır. Sonuç olarak sınıflandırmada %72,5 doğruluk elde edilmiştir. Konuşma içeriği çözümlemesi sonucunda manik ve depresif hastaların konuşmalarının yakın çevre ile etkileşim konusuna yoğunlaştığı, şizofrenilerin ise günlük yaşamdan uzak soyut konulara ve zihinsel süreçlere yöneldikleri görülmüştür.

Yukarıdaki çalışmada kullanılan örneklerle yapılan başka bir Türkçe çalışmada doğruluk oranı %48,33 olarak bildirilmiştir (Erol ve diğ., 1995). Bu araştırmada bilgisayarlı içerik çözümleme yönteminin tanı amaçlı genel kullanımı durumunda saptanacak doğru sınıflama oranı araştırılmıştır. Bu amaçla 20 depresif, 20 mani, 20 şizofreni ve 20 sağlıklı kontrolden oluşan deneklerin konuşmalarından üretilen diskriminant fonksiyonları kullanılarak Çapraz Geçerlilik

Programı (ÇGP) uygulanmıştır. 120 denek (30 mani hastası, 30 depresyon hastası, 30 şizofreni hastası ve 30 sağlıklı kontrol grubu) %48,33 oranında doğru sınıflandırılmıştır.

Yapılan bir yüksek lisans tez çalışmasında ise metinlerin analizi yapılarak Türkçe'nin kullanımı ile halkın psikolojik durumları arasındaki ilişki ele alındı. Depresif, baskıcı olmayan, endişeli ve endişeli olmayan insanlara bazı metinler yazdırılmış ve bu metinler Weka MÖ algoritmaları kullanılarak sınıflandırılmıştır (Albayrak, 2011).



3. MALZEME VE YÖNTEM

Bu bölümde çalışmamızda kullanılan psikolojik sözlükler, verilerin elde edilmesi için kullanılan psikiyatrik ölçekler, kullanılan makine öğrenmesi yöntemleri, veri kümelerinin tanıtımı ve özellik çıkarım metotları anlatılacaktır.

3.1. PSİKIYATRİK ÖLÇEKLER

3.1.1. Çocukluk Çağı Depresyon Ölçeği

Çocuklar için Depresyon Ölçeği-ÇDÖ (Children's Depression Inventory-CDI) çocukluk depresyonunun şiddetini belirlemek için geliştirilmiş, her maddede son iki haftanın değerlendirildiği 0'dan 2'ye değer alan üç cümlede bulunan, 27 maddelik bir ölçektir. Puanlar 0-54 arasındadır. Yüksek puan depresyonun şiddetini göstermektedir. Kovacs tarafından geliştirilen, 6-17 yaş çocuklarına uygulanabilen, öz bildirim ölçeğidir (Kovacs, 1981). Ülkemizde geçerlilik ve güvenilirlik çalışması (Öy, 1991) tarafından yapılmış ve patolojik kesim noktası 19 puan olarak saptanmıştır.

3.1.2. Tematik Algı Testi

Tematik Algı Testi (Thematic Apperception Test-TAT), sıklıkla kullanılan projektif bir kişilik testidir ve dolaylı testlerdendir (Tomkins ve Tomkins, 1947). Bu testlerin özelliği neyin ölçüleceğinin anlaşılmamasının amaçlanmış olmasıdır. Böylece kişinin tarafsız cevap vermesi sağlanır. Kişiye birtakım kartlar gösterilir. Her bir kart üzerinde bir resim veya fotoğraf vardır. Her kart için bir hikâye anlatması istenir. 30 karttan oluşur ancak hepsi gösterilmek zorunda değildir. Bu kartlar aslında ne anlattığı belirsiz resimlerden oluşmaktadır. Şekil 3.1'de kartlardan bir bölüm gösterilmektedir. Kullanılan kartlardan birisi güneş batarken tarlayı sürmekte olan bir adamı göstermektedir. Kişinin anlatacağı hikâyenin onun iç dünyası hakkında fikir vereceği düşünülmektedir (Baysal ve Tekarslan, 1996).



Şekil 3.1: Tematik algı testindeki fotoğraflardan bazıları.

3.1.3. Beier Cümle Tamamlama Testi

Beier Cümle Tamamlama Testinin (Beier Sentence Completion Test -BSCT) (Beier ve Hanfmann, 1956) dayandığı temel varsayım, diğer yansıtma tekniklerinininki gibidir. Birey, bu test aracılığıyla, bir ölçüde belirsiz bir uyarıcı karşısında kendi ilgilerini, duygularını tutumlarını, isteklerini, üzüntülerini ve kişiliğinin diğer önemli isteklerini yansıtmaktadır. Eksik cümleleri tamamlarken, genel durumunun yanı sıra, iç durumunu yansıtan tepkilerde bulunmaktadır. Test yapılırken verilen form Şekil 3.2’de görülmektedir.

BEİER CÜMLE TAMAMLAMA TESTİ FORM A		
OKULU :		
ADI SOYADI :		
SINIFI :	NUMARASI:	TARİH: ... / ... / 200...
Sevgili arkadaşlar, Aşağıda eksik cümleler verilmiştir. Lütfen bunları sırasıyla okuyun ve eksikleri tamamlayın. Cümleleri elinizden geldiği kadar çabuk tamamlamaya çalışın. Eksik cümleleri istediğiniz gibi tamamlayabilirsiniz.		
1. Büyüdüğüm zaman		
2. Arkadaşlar		
3. Yaramazlık yaptığım zaman		
4. Okulda		
5. Bir türlü unutmadığım		
6. İyi arkadaşlar		
7. Evimiz		
8. Annelerin en iyisi		

Şekil 3.2: Beier Cümle Tamamlama Testi'nden bir bölüm.

3.2. PSİKOLOJİK SÖZLÜKLER

GI, ABD Ulusal Bilim Vakfı, İngiltere ve Avustralya Araştırma Hibe Kurulları tarafından desteklenen bir içerik analizi yazılımıdır. Başlangıçta PL/1 programlama dilinde geliştirilmiştir (Stone ve diğ., 1966). Gallup Örgütü tarafından desteklenen yeniden programlanması önce TrueBasic'te Philip Stone ve daha sonra Java'da Vanja Buvac tarafından gerçekleştirilmiştir. Sistem "Harvard" ve "Lasswell" genel amaçlı sözlüklerini kullanarak İngilizce dilinde yazılmış metinlerde içerik analizi sağlar. Ticari olarak kullanılmamaktadır. Sistem 2000'li yıllarda akademik amaçlı olarak kullanıma açılmıştır (Hurwitz, 2002).

GI temelde bir eşleme aracıdır. Her metin dosyasındaki kelimeleri, sözlük tarafından sağlanan kategorilerle eşleştirerek sayar. Güncel olarak kullanılan sürüm, "Harvard-IV" sözlük içerik analizi kategorilerini, "Lasswell" sözlük içerik analizi kategorilerini (Namenwirth ve Weberi, 1987) ve Semin ve Fiedler'in sosyal biliş çalışmalarına (Semin ve Fiedler, 1988) dayanan beş kategoriye birleştirerek, 182 kategori oluşturur. Her kategori, kelimelerden oluşan bir listedir (Hurwitz, 2002).

Çalışmamızda Türkçeye daha önce çevrilmiş olan Harvard-III sözlüğünün Türkçe versiyonu (Mete ve diğ., 1993) kullanılmıştır. Daha sonra sonuçların nasıl değişeceğini görmek için GI programının yeni versiyonunda kullanılan sözlük Türkçeye çevrilerek kullanılmıştır. Bu yeni sözlükte Harvard-IV ve diğer sözlüklerden eklenmiş kategoriler bulunmaktadır.

3.2.1. Harvard-III Sözlüğü

Harvard-III psikolojik sözlüğü (Montague, 1971), General Inquirer (GI) projesinde yer alan Harvard Üniversitesi öğretim üyeleri ve yüksek lisans öğrencilerinin çalışmasının bir ürünüdür (Stone ve diğ., 1966). GI yazılımı bir İngilizce içerik analizi sistemidir. GI yazılımında kullanılan bu sözlüğün geliştirme aşamalarında Harvard Üniversitesi'nden P.S. Stone ve arkadaşları yer almış ve ayrıca Yale ve Amerikan Üniversitesi sosyoloji bölümünden iki kişi çalışmıştır. (Hurwitz, 2002) (Montague, 1971). Sözlüğün Mart 1963'de ilk sürümü ve 1971 de üçüncü sürümü hazırlanmıştır. Sözlükte 83 kategori vardır. Bu kategoriler iki sınıfta toplanmıştır. Birinci sınıf 55 kategoriden, ikinci sınıf ise 28 kategoriden oluşmaktadır. Sınıflar ve kategoriler aşağıda açıklanmıştır (Montague, 1971). Her kategorinin kategori numarası yanına yazılmıştır. Kategorilerin ayrıntılı açıklaması EK 2 de yapılmıştır.

1.Sınıf kategoriler (Varlıklar, Süreçler, Niteleyiciler)

❖ Varlıklar

➤ Sosyal Evren

▪ *Kişiler*

SELF: Kendi (1)

SELVES: Kendileri (2)

OTHER: Diğer kişiler (3)

▪ *Roller*

MALE ROLE: Eril rol (4)

FEMALE ROLE: Dişil rol (5)

NEUTER ROLE: Nötr rol (6)

JOB ROLE: İş rolü (7)

▪ *Topluluklar*

SMALL GROUP: Küçük grup (8)

LARGE GROUP: Büyük grup (9)

➤ Kültürel Evren

▪ *Kültürel varlıklar*

FOOD: Yiyecek (11)

CLOTHING: Kıyafet (12)

TOOL: Araç (13)

▪ *Kültürel ortamlar*

SOCIAL PLACE: Sosyal yer (20)

▪ *Kültürel örüntüler*

IDEAL VALUE: İdeal değerler (22)

DEVIATION VALUES: Sapma değerleri (23)

ACTION-NORM: Eylem normu (24)

MESSAGE-FORM: Mesaj-formu (25)

THOUGHT-FORM: Düşünce-formu (26)

NONSPECIFIC-OBJECTS: Belirsiz isimler (15)

➤ Doğal Evren

BODY PART: Beden ögesi (10)

NATURAL OBJECTS: Doğal varlıklar (14)

NATURAL WORLD: Doğal dünya (21)

▪ *Niteleyiciler*

SENSORY-REFERENCE (QUALITY): Duyusal-referans (16)

QUANTITY-REFERENCE: Miktar-Referans (19)

SPACE-REFERENCE: Uzay-Referans (18)

TIME-REFERENCE: Zaman-Referans (17)

❖ Süreçler

➤ Psikolojik süreçler

▪ *Duygular*

AROUSAL: Uyarılma (33)

URGE: Dürtüsel (28)

AFFECTION: Sevgi (29)

PLEASURE: Memnuniyet (30)

DISTRESS: Sıkıntı (31)

ANGER: Kızgınlık (32)

- *Düşünce*
 - SENSE: Duyu (27)
 - THINK: Düşünme (34)
 - IF: Eğer (35)
 - EQUAL: Eşit (36)
 - NOT: Değil (37)
 - CAUSE: Sebep (38)
- *Değerlendirme*
 - GOOD: İyi (40)
 - BAD: Kötü (41)
 - OUGHT: Zorunluluk (42)
- *Davranış Süreçleri*
 - *Sosyal duygusal eylemler*
 - COMMUNICATE: İletişim (43)
 - APPROACH: Yaklaşım, tutum (51)
 - GUIDE: Yönlendirme (45)
 - CONTROL: Kontrol (46)
 - ATTACK: Saldırmak (48)
 - AVOID: Kaçınmak (49)
 - FOLLOW: Takip etmek (47)
 - DEFENCE MECHANISM: Savunma Mekanizması (39)
 - *Araçsal eylemler: Bir amaca ulaşmak için yapılan eylemler*
 - ATTEMPT: Girişim (50)
 - WORK: İş, çalışma (54)
 - MOVE: Hareket (55)
 - GET: Almak (52)
 - POSSES: Sahiplenmek (44)
 - EXPEL: Defetmek, atmak (53)

2.Sınıf kategoriler (Kurumsal, Mevkiler, Psikolojik)

- ❖ Kurumsal alan: *Kurallar ve hareketlerin sosyal bağlamının tanımlaması*
 - ACADEMIC: Akademik (56)
 - ARTISTIC: Artistik (57)
 - COMMUNITY: Cemaat, halk (58)

ECONOMIC: Ekonomik (59)
 FAMILY: Aile (60)
 LEGAL: Yasal (61)
 MEDICAL: Tıbbi (62)
 MILITARY: Askeri (63)
 POLITICAL: Politik (64)
 RECREATIONAL: Eğlence, dinlenmeyle ilgili (65)
 RELIGIOUS: Dini (66)
 TECHNOLOGICAL: Teknolojik (67)

❖ Mevki çağrışımları

HIGHER-STATUS: Yüksek statü (68)
 PEER-STATUS: Eş-statü (69)
 LOWER-STATUS: Düşük statü (70)

❖ Psikolojik temalar

OVERSTATE: Abartmak (71)
 UNDERSTATE: Küçümsemek, tevazu (72)
 SIGN-STRONG: Güçlü belirtisi (73)
 SIGN-WEAK: Zayıf belirtisi (74)
 SIGN-ACCEPT: Kabul belirtisi (75)
 SIGN-REJECT: Ret belirtisi (76)
 MALE-THEME: Eril tema (77)
 FEMALE-THEME: Dişil tema (78)
 SEX-THEME: Cinsel tema (79)
 ASCEND-THEME: Yükselme teması (80)
 AUTHORITY-THEME: Otorite teması (81)
 DANGER-THEME: Tehlike teması (82)
 DEATH-THEME: Ölüm teması (83)

NULL: Kategoriye girmeyen kelimeler (84)

Harvard-III sözlüğünde “he”, “she” ve türevleri (her, his, him) gibi erkek-dişi farkı olan zamirler ayrı değerlendirilmeyip sonraki sürümlere bırakılmıştır (Montague, 1971). Türkçede zamirlerde cinsiyet ayrımı olmadığı için, böyle bir çalışma gerekli olmayacaktır.

Harvard-III sözlüğünün ilk çevirisi 1992 yılında benzer bir içerik analizi çalışmasında kullanılmak üzere Doç. Dr. Levent Mete tarafından yapılmıştır (Mete ve diğ., 1993). Kendisiyle

irtibata geçilip sözlük temin edilmiş ve çalışmamıza uygun hale getirmek için aşağıda bahsedilen işlemlerden geçirilip düzenlenmiştir. Tablo 3.1’de ilk çevirisinin örnek bir bölümü verilmiştir. Her satırda kelime ve kelimenin girdiği kategoriler listelenmiştir. Bir kelime en az bir, en fazla beş kategoriye girmektedir. Kodlama farkından dolayı Türkçe karakterler düzgün değildir. Her sözcüğün karşısında bulunan sayılar kategorilere işaret etmektedir. Örneğin ABES sözcüğünün karşısında bulunan 23 sayısı DEVIATION VALUES (Sapma Değerleri), 74 sayısı ise SIGN-WEAK (Zayıflık Belirtisi) kategorisini göstermektedir. Harvard-III sözlüğünde bu şekilde toplam 4919 sözcük bulunmaktadır.

Tablo 3.1: Harvard-III Türkçe Sözlüğü’nün düzenlenmeden önceki yapısı.

Kelime	Kategoriler
ABART	19
ABES	23 74
AB\$	04 68 60
A!ABEY	04 68 60
ABUK	23
A>	28
ACAY\$P	23
ACELE	31
ACI	31
A>IK	18
A>IKLA	43
A>IKLIY	43
A>IKLAMA	25 81
ACIKLI	31 74

Tablo 3.2: Harvard-III Sözlüğünün düzenlendikten sonraki yapısı.

Kelime	Kategori 1	Kategori 2	Kategori 3	Kategori 4	Kategori 5
ABART	19				
ABES	23	74			
ABİ	4	68	60		
ABLA	5	68	60		
ABUK	23				
ACABA	26	34	35		
ACAYİP	23				
ACELE	31				
ACI	31				
ACIK	28				
ACIKLI	31	74			

Türkçe karakterler düzeltilince aynı kelimeden birden fazla oluşmuştur. Bu kelimelerden birer tane bırakılınca Tablo 3.2'deki sözlük elde edilmiştir. Her kelime 5 kategoriye girmektedir. Tabloda boş olan kategoriler o kelimenin 5'den az kategorisi olduğu içindir. Birinci kategori boş olamaz. Diğerlerinin aralarında boşluk olamaz. Kategoriler arasında da herhangi bir önem derecesi yoktur. Hepsi aynı değerdedir. Örneğin ACABA kelimesi 26, 34 ve 35 numaralı kategorilere girmektedir. Bazı kategori numaraları ve isimleri Tablo 3.3'de verilmiştir. Bazı kategoriler ve altındaki kelimeler Şekil 3.3'de verilmiştir.

Tablo 3.3: Harvard-III sözlüğündeki bazı kategoriler ve numaraları.

Kategori No	Kategori adı
1	SELF
2	SELVES
3	OTHER
4	MALE ROLE
5	FEMALE ROLE
6	NEUTER ROLE
7	JOB ROLE
8	SMALL GROUP
9	LARGE GROUP
10	BODY PART
11	FOOD
12	CLOTHING
13	TOOL
14	NATURAL OBJECT
15	NON-SPECIFIC OBJECTS
16	SENSORY REFERENCE
17	TIME REFERENCE
18	SPACE REFERENCE
19	QUANTITY REFERENCE

**** 1 SELF ****
BANA
BEN
BENİM
BENİMKİ
BİZZAT
KENDİ
KENDİM
**** 2 SELVES ****
BİZ
BİZİM
BİZLER
KENDİMİZ
**** 3 OTHER ****
KENDİLERİ
ONLAR
SANA
SEN
**** 4 MALE ROLE ****
ABİ
AĞABEY
ADAM
AKTÖR
ALLAH
AMCA
AYLAK
AYYAŞ
BABA
BÜYÜKBABA
DEDE

Şekil 3.3: Harvard-III sözlüğündeki bazı kategoriler ve içerdği kelimeler.

3.2.2. Harvard-IV Sözlüğü ve Diğer Kategoriler

Harvard-IV sözlüğü, Harvard-III sözlüğünün daha genişletilmiş ve geliştirilmiş sürümüdür ve GI içerik analizi yazılımının yeni sürümünde Lasswell sözlüğü (Namenwirth ve Weberi, 1987) ile birlikte kullanılmıştır (Hurwitz, 2002).

GI ana sayfasının yetkilisi olan Roger Hurwitz ile irtibata geçilip sözlüğün Türkçeye çevirip kullanma izni alınmıştır. Sözlükte 11788 kelime bulunmaktadır. Web sayfalarında yayınlanmış olan Excel tablosu formatındaki sözlük, GI'in kullandığı formatın aynısıdır. Tablo 3.4'de orijinal sözlüğün bir parçası verilmiştir. Tüm sözlükte 11788 kelime bulunmaktadır. İlk sütun alfabetik sırada sözcükleri gösterir. İkinci sütun sözcüğün Harvard-IV sözlüğünde, Lasswell (Lvd) sözlüğünde veya her ikisinde de (H4Lvd) olup olmadığını gösterir. Daha sonraki 182 sütun o sözcüğün bulunduğu kategorileri gösterir (Hurwitz, 2002).

Tablo 3.4: Harvard-IV İngilizce original sözlüğü.

Sözcük	H4Lvd	Positiv	Negativ	Pstv	Affil	Ngvt	Hostile	Strong	Power	Weak	Submit	Active	Passive
ABANDON	H4Lvd		Negativ			Ngvt				Weak			
ABATE	H4Lvd		Negativ										Passive
ABATEMENT	Lvd												
ABDICATE	H4		Negativ							Weak	Submit		Passive
ABHOR	H4		Negativ				Hostile						Passive
ABIDE	H4	Positiv			Affil							Active	
ABILITY	H4Lvd	Positiv						Strong					
ABJECT	H4		Negativ							Weak	Submit		Passive

Tablo 3.5: Harvard-IV sözlüğünün Türkçeye çevrilmesinin ilk aşaması.

Sözcük	Anlam 1	Anlam 2	Anlam 3	Anlam 4	Anlam 5	H4Lvd	Positiv	Negativ	Pstv	Affil	Ngvt	Hostile	Strong
ABANDON	bırakmak	terk etm1	vazgeçmek			H4Lvd		Negativ			Ngvt		
ABATE	azaltmak	çekilmek	dindirmek	hafifletmek	yatıştırmak	H4Lvd		Negativ					
ABATEMENT	azaltma	hafifletme	indirme	iptal	indirim	Lvd							
ABDICATE	çekilmek	feragat etmek	istifa etmek	vazgeçmek	bırakmak	H4		Negativ					
ABHOR	tiksilmek	nefret etmek	iğrenmek			H4		Negativ				Hostile	
ABIDE	beklemek	çekmek	dayanmak	katlanmak	kurala uymak	H4	Positiv			Affil			
ABILITY	beceri	hüner	kabiliyet	yetenek	yeterlik	H4Lvd	Positiv						Strong
ABJECT	adi	alçak	iğrenç	rezil	aşağılık	H4		Negativ					

Çeviri üç aşamada yapılmış ve İngilizce-Türkçe sözlük olan Tureng kullanılmıştır. Online sözlük Tureng kategorilere ayrılmış 2 milyondan fazla sözcük ve sözcükleri içermektedir (Süyel, 2002).

İlk aşamada GI sitesinde (Hurwitz, 2002) Excel tablosu halinde yayınlanan sözlükte bulunan kelimelerin tek tek Türkçe karşılıkları alınıp yeni bir tabloya konulmuş ve karşılıklarına kategorileri kopyalanmıştır. Birden fazla Türkçe karşılığı olanların karşılıklarına tüm anlamları yazılmış ve aynı satıra kategoriler kopyalanmıştır (Tablo 3.5). İkinci aşamada yan yana aynı satırda olan anlamlar için ayrı ayrı satırlar açılmış ve kategorileri aynı şekilde bu satırlara kopyalanmıştır (Tablo 3.6). Bu aşamada aynı Türkçe kelimedenden birçok satır elde edildiğinden bunların elenmesi gerekmektedir. Ancak burada sorun aynı sözcüğe karşılık gelen kategorilerinin birbirinden farklı olmasıdır. Hangi kategori karşılıklarının alınması hangilerinin silinmesi gerektiğine kelimenin Türkçede kullanım şekline göre karar verilmiştir. Üçüncü aşamada anlamsız kalan satırlar tamamen silinmiş, diğer satırlardaki kelimelerin göre kategorilerin birleşim kümesi yazılmıştır (Tablo 3.7)

Örneğin “Funny” sözcüğünün Türkçede sözcük karşılıklarından biri de ‘komik’ sözcüğüdür. H4 kategorileri ise *Positiv*, *Pstv*, *Virtue*, *EVAL* olarak belirlenmiştir. Hazırladığımız sözlükte

‘komik’ sözcüğünün karşısına kategori olarak her ikisinde de geçen kategorilerin birleşimi Negativ, Ngvtv, Weak, Vice, Ovrst, RspLoss, RspTot, *Positiv*, *Pstv*, *Virtue*, *EVAl* yazılacaktır.

Tablo 3.6: Türkçe Harvard IV sözlüğünün Türkçeye çevrilmesinin ikinci aşaması.

<i>Sözcük</i>	<i>H4Lvd</i>	<i>Positiv</i>	<i>Negativ</i>	<i>Pstv</i>	<i>Affil</i>	<i>Ngvtv</i>	<i>Hostile</i>	<i>Strong</i>	<i>Power</i>	<i>Weak</i>
abartı	H4		Negativ							
abartı	H4Lvd	Positiv								
abartı	H4Lvd		Negativ			Ngvtv				
abartılı	H4Lvd		Negativ			Ngvtv				
abartma	H4		Negativ							
abartmak	H4Lvd									
abartmak	H4Lvd							Strong		
abartmak	H4Lvd							Strong		
abartmak	H4Lvd									

Tablo 3.7: Türkçe Harvard IV sözlüğünün Türkçeye çevrilmesinin üçüncü aşaması.

<i>Sözcük</i>	<i>H4Lvd</i>	<i>Positiv</i>	<i>Negativ</i>	<i>Pstv</i>	<i>Affil</i>	<i>Ngvtv</i>	<i>Hostile</i>	<i>Strong</i>	<i>Power</i>	<i>Weak</i>
abajur	H4Lvd									
abartı	H4Lvd	Positiv	Negativ			Ngvtv				
abartılı	H4Lvd		Negativ			Ngvtv				
abartma	H4		Negativ							
abartmak	H4Lvd							Strong		

Harvard IV Kategori Etiketleri;

GI bir içerik analizi programıdır. Toplamda 183 kategori vardır. Kullanılan kategori etiketleri 4 kaynaktan alınmıştır.

1. Harvard IV sözlüğü
2. Lasswell değer sözlüğü
3. Yeni oluşturulan bazı kategoriler
4. Önceden belirsizliği kaldırmak için geliştirilen fakat şimdi kullanıma açık olan etiketleme kategorileri

Yukarıda bahsedilen 4 kategorinin etiket isimleri ayırt etmek için harf büyüklüğüne duyarlı isimler konmuştur. Harvard IV kategorilerinin sonuna @ işareti gelir. Lasswell sözlüğünde (Namenwirth ve Weberi, 1987) olan kategori isimlerine Lw eklenmiştir. Bazı etiketleme kategorileri büyük harfle ve diğerleri büyük küçük karışık harflerle isimlendirilmiştir.

Lasswell sözlüğü Namenwirth ve Weber tarafından geliştirilmiştir. Daha sonra fazla değişmemiştir. Harvard IV sözlüğü ise 1998 senesinde genişletilmiş ve neredeyse tüm kelimeleri kapsamıştır. Fakat özel isimlerin hepsini içermemekle beraber sıklığı milyonda 4 ve üstü olanları kapsamıştır (Hurwitz, 2002). Kategori isimleri aşağıdaki gibidir.

1. En büyük iki kategori

Positiv, Negativ

2. Osgood'un üç anlamsal boyutu

- Pstv, Ngtev (Eski sürümdeki şekilleridir. Yeni sürümde eskileri de korunmuş ve kategoriye farklı kelimeler eklenerek *Positiv*, *Negativ* ismi verilmiştir),
- Strong, Weak
- Active, Passive

3. Zevk, acı, erdem ve ahlak sözcükleri

Pleasur (Zevk), Pain (Acı), Feel (Hissetme), Arousal (Uyarılma), EMOT (Heyecan), Virtue (Erdem), Vice (Kusur)

4. Abartı veya hafife almayı gösteren kelimeler

Ovrst (Abartı), Undrst (Küçümsemek)

5. Belirli bir kurum dilini yansıtan kelimeler.

Academ (Akademik), Doctrin (Öğreti), Econ@ (Ekonomik), Exprsv (İfade), Legal (Yasal), Milit (Askeri), Polit@ (Politika), Relig (Din)

6. Roller, toplumlar, ritüeller, insanlar arası ilişki çeşitleri gibi kurumsal bağlamlardan herhangi birini ifade eden kelimeler.

Rol (Rol), COLL (Toplumlar), Work (İş), Ritual (Ritüel, Ayin), SocRel (Sosyal ilişkiler)

7. Atfedici toplumsal kategoriler yanı sıra, insan ve hayvanı işaret eden genel ifadeler.

Race (İrk), Kin@ (Akrabalık), MALE (Erkek), Female (Bayan), NonAdlt (Yetişkin olmayan), HU (İnsan), ANI (Hayvan)

8. Yerleri ve aralarındaki güzergâhları ifade eden kelimeler.

Place (Yer), Social (Sosyal), Region (Bölge), Route (Rota), Aquatic (Su ile ilgili), Land (Kara), Sky (Gökyüzü)

9. Varlıklar için kullanılan kelimeler.

Object (Nesne), Tool (Araç), Food (Yiyecek), Vehicle (Araç), BldgPt (Bina öğeleri), ComnObj (İletişim nesneleri), NatObj (Doğal nesneler), BodyPt (Beden öğeleri)

10. İletişim süreçleri.

ComForm (İletişim şekli), Say (Söylemek)

11. Teşvik ile ilişkili kelimeler.

Need (İhtiyaç), Goal (Hedef), Try (Dene), Means (Usûl), Persist (İsrar), Complet (Tamamlama), Finish (Bitirme), Fail (Başarısızlık)

12. Diğer süreç ya da değişim kelimeleri.

NatrPro (Doğal süreçler), Begin (Başla), Vary (Değişim), Increas (Artma), Decrease (Azalma), Finish (Bitiş), Stay (Kalmak), Rise (Yükselmek), Exert (Sarf), Fetch (Almak), Travel (Seyahat), Fall (Düşüş)

13. Akli yönelim

Think (Düşünmek), Know (Bilmek), Causal (Nedensel), Ought (Gerekliklik), Perciev (Algılamak), Compare (Karşılaştırmak), Eval@ (Değerlendirmek), EVAL (Değerlendirmek), Solve (Çözmek), Abs@ (Soyut), Quality (Kalite), Quantity (Miktar), NUMB (Sayılar), ORD (Kaçıncı), CARD (Sayılar), FREQ (Sıklık), DIST (Mesafe), Time@ (Zaman), TIME (Zaman), Space (Mekânsal), Rel (İlişki), COLOR (Renk)

14. “Ben”, “Sen”, ”Biz”, “Siz” yönelimli zamirler ve isimler.

Self (Kendi), Our (Bizim), You (Siz, Sen), Name (İsim), Nation (Millet)

15. “Evet”, “Hayır”, inkâr ve ünlem ifade eden kelimeler.

Yes (Evet), No (Hayır), Negate (Olumsuz), Intrj (Argo)

16. Fiil türleri

IAV (Yorum fiilleri), DAV (Tarif fiilleri), SV (Hal fiilleri)

17. Sıfat türleri

IPadj (İlişki sıfatları), IndAdj (Bireysel sıfatlar)

18. Güç

PowGain (Güç kazanımı), PowLoss (Güç kaybı), PowEnds (Güç hedefi), PowAren (Güç bölgesi), PowCon (Güç çatışması), PowCoop (Güç iş birliği), PowAuPt (Yetkili güç katılımcıları), PowPt (Güç katılımcıları), PowDoct (Güç ilkesi), PowAuth (Güç yetkisi), PowOth (Güçle ilgili diğer kavramlar), PowTot (Güçle ilgili tüm kavramlar)

19. Dürüstlük ahlaki değerlerle ilgili kelimeler

RcEthic (Dürüstlük etiği), RcRelig (Dini dürüstlük), RcGain (Dürüstlük kazanımı), RcLoss (Dürüstlük kaybı), RcEnds (Dürüstlük hedefleri), RcTot (Dürüstlikle ilgili tüm kelimeler)

20. Konum, onur, tanınma ve itibardan gelen saygı kelimeleri.

RspGain (Saygı kazanımı), RspLoss (Saygı kaybı), RspOth (Diğer), RspTot (Toplam)

21. Sevgi, merhamet ve dostlukla ilgili yakınlıktır.

AffGain (Sevgi kazanımı), AffLoss (Sevgi kaybı), AffPt (Sevgi katılımcısı), AffOth (Diğer), AffTot (Toplam)

22. Servet, sahip olduğunun değerini bilmekle ilgili kelimeler

WltPt (Servet katılımcısı), WltTran (Servet transferi), WltOth (Diğer), WltTot (Toplam)

23. Canlının refahı ile ilgili kelimeler

WlbGain (Refah kazanımı), WlbLoss (Refah kaybı), WlbPhys (Refahın fiziksel yönü), WlbPsys (Refahın psikolojik yönü), WlbPt (Refah katılımcısı), WlbTot (Toplam)

24. Aydınlanma, bilgi, iç görü ve kişisel ve kültürel ilişkiler ile ilgili bilgilerdir.

EnlGain (Bilgi kazanımı), EnlLoss (Bilgi kaybı), EnlEnds (Bilgi edinme hedefi), EnlPt (Aydınlanma katılımcısı), EnlOth (Diğer), EnlTot (Toplam)

25. Sanat, estetik gibi alanlarda becerileri yansıtan kelimeler.

SklAsth (Estetik becerisi), SklPt (Beceri katılımcısı), SklOth (Diğer), SklTot (Toplam)

26. Değer alanlarından herhangi birine özgü olmayan sözlük kategorileri.

TrnGain (İşlem kazancı), TrnLoss (İşlem kaybı), TranLw (İşlem), MeansLw (Araçlar), EndsLw (Hedef), ArenaLw (Bölge), PtLw (Katılımcı), Nation (Ulus), Anomie (Tercih inkarı), NegAff (Olumsuz duygular), PosAff (Olumlu duygular), SureLw (Kesinlik), If (Eğer), NotLw (Ret), TimeSpc (Zaman-Mekân), FormLw (Biçim). HV-IV kategorilerinin geniş açıklamaları Ek 3’de verilmiştir.

3.3. MAKİNE ÖĞRENMESİ

3.3.1. Sınıflandırma Yöntemleri

Metin sınıflandırma bir dokümanın (d) verilen sınıflardan $C = \{c_1, c_2, \dots, c_j\}$ hangisine ait olduğunu bulunmasıdır.

Genellikle el ile yazılmış kurallar ve eğitici (supervised) MÖ yöntemleri birleştirerek uygulanır. El ile yazılan kurallar kelime kombinasyonlarına veya seçilmiş kelimelere bağlı olabilir. Örneğin istenmeyen mailleri belirlemek için “siz seçildiniz”, “TL kazandınız” veya mail kara listesi gibi kelimeler özellik (feature) olarak kullanılabilir. Kurallar dikkatlice belirlenirse doğruluk yükselir.

Metin sınıflandırması eğitici MÖ metodu ile yapıldığında sınıflandırıcı $\gamma : d \rightarrow c$, $c \in C$ 3.1'deki gibi tanımlanır.

Girdi verileri:

d : doküman

$C = \{c_1, c_2, \dots, c_j\}$: sınıflar

m : elle işaretlenmiş dokümanlar $(d_1, c_1), (d_2, c_2), \dots (d_m, c_m)$

Çıktı: Eğitilmiş sınıflandırıcı γ 'yı şu şekilde tanımlayabiliriz

$$\gamma : d \rightarrow c \text{ , } c \in C \quad (3.1)$$

Eğitici MÖ sınıflandırıcılar çeşitlidir.

- Naive Bayes
- Logistic Regression
- Support Vector Machines
- K-Nearest Neighbors
-

Kullanılan sınıflandırıcıdan bağımsız olarak metin sınıflandırıcı, önce girdi olarak bir metin dokümanı ve özellikleri (feature) alır. Dokümanı feature'ları kullanarak temsil eder. Sonra dokümanın hangi sınıfta olduğunu gösterir.

Bu çalışmada, deneklerin sınıflandırılması için MÖ algoritmalarından olan Navie Bayes, Destek Vektör Makinesi (Support Vector Machine-SVM) ve J48 gibi öğrenme algoritmaları kullanılmıştır.

3.3.1.1. Naive Bayes Sınıflandırıcı

Bayes kuralını temel alan bir sınıflandırma metodudur. Bu çalışmada dokümanın en basit şekilde temsili olan Kelime Çantası (Bag of Words (BOW) metodu kullanılmıştır. BOW yönteminde dokümanın temsili, bir dizi kelime ve onların frekanslarından oluşan bir vektör ile yapılır.

$$\gamma(d) = c \quad (3.2)$$

Burada amaç sınıflandırıcı (γ) yardımıyla metin dokümanının (d) sınıfının (c) pozitif veya negatif olarak belirlenmesidir (3.1 ve 3.2). Dokümanın temsili için metnin içindeki tüm kelimeler özellik olarak alınabileceği gibi bir kısmı seçilip alınabilir. Basitleştirmek için yapılan

bu doküman temsil yönteminde kelimeler konumundan ayrılıp kullanıldığı için kaybolan bilgiler olur. Örneğin kelimenin cümle içindeki yeri, zaman bilgisi, cümle öge bilgisi gibi bilgiler kaybolur.

Doküman ve sınıflara Bayes kuralının uygulanması:

d: Doküman

c: Sınıf

$P(c)$: Sınıfın tüm dokümanlar içindeki olasılığı (Bu sınıf hangi sıklıkta görülür)

$P(d|c)$: Verilen sınıf içindeki doküman olasılığı

$P(d)$: Dokümanın olasılığı

Verilen dokümanın sınıfı, 3.3 eşitliğindeki gibidir.

$$P(c|d) = \frac{P(d|c).P(c)}{P(d)} \quad (3.3)$$

Her sınıf için olasılık hesaplanır ve olasılığı en büyük olan alınır. Her sınıf için hesaplandığında $P(d)$ sabit olduğundan kesirin pay kısmının hesaplanması ve en büyüğünün alınması yeterli olacaktır (3.4).

$$\max(P(d|c).P(c)) \quad (3.4)$$

d dokümanı vektör $(x_1, x_2, \dots, x_n)$ olarak ifade edilince

$$c_{\text{map}} = \max_{c \in C} P(x_1, x_2, \dots, x_n | c) \cdot P(c) \quad (3.5)$$

3.5 eşitliğini yazarken basitleştirmek için yaptığımız kabuller şöyledir:

- Kelimenin metindeki yeri önemsizdir.
- $x_1, x_2, \dots, x_n$ özelliklerinin sınıf içindeki olasılıkları birbirinden bağımsızdır.

Bu kabuller ışığında 3.6 eşitliğini yazabiliriz.

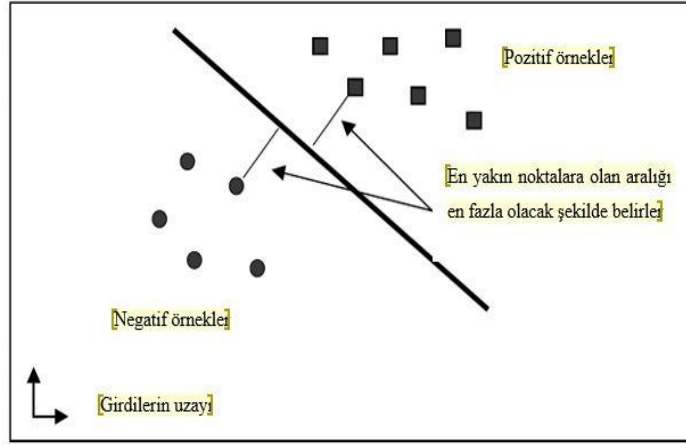
$$P((x_1, x_2, \dots, x_n)|c) = P(x_1|c) \cdot P(x_2|c) \dots P(x_n|c) \quad (3.6)$$

Sonuç olarak 3.6 eşitliği tüm sınıflar için hesaplanır. Sınıflandırılacak olan doküman maksimum değer elde edilen sınıfa atanır.

3.3.1.2. SMO Sınıflandırıcı

Sıralı Minimal Optimizasyon-SMO (Sequential Minimal Optimization): John Platt'ın ardışık minimum optimizasyon algoritması ile bir Destek Vektör Sınıflandırıcısı (Support Vector Classifier-SVC) eğitir. Bir SVC, bir grup veri alan bir sınıflandırıcıdır ve iki sınıf için, her bir verinin bu sınıflara üyeliğini belirler (Platt, 1999).

Destek vektör makineleri Vladimir Vapnik tarafından 1979'da icat edilmiştir. (Vapnik, 1982). En basit anlamda ve doğrusal formda SVM, bir dizi olumlu örnekleri maksimum aralık ile negatif örneklerden ayıran bir hiperdüzlemdir (Şekil 3.4). Lineer durumda, aralık, hiperdüzlemin pozitif ve negatif örneklerin en yakınına olan uzaklığı ile tanımlanır (Platt, 1998).



Şekil 3.4: Lineer bir SVM örneği (Platt, 1998)

3.3.1.3. J48 Sınıflandırıcı

WEKA programında J48 olarak yazılan C4.5 karar ağacı algoritmasıdır (Quinlan, 1993). Bilgi kazanımı (information gain) en yüksek olan özellik karar ağacının köküne konulur. Daha sonra diğer dallar, bilgi kazanımı en yüksek olanı alt dallar için kök düğüm olarak devam eder. Amaç sürekli olarak ağaçtaki verileri mümkün olan en büyük iki parçaya bölmektir. Ağaç bittikten sonra elemanı olmayan dallar budanır. Algoritma aşağıdaki gibi çalışır;

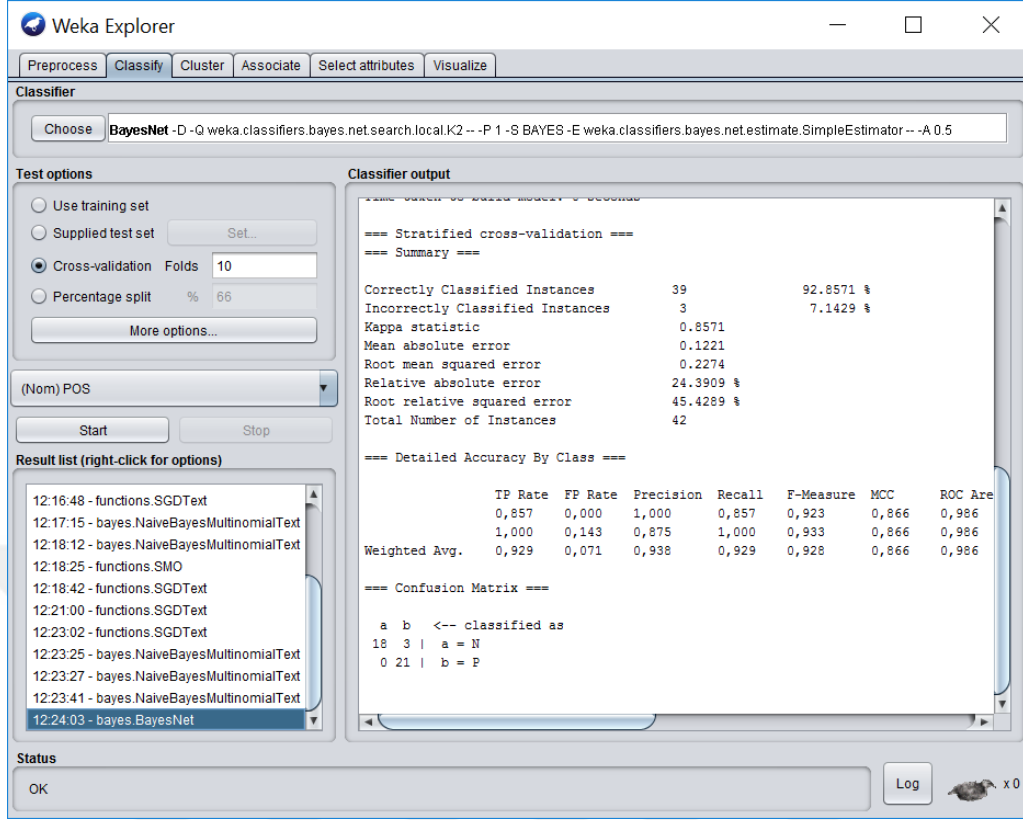
- Her adımda bütün özellikler kontrol edilir
- Her özelliğin normalize edilmiş bilgi kazanımı hesaplanır
- En iyi bilgi kazanımını veren özellik karar ağacında karar olarak taşınır

- Ardından bu yeni karar düğümünün altında bir alt liste oluşturularak alt karar ağacı inşa edilir (Kotsiantis, 2007) (Şeker, 2012)

Bilgi kazanımının anlamı, özellik ile sınıf arasında birebir bağlantıdır. Özellik, sınıflardan ne kadar bağımsızsa bilgi kazanımı da o kadar düşük çıkacaktır.

3.3.2. Weka Programı

Kişilerin kelime kullanım sonuçlarını değerlendirmek için bazı sınıflandırma yöntemleri uygulanmalıdır. Veriler, kelime kullanım analiz sonuçlarına bakılarak tanılarına göre sınıflandırılmıştır. Ardından psikolojik bozuklukların kelime kullanımı üzerindeki etkisini saptamak için sınıflandırma sonuçları incelenmiştir. Bu amaca ulaşmak için Weka adlı bir program (Witten ve Frank, 2005) kullanılır. Weka MÖ metotlarını birarada bulunduran veri madenciliği programıdır. Weka ayrıca açık kaynak kodlu bir programdır ve Java kütüphanelerine sahiptir. Bu kütüphaneleri çeşitli proje kodlarında kullanma fırsatı verir. Weka, çeşitli verilere uygulanabilen sınıflandırıcılar olarak birçok tanınmış MÖ algoritması sunmaktadır. Sınıflandırıcıları test etme ve eğitme imkânı vardır. Her bir eğitim ve test sürecinin sonuçlarını verir. Weka, özelliklerin (features) veya veri örneklerinin filtrelenmesine de izin verir. Şekil 3.5 Weka programının ekran görüntüsünü göstermektedir.



Şekil 3.5: Weka programının ekran görüntüsü.

Weka, birçok dosya formatının yanı sıra Virgülle Ayrılmış Değerler (Comma Separated Values- CSV) dosya biçimini ve kendi dosya formatı olan ARFF formatını da kullanır. CSV formatlı bir dosya, bir veri tablosu içerir. Veri tablosunun her satırı bir kaydı temsil eder ve her kaydın değeri virgülle ayrılır. CSV biçimli veri kümeleri Weka ile daha önce bahsedilen verileri eğitmek ve test etmek için oluşturulmuştur. CSV dosyaları Excel tarafından okunabilir.

Öznitelik İlişki Dosyası Biçiminde (Attribute Relationship File Format- ARFF) başlık ve veri bölümü vardır. Başlık bölümünde özellikler ve değer tipleri bulunur. Veri bölümünde ise özellik değerleri, satırlar halinde bulunmaktadır. Weka seçilen MÖ tekniğinde çapraz geçerlilik uygulamaya imkân sağlar. n-katlı çapraz geçerlilik durumunda veri rastgele n kümeye ayrılıp, n-1 küme eğitim, kalan 1 küme test için kullanılır. Bu işlem n kez değişik test verileri ile tekrarlanır. Sonuç olarak ise n sonucun ortalaması alınır.

3.4. MORFOLOJİK ANALİZ

Bu çalışmada, tr-tagter (Oflazer, 1994) isimli morfolojik analiz programı kullanılmıştır. Bu program Şekil 3.6'da gösterilen biçimde metin verisi alır.

```

<DOC>
<TITLE>
<S>
Ben
zarar
görüyorum
.
</S>
<S>
Canım
çok
yandı
,
artık
yanmasını
istemiyorum
.
</S>
</DOC>

```

Şekil 3.6: Morfolojik analiz programının aldığı veri yapısı.

Xerox Finite State yazılımının bir parçası olan "lookup.exe" dosyasını kullanır. Bu program, kelimeleri kök ve eklerine ayırır. Bir kelimenin birden fazla anlamı veya türü olabilir. Aşağıdaki Tablo 3.8, morfolojik analiz programının başlıca kelime türlerini göstermektedir. Tablo 3.9 morfolojik analiz programının örnek sözcük kelimelerini göstermektedir.

Tablo 3.8: Morfolojik analiz programının kelime türleri.

Kelime Türü	Açıklama
+Noun	İsim
+Adj	Sıfat
+Adv	Zarf
+Verb	Fiil
+Pron	Zamir

Tablo 3.9: Morfolojik analiz programının örnek anlam kelimeleri.

Kelimeler	Açıklama
+Conj	Bağlaç
+Inf	Mastar
+PastPar	Geçmiş zaman
+FutPart	Gelecek zaman
+Prop	Özel isim
+Zero	Yapım eki almamış isim

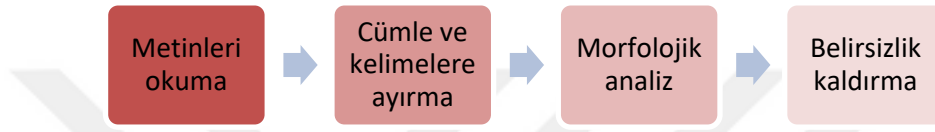
Dilde olabilecek anlam belirsizliğinden dolayı morfolojik analiz birden fazla sonuç çıkarabilir. Bu bir kelimenin birden fazla anlamı olabileceği ve dolayısıyla birden fazla analiz sonucuna sahip olabileceği anlamına gelir. Bu şekilde çok fazla kelime olabileceği için morfolojik analiz çıktısından en doğru sonucu seçmek önemli bir sorundur. Çözüm için bir belirsizlik giderici anlam ayrıştırıcı yapan araç program kullanılır. Anlam ayrıştırıcı ise bir kelimenin veya cümlelerin hangi dilsel bağlamda kullanıldığını inceleyerek o kelime veya cümlelerin amaçlanan anlamını bulmaya yarayan bir DDİ aracıdır. Anlam ayrıştırma işleminden sonra, en uygun sonuç bu araç tarafından seçilmektedir. Bu çalışmada, tr-disamb olarak bilinen Türkçe anlam ayrıştırıcı program kullanılmıştır (Ofłazer, 1994). Türkçe'nin kurallarına uygun olarak morfolojik veri çıktısından alınan metindeki en doğru çıktıya karar verir. Bu program %96 oranında doğru sonuç vermektedir. Tablo 3.10'de iki cümlelerin morfolojik analizi ve belirsizlik kaldırılması sonucu verilmiştir.

Tablo 3.10: Verinin morfolojik analizi ve belirsizlik kaldırılması.

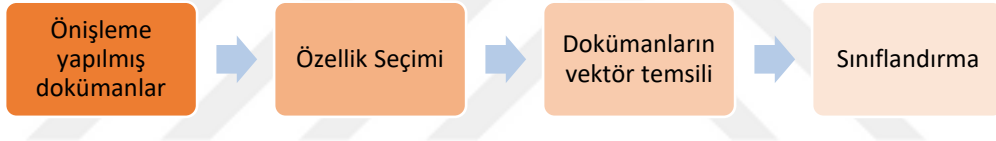
Orjinal veri/ Morfolojik analizin girdisi	Morfolojik analiz çıktısı/ Belirsizlik kaldırılmasının girdisi	Belirsizlik kaldırılmasının çıktısı
<DOC> <TITLE> <S> Ben zarar görüyorum . </S>	<DOC> <DOC> <TITLE> <TITLE> <S> <S> ben ben +Noun+A3sg+Pnon+Nom ben ben +Pron+Pers+A1sg+Pnon+Nom zarar zarar +Noun+A3sg+Pnon+Nom görüyorum gör+Verb+Pos+Prog1+A1sg .. +Punc </S> </S>	<DOC> <DOC> <TITLE> <TITLE> <S> <S> ben ben+Pron+Pers+A1sg+Pnon+Nom zarar zarar+Noun+A3sg+Pnon+Nom görüyorum gör+Verb+Pos+Prog1+A1sg .. +Punc </S> </S>
<S> Canım çok yandı , artık yanmasını istemiyorum . </S> </DOC>	<S> <S> Canım can +Noun+A3sg+P1sg+Nom Canım can +Noun+A3sg+Pnon+Nom^DB+Verb+Zero+Pres+A1sg Canım Canım +Noun+Prop+A3sg+Pnon+Nom çok çok +Det çok çok +Adverb çok çok +Adj çok çok +Postp+PCabl yandı yan +Noun+A3sg+Pnon+Nom^DB+Verb+Zero+Past+A3sg yandı yan +Verb+Pos+Past+A3sg , , +Punc artık artık +Adverb artık artık +Noun+A3sg+Pnon+Nom artık artık +Adj yanmasını yan+Verb+Pos^DB+Noun+Inf2+A3sg+P3sg+Acc istemiyorum iste+Verb+Neg+Prog1+A1sg .. +Punc </S> </S> </DOC> </DOC>	<S> <S> Canım can+Noun+A3sg+P1sg+Nom çok çok+Adverb yandı yan+Verb+Pos+Past+A3sg , , +Punc artık artık+Adverb yanmasını yan+Verb+Pos^DB+Noun+Inf2+A3sg+P3sg+Acc istemiyorum iste+Verb+Neg+Prog1+A1sg .. +Punc </S> </S> </DOC> </DOC>

Bu bölümde, veri kümelerini içeren sistemin deneysel kurulumu ve önışlemleri, özellik seçimi metodolojileri ve kullanılan sınıflandırma algoritmaları ayrıntılı olarak açıklanmıştır.

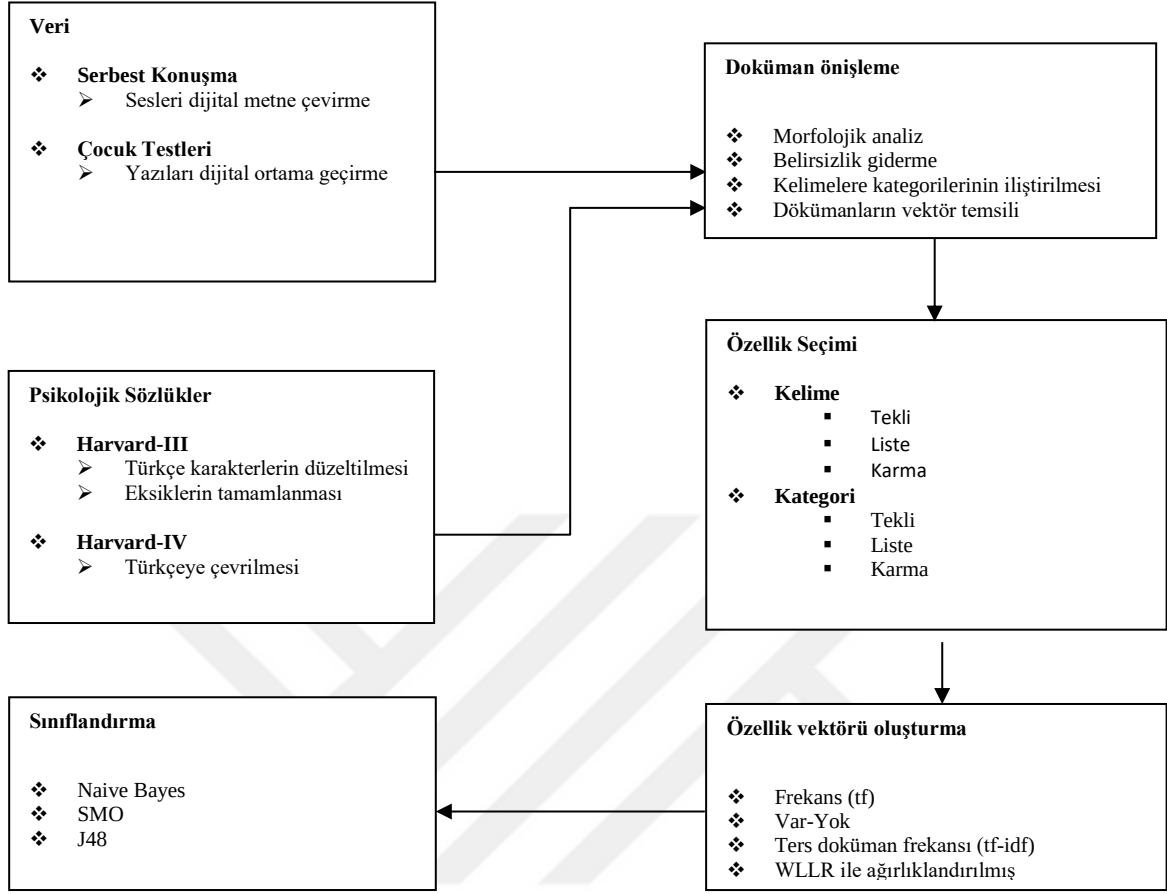
Alınan metinler Şekil 3.7’de görüldüğü gibi sistem tarafından ilgili klasörden okunur. Cümle ve kelimelere ayrılır. Morfolojik analizi yapılır ve morfolojik analiz sonuçları belirsizlik analizinden geçirilir. Belirsizlik analizi yapılmış metinler Şekil 3.8’deki gibi özellik seçimine göre her bir metin bir vektörle temsil edilir. Bir vektör bir satıra denk gelecek şekilde eğitim tabloları hazırlanır. Şekil 3.9 ise bu işlem aşamalarını bir sistem şeması halinde göstermektedir.



Şekil 3.7: Doküman önışlem aşamaları.



Şekil 3.8: Doküman sınıflandırma aşamaları.



Şekil 3.9: Sistem şeması.

3.5. VERİ KÜMESİ

Sağlıklı ve psikiyatrik rahatsızlığı olan deneklerden toplanmış veriler üzerinde incelemeler yapılmıştır. Veriler çocuk, yetişkin serbest konuşma ve yetişkin soru-cevap olarak üç grupta toplanmıştır. Çocuk verisi Beier cümle tamamlama testi ve TAT testi metinlerinden oluşmaktadır. Tüm veri gruplarında teşhislerine göre sağlıklı-hasta sınıflandırması yapılmıştır. Çeşitli DDİ özellik seçim yöntemleri kullanılarak sınıflandırmalar yapılmıştır. Tüm veriler çocuk ve yetişkin olarak iki ayrı grupta cinsiyet sınıflandırması yönünden de incelenmiştir.

3.5.1. Çocuk Verisi (Testler)

Bu çalışmada kullanılan çocuk verileri bir NPIstanbul Nöropsikiyatri Kliniğinin Çocuk Ergenlik Ünitesinde psikiyatrik tedavi gören 140 çocuktan alınmıştır. Yaşları okul çağında

bulunan bu çocuklardan 102'si erkek 38'i kızdır. Bir kısmında Depresyon (DEP), Dikkat Eksikliği Hiperaktivite Bozukluğu (DEHB) ve Anksiyete (ANK) gibi psikolojik rahatsızlıklar görülmektedir. Herhangi bir teşhis konulmamış çocuklar ise sağlıklıdır (SAĞ). Aynı anda depresyon ve DEHB görülen 12 kişi bulunmaktadır. Bazı katılımcılara konulan tanılar konumuzun dışında psikolojik rahatsızlıklar olması sebebi ile çalışmaya dahil edilmemiştir. Çocuk verilerinin teşhislere ve cinsiyete göre dağılımı Tablo 3.11'de verilmiştir.

Tablo 3.11: Çocuk deneklerin bilgileri.

Teşhis	Cinsiyet		Toplam
	E	K	
DEP	27	6	33
ANX	9	9	18
DEHB	21	8	29
Diğer	33	10	43
SAĞ	22	7	29

Cümle tamamlama testleri tipik olarak kullanılan cümlelerin başlangıçlarını katılımcılara verir ve geri kalan kısımlarının tamamlanmasını ister. Katılımcı da geri kalan kısmı kendine en uygun bir biçimde doldurur. Verilen yanıtlar bu katılımcının tutumu, inançları, motivasyon durumu ve diğer ruhsal durumları ile ilgili belirtileri ortaya koymaktadır. Beier Cümle Tamamlama Testi ağırlıklı olarak kişilik analizi, kliniksel uygulamalar, tutum değerlendirme ve psikoloji, yönetim ve eğitim gibi çeşitli alanlarda başarı motivasyonunun ölçümünü içerir (Beier ve Hanfmann, 1956).

Katılımcılara uygulanan diğer bir test ise TAT'dir. Birtakım kartların katılımcılara gösterilip sonrasında katılımcılardan her bir kart ile ilgili hikâye anlatması istenir. Bu kartların üzerinde çeşitli durum ve konularla ilgili, üzerinde insan figürleri içeren resimler bulunur. Testi yapan kişi her söyleneni yazar. Katılımcının anlattığı hikayelerin içeriği ve yapısı o katılımcının ruhsal durumu hakkında ipuçları vermektedir (Tomkins ve Tomkins, 1947).

Katılımcıların cevapları elektronik ortama geçirilerek imla ve kelime kontrollerinden sonra kullanılır. Aşağıdaki Şekil 3.10'de bir kişiye ait örnek veri görülmektedir. Her katılımcıya ait bir metin dosyası bulunmaktadır. Bu dosyada metinler "/" işareti ile bölmelere ayrılmıştır. Öncelikle katılımcının verilen kimlik numarası, cinsiyeti ve yaşını içermektedir. Sonrasındaki

üç bölme ise, katılımcıya doktorlar tarafından konulan depresyon, anksiyete, öğrenme bozuklukları, dikkat eksikliği gibi tanıları içermektedir. Daha sonra gelen 20 bölme ise Beier cümle tamamlama testi ile ilgilidir ve devamında da TAT testi sırasında kartlara verilen cevapları içermektedir.

36151/E/11/Depresyon/dikkat eksikliği/ Kronik Tik, Maternal Depresyon/////Büyüdüğüm zaman mimar olmak istiyorum/Arkadaşlarımla çok iyi geçiniyorum/Yaramazlık yaptığım zaman annemden özür diliyorum/Okulda öğretmenimi dinliyorum/Evimizi çok seviyorum/İyi arkadaşlar her zaman iyi olmalı/Annelerin en iyisi benim annem/Benim en iyi yanımla yaramaz olmamamdır/Öğretmenler bizim başarılı olmamız için her şeyi yapar/Annemi severim ama aynı zamanda babamı da çok severim/Canım sıkılınca kardeşimle oynarım/Bazen babam eve geldiğinde sinirli olur ama sonra düzelirdi./En büyük üzüntüm kardeşimin boğazına mıknaş kaçmasıydı/Kardeşim çok sevimlidir ama aynı zamanda yaramazdır/En çok istediğim şey yeni bir oda/Cezalar ve yasaklar her zaman iyilik içindir/Evde her şey güzel olduğunda mutlu olurum/Çocuklar oyun oynamayı çok severler/Korktuğum şey ailemizden birinin hastalanması/Büyüklükler hem kızar hem de sever/Babam ve ben birbirimizi çok severiz/Yalnız kalınca başıma kötü şeyler geleceğini zannederdim./ Çocuk keman çalmak istiyor. Keşke benimde bir kemanın olsa diyor. Düşünüyör. Aklına fikirler almaya getiriyor./Kız kitap okumayı çok seviyor. Gezerken bile okuyor. Başkaları ona hayran kalıyor. Diğerleri kendi işlerine bakıyor.// Dilenci mi? Bir gün dilenci para istemek için çıkıp. Bulamamış.//Bir ailede iki tane çocuk varmış. Baba işte anne ve çocuklar evde. Ama çocukları tembillemiş sakın salona girmeyin. Ben markete gideceğim demiş. Çocuklar annelerini dinlemiş. Geldiğinde salona bakıp doğru olduğunu görünce çocuk daha çok sevmiş. ////Birgün, işle ilgiliymiş, işe giderken geç kalmış. Otobüse binerek gitmiş. Koşmuş Patronunu görmüş. Görmeden işinin başına gitmek istemiş. Patronu bir şey dememiş işine verdiği öneminden dolayı birşey dememiş.//Bir kişi yolda giderken kötü kişilere rastlamış.yaralanmış. Etrafına başkaları gelip onu hastaneye götürmüşler. kurşunu almak için karnını yarmışlar. Bayıltmışlar. Kurşunu almaya çalışmışlar. Kurşunu çıkarmışlar. çocuğu eski haline getirmişler. yürüyemez haldeymiş.o kötü adamda yakalayıp hapse atmışlar.///////ev iki katlı, kenarında bahçesi ağaçlar ve çiçekler var. Böcekler dolaşacak arılar polenleri alıp bal yapacaklar. Manzara.

Şekil 3.10: Bir denneğin verisi.

Çocukların sağlıklı (SAĞ) veya herhangi bir psikolojik hastalıkları olup olmadıklarını anlamak için kullanılan metinlerin dil analizi için Türkçe morfolojik analiz (Ofłazer, 1994) ve belirsizlik giderici programları (Ofłazer, 1994) ve MÖ tekniklerinin uygulaması için Weka (Witten ve Frank, 2005) programı kullanılmıştır.

3.5.2. Yetişkin Verisi

Çalışmamızda alınacak metnin elde edilmesi için deneklere o an herhangi bir konuda aklına gelenlerin yazılması istenmiştir. Denekler psikiyatri uzmanı doktorlar tarafından teşhisi yapılmış kişilerdir.

Serbest konuşma verilerinde deneklerin hepsi yetişkin olup Depression (DEP), Mani (MAN) ve Kontrol Sağlıklı grubu (SAĞ) olarak üç grup veri üzerinde çalışılmıştır. DEP ile SAĞ ve MAN ile SAĞ. Verilerin tümü psikiyatrist Dr. Levent Mete yardımıyla İzmir Atatürk Eğitim ve Araştırma Hastanesi'nden alınmıştır.

Veriler etik onay raporu da alınarak 90 yetişkinden alınmıştır. Her gruptan 30 kişi ile görüşülüp 10 dakikalık serbest konuşmaları kaydedilmiştir. Herhangi bir yönlendirme yapılmamıştır ve soru sorulmamıştır. Katılımcıların cinsiyetleri Tablo 3.12'deki gibidir. Konuşmalar metine dönüştürülüp her bir gruptan 21 deneğin metni sistem eğitimi ve 9 deneğin metni ise test için kullanılmıştır.

Tablo 3.12: Serbest konuşma deneklerin istatistik bilgileri.

Teşhis	Cinsiyet		Toplam
	E	K	
DEP	13	17	30
MANİ	13	17	30
SAĞ	13	17	30

Toplanan veriler üzerinde cinsiyet sınıflandırması yapılmıştır. Yetişkin ve çocuk verisi olarak iki grupta yapılmıştır. Yetişkin verisi Tablo 3.13'deki gibidir. Çocuk verisi ise Tablo 3.14'deki gibidir.

Tablo 3.13: Cinsiyet sınıflandırması için yetişkin verisi.

Yetişkin Cinsiyet Verisi	
Erkek	37
Kadın	53
Toplam	90

Tablo 3.14: Cinsiyet sınıflandırması için çocuk verisi.

Çocuk Cinsiyet Verisi	
Erkek	102
Kadın	38
Toplam	140

3.6. ÖZELLİK ÇIKARIMI

Sınıflandırmanın başarılı olması için veriyi en iyi temsil eden özelliklerin bulunması gerekir. Bu çalışmamızda dokümanı en iyi temsil eden özellik vektörü bulunmaya çalışılmıştır. Metin sınıflandırmada eğitim ve test verileri bir dizi özellik vektöründen oluşmaktadır. Genellikle

özellik kümesi olarak tüm verideki ayrıık kelimeler seçilir. Bu şekilde bir metin, bir grup kelime ile temsil edilir (BOW). Başka bir yaklaşım ise yan yana gelen kelimelerin özellik olarak seçilmesidir. Bu özellik seçimine n-gram denir.

DA alanında başarı büyük oranda özellik seçimine bağlıdır (Li ve diğ., 2009). Amaç dokümanları temsil eden en iyi özellik vektörünü bulmaktır. Araştırmamızda mümkün olduğu kadar farklı özellik seçimleri denenmeye çalışılmıştır. Temel olarak kelime veya kategorinin olup olmaması (Pang ve diğ., 2002) ve frekansı ele alındı. Başka çalışmalarda kullanılmış olan ters doküman frekansı (idf) ve ağırlıklı logaritmik benzerlik oranı (Weighted Logarithmic Likelihood Ratio-WLLR) ile ağırlıklandırılmış frekans denemeleri yapılmıştır. Bu şekilde 28 unigram, 28 bigram özellik seçimi yöntemi uygulanmıştır. Genel olarak en iyi sonuçlar WLLR ağırlıklı yöntemde alınmıştır.

Deneklerin metinlerinde kullandıkları kelimeler morfolojik analiz ve belirsizlik kaldırma programları ile analiz edilmiş, bu yolla kelimelerin kökleri bulunmuştur. Kategorilerin bulunması için ise sözlükten köklerin dahil olduğu kategoriler bulunmuştur.

3.6.1. n-gram

n-gram, yanyana gelen n tane kelimededen oluşur. Özellik olarak tek tek kelimeler (Unigram) söz konusu olduğunda kelimelerin sırası önemli değildir ve birbirlerinden bağımsız oldukları düşünülür. Bigram, ikili kelime gruplarından, Trigram, üçlü kelime gruplarından oluşur (Tablo 3.15). Kelimelerin sırası önemlidir. Örneğin “Küçük çocuk okula gitti.” ile “Çocuk küçük okula gitti.” şeklinde iki cümle, BOW’da aynı temsil edilirken Bigram’da farklı temsil edilir.

Tablo 3.15: Örnek bir cümlede unigram, bigram ve trigram öğeler.

Cümle	İnsanın özü ne ise gözü de onu görür.
Unigram	[İnsanın] [özü] [ne] [ise] [gözü] [de] [onu] [görür]
Bigram	[İnsanın özü] [özü ne] [ne ise] [ise gözü] [gözü de] [de onu] [onu görür]
Trigram	[İnsanın özü ne] [özü ne ise] [ne ise gözü] [ise gözü de] [gözü de onu] [de onu görür]

n-gram kullanımı metin sınıflandırmada bilgi kaybını azalttığı düşünülür ancak bazı çalışmalarda başarıyı arttırmadığı gözlenmiştir. n sayısı arttıkça başarı azalmaktadır. (Kim ve Hovy, 2007) (Fürnkranz, 1998) (Mladenec ve Grobelnik, 1998).

Her kelime birden fazla kategoriye sahip olabilir. Bigram durumunda özellik olarak kelimelerin kategorileri seçildiğinde kelimelere ait kategorilerin ilkleri alınır.

3.6.2. Ters Doküman Frekansı

Özellik seçiminde kullanılan yöntemlerden birisi ters doküman frekansıdır. Terim frekansı, değerin dokümandaki özelliğin sıklık sayısı olduğu bir yöntemdir. Terim frekansı (tf) değerlendirilirken tüm terimler eşit derecede önemli kabul edilir. Aslında belirli terimlerin çok az ya da hiç ayırıcı özelliği yoktur. Terimlerin ayırıcı özelliğinin belirlenmesi için terim frekansı, ters doküman frekansı (idf) ile ağırlıklandırılır (Pang ve diğ., 2002). Formüle göre az dokümanda görülen bir terimin idf değerinin yüksek olacağını ancak daha çok dokümanda görülen bir terimin idf değerinin (ortak terim) düşük olacağını söyleyebiliriz (3.7). Bir terimin frekansının idf ile ağırlıklandırılmış formülü 3.8’de görülmektedir.

$$idf_t = \log \frac{N_{all}}{df_t} \quad (3.7)$$

$$tf_idf_t = tf_t * \log \frac{N_{all}}{df_t} \quad (3.8)$$

N_{all} : eğitim verisindeki doküman sayısı

df_t : t terimin görüldüğü doküman sayısı

idf_t : t teriminin ters doküman frekansı

tf_t : t teriminin doküman içindeki frekansı

3.6.3. Ağırlıklı Logaritmik Benzerlik Oranı

En iyi özellikleri seçme konusunda Nigam (Nigam ve diğ., 2000), WLLR ile sıralama özelliklerinin iyi sonuçlar verdiğini göstermiştir. Her bir özellik için 3.9 eşitliğinde olduğu gibi hesaplanır ve en yüksek puanlara sahip olan özellikler, sınıfların en ayırt edici özellikleridir. (Li ve diğ., 2009) (Ng ve diğ., 2006).

$$WLLR(w_t, c_j) = P(w_t|c_j) * \log\left(\frac{P(w_t|c_j)}{P(w_t|\neg c_j)}\right) \quad (3.9)$$

w_t : t.nci terim

c_j : j.nci sınıf

$P(w_t | c_j)$: t.nci terimin j.nci sınıfta olma olasılığı (c_j sınıfındaki frekansı/ c_j -de geçen terimlerin toplam frekansı)

$P(w_t | \neg c_j)$: t.nci terimin j.nci sınıfın dışındaki sınıflarda olma olasılığı (diğer sınıflardaki toplam frekansı/ diğer sınıflarda geçen tüm terimlerin toplam frekansı)

w_t 'nin, c_j sınıfında sık görüldüyse ve diğer sınıflarda nadir görüldüyse WLLR'si yüksek olacaktır.

3.7. ÖZELLİK VEKTÖRÜ OLUŞTURMA

Yukarıda bahsedildiği gibi veriler özellik vektörü olarak temsil edilir. Bunun anlamı dokümanların her biri bir sayısal satır olarak temsil edilir. Yukarıda bahsedilen özellikler için hesaplanan değerler üç temel yöntemle oluşturulmuştur. Birincisi Tekli yöntem bir BOW yöntemidir. Tüm kelime ve kategorileri özellik olarak alır. Liste yönteminde her grup için sık kullanılan kelime / kategori listeleri oluşturulur. Deneklerin bu listelerden kaçar tane kullandığı sayılır. Üçüncü yöntem Karma yöntemdir. Frekans listelerindeki kelime / kategorileri özellik olarak alan bir BOW yöntemidir.

3.7.1. Tekli Yöntem

3.7.1.1. Kelime/Kategori Frekansı

Özellik olarak tek tek kelimelerin/kategorilerin kullanıldığı yöntemdir. Bu yöntemde her denek için tek tek kelime/kategori özellik olarak kullanılır. Her bir kelime/kategori için kullanım sayıları (frekansı) çıkarılıp, metin uzunluğuna bölünür. Sonuç küçük bir reel sayı olabileceği için bir çarpan ile çarpılıp tam sayıya çevrilmiştir. Örnek olarak 125 numaralı denekğin kelime frekans değerleri Tablo 3.16'de gösterilmiştir. W kısaltması kelimeleri göstermektedir.

Tablo 3.16: Bir denek için frekans tablosu

No	Cinsiyet	w_1	w_2	w_3	w_4	...	w_n	Sınıf
125	M	0	20	45	0	...	22	N

Her bir satır bir deneye ait olacak şekilde tablolar hazırlanır. Sadece sistem eğitiminde kullanılan denekler yoluyla eğitim veri tablosu oluşturulur. Test denekleri için yapıldığında test veri tablosu oluşur. Tablo 3.17'de kelime frekans metodu (KelimeTekliAğırlıklı) ve kategori frekans metodu (KategoriTekliAğırlıklı) yöntemlerinin eğitim tablolarından birer bölüm gösterilmiştir.

Tablo 3.17: Sınıflandırma için eğitim veri tablosundan bir bölüm.

	Kelime (KelimeTekliAğırlıklı)									Kategori(KategoriTekliAğırlıklı)								
	No	Cinsiyet	W ₁	W ₂	W ₃	W ₄	...	W _n	Sınıf	No	Cinsiyet	C ₁	C ₂	C ₃	C ₄	...	C _n	Sınıf
Frekans (kelime veya kategori frekansı)	125	M	0	20	45	0	...	22	N	125	M	0	45	0	0	...	73	N
	41	F	15	0	0	105	...	0	P	41	F	34	0	12	0	...	67	P
	134	M	0	100	35	0	...	43	P	134	M	90	0	0	23	...	0	P

3.7.1.2. Kelime/Kategori varlığı

Yukarıdaki yöntemle benzer şekilde, frekans değerleri yerine varlık/yokluk bilgileri kullanılırsa Kelime/Kategori varlığı yöntemi elde edilir. Kullanılan kelimeler için kullanım sayısı değil de 1 değeri alınır. Kullanılmayan kelimeler için 0 değeri alınır. Böylece 1 ve 0 değerlerinden oluşan eğitim ve test tabloları elde edilir. Tablo 3.18 de kategori ve kelime yöntemleri için kullanılan tablolardan bölümler görülmektedir. Kategorilerin herbiri için varlık/yokluk (*KategoriTekliVarYok*), kelimelerin varlığı/yokluğu metodu için (*KelimeTekliVarYok*) kısaltmaları kullanılmıştır.

Tablo 3.18: Sınıflandırma için verilerin kelime ve kategori varlık/yokluk tabloları.

	Kelime									Kategori								
	No	Gender	W ₁	W ₂	W ₃	W ₄	...	W _n	Class	No	Gender	C ₁	C ₂	C ₃	C ₄	...	C _n	Class
Varlık/Yokluk (kelime veya kategori varlığı)	125	M	0	1	1	0	...	1	N	125	M	0	1	0	0	...	1	N
	41	F	1	0	0	1	...	0	P	41	F	1	0	0	1	...	1	P
	134	M	0	1	1	0	...	1	P	134	M	1	0	0	1	...	0	P

3.7.2. Liste Yöntemi

Liste yöntemi, en genel ifadeyle, tüm kelimeleri veya tüm kategorileri kullanmak yerine hastalığa özel olarak sıkça kullanılan kelimelerin veya kategorilerin belirlenmesi ve daha sonra deneklerin kullanım frekanslarına göre sınıflandırılmasıdır. Bu da her bir denneğin hangi gruptan toplamda kaçar tane kullandığına dayanır.

Hasta ve sağlıklı gruplarının sık kullanım listeleri oluşturulur. Listelerde kelimeler ve kullanım sayıları bulunur. Her iki grupta da aynı kelime varsa ve yakın frekanslarda ise bu kelime listelerden silinir. Çünkü ayırıcı özelliği yoktur.

Silme işlemi için iki yöntem uygulanır. Birincisinde, eşik değer kadar birbirine yakın olan kelimeler her iki listeden silinir (Herikisini-Sil (HS)), fakat diğer yöntemde yüksek olan silinmez ve değeri, aralarındaki fark kadar bırakılır (Küçüğü-Sil (KS)). Ek 4’de metotlar ve kısaltmaları özet halinde gösterilmiştir.

Liste ve karma yöntemlerinin hepsinde yukarıdaki eşik değer yöntemleri (HS-KS) uygulanmıştır. Buna göre aynı kelimenin/kategorinin birden fazla listede olması durumunda silinecek kelimeler/kategoriler, eşik değerinin içinde olmasına göre karar verilir. Örnek olarak eşik değer %25 ise ve sağlıklı listesinde “okul” kelimesi frekansı 1000 ve hasta listesinde 900 ise kelimenin ayırıcı özelliğini belirlemek için $(1000-900)/(1000+900)$ hesaplaması sonucu olarak 0,05 bulunur. Eşik değer olan 0,25 den küçük olduğu için ayırıcı özelliği anlamlı değildir. Bu durumda HS yöntemine göre “okul” kelimesi listelerden çıkarılır veya KS’ye göre 100 değerinde olarak sağlıklı listesinde bırakılır. Tüm denemelerde seçilen eşik değerler yüzde olarak 50, 25, 5, 0 olarak belirlenmiştir.

Daha sonra her bir listedeki tüm kelimelerin frekansları toplanıp kendi grubundaki denek sayısına bölünür. Çok küçük reel sayılardan kurtulmak için yeniden sabit bir sayı ile çarpılır. Her iki grupta da aynı işlemler uygulanır. Listeler bu şekilde hazırlanır. Bundan sonra bu listelerden frekans listeleri adıyla bahsedilecektir.

Frekans listelerine dayanan bir kaç özellik seçim metodu kullanılmıştır ve isimleri kelime listesi (*KelimeListe*), ağırlıklı kelime listesi (*KelimeListeAğırlıklı*), kategori listesi (*KategoriListe*), ağırlıklı kategori listesi (*KategoriListeAğırlıklı*) olarak adlandırılmıştır. Eğitim tabloları için her bir deneğin listelerden skorları bir kaç yöntemle göre bulunur;

Kelime listesi (*KelimeListe*) ve kategori listesi (*KategoriListe*) için bu skorlar hesaplanırken sadece frekans listesinden kaçar kelime kullanıldığı (tf) alınır (kısa veya uzun metinler olabileceği için normalizasyon) ve metin uzunluğuna (textSize) bölünür. Burada metin uzunluğu kullanılan kelime sayısıdır. Örneğin denek metninde sağlıklı frekans listesinden toplam 25 kelime varsa ve metin uzunluğu 100 kelime ise sağlıklı skoru 25/100’dür.

Kelime frekans listesi (*KelimeListeAğırlıklı*) metotları arasındaki fark şudur: birincisinde sadece kelimeler sayılır sonra toplanır ve deneğe ait pozitif ve negatif skorlar ortaya çıkar. İkincisinde ise her bir pozitif veya negatif kelimenin sayısı frekans listesindeki kelimelerin frekans değerleriyle(listFreq) çarpılır, bu değerler toplanır.

WordList, CateList yöntemlerinde denek skoru, pozitif ve negatif için ayrı ayrı olmak üzere 3.10'daki gibi hesaplanır;

$$\sum_{i=1}^n coeff * \frac{tf_i}{textSize} \quad (3.10)$$

n: Frekans listesindeki kelime sayısı

w_i: i'nci terim

tf_i: i'nci terim frekansı

textSize: metin uzunluğu

coeff: katsayı

WordListFreq, CateListFreq yöntemlerinde denek skoru, pozitif ve negatif için ayrı ayrı olmak üzere 3.11'deki gibi hesaplanır;

$$\sum_{i=1}^n coeff * \frac{tf_i * listFreq_i}{textSize} \quad (3.11)$$

listFreq_i : w_i teriminin listedeki frekansı

Yukarıdaki her yöntemde iki sözlük (Harvard-III, Harvard-IV) ayrı ayrı kullanılarak kategoriler bulunmuştur.

Tablo 3.19: Sık kullanılan kelime listelerinden örnek bir bölüm.

SAĞ	Frekans	DEP	Frekans
memnun	40	iste-	125
iste	35	hiç	96
sev	35	yemek	33
el	30	artık	32
uygula	28	ben	31
okul	23	dayan-	28
izle	22	bunal	27

Yöntemi bir örnek ile somutlaştırmak gerekirse, örneğin sağlıklı ve depresyon sınıflandırılması yapılacaksa sağlıklı (SAĞ) ve depresyon (DEP) frekans listeleri Tablo 3.19'daki gibi oluşturulur.

Şekil 3.11 bir denek metnini göstermektedir. SAĞ, DEP, kelimeleri sırasıyla SAĞ, DEP, şeklinde alt çizgilerle işaretlenmiştir. Metnin özellik tablosu yukarıda bahsedilen

hesaplamalardan sonra Tablo 3.20'deki gibi çıkarılmıştır. 91 numaralı bu denek 9 DEP kelimesi, 4 SAĞ kelimesi kullanılmıştır. Dolayısıyla *KelimeListe* yönteminde SAĞ skoru $(5/32)*1000=156$, DEP skoru ise $(9/32)*1000=281$ olarak hesaplanır. *KelimeListeAğırlıklı* yöntemine göre ise her kelimenin frekansı listedeki ağırlıklarıyla çarpılır. 2 “memnun” kelimesi, 1 “iste” kelimesi, 1 “uygula” kelimesi, 1 “izle” kelimesi kullanmıştır. SAĞ skoru $((2*40+1*35+1*28+1*22)/32*1000=5156$ olarak bulunur.

91 / K / 29 / yüksekokul / Depresyon / Hayatımdan memnun hiç memnun değilim. Değiştirmek istiyorum, eşim beni çok üzdü, artık dayanamıyorum bu hayata. Çekmek istemiyorum artık. Türkiye’de yaşamaya karar verdim. Uygulayacağım. / Evde oturmak. TV izlemek. Yemek yemek. Tartışmak. Yalnızlıktan bunalmak.

Şekil 3.11: Örnek bir denek verisi. Sağlıklı ve Depresyon alt çizgilerine göre.

Tablo 3.20: Örnek deneğin KelimeListe yöntemi için metin özelliklerinin eğitim verisine aktarılması.

No	Cins	Yaş	MU	SAĞ	DEP	Sınıf (N:Neg P:Poz)
91	K	29	32	138	276	N

Her denek için bu satırlar oluşturulur ve Tablo 3.21 elde edilir. Train ve test denekleri için ayrı tablolar hazırlanıp MÖ yöntemleri uygulanır.

Tablo 3.21: Sistem eğitim veri tablosundan örnek bir bölüm.

No	Cins	Yaş	MU	SAĞ kelimeleri	DEP kelimeleri	Sınıf (N:Neg P:Poz)
125	E	26	85	840	300	P
91	K	29	29	138	276	N
134	E	38	112	160	490	N

3.7.3. Karma yöntem

Karma yöntem Liste ve Tekli yöntemlerin birleşmesiyle oluşur. Liste yönteminde hazırlanmış olan frekans listelerinin her bir elemanı özellik olarak kullanılır. Amaç tüm kelimeler veya kategoriler yerine, seçilmiş kelime veya kategorileri kullanmaktır. Sınıflandırma yapılacak iki grubun (hasta ve sağlıklı) kelime frekans listeleri birleştirilir. Bu birleştirmede aynı olan

elemanlar olabileceği için bunlardan birinin alınması gerekir. Aynı olan elemanın değerleri de aynı ise iki grupta da aynı sayıda kullanıldığından ayırt edici etkisi olmayacağı düşünülerek birleştirilmiş listeye alınmaz. Değerleri farklı ise aradaki fark yeni değer olarak alınır. Böylece ayırt edici özelliği yüksek olan elemanların frekans değerlerinin de yüksek olması sağlanır. Yeni listelerle eğitim ve test verisi hazırlamak için varlık-yokluk (*KelimeKarma*, *KategoriKarmaTekli*) ve ağırlıklı frekans (*KelimeKarmaAğırlıklı*, *KategoriKarmaAğırlıklı*) yöntemleri uygulanır. Yine depresyon ve sağlıklı gruplar örnek olarak alınırsa, Tablo 3.22’te her iki gruptan çıkan örnek frekans listeleri ve Tablo 3.23’te birleştirilmiş frekans listesi örneği bulunmaktadır.

Tablo 3.22: İlk aşamada çıkarılan frekans listeleri örneği.

SAĞ	Frekans	DEP	Frekans
al	40	al	125
iste	35	hiç	96
artık	35	yemek	33
el	30	artık	30
uygula	28	ben	31
okul	23	uygula	28
hiç	22	bunal	27

Tablo 3.23: Birleştirilmiş frekans listesi örneği.

PosNegList	Frekans
al	85
iste	35
artık	5
el	30
yemek	33
okul	23
hiç	74
ben	31
bunal	27

Eğitim ve test tabloları Tablo 3.17 ve Tablo 3.18’deki gibi hazırlanır. Özellik seçimi için yöntemlerin özeti aşağıdadır.

❖ Kelime / Kategori için

➤ Tekli yöntem

- Var-Yok (*KelimeTekliVarYok*)
- WRLl Var-Yok (*WRLlKelimeTekliVarYok*)
- Ağırlıklı Frekans (tf) (*KelimeTekliAğırlıklı*)
- Ters doküman frekansı (tf-idf) (*IDFKelimeTekliAğırlıklı*)

- WRLl ağırlıklı frekans (WRLlKelimeTekliAğırlıklı)
- Liste yöntemi
 - Frekans (KelimeListe)
 - Ağırlıklı Frekans (tf) (KelimeListeAğırlıklı)
 - Ters doküman frekansı (tf-idf) (IdfKelimeListe)
 - WRLl ağırlıklı frekans (WRLlKelimeListe)
- Karma yöntem
 - Var-Yok (KelimeKarmaVarYok)
 - Frekans (tf) (KelimeKarmaAğırlıklı)
 - Ters doküman frekansı (tf-idf) (IdfKelimeKarmaAğırlıklı)
 - WRLl ağırlıklı frekans (WRLlKelimeKarmaAğırlıklı)
 - WRLl Var-Yok (WRLlKelimeKarmaVarYok)

Tüm yöntemlerin isimleri ve kısaltmaları Ekler bölümünde Ek 4’de verilmiştir.

4. BULGULAR

Bu çalışmada psikiyatrik rahatsızlığı olan yetişkin ve çocukların konuşma metinleri üzerinde çalışılmıştır. Hasta sağlıklı ikili sınıflandırması yapıldığı gibi iki hasta ve bir sağlıklı olarak üçlü sınıflandırmalar da yapılmıştır. İki tip veri bulunmaktadır. Bunlar yetişkin serbest konuşma ve yönlendirilmiş çocuk verisidir. Çocuk verisi uygulanan psikolojik ölçek metinlerinden oluşmaktadır. Veriler sınıflandırılırken özellik olarak kelimeler veya kategoriler seçilmiştir. Kategori özellik seçildiğinde kullanılan kategoriler bir psikolojik sözlükten alınmıştır. Bölüm 3.2’de bahsedilen HV-III ve HV-IV olarak iki sözlük vardır. Her iki sözlükle aynı veriler üzerinde ayrı ayrı sınıflandırmalar yapılmıştır. Yapılan sınıflandırmaların sonuçları bu bölümde ele alınmıştır. Bulgular her grup için ayrı başlık altında incelenecektir. Veriler çocuk ve yetişkin olarak iki grupta toplanmıştır. Bu iki grup içinde ikişer hasta ve birer sağlıklı alt grup vardır. Bu şekilde 6 alt grup oluşturulmuştur. Her alt gruptan 9’ar denek metni test için ayrılmıştır. Test metinleri 3 kez değiştirilip sınıflandırmalar yapılarak (üçlü çapraz geçerlilik uygulanmış) sonuçların ortalaması alınmıştır.

Bu bölümde sonuçlar verilirken önce ikili sınıflandırma sonuçları, HV-III veya HV-IV kullanılmasına göre sırasıyla Çocuk - (HV-III ve HV-IV), Yetişkin - (HV-III ve HV-IV) olarak verilmiştir. Daha sonra üçlü sınıflandırma sonuçları aynı sırada verilmiştir.

Tekli, liste ve karma özellik çıkarımı olmak üzere üç tip özellik çıkarım yöntemi belirlenmiştir. Bunlardan tekli ve karma yöntemlerine ikili (var-yok) türleri ve liste yönteminin ağırlıklandırılmış türleri de eklenmiştir. Bu şekilde toplamda 14 özellik çıkarım metodu kategori ve kelime için ayrı ayrı uygulanmıştır.

Tekli yöntemde her bir kelime veya kategori özellik olarak alınmıştır. Veriler bölüm 3.4’de anlatıldığı gibidir. Önceki metin sınıflandırma çalışmalarına göre tf-idf ile ağırlıklandırılmış sınıflandırmaların tf’e göre daha iyi sonuç vermesi beklenmekteydi fakat burada öyle olmamıştır. WLLR ise belirgin bir fark oluşturmamıştır.

Bazı tablolarda boş olan satırlar ilgili yöntemde anlamlı olmadığı için yapılmamıştır. Örneğin *KategoriListeAğırlıklı* ve *KelimeListeAğırlıklı* yöntemlerinde frekans listesindeki ağırlık değeriyle doküman frekansı çarpılarak vektör değeri elde edilmiştir. Oysa listedeki ağırlık değeri tf-idf ve WLLR için kullanılmaz. Çünkü *KategoriListe* ve *KelimeListe* yönteminde tf-

idf için tf değeri, idf ile çarpılır, listedeki değeri ile çarpılmaz. WLLR yönteminde ise, WLLR skoru sıfırdan farklı olarak hesaplanan terimler özellik olarak seçilmiştir.

4.1. ÇOCUK VERİSİ İKİLİ SINIFLANDIRMA BULGULARI

Bu bölümde çocuk verisi sınıflandırmaları HV-III sözlüğü ve HV-IV sözlüğü ile olmak üzere iki şekilde yapıp sonuçları verilmiştir. Veri dağılımı tablo'daki gibidir.

4.1.1. HV-III ile Çocuk Verisi İçin Bulgular

Çocuk verileri Depresyon x Sağlıklı ve ADHD x Sağlıklı olarak ayrı ayrı sınıflandırılmıştır. Her iki grubun da test verileri üç kere değiştirilerek sonuçlarının ortalamaları alınmıştır. Sınıflandırma işlemi veri madenciliğinde yaygın olarak kullanılan Weka programında MÖ metotları kullanılarak yapılmıştır (Witten ve Frank, 2005). MÖ metotlarından Navie Bayes (NB), karar ağacı algoritması olan J48 ve Sequential Minimal Optimization (SMO) kullanılmıştır.

Tekli (BOW), Liste ve Karma yöntemlerinin NB, SMO, J48 makine öğrenmesi teknikleri uygulanarak elde edilen doğruluk sonuçları tablolar halinde verilmiştir. Her bir sınıflandırmada özellik vektörünün oluşturma şekline göre tf, tf-Idf ve WLLR olarak sonuçlar düzenlenmiştir. Tüm sonuçlarda üçlü çapraz geçerlilik ortalamalar kullanılmış ve doğruluk yüzdeleri yuvarlatılarak verilmiştir.

4.1.1.1. HV-III ile Depresyon x Sağlıklı Sınıflandırması Sonuçları

- Tekli Yöntem Sonuçları

Bu veri grubunda tf için KategoriTekliVarYok ve KategoriTekliAğırlıklı'te NB %78 ile ve J48 metodu KategoriTekliVarYok ve KelimeTekliVarYok özellik seçiminde sırasıyla %73 ve %75 ile diğer özellik seçimlerine göre daha iyi sonuç vermiştir. WLLR kullanıldığında ise TekliVarYok metotları düşük sonuçlar vermiştir. TekliAğırlıklı metotlar ise SMO (% 73) ve NB'de (%76) iyi sonuç vermiştir. Tablo 3.1'de yüksek değerler koyu yazılmıştır. Sonuçlar yüzde olarak verilmiştir.

Tablo 4.1: Çocuk DEP x SAĞ sınıflandırmasının NB, SMO, J48 teknikleri uygulanarak elde edilen Tekli yöntem doğruluk sonuçları.

DEP	tf				tf-idf				WLLR			
Feature	NB	SMO	J48	Ort	NB	SMO	J48	Ort	NB	SMO	J48	Ort
KelimeTekliVarYok	69	58	73	67	69	58	73	67	78	67	71	72
KelimeTekliAğırlıklı	66	69	62	66	63	45	62	57	73	62	62	66
KategoriTekliVarYok	78	58	75	70	78	58	75	70	82	58	75	71
KategoriTekliAğırlıklı	78	71	62	70	47	51	55	51	74	74	60	69

- Liste ve Karma Yöntem Sonuçları

Tüm sonuçlar Ek 5’de verilmesiyle birlikte eşik değerlerin ortalaması alınarak özet tablo çıkarılmak istendiğinde, tüm ağırlıklı, ağırlıksız olarak liste ve karma yöntemlerinin de ortalaması alınınca Tablo 4.2 elde edilir. Buna göre liste yönteminin bu veri grubunda daha başarılı olduğu gözlenir tf’de KelimeListe SMO’da %72, tf-Idf’de KategoriListe %77 ile J48’de ve WLLR’de %77 ile SMO’da en iyi değerler alınmıştır.

Tablo 4.2: Çocuk DEP x SAĞ Sınıflandırmasında Karma, Liste ve Tekli yöntemlerinin kategori ve kelime olarak doğruluk sonuçlarının ortalamaları.

DEP	tf				tf-idf				WLLR			
Liste/Karma	NB	SMO	J48	Ort	NB	SMO	J48	Ort	NB	SMO	J48	Ort
Kategori Tüm Liste	72	72	70	71	70	70	77	72	73	74	69	72
Kelime Tüm Liste	73	76	62	70	69	68	52	63	75	77	64	72
Kategori Tüm Karma	67	62	67	65	73	70	67	70	75	67	67	69
Kelime Tüm Karma	66	59	67	64	66	59	70	65	72	64	67	68
Karma, Liste, Tekli Ortalama	71	66	67	68	67	60	66	64	75	68	67	70

Eşik Değer ortalamalarına bakıldığında Tablo 4.3’de görüldüğü gibi Eşik Değer 5 her 3 özellik seçim metodu için %68, %68, %71 ile en iyi sonucu vermiştir.

Tablo 4.3: DEP x SAĞ Sınıflandırmasında Karma, Liste yöntemlerinin doğruluk sonuçlarının Eşik değerlere göre genel ortalamaları.

DEP												
Eşik Değer	tf				tf-idf				WLLR			
	NB	SMO	J48	Ort	NB	SMO	J48	Ort	NB	SMO	J48	Ort
Eşik Değer 0	68	66	66	67	69	66	68	68	75	70	67	71
Eşik Değer 5	70	67	67	68	70	67	67	68	76	69	67	71
Eşik Değer 25	72	68	65	68	69	67	65	67	74	69	67	70
Eşik Değer 50	68	68	66	67	70	67	64	67	70	67	66	68
Ortalama	69	67	66	68	70	67	66	68	74	69	67	70

KS ve HS sonuçlarını değerlendirmek için Tablo 4.4 özet tablosu çıkarılmıştır. Buna göre bu veri grubu için HS ve KS arasında ortalamalara bakıldığında fark görünmemektedir (Tablo 4.5). Her ikisi için de (%68, %68, %69-%70) ortalama %69 olmuştur.

Tablo 4.4: Çocuk DEP x SAĞ Sınıflandırmasında Karma ve Liste yöntemlerinin kategori ve kelime olarak doğruluk sonuçlarının HS ve KS yöntemlerine göre karşılaştırması.

DEP	tf				tf-idf				WLLR			
HS	NB	SMO	J48	Ort	NB	SMO	J48	Ort	NB	SMO	J48	Ort
KategoriListe_HS	69	71	71	70	70	71	76	72	73	75	69	72
KategoriListeAğırlıklı_HS	71	71	69	70								
KategoriKarmaFrek_HS	65	68	58	64	74	70	66	70	72	70	61	68
KategoriKarmaVarYok_HS	76	60	71	69	76	60	71	69	77	63	71	70
KelimeListe_HS	72	75	55	67	69	68	51	63	76	77	65	72
KelimeListeAğırlıklı_HS	75	78	69	74								
KelimeKarmaFrek_HS	64	59	59	61	66	60	70	65	67	60	62	63
KelimeKarmaVarYok_HS	67	59	74	67	67	59	74	67	72	64	72	69
Ortalama	70	68	66	68	70	65	68	68	73	68	67	69
KS	NB	SMO	J48	Ort	NB	SMO	J48	Ort	NB	SMO	J48	Ort
KategoriListe_KS	75	74	70	73	70	70	78	72	74	74	69	72
KategoriListeAğırlıklı_KS	72	72	71	72								
KategoriKarmaFrek_KS	49	69	64	61	71	66	73	70	71	73	60	68
KategoriKarmaVarYok_KS	78	52	75	68	78	52	75	68	82	60	75	72
KelimeListe_KS	69	75	54	66	69	67	52	63	75	77	63	72
KelimeListeAğırlıklı_KS	75	78	67	73								
KelimeKarmaFrek_KS	64	60	62	62	67	58	69	65	71	64	62	66
KelimeKarmaVarYok_KS	69	58	73	67	69	58	73	67	78	69	71	73
Ortalama	69	67	67	68	71	62	70	68	75	69	67	70

4.1.1.2. HV-III ile ADHD x Sağlıklı Sınıflandırması Sonuçları

- Tekli Yöntem Sonuçları

Bu veri grubunda Tekli yöntemlerde tf için en iyi sonuc %72 olarak J48 de alınmıştır. tf-Idf için %78 NB sonucu ve WLLR için %68 NB sonucu ile KategoriTekliVarYok en iyi değerleri almıştır.

Tablo 4.5: ADHD x SAĞ sınıflandırmasında Tekli yöntem sonuçları.

ADHD	tf				tf-idf				WLLR			
Feature	NB	SMO	J48	Ort	NB	SMO	J48	Ort	NB	SMO	J48	Ort
KelimeTekliVarYok	59	59	50	56	59	59	50	56	66	62	64	64
KelimeTekliAğırlıklı	52	54	60	55	44	50	43	46	63	57	48	56
KategoriTekliVarYok	72	57	50	60	72	57	50	60	68	59	46	58
KategoriTekliAğırlıklı	56	52	54	54	52	46	39	46	57	53	51	54

- Liste ve Karma Yöntem Sonuçları

Bu veri grubu için yapılan sınıflandırmada yine bölüm 4.3.1’de anlatıldığı gibi listeler çıkarılmış, eşik değer uygulanmıştır. Alınan sonuçlar Ek 6’da verilmiştir. Tablo 4.6, Tablo 4.7 ve Tablo 4.8 çıkarılmış özet tablolarıdır.

Tüm ağırlıklı, ağırlıksız olarak liste ve karma yöntemlerinin de ortalaması alınınca Tablo 4.6 elde edilir. Buna göre sonuçlar tf’de KelimeListe SMO %63, tf-idf’de KategoriKarma NB %66 ve WLLR’de KategoriListe J48 %66 ile diğerlerine göre daha iyi sonuçlar vermiştir. KelimeKarma sonuçları bu veri grubunda düşük çıkmıştır.

Tablo 4.6: ADHD x SAĞ Sınıflandırmasında Karma, Liste ve Tekli yöntemlerinin kategori ve kelime olarak doğruluk sonuçlarının ortalamaları.

ADHD	tf				tf-idf				WLLR			
Liste/Karma	NB	SMO	J48	Ort	NB	SMO	J48	Ort	NB	SMO	J48	Ort
Kategori Tüm Liste	54	55	51	54	50	57	54	54	60	60	66	62
Kelime Tüm Liste	61	63	56	60	50	55	58	54	58	59	55	57
Kategori Tüm Karma	62	57	52	57	66	64	47	59	57	55	50	54
Kelime Tüm Karma	57	56	60	58	55	56	53	55	60	53	57	56
Karma, Liste, Tekli Ortalama	58	51	54	54	58	56	55	56	61	57	55	58

Eşik Değer ortalamalarına bakıldığında Tablo 4.7’de görüldüğü gibi sonuçlar genel olarak iyi olmamakla birlikte Eşik Değer 25 her 3 özellik seçim metodu için (%58, %55, %58) % 57 ortalama ile en iyi sonucu vermiştir.

Tablo 4.7: Çocuk ADHD x SAĞ Sınıflandırmasında Karma ve Liste yöntemlerinin doğruluk sonuçları.

ADHD												
	tf				tf-idf				WLLR			
Eşik Değer	NB	SMO	J48	Ort	NB	SMO	J48	Ort	NB	SMO	J48	Ort
Eşik Değer 0	59	57	54	57	55	58	51	55	58	54	56	56
Eşik Değer 5	58	59	55	57	54	56	51	54	60	56	55	57
Eşik Değer 25	59	57	56	58	57	57	52	55	59	58	57	58
Eşik Değer 50	57	58	54	56	55	57	52	54	58	54	55	56
Ortalama	59	58	55	57	55	57	51	55	59	56	56	57

KS ve HS sonuçlarını değerlendirmek için Tablo 4.8 özet tablosu çıkarılmıştır. Buna göre bu veri grubu için HS (%56, %56, %57) ve KS (%56, %55, %57) arasında belirgin bir fark görünmemektedir. Her ikisi de ortalama %74 çıkmıştır.

Tablo 4.8: ADHD x SAĞ Sınıflandırmasında Karma ve Liste yöntemlerinin kategori ve kelime olarak doğruluk sonuçlarının HS ve KS yöntemlerine göre karşılaştırması.

ADHD	tf				tf-idf				WLLR			
HS	NB	SMO	J48	Ort	NB	SMO	J48	Ort	NB	SMO	J48	Ort
KategoriListe_HS	51	52	50	51	52	56	54	54	58	57	61	59
KategoriListeAğırlıklı_HS	51	54	51	52								
KategoriKarmaFrek_HS	60	60	54	58	66	64	50	60	57	57	52	55
KategoriKarmaVarYok_HS	60	54	52	55	60	54	52	55	61	55	49	55
KelimeListe_HS	58	60	50	56	48	52	58	52	59	58	56	58
KelimeListeAğırlıklı_HS	62	64	58	61								
KelimeKarmaFrek_HS	55	53	55	54	58	56	52	55	55	54	51	53
KelimeKarmaVarYok_HS	60	58	64	60	60	58	64	60	66	56	62	61
Ortalama	57	57	54	56	57	57	55	56	59	56	55	57
KS	tf				tf-idf				WLLR			
KS	NB	SMO	J48	Ort	NB	SMO	J48	Ort	NB	SMO	J48	Ort
KategoriListe_KS	59	57	51	56	48	57	55	53	62	62	71	65
KategoriListeAğırlıklı_KS	57	58	55	56								
KategoriKarmaFrek_KS	58	52	54	46	67	63	44	58	46	48	55	50
KategoriKarmaVarYok_KS	69	61	50	48	69	61	50	60	64	60	46	57
KelimeListe_KS	64	65	55	61	48	52	53	51	57	60	53	57
KelimeListeAğırlıklı_KS	62	64	59	62								
KelimeKarmaFrek_KS	56	58	61	58	56	55	46	52	55	48	48	50
KelimeKarmaVarYok_KS	58	54	61	57	58	54	61	57	65	53	66	61
Ortalama	60	59	56	56	58	57	52	55	58	55	57	57

4.1.1.3. HV-III ile Çocuk Verisinde Cinsiyet Sınıflandırması İçin Bulgular

Bu bölümde DEP, ADHD ve SAĞ verilerinin tümü kullanılarak yapılan Kadın x Erkek (K x E) sınıflandırılmasının sonuçları verilmiştir.

- Tekli Yöntem Sonuçları

Bu veri grubunda tf için KelimeTekliVarYok için SMO'da %87 ile ve KelimeTekliAğırlıklı için SMO'da %80 ile en iyi sonuçları alınmıştır. tf-idf olarak KelimeTekliAğırlıklı için J48'de %76 ile, WLLR için KelimeTekliAğırlıklı yönteminde %68 ile J48'de en iyi sonuç vermiştir (Tablo 4.9).

Tablo 4.9: Çocuk K x E sınıflandırmasında Tekli yöntem sonuçları.

Cinsiyet	tf				tf-idf				WLLR			
Feature	NB	SMO	J48	Ort	NB	SMO	J48	Ort	NB	SMO	J48	Ort
KelimeTekliVarYok	76	87	58	74	59	59	50	56	50	64	64	59
KelimeTekliAğırlıklı	78	80	53	70	49	73	76	66	53	51	68	58
KategoriTekliVarYok	76	73	64	71	72	57	50	60	56	54	53	54
KategoriTekliAğırlıklı	69	76	58	68	36	73	53	54	60	54	61	58

- *Liste ve Karma Yöntem Sonuçları*

Cinsiyet grubu için yapılan sınıflandırmada yine bölüm 4.3.1’de anlatıldığı gibi listeler çıkarılmış, eşik değer uygulanmıştır. Buna göre alınan sonuçlar Ek 7’de verilmiştir. Tablo 4.10, Tablo 4.11 ve Tablo 4.12 çıkarılmış özet tablolardır.

Eşik değerlerin ve tüm ağırlıklı, ağırlıksız olarak liste ve karma yöntemlerinin de ortalaması alınınca Tablo 4.10 elde edilir. Buna göre tf için KelimeKarma NB ve SMO da %77 ile en iyi sonucu, tf-idf için KelimeKarma SMO’da %74 ile iyi sonuç vermiştir. WLLR sonuçları bu veri grubunda genel olarak düşük çıkmıştır.

Tablo 4.10: Çocuk K x E Sınıflandırmasında Karma, Liste ve Tekli yöntemlerinin kategori ve kelime olarak doğruluk sonuçlarının ortalamaları.

Cinsiyet	tf				tf-idf				WLLR			
Liste/Karma	NB	SMO	J48	Ort	NB	SMO	J48	Ort	NB	SMO	J48	Ort
Kategori Tüm Liste	73	73	73	73	70	73	73	72	53	50	53	52
Kelime Tüm Liste	73	72	68	71	72	73	73	73	52	50	53	52
Kategori Tüm Karma	68	74	65	69	59	69	62	63	56	52	55	54
Kelime Tüm Karma	77	77	54	69	70	74	73	72	52	53	64	57
Karma, Liste, Tekli Ortalaması	74	77	62	71	63	69	70	67	54	54	59	55

Eşik Değer ortalamalarına bakıldığında Tablo 4.11’de görüldüğü gibi Eşik Değer 25 ve Eşik Değer 50 her 3 özellik seçim metodu için tf, tf-Idf ve WLLR için sırasıyla %72, %72, %54 ile en iyi sonucu vermiştir.

Tablo 4.11: Çocuk K x E Sınıflandırmasında Karma ve Liste yöntemlerinin doğruluk sonuçlarının Eşik değerlere göre genel ortalamaları.

Cinsiyet												
Eşik Değer	tf				tf-idf				WLLR			
	NB	SMO	J48	Ort	NB	SMO	J48	Ort	NB	SMO	J48	Ort
Eşik Değer 0	71	74	60	68	67	72	70	68	52	52	57	54
Eşik Değer 5	73	71	66	70	68	72	72	70	54	52	58	55
Eşik Değer 25	73	75	67	72	67	73	70	72	54	51	57	54
Eşik Değer 50	74	76	67	72	69	73	71	72	54	51	57	54
Ortalama	73	74	65	71	73	74	65	71	54	52	57	54

HS ve KS sonuçlarını değerlendirmek için Tablo 4.12 özet tablosu çıkarılmıştır. Buna göre bu veri grubu için HS (%71, %71, %54) ortalama %65 ve KS (%71, %70, %54) ortalama %65 ile HS aynı sonucu vermiştir (Tablo 4.12).

Tablo 4.12: Çocuk K x E sınıflandırmasında Karma ve Liste yöntemlerinin kategori ve kelime olarak doğruluk sonuçlarının HS ve KS yöntemlerine göre karşılaştırması.

Cinsiyet	tf				tf-idf				WLLR			
HS	NB	SMO	J48	Ort	NB	SMO	J48	Ort	NB	SMO	J48	Ort
KategoriListe_HS	68	73	73	71	70	73	73	72	52	50	54	52
KategoriListeAğırlıklı_HS	73	73	73	73								
KategoriKarmaFrek_HS	66	73	67	68	61	72	71	68	58	51	56	55
KategoriKarmaVarYok_HS	73	74	70	72	73	74	70	72	55	54	52	54
KelimeListe_HS	71	70	60	67	72	73	69	71	53	50	54	53
KelimeListeAğırlıklı_HS	78	75	76	76								
KelimeKarmaFrek_HS	78	65	54	66	70	76	74	73	55	50	65	57
KelimeKarmaVarYok_HS	77	81	56	71	77	81	56	71	50	55	64	56
Ortalama	73	73	66	71	71	75	69	71	54	52	57	54
KS	NB	SMO	J48	Ort	NB	SMO	J48	Ort	NB	SMO	J48	Ort
KategoriListe_KS	76	73	73	74	71	73	74	72	54	50	52	52
KategoriListeAğırlıklı_KS	75	73	72	73								
KategoriKarmaFrek_KS	64	73	58	65	56	66	53	58	57	51	57	55
KategoriKarmaVarYok_KS	71	76	67	71	71	76	67	71	55	51	54	54
KelimeListe_KS	69	70	62	67	72	73	77	74	50	50	51	50
KelimeListeAğırlıklı_KS	75	73	76	75								
KelimeKarmaFrek_KS	78	73	53	68	69	73	73	72	53	51	65	56
KelimeKarmaVarYok_KS	76	87	53	72	76	87	53	72	50	57	63	56
Ortalama	73	75	64	71	69	75	66	70	53	52	57	54

Bu bölümde okul çağındaki çocukların zihinsel bozukluklarının saptanması için bir metin analiz yapılmıştır. Depresyon, ADHD ve Kontrol grubunda teşhisi psikiyatri uzmanları tarafından yapılan deneklerin metinleri üzerinde çalışılmıştır. Deneklerin metinlerinden elde edilen farklı kelimeler / kategoriler ve bunların varlık veya sıklık bilgisi, idf veya WLLR ile belirlenerek NB, SMO ve J48 metotları kullanılarak sınıflandırılmıştır. DEP (hasta x sağlıklı), ADHD (hasta x sağlıklı) ve cinsiyet (K x E) olarak ikili sınıflandırmaların sonuçları verilmiştir.

HV-III sözlüğü kullanılarak yapılan çocuk verisinin sonuçları değerlendirildiğinde sonuçların DEP grubunda ADHD'den daha iyi olduğu gözlemlenmiştir. DEP x SAĞ sınıflandırmasında genel ortalamalar tf için %68, tf-idf için 64 ve WLLR için %70 dir (Tablo 4.2). Eşik Değer 5 bu veri grubunda %69 ortalama ile en iyi sonucu vermiştir (Tablo 4.3).

ADHD ise tüm yöntemlerde DEP grubunun gerisinde kalmıştır. Genel ortalama sonuçlara bakılacak olursa tf için %54, tf-idf için %56 ve WLLR için %58 olduğu görülmüştür (Tablo 4.6). Buradan ADHD rahatsızlıklarının metin düzeyinde etkisinin belirgin olmadığı sonucuna varılabilir. Ancak elimizdeki ADHD rahatsızlığı olan verilerde genellikle başka rahatsızlıkların

da teşhisi yapılmış olduğu görülmüştür. Bu durum ADHD rahatsızlığının belirtilerini gölgelemiş olabilir. Bu grupta Tresh 25 %57 ile en iyi sonucu vermiştir.

Cinsiyet verisi beklendiği gibi daha iyi sonuçlar vermiştir. Çünkü hasta teşhisi uzman doktorlar tarafından yapılmış olsa da, hata payı ve diğer hastalıkların etkilemesine bağlı olduğu için sınıflandırma sonuçlarının düşük olması beklenir. Cinsiyet sınıflandırmasında genel ortalamalar tf için %71, tf-idf için %67 ve WLLR için %55 doğruluk değerleri gözlenmektedir (Tablo 4.10). En iyi Eşik Değer % 25 ve %50'dir, %66 ortalama sonucu vermiştir (Tablo 4.11).

4.1.2. HV-IV ile Çocuk Verisi İçin Bulgular

Çocuk verileri HV-IV kullanılarak Depresyon x Sağlıklı ve ADHD x Sağlıklı olarak ayrı ayrı sınıflandırılmıştır. Her iki grubun da test verileri üç kere değiştirilerek sonuçlarının ortalamaları alınmıştır. Sınıflandırma işlemi veri madenciliğinde yaygın olarak kullanılan Weka programında MÖ metotları kullanılarak yapılmıştır (Witten ve Frank, 2005). MÖ metotlarından Navie Bayes (NB), karar ağacı algoritması olan J48 ve Sequential Minimal Optimization (SMO) kullanılmıştır.

4.1.2.1. HV-IV ile Depresyon x Sağlıklı Sınıflandırması Sonuçları

Tekli (BOW), Liste ve Karma yöntemlerinin NB, SMO, J48 makine öğrenmesi teknikleri uygulanarak elde edilen doğruluk sonuçları tablolar (Tablo 4.13, Ek 8) halinde verilmiştir. Her bir sınıflandırmada özellik vektörünün oluşturma şekline göre tf, tf-Idf ve WLLR olarak sonuçlar düzenlenmiştir. Sonuçlarda eşik değerlerine göre alınan ortalamalar kullanılmıştır. Sonuçlar yüzde olarak ve yuvarlatılarak verilmiştir.

- Tekli Yöntem Sonuçları

Bu veri grubunda tf için KelimeTekliVarYok ve KelimeTekliAğırlıklı %70 ortalama ile ve tf-idf için KategoriTekliVarYok ve KelimeTekliVarYok özellik seçimi %70 ortalama ile diğer özellik seçimlerine göre daha iyi sonuç vermiştir (Tablo 4.13). WLLR'de ise KelimeTekliVarYok ve KategoriTekliAğırlıklı özellik seçimlerinde %72 ortalama ile en iyi sonucu vermiştir.

Tablo 4.13: HV-IV sözlüğü kullanılarak çocuk DEP x SAĞ sınıflandırmasının NB, SMO, J48 teknikleri uygulanarak elde edilen Tekli yöntem doğruluk sonuçları.

DEP	tf				tf-idf				WLLR			
Feature	NB	SMO	J48	Ort	NB	SMO	J48	Ort	NB	SMO	J48	Ort
KelimeTekliVarYok	71	72	69	70	70	69	71	70	78	67	71	72
KelimeTekliAğırlıklı	73	76	62	70	69	68	52	63	73	62	62	66
KategoriTekliVarYok	64	56	64	61	65	63	72	70	51	53	53	52
KategoriTekliAğırlıklı	66	59	67	64	66	59	70	65	71	73	71	72

- *Liste ve Karma Yöntem Sonuçları*

Tüm sonuçlar Ek 8’de verilmesiyle birlikte eşik değerlerin ortalaması alınarak özet tablo çıkarılmak istendiğinde, tüm ağırlıklı, ağırlıksız olarak liste ve karma yöntemlerinin de ortalaması alındığında Tablo 4.14 elde edilir. Buna göre bu veri grubunda tf ve WLLR için liste yönteminin, tf-Idf için KategoriListe ve KategoriKarma yöntemlerinin daha başarılı olduğu gözlenir.

Tablo 4.14: DEP x SAĞ Sınıflandırmasında Karma, Liste ve Tekli yöntemlerinin kategori ve kelime olarak HV-IV ile doğruluk sonuçlarının ortalamaları.

DEP	tf				tf-idf				WLLR			
Liste/Karma	NB	SMO	J48	Ort	NB	SMO	J48	Ort	NB	SMO	J48	Ort
Kategori Tüm Liste	71	72	69	70	70	69	71	70	76	75	69	73
Kelime Tüm Liste	73	76	62	70	69	68	52	63	69	66	64	66
Kategori Tüm Karma	64	56	64	61	65	63	72	70	65	64	61	64
Kelime Tüm Karma	66	59	67	64	66	59	70	65	67	64	63	65
Karma, Liste, Tekli Ortalama	69	67	66	67	66	59	66	63	69	66	64	66

Eşik Değer ortalamalarına bakıldığında Tablo 4.15’de görüldüğü gibi 3 özellik seçim metodu için de iyi olan Eşik Değer yoktur. Eşik Değer 0 için tf-idf (%67) ve WLLR’de (%68) en iyi sonucu vermiştir. tf’de ise Eşik Değer 25 % 68 ile iyi sonuç vermiştir.

Tablo 4.15: Çocuk DEP x SAĞ Sınıflandırmasında Karma ve Liste yöntemlerinin HV-IV ile doğruluk sonuçlarının Eşik değerlere göre en genel ortalamaları.

DEP												
Eşik Değer	tf				tf-idf				WLLR			
	NB	SMO	J48	Ort	NB	SMO	J48	Ort	NB	SMO	J48	Ort
Eşik Değer 0	68	63	65	65	69	65	68	67	71	69	63	68
Eşik Değer 5	68	66	65	66	68	64	66	66	69	66	64	66
Eşik Değer 25	70	67	66	68	67	66	65	66	69	66	64	66
Eşik Değer 50	69	66	65	67	66	64	65	65	65	64	63	64
Ortalama	69	66	65	66	67	65	66	66	69	66	63	66

KS ve HS sonuçlarını karşılaştırdığımızda Tablo 4.16'a göre bu veri grubu için genel ortalama HS (67, 66, 65) için %66 ve KS (66, 66, 67) için %66 olarak birbirine yakın sonuç vermiştir.

Tablo 4.16: HV-IV sözlüğü kullanılarak DEP x SAĞ Sınıflandırmasında Karma ve Liste yöntemlerinin kategori ve kelime olarak doğruluk sonuçlarının HS ve KS yöntemlerine göre karşılaştırması.

DEP	tf				tf-idf				WLLR			
HS	NB	SMO	J48	Ort	NB	SMO	J48	Ort	NB	SMO	J48	Ort
KategoriListe_HS	70	71	70	70	69	69	71	69	76	74	70	73
KategoriListeAğırlıklı_HS	70	72	70	71								
KategoriKarmaFrek_HS	62	56	67	62	62	62	67	64	63	60	60	61
KategoriKarmaVarYok_HS	70	64	66	67	70	64	66	67	51	52	52	51
KelimeListe_HS	72	75	55	67	69	68	51	63	76	77	65	72
KelimeListeAğırlıklı_HS	75	77	69	74								
KelimeKarmaFrek_HS	64	59	59	61	66	60	70	65	67	60	62	63
KelimeKarmaVarYok_HS	67	59	74	67	67	59	74	67	72	64	72	69
Ortalama	69	67	66	67	67	64	67	66	67	64	63	65
KS	NB	SMO	J48	Ort	NB	SMO	J48	Ort	NB	SMO	J48	Ort
KategoriListe_KS	72	71	70	71	71	69	70	70	76	76	69	74
KategoriListeAğırlıklı_KS	73	73	64	70								
KategoriKarmaFrek_KS	57	49	64	56	67	64	76	69	69	71	64	68
KategoriKarmaVarYok_KS	68	53	62	61	68	53	62	61	51	53	53	52
KelimeListe_KS	69	75	54	66	69	67	52	63	75	77	63	72
KelimeListeAğırlıklı_KS	75	78	67	73								
KelimeKarmaFrek_KS	64	60	62	62	67	58	69	65	71	64	62	66
KelimeKarmaVarYok_KS	69	58	73	67	69	58	73	67	78	69	71	73
Ortalama	68	65	64	66	69	62	67	66	70	68	64	67

4.1.2.2. HV-IV ile ADHD x Sağlıklı Sınıflandırması Sonuçları

Tekli (BOW), Liste ve Karma yöntemlerinin NB, SMO, J48 makine öğrenmesi teknikleri uygulanarak elde edilen doğruluk sonuçları tablolar (Tablo 4.17, Ek 9) halinde verilmiştir. Her bir sınıflandırmada özellik vektörünün oluşturma şekline göre tf, tf-Idf ve WLLR olarak sonuçlar düzenlenmiştir. Sonuçlarda eşik değerlerine göre alınan ortalamalar kullanılmıştır.

- Tekli Yöntem Sonuçları

Bu veri grubunda WLLR için KelimeTekliVarYok %64 ortalama ile diğer özellik seçimlerine göre daha iyi sonuç vermiştir (Tablo 4.17). tf için KategoriTekliAğırlıklı SMO'da %61 ile iyi sonuç vermiştir.

Tablo 4.17: ADHD x SAĞ sınıflandırmasında Tekli yöntem sonuçları.

ADHD	tf				tf-idf				WLLR			
Feature	NB	SMO	J48	Ort	NB	SMO	J48	Ort	NB	SMO	J48	Ort
KelimeTekliVarYok	59	59	50	56	59	59	50	56	66	62	64	64
KelimeTekliAğırlıklı	52	54	60	55	44	50	43	46	63	57	48	56
KategoriTekliVarYok	57	46	56	53	57	46	56	53	50	50	50	50
KategoriTekliAğırlıklı	57	61	33	51	50	50	50	50	62	59	48	56

- *Liste ve Karma Yöntem Sonuçları*

ADHD grubu için yapılan sınıflandırmada yine bölüm 4.3.1’de anlatıldığı gibi listeler çıkarılmış, eşik değer uygulanmıştır. Buna göre alınan sonuçlar Ek 9’da verilmiştir.

Eşik değerlerin ortalaması alınarak özet tablo çıkarılmak istendiğinde, tüm ağırlıklı, ağırlıksız olarak liste ve karma yöntemlerinin de ortalaması alınınca Tablo 4.18 elde edilir. Buna göre tf ve tf-idf için en iyi sonuç %61 olarak kelime liste özelliğindeki NB sonucudur. WLLR için en iyi sonuç %60 ile KelimeKarma ve NB uygulandığında alınmıştır.

Tablo 4.18: HV-IV kullanılarak çocuk ADHD x SAĞ sınıflandırmasında Karma, Liste ve Tekli yöntemlerinin kategori ve kelime olarak doğruluk sonuçlarının ortalamaları.

ADHD	tf				tf-idf				WLLR			
Liste/Karma	NB	SMO	J48	Ort	NB	SMO	J48	Ort	NB	SMO	J48	Ort
Kategori Tüm Liste	55	57	51	54	50	52	52	51	57	58	59	58
Kelime Tüm Liste	61	63	56	60	48	52	56	52	58	59	55	57
Kategori Tüm Karma	50	47	49	49	58	59	54	57	54	52	49	52
Kelime Tüm Karma	57	56	60	58	57	55	49	54	60	52	57	56
Karma, Liste, Tekli Ortalama	56	55	52	54	53	53	51	52	59	56	54	56

Eşik Değer ortalamalarına bakıldığında Tablo 4.19’de görüldüğü gibi 3 özellik seçim metodu için de iyi olan Eşik Değer yoktur. tf için Eşik Değer 5 (%57), tf-idf için Eşik Değer 0 ve Eşik Değer 25 %54 ile, ve WLLR’de ise Eşik Değer 0 %56 ile en iyi sonucu vermiştir.

Tablo 4.19: ADHD x SAĞ Sınıflandırmasında Karma ve Liste yöntemlerinin doğruluk sonuçlarının Eşik değerlere göre en genel ortalamaları.

ADHD												
	tf				tf-idf				WLLR			
Eşik Değer	NB	SMO	J48	Ort	NB	SMO	J48	Ort	NB	SMO	J48	Ort
Eşik Değer 0	56	56	53	55	53	56	52	54	60	55	54	56
Eşik Değer 5	57	59	55	57	53	55	53	53	57	53	54	55
Eşik Değer 25	57	55	55	56	53	53	54	54	55	55	55	55
Eşik Değer 50	54	54	53	54	53	54	51	53	57	54	54	55
Ortalama	56	56	54	55	53	55	53	53	57	54	54	55

Tablo 4.20'e göre bu veri grubu için HS ve KS arasında belirgin bir fark görünmemektedir. Genel ortalamada HS (55, 55, 55) için %55 ve KS (56, 53, 56) için %55 olarak KS birbirine yakın sonuç vermiştir.

Tablo 4.20: HV-IV kullanılarak ADHD x SAĞ sınıflandırmasında Karma ve Liste yöntemlerinin kategori ve kelime olarak doğruluk sonuçlarının HS ve KS yöntemlerine göre karşılaştırması.

ADHD	tf				tf-idf				WLLR			
HS	NB	SMO	J48	Ort	NB	SMO	J48	Ort	NB	SMO	J48	Ort
KategoriListe_HS	52	57	49	53	51	52	52	52	54	57	56	56
KategoriListeAğırlıklı_HS	55	56	50	54								
KategoriKarmaFrek_HS	52	51	46	50	56	58	54	56	53	51	49	51
KategoriKarmaVarYok_HS	55	47	53	52	55	47	53	52	50	50	50	50
KelimeListe_HS	58	60	50	56	48	52	58	53	59	58	56	58
KelimeListeAğırlıklı_HS	62	64	58	61								
KelimeKarmaFrek_HS	55	53	55	54	58	56	52	55	55	54	51	53
KelimeKarmaVarYok_HS	60	58	64	60	60	58	64	60	66	56	62	61
Ortalama	56	56	53	55	55	54	56	55	56	54	54	55
KS	tf				tf-idf				WLLR			
KS	NB	SMO	J48	Ort	NB	SMO	J48	Ort	NB	SMO	J48	Ort
KategoriListe_KS	56	59	52	56	49	51	52	51	59	60	62	60
KategoriListeAğırlıklı_KS	58	58	52	56								
KategoriKarmaFrek_KS	44	49	46	46	59	60	54	58	62	58	48	56
KategoriKarmaVarYok_KS	50	42	52	48	50	42	52	48	50	50	50	50
KelimeListe_KS	64	65	55	61	48	53	53	51	57	60	53	57
KelimeListeAğırlıklı_KS	62	65	59	62								
KelimeKarmaFrek_KS	56	58	61	58	56	55	46	52	55	47	48	50
KelimeKarmaVarYok_KS	58	54	61	57	58	54	61	57	65	53	66	61
Ortalama	56	56	55	56	53	53	53	53	58	55	55	56

4.1.2.3. HV-IV ile Çocuk Verisinde Cinsiyet Sınıflandırması İçin Bulgular

Bu bölümde DEP, ADHD ve SAĞ verilerinin Kadın x Erkek (K x E) sınıflandırılmasının sonuçları verilmiştir. Kelime kategorileri HV-IV sözlüğü kullanılarak belirlenmiştir.

- Tekli Yöntem Sonuçları

Bu veri grubunda tf ve idf için KelimeTekliVarYok %58 ortalama ile ve WLLR'de ise yine KelimeTekliVarYok %59 ile diğer özellik seçimlerine göre daha iyi sonuç vermiştir (Tablo 4.21).

Tablo 4.21: Çocuk K x E sınıflandırmasında Tekli yöntem sonuçları.

Cinsiyet	tf				tf-idf				WLLR			
Feature	NB	SMO	J48	Ort	NB	SMO	J48	Ort	NB	SMO	J48	Ort
KelimeTekliVarYok	51	60	61	58	51	60	61	58	50	64	64	59
KelimeTekliAğırlıklı	54	50	64	56	46	51	65	54	53	53	68	58
KategoriTekliVarYok	50	57	54	54	50	57	54	54	50	50	50	50
KategoriTekliAğırlıklı	56	49	49	51	50	50	50	50	50	50	50	50

- *Liste ve Karma Yöntem Sonuçları*

Çocuk cinsiyet grubu için yapılan sınıflandırmada yine bölüm 4.3.1’de anlatıldığı gibi listeler çıkarılmış, eşik değeri uygulanmıştır. Buna göre alınan sonuçlar Ek 10’da verilmiştir. Tablo 4.23 ve Tablo 4.22 çıkarılmış özet tablolardır. Buna göre en iyi ortalamalar KelimeKarma kullanıldığında alınmıştır. KelimeKarma tf için %55, tf-Idf için %60, WLLR için %54 ile en iyi sonucu vermiştir.

Tablo 4.22: Çocuk K x E Sınıflandırmasında Karma, Liste ve Tekli yöntemlerinin kategori ve kelime olarak doğruluk sonuçlarının ortalamaları.

Cinsiyet	tf				tf-idf				WLLR			
Liste/Karma	NB	SMO	J48	Ort	NB	SMO	J48	Ort	NB	SMO	J48	Ort
Kategori Tüm Liste	50	50	50	50	52	50	50	51	50	50	50	50
Kelime Tüm Liste	51	51	52	51	52	50	56	53	52	50	53	52
Kategori Tüm Karma	52	53	53	53	54	51	50	52	50	50	50	50
Kelime Tüm Karma	55	52	60	55	58	57	66	60	51	52	59	54
Karma, Liste, Tekli Ortalama	52	53	55	53	52	53	57	54	51	52	56	53

Eşik Değer ortalamalarına bakıldığında Tablo 4.23’de görüldüğü gibi tf için Eşik Değer 50 (%53), tf-idf için Eşik Değer 25 (%55) ve WLLR için Eşik Değer 50 %53 ile en iyi sonucu vermiştir.

Tablo 4.23: Çocuk K x E Sınıflandırmasında Karma ve Liste yöntemlerinin doğruluk sonuçlarının Eşik değerlere göre genel ortalamaları.

Cinsiyet												
Eşik Değer	tf				tf-idf				WLLR			
	NB	SMO	J48	Ort	NB	SMO	J48	Ort	NB	SMO	J48	Ort
Eşik Değer 0	52	52	53	52	54	51	55	53	51	51	55	52
Eşik Değer 5	52	51	53	52	54	53	55	54	52	51	55	52
Eşik Değer 25	52	51	54	52	54	52	57	55	51	50	55	52
Eşik Değer 50	52	52	55	53	53	52	55	53	51	51	56	53
Ortalama	52	51	54	52	54	52	55	54	51	51	55	52

KS ve HS sonuçlarını değerlendirmek için Tablo 4.24'e bakıldığında bu veri grubu için HS (%52, %55, %53) ortalama %53 ve KS (%53, %56, %52) ortalama %54 arasında belirgin bir fark görünmemektedir.

Tablo 4.24: HV-IV sözlüğü kullanılarak yapılan K x E sınıflandırmasında Karma ve Liste yöntemlerinin kategori ve kelime olarak doğruluk sonuçlarının HS ve KS yöntemlerine göre karşılaştırması.

Cinsiyet	tf				tf-idf				WLLR			
HS	NB	SMO	J48	Ort	NB	SMO	J48	Ort	NB	SMO	J48	Ort
KategoriListe_HS	50	50	50	50	51	50	50	50	51	50	50	50
KategoriListeAğırlıklı_HS	50	50	50	50								
KategoriKarmaFrek_HS	54	50	51	52	54	51	51	52	50	50	50	50
KategoriKarmaVarYok_HS	50	53	52	52	50	53	52	52	50	50	50	50
KelimeListe_HS	51	51	51	51	52	50	55	52	53	50	54	53
KelimeListeAğırlıklı_HS	52	51	53	52								
KelimeKarmaFrek_HS	57	48	61	55	59	61	65	62	55	50	65	57
KelimeKarmaVarYok_HS	52	56	59	56	52	56	59	56	50	55	64	56
Ortalama	52	51	54	52	54	55	57	55	52	51	56	53
KS	NB	SMO	J48	Ort	NB	SMO	J48	Ort	NB	SMO	J48	Ort
KategoriListe_KS	50	50	50	50	52	50	50	51	50	50	50	50
KategoriListeAğırlıklı_KS	50	50	50	50								
KategoriKarmaFrek_KS	54	50	51	52	54	51	49	51	51	50	50	50
KategoriKarmaVarYok_KS	49	60	57	55	49	60	57	55	50	50	50	50
KelimeListe_KS	51	51	52	51	53	50	57	53	50	50	51	50
KelimeListeAğırlıklı_KS	50	50	52	52								
KelimeKarmaFrek_KS	57	44	57	53	57	54	67	59	53	51	65	56
KelimeKarmaVarYok_KS	53	58	62	58	53	58	62	58	50	56	63	56
Ortalama	52	52	54	53	54	54	59	56	51	51	55	52

Bu bölümde, okul çağındaki çocukların zihinsel bozukluklarının saptanması için bir metin analizi yapılmıştır. Okul çağı çocuklarının yönlendirilmiş metinleri üzerinde çalışılmıştır. Deneklerin metinlerinden elde edilen farklı kelimeler / kategoriler ve bunların varlık veya sıklık bilgisi, idf ile ağırlıklandırılarak ve WLLR ile seçilerek NB, SMO ve J48 metotları ile sınıflandırılmıştır. DEP (hasta x sağlıklı), ADHD (hasta x sağlıklı) ve cinsiyet (K x E) olarak ikili sınıflandırılmıştır. Kelimelerin ait olduğu duygu kategorilerinin bulunmasında HV-IV sözlüğünden faydalanılmıştır.

DEP grubunda ADHD'den ve cinsiyet sınıflandırılmasından daha iyi sonuç alındığı gözlemlenmiştir.

Çocuk verileri için HV-IV ile yapılan sınıflandırmalar ile HV-III ile yapılanlar karşılaştırıldığında ise HV-III sözlüğünde daha iyi sonuçlar alındığı gözlenmiştir (Tablo 4.25). Çocuk verisi doğruluk sonuçları MÖ sınıflandırıcıları yönünden karşılaştırıldığında tf için en iyi sınıflandırıcı NB (%68, %59, %67) olmuştur. En düşük sonuç ise J48'de (%61, %58, %63) alınmıştır. tf-idf'de J48 (%64, %58, %65) ile en iyi sonucu almıştır. WLLR'de ise NB (%64, %57, %63) ile en iyi sonucu vermiştir.

Tablo 4.25: Çocuk verisinde HV-III, HV-IV ve kelime karşılaştırması.

	tf				tf-idf				WLLR			
HV-III	NB	SMO	J48	Ort	NB	SMO	J48	Ort	NB	SMO	J48	Ort
Dep Çocuk	71	66	67	68	67	60	66	64	76	68	68	71
ADHD Çocuk	58	51	54	54	58	56	55	56	61	57	53	57
Cinsiyet Çocuk	74	77	62	71	63	69	70	67	56	52	55	55
Ortalama	68	65	61	64	63	62	64	62	64	59	59	61
HV-IV	NB	SMO	J48	Ort	NB	SMO	J48	Ort	NB	SMO	J48	Ort
Dep Çocuk	69	67	66	67	66	59	66	63	66	66	64	65
ADHD Çocuk	56	55	52	54	53	53	51	52	56	55	52	54
Cinsiyet Çocuk	52	53	55	53	52	53	57	54	50	50	50	50
Ortalama	59	58	58	58	57	55	58	56	57	57	55	56
Kelime	NB	SMO	J48	Ort	NB	SMO	J48	Ort	NB	SMO	J48	Ort
Dep Çocuk	69	66	66	67	67	58	64	62	75	68	66	69
ADHD Çocuk	56	52	64	57	55	55	57	56	62	58	56	58
Cinsiyet Çocuk	76	79	58	71	65	70	74	70	52	55	62	56
Ortalama	67	65	63	65	62	61	65	62	63	60	61	61

4.2. YETİŞKİN VERİSİ İKİLİ SINIFLANDIRMA BULGULARI

4.2.1. HV-III ile Yetişkin Verisi İçin Bulgular

Yetişkin verileri Depresyon x Sağlıklı ve Mani x Sağlıklı olarak ayrı ayrı sınıflandırılmıştır. Her iki grubun da test verileri üç kere değiştirilerek sonuçlarının ortalamaları alınmıştır. Sınıflandırma işlemi veri madenciliğinde yaygın olarak kullanılan Weka programında MÖ metotları kullanılarak yapılmıştır (Witten ve Frank, 2005). MÖ metotlarından Navie Bayes (NB), J48 ve SMO kullanılmıştır.

Tekli (BOW), Liste ve Karma yöntemlerinin NB, SMO, J48 makine öğrenmesi teknikleri uygulanarak elde edilen doğruluk sonuçları verilmiştir. Her bir sınıflandırmada özellik vektörünün oluşturma şekline göre tf, tf-Idf ve WLLR olarak sonuçlar düzenlenmiştir.

4.2.1.1. HV-III ile Yetişkin Depresyon x Sağlıklı Sınıflandırması Sonuçları

- Tekli Yöntem Sonuçları

Bu veri grubunda WLLR'den sonra tf için KelimeTekliVarYok ve KelimeTekliAğırlıklı'te ortalama %86, %81 ile ve tf-idf için %86, %73 ile diğer özellik seçimlerine göre daha iyi sonuç vermiştir (Tablo 4.26).

Tablo 4.26: Yetişkin DEP x SAĞ sınıflandırmasında HV-III ve NB, SMO, J48 yöntemleri uygulanarak elde edilen Tekli yöntem doğruluk sonuçları.

DEP	tf				tf-idf				WLLR			
Feature	NB	SMO	J48	Ort	NB	SMO	J48	Ort	NB	SMO	J48	Ort
KelimeTekliVarYok	94	92	70	86	94	92	70	86	83	75	67	75
KelimeTekliAğırlıklı	85	93	65	81	79	76	65	73	82	83	58	74
KategoriTekliVarYok	74	59	72	69	74	59	72	69	77	65	65	69
KategoriTekliAğırlıklı	81	81	63	75	50	50	52	51	78	78	65	74

- Liste ve Karma Yöntem Sonuçları

Alınan tüm sonuçlar (Tablo 4.26, Ek 11)'de verilmiştir. Liste ve karma yöntemlerinde ve HS-KS yöntemlerinin sonuçları karşılaştırmak için eşik değerlerin ortalaması alınarak özet tablolar çıkarılmıştır. Eşik Değer ortalamalarına bakıldığında Tablo 4.27'de görüldüğü gibi tf için Eşik Değer 25 (%73), tf-idf için Eşik Değer 25 ve Eşik Değer 50 (%70) ve WLLR için Eşik Değer 25 (%70) en iyi sonucu vermiştir.

Tablo 4.27: Yetişkin DEP x SAĞ sınıflandırmasında Karma ve Liste yöntemlerinin doğruluk sonuçlarının Eşik değerlere göre genel ortalamaları.

	DEP											
	tf				tf-idf				WLLR			
Eşik Değer	NB	SMO	J48	Ort	NB	SMO	J48	Ort	NB	SMO	J48	Ort
Eşik Değer 0	66	64	60	63	67	73	60	67	71	70	59	66
Eşik Değer 5	70	71	62	68	66	74	59	66	74	72	62	69
Eşik Değer 25	76	76	66	73	70	74	67	70	74	73	64	70
Eşik Değer 50	76	70	68	71	70	74	66	70	71	69	62	67
Ortalama	72	71	64	69	68	74	63	68	72	71	62	68

Tüm ağırlıklı, ağırlıksız olarak liste ve karma yöntemlerinin de ortalaması alınınca Tablo 4.28 elde edilir. Buna göre kelime-karma yönteminin tf için %77, tf-idf için %80 ve WLLR için %74 olarak bu veri grubunda daha başarılı olduğu gözlenir.

Tablo 4.28: Yetişkin DEP x SAĞ Sınıflandırmasında Karma, Liste ve Tekli yöntemlerinin kategori ve kelime olarak doğruluk sonuçlarının ortalamaları.

DEP	tf				tf-idf				WLLR			
Liste/Karma	NB	SMO	J48	Ort	NB	SMO	J48	Ort	NB	SMO	J48	Ort
Kategori Tüm Liste	64	66	65	65	57	60	61	59	70	69	65	68
Kelime Tüm Liste	68	76	62	69	66	69	57	64	70	69	65	68
Kategori Tüm Karma	68	65	62	65	67	75	67	70	65	64	56	62
Kelime Tüm Karma	87	75	67	77	83	91	66	80	81	79	64	74
Karma, Liste, Tekli Ortalama	78	76	66	73	71	72	64	69	76	73	63	71

Tablo 4.29 tablosuna göre bu veri grubu için HS (%67, %70, %68) ve KS (%70, %70, %69) karşılaştırıldığında sırasıyla ortalama %68 ve %70 olarak %2 farkla KS daha iyi sonuç vermiştir.

Tablo 4.29: Yetişkin DEP x SAĞ Sınıflandırmasında Karma ve Liste yöntemlerinin kategori ve kelime olarak doğruluk sonuçlarının HS ve KS yöntemlerine göre karşılaştırması.

DEP	tf				tf-idf				WLLR			
HS	NB	SMO	J48	Ort	NB	SMO	J48	Ort	NB	SMO	J48	Ort
KategoriListe_HS	61	64	63	63	56	58	62	59	65	66	60	64
KategoriListeAğırlıklı_HS	57	59	61	59								
KategoriKarmaFrek_HS	65	67	57	63	67	73	66	69	68	69	60	66
KategoriKarmaVarYok_HS	64	60	63	62	64	60	63	62	61	57	56	58
KelimeListe_HS	68	80	57	68	69	67	55	64	72	75	65	71
KelimeListeAğırlıklı_HS	69	71	67	69								
KelimeKarmaFrek_HS	84	64	65	71	85	91	71	82	81	82	61	75
KelimeKarmaVarYok_HS	93	91	69	84	93	91	69	84	81	74	68	74
Ortalama	70	70	63	67	72	73	64	70	71	70	62	68
KS	NB	SMO	J48	Ort	NB	SMO	J48	Ort	NB	SMO	J48	Ort
KategoriListe_KS	71	70	71	71	56	61	59	59	74	71	71	72
KategoriListeAğırlıklı_KS	67	71	64	67								
KategoriKarmaFrek_KS	74	76	68	73	67	78	69	71	73	73	57	68
KategoriKarmaVarYok_KS	69	59	59	62	69	59	59	62	58	56	50	55
KelimeListe_KS	65	79	59	68	63	71	60	65	73	71	66	70
KelimeListeAğırlıklı_KS	69	74	65	69								
KelimeKarmaFrek_KS	79	55	65	66	81	91	61	78	83	83	58	75
KelimeKarmaVarYok_KS	94	90	70	85	94	90	70	85	78	75	67	73
Ortalama	74	72	65	70	72	75	63	70	73	72	61	69

4.2.1.2. HV-III ile Yetişkin Mani x Sağlıklı Sınıflandırması Sonuçları

- Tekli Yöntem Sonuçları

Bu veri grubunda tf için KelimeTekliAğırlıklı'da ortalama %73 ile, tf-idf için KelimeTekliVarYok %66 ile ve WLLR için KelimeTekliAğırlıklı'da %72 doğruluk sonucu ile

diğer özellik seçimlerine göre daha iyi sonuç vermiştir (Tablo 4.30).

Tablo 4.30: Yetişkin MAN x SAĞ sınıflandırmasında NB, SMO, J48 yöntemleri uygulanarak elde edilen Tekli yöntem doğruluk sonuçları.

MAN	tf				tf-idf				WLLR			
Feature	NB	SMO	J48	Ort	NB	SMO	J48	Ort	NB	SMO	J48	Ort
KelimeTekliVarYok	65	80	53	66	65	80	53	66	75	70	65	70
KelimeTekliAğırlıklı	72	75	73	73	52	65	53	57	77	77	63	72
KategoriTekliVarYok	67	60	58	62	67	60	58	62	63	58	58	60
KategoriTekliAğırlıklı	70	72	58	67	50	50	50	50	70	73	60	68

- *Liste ve Karma Yöntem Sonuçları*

Eşik Değer ortalamalarına bakıldığında Tablo 4.31’de görüldüğü gibi tf için Eşik Değer 25 (%68), tf-idf için Eşik Değer 25 (%66) ve WLLR için Eşik Değer 0, 5 ve 25 (%69) en iyi sonucu vermiştir.

Tablo 4.31: Yetişkin MAN x SAĞ sınıflandırmasında Karma ve Liste yöntemlerinin doğruluk sonuçlarının Eşik değerlere göre genel ortalamaları.

MAN												
Eşik Değer	tf				tf-idf				WLLR			
	NB	SMO	J48	Ort	NB	SMO	J48	Ort	NB	SMO	J48	Ort
Eşik Değer 0	65	63	58	62	64	68	63	65	70	72	64	69
Eşik Değer 5	66	65	61	64	61	65	59	62	71	72	65	69
Eşik Değer 25	70	68	66	68	65	69	64	66	71	71	66	69
Eşik Değer 50	69	69	65	67	65	65	61	64	66	68	63	65
Ortalama	67	66	62	65	64	67	62	64	69	70	64	68

Tüm liste ve karma sonuçları Ek 12’de verilmesiyle birlikte eşik değerlerin ortalaması alınarak özet tablo çıkarılmak istendiğinde, tüm ağırlıklı, ağırlıksız olarak liste ve karma yöntemlerinin sonucunun da ortalamaları alınınca Tablo 4.32 elde edilir.

Buna göre tf için KelimeListe yöntemi %68, tf-idf için KelimeKarma yöntemi %69 ve WLLR için KelimeListe yöntemi %77 olarak bu veri grubunda daha başarılı olduğu gözlenir (Tablo 4.32).

Tablo 4.32: Yetişkin MAN x SAĞ Sınıflandırmasında Karma, Liste ve Tekli yöntemlerinin kategori ve kelime olarak doğruluk sonuçlarının ortalamaları.

MAN	tf				tf-idf				WLLR			
Liste/Karma	NB	SMO	J48	Ort	NB	SMO	J48	Ort	NB	SMO	J48	Ort
Kategori Tüm Liste	70	69	61	67	64	61	57	61	76	79	71	75
Kelime Tüm Liste	68	71	65	68	61	60	64	62	80	77	74	77
Kategori Tüm Karma	63	58	61	60	65	68	64	66	58	63	60	60
Kelime Tüm Karma	69	67	63	66	65	79	62	69	73	70	61	68
Karma, Liste, Tekli Ortalama	68	69	62	66	61	65	58	61	71	71	64	69

Tablo 4.33'e göre bu veri grubunda HS ve KS karşılaştırıldığında en genel ortalama HS (64, 63, 67) için %65 ve KS (66, 65, 69) için %67 olarak %2 KS daha iyi sonuç vermiştir.

Tablo 4.33: Yetişkin MAN x SAĞ Sınıflandırmasında Karma ve Liste yöntemlerinin kategori ve kelime olarak doğruluk sonuçlarının HS ve KS yöntemlerine göre karşılaştırması.

MAN	tf				tf-idf				WLLR			
HS	NB	SMO	J48	Ort	NB	SMO	J48	Ort	NB	SMO	J48	Ort
KategoriListe_HS	63	59	52	58	63	60	55	54	73	77	70	73
KategoriListeAğırlıklı_HS	67	69	61	66								
KategoriKarmaFrek_HS	63	58	60	60	64	65	63	64	63	66	60	63
KategoriKarmaVarYok_HS	65	63	63	63	65	63	63	63	56	55	55	56
KelimeListe_HS	63	65	60	63	61	58	64	61	79	75	73	76
KelimeListeAğırlıklı_HS	71	77	70	73								
KelimeKarmaFrek_HS	67	58	68	64	65	76	61	67	73	71	62	69
KelimeKarmaVarYok_HS	68	76	54	66	68	76	54	66	71	68	63	67
Ortalama	66	66	61	64	64	66	60	63	69	69	64	67
	tf				tf-idf				WLLR			
KS	NB	SMO	J48	Ort	NB	SMO	J48	Ort	NB	SMO	J48	Ort
KategoriListe_KS	75	72	62	70	65	61	59	62	78	81	73	78
KategoriListeAğırlıklı_KS	77	75	71	74								
KategoriKarmaFrek_KS	55	55	60	57	67	70	65	67	58	73	67	66
KategoriKarmaVarYok_KS	67	55	60	61	67	55	60	61	53	57	58	56
KelimeListe_KS	63	64	64	63	61	62	64	62	80	78	75	78
KelimeListeAğırlıklı_KS	73	79	68	73								
KelimeKarmaFrek_KS	65	60	73	66	65	82	63	70	78	75	60	71
KelimeKarmaVarYok_KS	75	75	55	68	75	75	55	68	70	67	58	65
Ortalama	69	67	64	66	67	68	61	65	70	72	65	69

4.2.1.3. HV-III ile Yetişkin Cinsiyet Sınıflandırması Sonuçları

Bu bölümde yetişkin serbest konuşma verilerinin Kadın x Erkek (K x E) sınıflandırılması yapılmıştır.

- *Tekli Yöntem Sonuçları*

Bu veri grubunda tf için KategoriTekliAğırlıklı'da ortalama %70 ile, tf-Idf için KelimeTekliVarYok %64 ile ve WLLR'de KelimeTekliAğırlıklı'da %71 ile diğer özellik seçimlerine göre daha iyi sonuç vermiştir (Tablo 4.34).

Tablo 4.34: Yetişkin K x E sınıflandırmasında NB, SMO, J48 yöntemleri uygulanarak elde edilen tekli yöntem doğruluk sonuçları.

Cinsiyet	tf				tf-idf				WLLR			
Feature	NB	SMO	J48	Ort	NB	SMO	J48	Ort	NB	SMO	J48	Ort
KelimeTekliVarYok	62	83	48	64	62	83	48	64	53	72	58	61
KelimeTekliAğırlıklı	59	66	41	55	69	59	45	58	78	61	75	71
KategoriTekliVarYok	59	48	59	55	59	48	59	55	53	58	50	54
KategoriTekliAğırlıklı	72	72	66	70	62	59	41	54	88	88	32	69

- *Liste ve Karma Yöntem Sonuçları*

Eşik Değer ortalamalarına bakıldığında Tablo 4.35'de görüldüğü gibi tf için Eşik Değer 5 ve 25 (%60), tf-idf için Eşik Değer 25 (%56) ve WLLR için Eşik Değer 5 (%68) en iyi sonucu vermiştir.

Tablo 4.35: Yetişkin K x E sınıflandırmasında Karma ve Liste yöntemlerinin doğruluk sonuçlarının Eşik değerlere göre genel ortalamaları.

	Cinsiyet											
	tf				tf-idf				WLLR			
Eşik Değer	NB	SMO	J48	Ort	NB	SMO	J48	Ort	NB	SMO	J48	Ort
Eşik Değer 0	56	60	52	56	46	56	56	53	66	62	60	63
Eşik Değer 5	61	64	55	60	46	57	56	53	71	70	63	68
Eşik Değer 25	59	62	59	60	51	59	60	56	69	66	64	66
Eşik Değer 50	59	61	52	57	51	58	53	54	62	63	62	62
Ortalama	59	62	54	58	48	58	56	54	67	65	62	65

Tüm liste ve karma sonuçları Ek 13'de verilmesiyle birlikte eşik değerlerin ortalaması alınarak özet tablo çıkarılmak istendiğinde, tüm ağırlıklı, ağırlıksız olarak liste ve karma yöntemlerinin de ortalaması alınınca Tablo 4.36 elde edilmiştir.

tf için KelimeListe yöntemi %64, tf-idf için KelimeKarma yöntemi %64 ve WLLR için KelimeListe yöntemi %72 olarak bu veri grubunda daha başarılı olduğu gözlenir (Tablo 4.36).

Tablo 4.36: Yetişkin K x E sınıflandırmasında Karma, Liste ve Tekli yöntemlerinin kategori ve kelime olarak doğruluk sonuçlarının ortalamaları.

Cinsiyet	tf				tf-idf				WLLR			
Liste/Karma	NB	SMO	J48	Ort	NB	SMO	J48	Ort	NB	SMO	J48	Ort
Kategori Tüm Liste	58	60	60	59	41	57	56	51	69	68	71	69
Kelime Tüm Liste	60	70	62	64	52	59	52	55	73	67	75	72
Kategori Tüm Karma	57	49	52	53	43	55	42	46	64	61	49	58
Kelime Tüm Karma	61	68	44	57	57	60	75	64	66	68	65	66
Karma, Liste, Tekli Ortalama	61	65	54	60	54	58	52	55	68	68	59	65

Tablo 4.37'e göre bu veri grubu için HS ve KS karşılaştırıldığında en genel ortalama HS (58, 56, 63) için %59 ve KS (58, 54, 67) için %60 olarak %1 farkla KS daha iyi sonuç vermiştir.

Tablo 4.37: Yetişkin K x E sınıflandırmasında Karma ve Liste yöntemlerinin kategori ve kelime olarak doğruluk sonuçlarının HS ve KS yöntemlerine göre karşılaştırması.

Cinsiyet	tf				tf-idf				WLLR			
HS	NB	SMO	J48	Ort	NB	SMO	J48	Ort	NB	SMO	J48	Ort
KategoriListe_HS	53	59	58	57	42	58	55	52	62	61	62	61
KategoriListeAğırlıklı_HS	54	59	61	58								
KategoriKarmaFrek_HS	54	54	53	54	48	57	46	50	75	74	51	67
KategoriKarmaVarYok_HS	53	50	56	53	53	50	56	53	50	50	51	50
KelimeListe_HS	53	71	52	58	52	60	51	54	73	63	76	71
KelimeListeAğırlıklı_HS	63	68	66	66								
KelimeKarmaFrek_HS	63	52	41	52	56	61	74	63	79	74	78	77
KelimeKarmaVarYok_HS	63	81	46	63	63	81	46	63	50	50	54	51
Ortalama	57	62	54	58	52	61	55	56	65	62	62	63
KS	NB	SMO	J48	Ort	NB	SMO	J48	Ort	NB	SMO	J48	Ort
KategoriListe_KS	61	60	61	60	39	56	58	51	76	74	81	77
KategoriListeAğırlıklı_KS	65	64	59	62								
KategoriKarmaFrek_KS	66	55	41	54	38	52	38	43	82	69	44	65
KategoriKarmaVarYok_KS	55	38	59	51	55	38	59	51	50	50	51	50
KelimeListe_KS	60	70	57	58	53	59	54	55	73	72	74	73
KelimeListeAğırlıklı_KS	64	72	72	66								
KelimeKarmaFrek_KS	59	59	41	53	59	59	76	65	85	89	79	84
KelimeKarmaVarYok_KS	59	79	48	62	59	79	48	62	50	57	49	52
Ortalama	61	62	55	58	50	57	55	54	69	68	63	67

Yetişkin veri grubunda, özellik seçimi yöntemlerinde eşik değerlerine göre alınan ortalamalar ve kelime/kategori varlık/yokluk ve frekans, idf ağırlıklı ve WLLR ile seçilen özellik yöntemlerine göre sınıflandırmalar yapılmıştır. Yapılan DEP x SAĞ, MAN x SAĞ ve cinsiyet

sınıflandırmalarının NB, SMO, J48 olarak uygulanan MÖ teknikleri ile elde edilen doğruluk sonuçları verilmiştir. Genel olarak özet tablolarındaki en genel ortalamalar karşılaştırıldığında depresyon verisinin sonuçları mani verisinden daha iyidir, mani verisinin sonuçları ise cinsiyetten daha iyi çıkmıştır (Tablo 4.28, Tablo 4.32, Tablo 4.36). HS ve KS yöntemleri arasında ise birbirine yakın sonuçlar elde edilmiştir. KS yöntemi DEP ve MAN'da %2 cinsiyet verisinde %1 farkla daha iyi sonuç vermiştir.

4.2.2. HV-IV ile Yetişkin Verisi İçin Bulgular

Yetişkin verileri Depresyon x Sağlıklı ve Mani x Sağlıklı olarak ayrı ayrı sınıflandırılmıştır. Her iki grubun da test verileri üç kere değiştirilerek sonuçlarının ortalamaları alınmıştır. Sınıflandırma işlemi veri madenciliğinde yaygın olarak kullanılan Weka programında MÖ metotları kullanılarak yapılmıştır (Witten ve Frank, 2005). MÖ metotlarından Navie Bayes (NB), karar ağacı algoritması olan J48 ve Sequential Minimal Optimization (SMO) kullanılmıştır.

Tekli (BOW), Liste ve Karma (BOW) yöntemlerinin NB, SMO, J48 makine öğrenmesi teknikleri uygulanarak elde edilen doğruluk sonuçları verilmiştir. Her bir sınıflandırmada özellik vektörünün oluşturma şekline göre tf, tf-Idf ve WLLR olarak sonuçlar düzenlenmiştir. Sonuçlarda eşik değerlerine göre alınan ortalamalar kullanılmıştır. Sonuçlar yüzde olarak ve yuvarlatılarak verilmiştir.

4.2.2.1. HV-IV ile Yetişkin Depresyon x Sağlıklı Sınıflandırması Sonuçları

- Tekli Yöntem Sonuçları

Bu veri grubunda tf için KelimeTekliVarYok yöntemi ortalama %81 ile, tf-Idf için %81 ile ve WLLR için %75 ile diğer özellik seçimlerine göre daha iyi sonuç vermiştir (Tablo 4.38).

Tablo 4.38: Yetişkin DEP x SAĞ sınıflandırmasında HV-IV sözlüğü kullanılarak elde edilen Tekli yöntem doğruluk sonuçları.

DEP	tf				tf-idf				WLLR			
Feature	NB	SMO	J48	Ort	NB	SMO	J48	Ort	NB	SMO	J48	Ort
KelimeTekliVarYok	92	92	58	81	92	92	58	81	83	75	67	75
KelimeTekliAğırlıklı	88	85	62	78	75	78	60	71	82	83	58	74
KategoriTekliVarYok	65	53	63	61	65	53	63	61	50	50	50	50
KategoriTekliAğırlıklı	58	80	72	70	50	50	50	50	58	75	78	71

- *Liste ve Karma Yöntem Sonuçları*

Alınan tüm sonuçlar Ek 14’de verilmiştir. Liste ve karma yöntemlerinde ve HS-KS yöntemlerinin sonuçları karşılaştırmak için eşik değerlerin ortalaması alınarak özet tablolar çıkarılmıştır.

Eşik Değer ortalamalarına bakıldığında Tablo 4.39’de görüldüğü gibi tf için Eşik Değer 25 (%72), tf-idf için Eşik Değer 50 (%63) ve WLLR için Eşik Değer 25 (%71) en iyi sonucu vermiştir.

Tablo 4.39: Yetişkin DEP x SAĞ sınıflandırmasında HV-IV kullanılarak Karma ve Liste yöntemlerinin doğruluk sonuçlarının Eşik değerlere göre genel ortalamaları.

DEP	tf				tf-idf				WLLR			
	NB	SMO	J48	Ort	NB	SMO	J48	Ort	NB	SMO	J48	Ort
Eşik Değer 0	65	64	57	62	59	61	58	59	69	71	65	69
Eşik Değer 5	70	66	60	65	60	62	58	60	71	71	65	69
Eşik Değer 25	76	74	65	72	59	62	58	60	72	72	67	71
Eşik Değer 50	75	68	65	69	63	67	59	63	68	68	63	67
Ortalama	71	68	62	67	60	63	58	60	70	71	65	69

Tüm ağırlıklı, ağırlıksız olarak liste ve karma yöntemlerinin de ortalaması alınınca Tablo 4.40 elde edilir. KelimeKarma yönteminin tf için %75, tf-idf için %74 ve WLLR için KategoriListe yönteminin %75 olarak bu veri grubunda daha başarılı olduğu gözlenir.

Tablo 4.40: Yetişkin DEP x SAĞ Sınıflandırmasında HV-IV kullanılarak Karma, Liste ve Tekli yöntemlerinin kategori ve kelime olarak doğruluk sonuçlarının ortalamaları.

DEP	tf				tf-idf				WLLR			
Liste/Karma	NB	SMO	J48	Ort	NB	SMO	J48	Ort	NB	SMO	J48	Ort
Kategori Tüm Liste	68	70	63	67	42	42	50	45	76	74	74	75
Kelime Tüm Liste	68	72	65	68	62	64	63	63	72	73	66	70
Kategori Tüm Karma	61	53	59	58	61	53	59	58	55	60	62	59
Kelime Tüm Karma	87	77	60	75	73	90	60	74	81	79	64	74
Karma, Liste, Tekli Ortalama	73	73	63	70	65	65	58	63	70	71	65	69

Tablo 4.41 tablosuna göre bu veri grubu için HS (66, 62, 68) ve KS (67, 63, 69) karşılaştırıldığında sırasıyla %65 ve %66 olarak %1 farkla KS daha iyi sonuç vermiştir.

Tablo 4.41: Yetişkin DEP x SAĞ Sınıflandırmasında HV-IV kullanılarak elde edilen doğruluk sonuçlarının HS ve KS yöntemlerine göre karşılaştırılması.

DEP	tf				tf-idf				WLLR			
HS	NB	SMO	J48	Ort	NB	SMO	J48	Ort	NB	SMO	J48	Ort
KategoriListe_HS	64	66	54	61	41	43	50	45	74	70	70	72
KategoriListeAğırlıklı_HS	68	69	60	66								
KategoriKarmaFrek_HS	63	58	66	62	55	51	58	55	63	68	70	67
KategoriKarmaVarYok_HS	61	53	54	56	61	53	54	56	50	50	50	50
KelimeListe_HS	65	74	56	65	63	68	62	64	72	73	65	70
KelimeListeAğırlıklı_HS	71	70	73	71								
KelimeKarmaFrek_HS	79	63	62	68	71	81	59	70	81	82	61	75
KelimeKarmaVarYok_HS	89	91	58	79	89	91	58	79	81	74	68	74
Ortalama	70	68	60	66	63	65	57	62	70	70	64	68
KS	NB	SMO	J48	Ort	NB	SMO	J48	Ort	NB	SMO	J48	Ort
KategoriListe_KS	68	73	63	68	42	42	50	45	79	77	78	78
KategoriListeAğırlıklı_KS	71	72	71	71								
KategoriKarmaFrek_KS	60	47	72	59	62	53	58	58	58	74	78	70
KategoriKarmaVarYok_KS	62	53	47	54	62	53	47	54	50	50	50	50
KelimeListe_KS	64	70	58	65	64	66	62	64	73	72	66	70
KelimeListeAğırlıklı_KS	72	73	73	71								
KelimeKarmaFrek_KS	88	63	62	71	73	89	60	74	83	83	58	75
KelimeKarmaVarYok_KS	92	90	58	80	92	90	58	80	78	75	67	73
Ortalama	72	68	63	67	66	66	56	63	70	72	66	69

4.2.2.2. HV-IV ile Yetişkin Mani x Sağlıklı Sınıflandırması Sonuçları

- Tekli Yöntem Sonuçları

Bu veri grubunda tf için KelimeTekliAğırlıklı'da ortalama %76 ile, tf ve tf-idf için

KategoriTekliVarYok %76 ile ve WLLR için KelimeTekliAğırlıklı'da %72 ile diğer özellik seçimlerine göre daha iyi sonuç vermiştir (Tablo 4.42).

Tablo 4.42: Yetişkin MAN x SAĞ sınıflandırmasında HV-IV sözlüğü kullanılarak elde edilen Tekli yöntem doğruluk sonuçları.

MAN	tf				tf-idf				WLLR			
Feature	NB	SMO	J48	Ort	NB	SMO	J48	Ort	NB	SMO	J48	Ort
KelimeTekliVarYok	65	72	67	68	65	72	67	68	75	70	65	70
KelimeTekliAğırlıklı	75	77	75	76	62	65	58	62	77	77	63	72
KategoriTekliVarYok	80	82	65	76	80	82	65	76	50	50	50	50
KategoriTekliAğırlıklı	65	88	70	74	52	55	55	54	60	70	63	64

- *Liste ve Karma Yöntem Sonuçları*

Eşik Değer ortalamalarına bakıldığında Tablo 4.43'de görüldüğü gibi tf için Eşik Değer 5 ve Eşik Değer 25 (%68), tf-idf için Eşik Değer 50 (%65) ve WLLR için Eşik Değer 25 (%68) en iyi sonucu vermiştir.

Tablo 4.43: Yetişkin MAN x SAĞ sınıflandırmasında HV-IV kullanarak Karma ve Liste yöntemlerinin doğruluk sonuçlarının Eşik değerlere göre genel ortalamaları.

MAN												
	tf				tf-idf				WLLR			
Eşik Değer	NB	SMO	J48	Ort	NB	SMO	J48	Ort	NB	SMO	J48	Ort
Eşik Değer 0	69	68	63	67	63	65	56	62	68	68	60	66
Eşik Değer 5	69	70	65	68	60	66	56	61	68	68	63	66
Eşik Değer 25	69	69	66	68	61	68	58	62	70	68	66	68
Eşik Değer 50	68	68	65	67	65	69	61	65	65	65	61	64
Ortalama	69	69	65	67	62	67	58	62	68	67	63	66

Tüm liste ve karma sonuçları Ek 15'de verilmesiyle birlikte eşik değerlerin ortalaması alınarak özet tablo çıkarılmak istendiğinde, tüm ağırlıklı, ağırlıksız olarak liste ve karma yöntemlerinin de ortalaması alınınca Tablo 4.44 elde edilmiştir.

Buna göre tf için KategoriListe yöntemi %74, tf-idf için KelimeKarma yöntemi %68 ve WLLR için KategoriListe ve KelimeListe yöntemi %77 olarak bu veri grubunda daha başarılı olduğu gözlenir.

Tablo 4.44: Yetişkin MAN x SAĞ Sınıflandırmasında HV-IV kullanarak Karma, Liste ve Tekli yöntemlerinin kategori ve kelime olarak doğruluk sonuçlarının ortalamaları.

MAN	tf				tf-idf				WLLR			
Liste/Karma	NB	SMO	J48	Ort	NB	SMO	J48	Ort	NB	SMO	J48	Ort
Kategori Tüm Liste	76	75	70	74	52	55	51	53	77	79	76	77
Kelime Tüm Liste	68	72	64	68	64	67	63	65	80	77	74	77
Kategori Tüm Karma	65	62	62	63	66	67	54	62	52	54	52	53
Kelime Tüm Karma	68	67	64	66	65	78	61	68	73	70	61	68
Karma, Liste, Tekli Ortalama	70	74	67	70	63	68	59	64	68	68	63	66

Tablo 4.45 HS ve KS karşılaştırmasını göstermektedir. Buna göre en genel ortalama HS (67, 62, 64) için %64 ve KS (68, 63, 65) için %65 olarak %1 farkla KS daha iyi sonuç vermiştir.

Tablo 4.45: Yetişkin MAN x SAĞ Sınıflandırmasında HV-IV kullanarak yapılan Karma ve Liste yöntemlerinin kategori ve kelime olarak doğruluk sonuçlarının HS ve KS yöntemlerine göre karşılaştırması.

MAN	tf				tf-idf				WLLR			
HS	NB	SMO	J48	Ort	NB	SMO	J48	Ort	NB	SMO	J48	Ort
KategoriListe_HS	75	78	67	73	47	52	48	49	72	74	72	73
KategoriListeAğırlıklı_HS	75	74	71	74								
KategoriKarmaFrek_HS	59	59	56	58	62	64	55	60	57	60	58	58
KategoriKarmaVarYok_HS	64	65	63	64	64	65	63	64	50	50	50	50
KelimeListe_HS	64	67	60	63	63	67	65	65	73	71	62	69
KelimeListeAğırlıklı_HS	71	77	64	71								
KelimeKarmaFrek_HS	65	59	68	64	65	75	62	67	73	71	62	69
KelimeKarmaVarYok_HS	69	76	58	68	69	76	58	68	71	68	63	67
Ortalama	68	69	63	67	62	67	59	62	66	66	61	64
KS	NB	SMO	J48	Ort	NB	SMO	J48	Ort	NB	SMO	J48	Ort
KategoriListe_KS	75	73	70	73	56	56	53	55	82	83	80	82
KategoriListeAğırlıklı_KS	77	75	71	74								
KategoriKarmaFrek_KS	63	53	58	58	70	70	53	64	50	57	50	52
KategoriKarmaVarYok_KS	74	73	71	73	74	73	71	73	50	50	50	50
KelimeListe_KS	65	65	62	63	66	67	61	65	78	75	60	71
KelimeListeAğırlıklı_KS	72	78	70	71								
KelimeKarmaFrek_KS	65	57	73	65	65	81	61	69	78	75	60	71
KelimeKarmaVarYok_KS	72	76	56	68	72	76	56	68	70	67	58	65
Ortalama	70	69	66	68	64	68	57	63	68	68	60	65

4.2.2.3. HV-IV ile Yetişkin Cinsiyet Sınıflandırması Sonuçları

Bu bölümde yetişkin serbest konuşma verilerinin Kadın x Erkek (K x E) sınıflandırılması HV-

IV sözlüğü kullanılarak yapılmıştır. Tekli, Liste ve Karma olmak üzere üç farklı özellik seçimi uygulanmıştır.

- *Tekli Yöntem Sonuçları*

Bu veri grubunda tf için KategoriTekliAğırlıklı ve KelimeTekliAğırlıklı'da ortalama %73 ile, tf-idf için KelimeTekliVarYok %69 ile ve WLLR için KelimeTekliAğırlıklı'da %75 ile diğer özellik seçimlerine göre daha iyi sonuç vermiştir (Tablo 4.46).

Tablo 4.46: Yetişkin K x E sınıflandırmasında HV-IV sözlüğü kullanılarak elde edilen tekli yöntem doğruluk sonuçları.

Cinsiyet	tf				tf-idf				WLLR			
Feature	NB	SMO	J48	Ort	NB	SMO	J48	Ort	NB	SMO	J48	Ort
KelimeTekliVarYok	67	75	67	69	67	75	67	69	50	76	61	62
KelimeTekliAğırlıklı	74	74	71	73	60	72	53	62	80	63	82	75
KategoriTekliVarYok	53	56	53	54	53	56	53	54	50	50	50	50
KategoriTekliAğırlıklı	61	78	81	73	49	55	51	52	39	77	58	58

- *Liste ve Karma Yöntem Sonuçları*

Ek 16'da tüm sonuçlar verilmiştir. Eşik Değer ortalamalarına bakıldığında Tablo 4.47'de görüldüğü gibi tf için Eşik Değer 25 (%63), tf-idf için Eşik Değer 50 (%59) ve WLLR için Eşik Değer 5 (%62) en iyi sonucu vermiştir.

Tablo 4.47: Yetişkin K x E sınıflandırmasında HV-IV kullanılarak Karma ve Liste yöntemlerinin doğruluk sonuçlarının Eşik Değerlere göre genel ortalamaları.

Cinsiyet												
Eşik Değer	tf				tf-idf				WLLR			
	NB	SMO	J48	Ort	NB	SMO	J48	Ort	NB	SMO	J48	Ort
Eşik Değer 0	53	59	61	57	44	59	56	53	60	56	61	59
Eşik Değer 5	60	59	60	60	50	58	53	54	62	60	63	62
Eşik Değer 25	62	62	65	63	54	59	60	58	61	59	62	61
Eşik Değer 50	61	63	63	62	58	59	59	59	58	58	61	59
Ortalama	59	61	62	61	52	59	57	56	60	58	62	60

Tüm ağırlıklı, ağırlıksız olarak liste ve karma yöntemlerinin de ortalaması alınınca Tablo 4.48 elde edilir. Buna göre tf için KelimeListe ve KelimeKarma yöntemi %67, tf-idf için KelimeKarma yöntemi %71 ve WLLR için KelimeListe yöntemi %72 olarak bu veri grubunda daha başarılı olduğu gözlenir.

Tablo 4.48: Yetişkin K x E sınıflandırmasında HV-IV sözlüğü kullanılarak Karma, Liste ve Tekli yöntemlerinin kategori ve kelime olarak doğruluk sonuçlarının ortalamaları.

Cinsiyet	tf				tf-idf				WLLR			
Liste/Karma	NB	SMO	J48	Ort	NB	SMO	J48	Ort	NB	SMO	J48	Ort
Kategori Tüm Liste	51	53	56	54	36	49	46	44	58	56	62	58
Kelime Tüm Liste	67	65	69	67	54	55	44	51	73	68	75	72
Kategori Tüm Karma	51	55	60	55	45	61	66	57	50	50	51	50
Kelime Tüm Karma	68	70	63	67	71	71	71	71	64	63	66	64
Karma, Liste, Tekli Ortalama	61	66	65	64	54	62	56	58	58	63	63	61

Bu veri grubu için HS ve KS karşılaştırıldığında Tablo 4.49'a göre en genel ortalamada HS (59, 56, 59) için %58 ve KS (62, 55, 61) için %59 olarak %1 farkla KS daha iyi sonuç vermiştir.

Tablo 4.49: Yetişkin K x E sınıflandırmasında HV-IV kullanılarak Karma ve Liste yöntemlerinin kategori ve kelime olarak doğruluk sonuçlarının HS ve KS yöntemlerine göre karşılaştırması.

Cinsiyet	tf				tf-idf				WLLR			
HS	NB	SMO	J48	Ort	NB	SMO	J48	Ort	NB	SMO	J48	Ort
KategoriListe_HS	51	51	52	51	45	50	44	46	54	55	59	56
KategoriListeAğırlıklı_HS	54	52	59	55								
KategoriKarmaFrek_HS	52	54	63	56	53	59	60	57	49	48	52	50
KategoriKarmaVarYok_HS	53	55	51	53	53	55	51	53	50	50	50	50
KelimeListe_HS	53	52	60	55	51	50	51	51	73	63	76	71
KelimeListeAğırlıklı_HS	72	67	71	70								
KelimeKarmaFrek_HS	71	66	72	70	70	71	70	70	76	66	79	74
KelimeKarmaVarYok_HS	63	74	61	66	63	74	61	66	50	51	54	52
Ortalama	59	59	61	59	55	57	57	56	59	55	62	59
KS	NB	SMO	J48	Ort	NB	SMO	J48	Ort	NB	SMO	J48	Ort
KategoriListe_KS	47	56	53	52	27	49	47	41	61	57	65	61
KategoriListeAğırlıklı_KS	53	54	62	56								
KategoriKarmaFrek_KS	46	56	75	59	38	64	71	57	53	53	51	52
KategoriKarmaVarYok_KS	53	54	53	53	53	54	53	53	50	50	50	50
KelimeListe_KS	66	69	72	69	57	59	37	51	73	72	74	73
KelimeListeAğırlıklı_KS	75	74	73	74								
KelimeKarmaFrek_KS	71	65	71	69	72	71	72	72	80	79	81	80
KelimeKarmaVarYok_KS	67	72	49	63	67	72	49	63	50	57	49	52
Ortalama	60	63	63	62	48	61	57	55	61	61	62	61

Yetişkin veri grubunda, özellik seçimi yöntemlerinde eşik değerlerine göre alınan ortalamalar ve kelime/kategori varlık/yokluk ve frekans, idf ile ağırlıklandırılmış, WLLR yöntemlerine göre sınıflandırmalar yapılmıştır. Yapılan DEP x SAĞ, MAN x SAĞ ve cinsiyet sınıflandırmalarının NB, SMO, J48 olarak uygulanan MÖ teknikleri ile elde edilen doğruluk

sonuçları sırasıyla Ek 14, Ek 15, Ek 16 Tablo 4.48 ve Tablo 4.49’de gösterilmiştir. Genel olarak özet tablolarındaki en genel ortalamalar karşılaştırıldığında HV-IV kullanıldığında yetişkin depresyon verisinin sonuçları (%70, %63, %64) %66, mani verisi (%70, %64, %61) %65’den daha iyidir, mani verisinin sonuçları ise %59 olan cinsiyet (%64, %58, %54)’ten daha iyi çıkmıştır (Tablo 4.50). HS ve KS yöntemleri arasında ise birbirine yakın sonuçlar elde edilmiştir. KS yöntemi DEP x SAĞ ve MAN x SAĞ sınıflandırmasında %2 ve cinsiyet sınıflandırmasında %1 farkla daha iyi sonuç vermiştir.

Tablo 4.50: Yetişkin verisinde HV-III ve HV-IV karşılaştırması.

	tf				tf-idf				WLLR			
HV-III	NB	SMO	J48	Ort	NB	SMO	J48	Ort	NB	SMO	J48	Ort
Dep Yetişkin	78	76	66	73	71	72	64	69	72	69	63	68
Mani Yetişkin	68	69	62	66	61	65	58	61	67	68	62	66
Cinsiyet Yetişkin	61	65	54	60	54	58	52	55	65	62	55	61
Ortalama	69	70	61	66	62	65	58	62	68	67	60	65
HV-IV	NB	SMO	J48	Ort	NB	SMO	J48	Ort	NB	SMO	J48	Ort
Dep Yetişkin	73	73	63	70	65	65	58	63	60	65	66	64
Mani Yetişkin	70	74	67	70	63	68	59	64	60	63	60	61
Cinsiyet Yetişkin	61	66	65	64	54	62	56	58	49	58	55	54
Ortalama	68	71	65	68	61	65	58	62	56	62	61	60
Kelime	NB	SMO	J48	Ort	NB	SMO	J48	Ort	NB	SMO	J48	Ort
Dep Yetişkin	84	84	66	78	81	82	65	76	79	76	63	73
Mani Yetişkin	69	73	64	68	61	71	58	64	76	73	66	72
Cinsiyet Yetişkin	61	72	49	60	60	65	55	60	67	67	68	68
Ortalama	71	76	59	69	67	73	59	67	74	72	66	71

Yetişkin verisi doğruluk sonuçları MÖ sınıflandırıcıları yönünden karşılaştırıldığında tf için en iyi sınıflandırıcı SMO (%70, %71, %76) olmuştur. En düşük sonuç ise J48’de (%61, %65, %59) alınmıştır. tf-idf için SMO (%65, %65, %73) en iyi sınıflandırıcı olmuştur. WLLR’de ise SMO (%67, %62, %72) en iyi sonucu vermiştir (Tablo 4.50).

Bu çalışmada kullanılan tüm verilerin ve sınıflandırmaların Eşik değerlere göre ortalaması alındığında Eşik Değer 0 (%61, %60, %62) %61, Eşik Değer 5 (%63, %60, %61) %61, Eşik Değer 25 (%65, %62, %59) %62 ve Eşik Değer 50 (%64, %62, %61) % 62 sonucu vermiştir (Tablo 4.51). Buna göre Eşik Değer 25 ve 50 %62 ile Eşik Değer 0 ve 5 %61 ile birbirine yakın sonuçlar vermiştir.

Tablo 4.51: HV-III ve HV-IV ile yapılan sınıflandırmalarda tüm veriler için genel Eşik Değer ortalamaları.

	tf				tf-idf				WLLR			
	0	5	25	50	0	5	25	50	0	5	25	50
HV-III	62	65	67	65	63	62	64	64	65	62	60	62
HV-IV	60	61	63	62	58	58	59	60	59	60	59	59
Ortalama	61	63	65	64	60	60	62	62	62	61	59	61

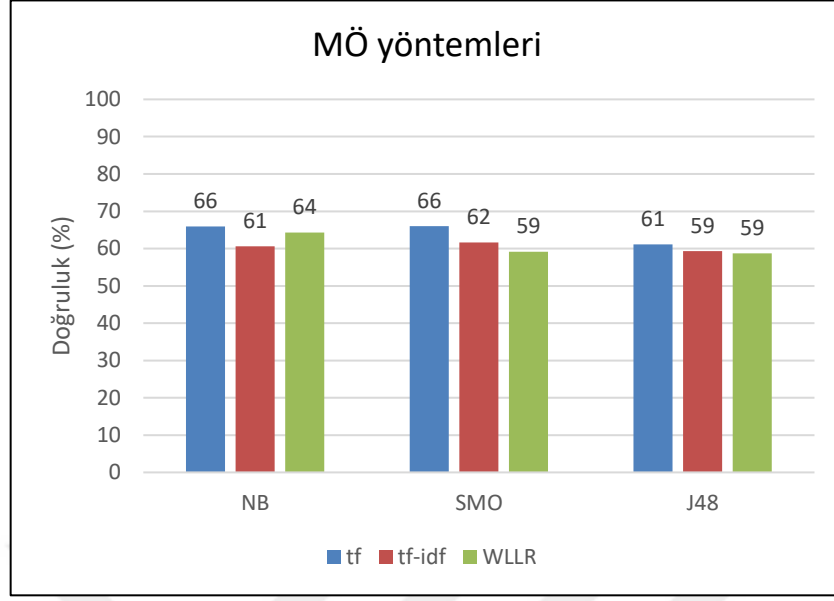
HS-KS karşılaştırmalarında ise genellikle KS yöntemi %1-%2 daha iyi sonuç vermiş ya da aynı kalmıştır (Tablo 4.52).

Tablo 4.52: HV-III ve HV-IV ile yapılan sınıflandırmalarda tüm veriler için genel HS ve KS ortalamaları.

Veri Türü	HS	KS
HV-III Çocuk	63	64
HV-IV Çocuk	58	59
HV-III Yetişkin	64	66
HV-IV Yetişkin	62	67
Ortalama	67.7	68.2

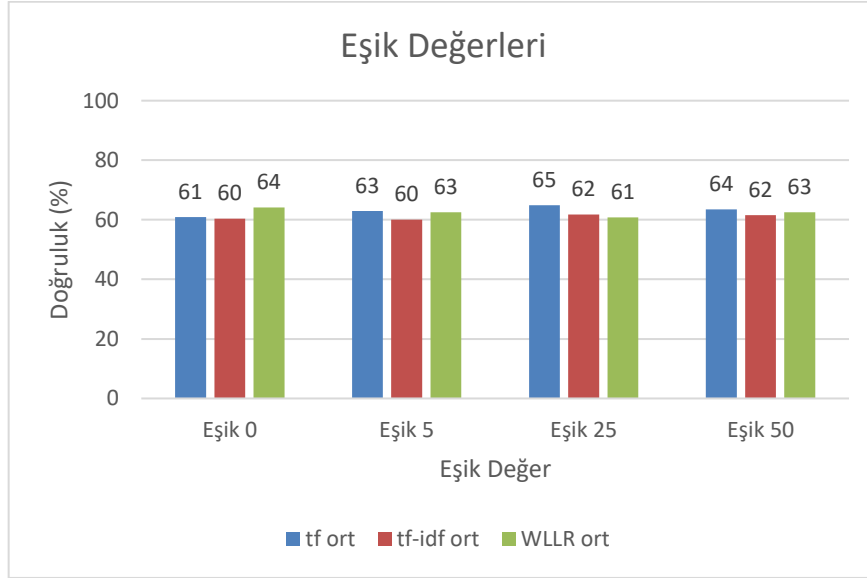
Sınıflandırma yapılırken eğitim verisi için ön işleme bölümünde her bir deneğe ait metinlerin morfolojik analizleri ve belirsizlik giderme işlemleri yapılır. Özellik seçimi için dokümanlar birer vektör halinde temsil edilir. Bu aşamada kelime veya kategori özellik olarak alınmış, var-yok bilgisi, frekansları (tf) ve ağırlıklı frekansları, vektör oluşturmada kullanılmıştır. Ağırlıklı frekansları hesaplanırken tf-idf hesaplamaları kullanılmıştır. WLLR ile ayrıca özellik seçimi yapılarak sınıflandırmalar yapılmıştır. Özellik metoduna göre oluşturulan bu vektörler biraraya geldiğinde eğitim verisi tablosu oluşmaktadır. Eğitim verisine katılmayan diğer metinler ise aynı vektör temsili ile test verisini oluşturur. Sınıflandırmalarda Weka programı kullanılmış ve NB, SMO, J48 olmak üzere 3 farklı MÖ algoritması uygulanmıştır. Şekil 4.1 MÖ yöntemlerine göre değişen doğrulukları göstermektedir. Buna göre NB en iyi sonuçları vermiştir. SMO ise tf ve tf-Idf'de NB ile birbirine yakın sonuç vermiştir ancak WLLR'de düşük sonuç vermiştir. J48 ise daha düşük değerler almıştır. Metin sınıflandırmada SMO ve NB'in daha iyi sonuç verdiği önceki metin sınıflandırma çalışmalarında da bahsedilmiştir.

Veri gruplarına bakıldığında tf kullanıldığında doğruluk sonuçları daha yüksek çıkmaktadır. WLLR ve tf-idf karşılaştırıldığında WLLR sonucu NB'de daha iyi çıkmıştır. tf-Idf sonucu SMO'da daha iyi çıkmıştır (Şekil 4.1).



Şekil 4.1: MÖ yöntemlerine göre doğruluk sonuçları.

Eşik değer yönünden bakılırsa tf kullanıldığında en iyi eşik değeri %25 görülmektedir. WLLR için en iyi eşik değeri 0 ve tf-idf için ise %25 ile %50 daha uygun değerlerdir (Şekil 4.2). Buna göre en iyi yöntem tf ve eşik değeri %25 olmuştur. Liste eleme yöntemi olarak ise KS seçilmelidir Tablo 4.52.

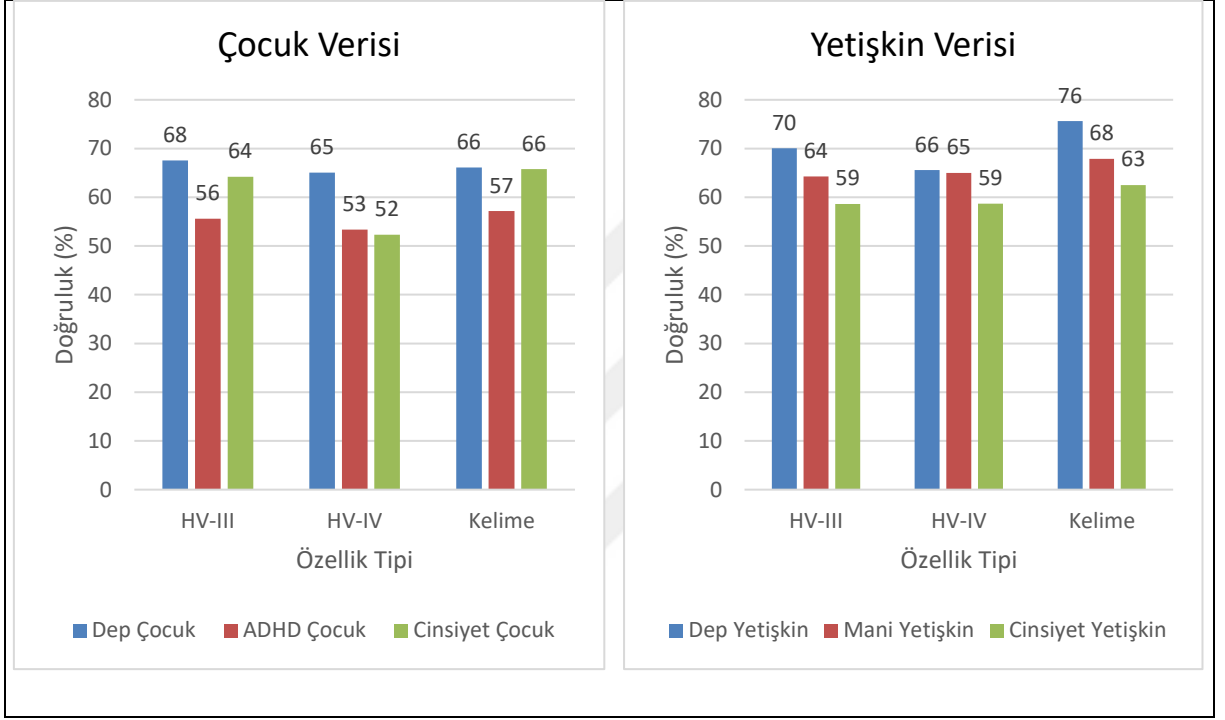


Şekil 4.2: Eşik değerlere göre doğruluk sonuçları.

Veri tiplerinde sözlüklere göre doğruluk sonuçlarına bakıldığında genel olarak HV-III kullanımı daha iyi sonuç vermiştir (Şekil 4.3). Yetişkin verisinde çocuk verisine oranla her iki

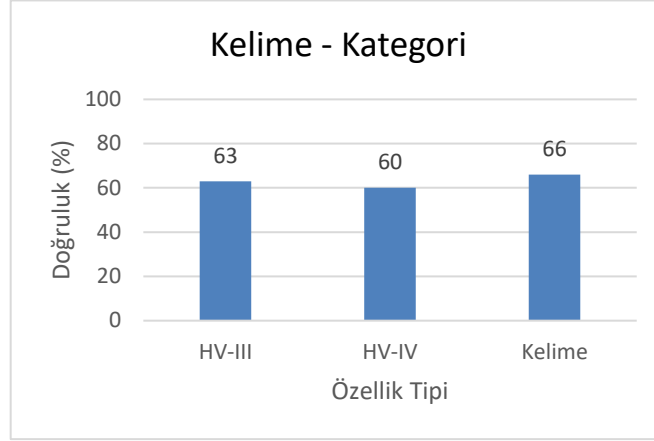
sözlükte de daha iyi sonuç alınmıştır. Çocuk cinsiyet sınıflandırması HV-III kullanıldığında çocuk verisi %9 farkla daha iyidir (Şekil 4.3).

Hastalıklara göre karşılaştırıldığında depresyon verisi yetişkinde ve çocukta her zaman en iyi sonucu vermiştir. Mani verisi de ona yakındır. En düşük sonucu ADHD verisi vermiştir (Şekil 4.3).



Şekil 4.3: Veri tiplerinde özelliklere göre doğruluk sonuçları.

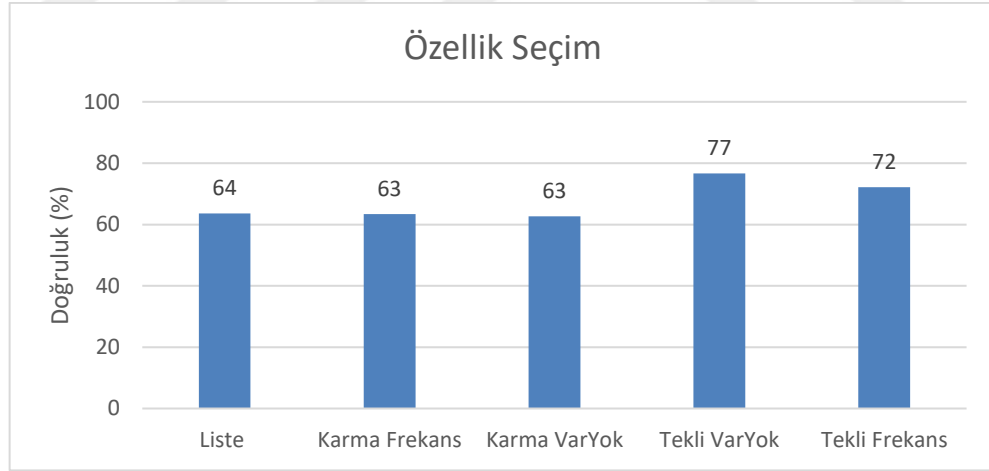
Kelime ve Kategori yöntemleri karşılaştırıldığında, tüm yöntemlerin ortalaması alınınca Şekil 4.4'deki grafik elde edilmiştir. Buna göre kelime yöntemi daha iyi sonuç vermiştir. Kategori özellik seçildiğinde ise HV-III daha iyi sonuç vermiştir.



Şekil 4.4: Kelime ve Kategori karşılaştırması.

Özellik seçim yöntemleri karşılaştırıldığında Şekil 4.5’de görüldüğü gibi tekli yöntem diğer yöntemlere göre daha iyi sonuç vermiştir.

Frekans ve var-yok olarak karşılaştırılırsa var-yok frekansa göre daha iyi sonuç vermiştir. Bunun sebebi olarak tekli sistemde daha fazla özellik kullanılmasının doğruluğun arttırmasına etkisi olduğu düşünülebilir.



Şekil 4.5: Özellik seçim metotlarının doğruluk sonucu karşılaştırmaları

4. 3. ÜÇLÜ SINIFLANDIRMA BULGULARI

Yetişkin verileri Depresyon x Mani x Sağlıklı olarak üçlü sınıflandırılmıştır. Her üç grubun da test verileri üç kere değiştirilerek sonuçlarının ortalamaları alınmıştır. Sınıflandırma işlemi veri madenciliğinde yaygın olarak kullanılan Weka programında MÖ metotları kullanılarak yapılmıştır (Witten ve Frank, 2005). MÖ metotlarından NB, J48 ve SMO kullanılmıştır.

Veri özellikleri Tablo 3.11 ve Tablo 3.12’de verilmiştir. Her gruptan 9’ar denek metni test için ayrılmıştır. Test metinleri 3 kez değiştirilip sınıflandırmalar yapılarak 3’lü çapraz geçerlilik uygulanmıştır. Sonuç olarak bu sınıflandırmaların ortalaması alınmıştır.

Tekli (BOW), Liste ve Karma (BOW) yöntemlerinin NB, SMO, J48 makine öğrenmesi teknikleri uygulanarak elde edilen doğruluk sonuçları tablolar halinde verilmiştir. Her bir sınıflandırmada özellik vektörünün oluşturma şekline göre tf, tf-Idf ve WLLR olarak sonuçlar düzenlenmiştir. Frekans listeleri üç tane olacağı için eşik değer ve KS-HS yöntemleri kullanılmamış, onun yerine listelerde aynı olan kelimeler varsa sadece frekans değeri büyük olan listesinde bırakılmış, diğer iki listeden bu kelime çıkarılmıştır. Buna göre alınan sonuçlar Yetişkin ve Çocuk olarak bu bölümde verilmiştir. Sonuçlar yüzde olarak ve yuvarlatılarak verilmiştir.

4.3.1. HV-III ile Yetişkin Üçlü Sınıflandırması Sonuçları

- Tekli Yöntem Sonuçları

Bu veri grubunda genel olarak en iyi sonuçları özellik olarak KelimeTekliVarYok ve KategoriTekliAğırlıklı, NB ve SMO MÖ tekniklerinde vermiştir. tf, tf-Idf ve WLLR için KelimeTekliVarYok yöntemi SMO’da %70 ile iyi sonuç vermiştir (Tablo 4.53). KelimeTekliAğırlıklı ise tf ve WLLR özellik seçimi kullanıldığında SMO’da %71 ile en iyi sonucu vermiştir.

Tablo 4.53: Yetişkin DEP x MAN x SAĞ sınıflandırmasında HV-III sözlüğü kullanılarak elde edilen Tekli yöntem doğruluk sonuçları.

Yetişkin	tf				tf-idf				WLLR			
Özellik	NB	SMO	J48	Ort	NB	SMO	J48	Ort	NB	SMO	J48	Ort
KelimeTekliVarYok	68	70	40	59	68	70	40	59	68	70	40	59
KelimeTekliAğırlıklı	57	71	42	57	58	59	37	51	55	71	43	56
KategoriTekliVarYok	51	47	49	49	51	47	49	49	51	47	49	49
KategoriTekliAğırlıklı	62	63	37	54	33	33	32	33	68	64	44	59

- Liste ve Karma Yöntem Sonuçları

Liste ve karma yöntemlerinde tüm ağırlıklı, ağırlıksız olarak liste ve karma yöntemlerinin de ortalaması alınınca Tablo 4.54 elde edilir. tf ve WLLR için KelimeListe yöntemi %70 ve tf-Idf için KelimeListe yöntemi %63 ortalama ile bu veri grubunda daha başarılı olduğu gözlenir.

Tablo 4.54: Yetişkin DEP x MAN x SAĞ Sınıflandırmasında HV-III kullanılarak Karma, Liste ve Tekli yöntemlerinin kategori ve kelime olarak doğruluk sonuçlarının ortalamaları.

Yetişkin	tf				tf-idf				WLLR			
Liste/Karma	NB	SMO	J48	Ort	NB	SMO	J48	Ort	NB	SMO	J48	Ort
Kategori Tüm Liste	70	70	69	70	57	53	47	52	70	70	69	70
Kelime Tüm Liste	66	67	56	63	63	68	57	63	56	48	34	46
Kategori Tüm Karma	47	48	40	45	57	58	44	53	50	48	44	47
Kelime Tüm Karma	57	55	37	49	52	69	38	53	54	62	37	51
Karma, Liste, Tekli Ortalama	60	61	46	56	55	57	43	52	59	60	45	55

Tablo 4.55'e göre bu veri grubu için KategoriListe ve KategoriListeAğırlıklı yöntemleri tf ve WLLR için en iyi sonucu vermiştir.

Tablo 4.55: Yetişkin DEP x MAN x SAĞ Sınıflandırmasında HV-III kullanılarak elde edilen doğruluk sonuçları.

Yetişkin	tf				tf-idf				WLLR			
	NB	SMO	J48	Ort	NB	SMO	J48	Ort	NB	SMO	J48	Ort
KategoriListe	71	69	69	70	57	53	47	52	70	70	69	70
KategoriListeAğırlıklı	69	70	69	70								
KategoriKarmaFrek	46	53	35	44	57	58	44	53	52	53	43	49
KategoriKarmaVarYok	48	43	44	45	48	43	44	45	48	43	44	45
KelimeListe	64	67	60	64	63	68	57	63	56	48	34	46
KelimeListeAğırlıklı	68	66	52	62								
KelimeKarmaFrek	48	41	41	43	52	69	38	53	43	57	40	46
KelimeKarmaVarYok	65	68	33	56	65	68	33	56	65	68	33	56
Ortalama	60	60	50	57	57	60	44	54	56	57	44	52

4.3.2. HV-IV ile Yetişkin Üçlü Sınıflandırması Sonuçları

- Tekli Yöntem Sonuçları

Bu veri grubunda tf için KelimeTekliVarYok yöntemi NB ve SMO'da diğer özellik seçimlerine göre daha iyi sonuç vermiştir (Tablo 4.56).

Tablo 4.56: Yetişkin DEP x MAN x SAĞ sınıflandırmasında HV-IV sözlüğü kullanılarak elde edilen Tekli yöntem doğruluk sonuçları.

Yetişkin	tf				tf-idf				WLLR			
Özellik	NB	SMO	J48	Ort	NB	SMO	J48	Ort	NB	SMO	J48	Ort
KelimeTekliVarYok	68	69	40	59	68	70	40	59	68	70	40	59
KelimeTekliAğırlıklı	57	73	42	57	58	59	37	51	55	71	43	56
KategoriTekliVarYok	53	53	48	52	51	47	49	49	51	47	49	49
KategoriTekliAğırlıklı	50	56	42	49	33	33	33	33	68	64	44	59

- *Liste ve Karma Yöntem Sonuçları*

Liste ve karma yöntemlerinde tüm ağırlıklı, ağırlıksız olarak liste ve karma yöntemlerinin de ortalaması alınınca Tablo 4.57 elde edilmiştir. KelimeListe yönteminde tf ve tf-Idf için NB ve SMO'nun daha başarılı olduğu gözlenir.

Tablo 4.57: Yetişkin DEP x MAN x SAĞ Sınıflandırmasında HV-IV kullanılarak Karma, Liste ve Tekli yöntemlerinin kategori ve kelime olarak doğruluk sonuçlarının ortalamaları.

Yetişkin	tf				tf-idf				WLLR			
Liste/Karma	NB	SMO	J48	Ort	NB	SMO	J48	Ort	NB	SMO	J48	Ort
Kategori Tüm Liste	58	60	59	59	38	41	43	41	57	62	59	59
Kelime Tüm Liste	66	66	58	63	62	68	57	62	56	46	33	45
Kategori Tüm Karma	42	38	39	40	45	45	28	39	37	38	37	37
Kelime Tüm Karma	60	52	38	50	57	67	36	53	56	61	35	50
Karma, Liste, Tekli Ortalama	57	58	46	54	52	54	40	48	56	57	43	52

Tablo 4.58'e göre bu veri grubu için tf ve tf-Idf'de KelimeListe ve KelimeListeAğırlıklı yöntemleri NB ve SMO'da iyi sonuçlar vermiştir. KelimeKarmaVarYok metodu da tf, tf-Idf ve WLLR için NB ve SMO'da iyi sonuçlar vermiştir.

Tablo 4.58: Yetişkin DEP x MAN x SAĞ Sınıflandırmasında HV-IV kullanılarak elde edilen doğruluk sonuçları.

Yetişkin	tf				tf-idf				WLLR			
	NB	SMO	J48	Ort	NB	SMO	J48	Ort	NB	SMO	J48	Ort
KategoriListe	57	59	53	56	38	41	43	41	57	62	59	59
KategoriListeAğırlıklı	59	60	64	61								
KategoriKarmaFrek	44	32	38	38	45	45	28	39	35	32	33	33
KategoriKarmaVarYok	39	44	41	41	48	43	44	45	39	44	41	41
KelimeListe	66	68	62	65	62	68	57	62	56	46	33	45
KelimeListeAğırlıklı	67	64	54	62								
KelimeKarmaFrek	52	41	43	45	57	67	36	53	43	59	37	46
KelimeKarmaVarYok	68	63	32	54	65	68	33	56	68	63	32	54
Ortalama	57	54	48	53	53	55	40	49	50	51	39	47

4.3.3. HV-III ile Çocuk Üçlü Sınıflandırması Sonuçları

- *Tekli Yöntem Sonuçları*

Bu veri grubunda genel olarak sonuçlar düşük çıkmıştır. En iyi sonuçları özellik olarak KategoriTekliVarYok için %54 doğruluk sonucuyla NB metodu vermiştir. tf, tf-Idf ve WLLR için aynıdır (Tablo 4.59).

Tablo 4.59: Yetişkin DEP x ADHD x SAĞ sınıflandırmasında HV-III sözlüğü kullanılarak elde edilen Tekli yöntem doğruluk sonuçları.

Çocuk	tf				tf-idf				WLLR			
Özellik	NB	SMO	J48	Ort	NB	SMO	J48	Ort	NB	SMO	J48	Ort
KelimeTekliVarYok	45	44	34	41	45	44	34	41	45	44	34	41
KelimeTekliAğırlıklı	37	41	38	39	42	29	33	35	43	39	37	40
KategoriTekliVarYok	54	30	35	39	54	30	35	39	54	30	35	39
KategoriTekliAğırlıklı	49	42	28	40	40	38	38	39	48	42	39	43

- *Liste ve Karma Yöntem Sonuçları*

Liste ve karma yöntemlerinde tüm ağırlıklı, ağırlıksız olarak ortalaması alınınca Tablo 4.60 elde edilir. Sonuçlar genel olarak düşük çıkmıştır. Bu veri grubunda en iyi sonucu tf için KelimeListe yöntemi SMO sınıflandırıcı kullanıldığında %48 ile almıştır.

Tablo 4.60: Çocuk DEP x ADHD x SAĞ sınıflandırmasında HV-III kullanılarak Karma, Liste ve Tekli yöntemlerinin kategori ve kelime olarak doğruluk sonuçlarının ortalamaları.

Çocuk	tf				tf-idf				WLLR			
Liste/Karma	NB	SMO	J48	Ort	NB	SMO	J48	Ort	NB	SMO	J48	Ort
Kategori Tüm Liste	44	43	41	43	46	42	38	42	39	45	46	43
Kelime Tüm Liste	47	48	31	42	39	39	32	37	34	42	32	36
Kategori Tüm Karma	38	37	31	35	45	39	38	41	43	34	36	38
Kelime Tüm Karma	39	41	35	38	42	34	38	38	43	39	36	39
Karma, Liste, Tekli Ortalama	44	41	34	40	44	37	36	39	43	39	37	40

Tablo 4.61'e göre bu veri grubu için KelimeListeAğırlıklı yöntemi tf için SMO'da %48 ile en iyi sonucu vermiştir.

Tablo 4.61: Çocuk DEP x ADHD x SAĞ Sınıflandırmasında HV-III kullanılarak elde edilen doğruluk sonuçları.

Çocuk	tf				tf-idf				WLLR			
	NB	SMO	J48	Ort	NB	SMO	J48	Ort	NB	SMO	J48	Ort
KategoriListe	44	44	41	43	46	42	38	42	39	45	46	43
KategoriListeAğırlıklı	44	41	42	42								
KategoriKarmaFrek	33	43	28	35	45	39	38	41	43	38	38	40
KategoriKarmaVarYok	43	30	35	36	43	30	35	36	43	30	35	36
KelimeListe	47	47	26	40	39	39	32	37	34	42	32	36
KelimeListeAğırlıklı	47	48	37	44								
KelimeKarmaFrek	34	37	36	36	42	34	38	38	42	34	38	38
KelimeKarmaVarYok	43	44	34	41	43	44	34	41	43	44	34	41
Ortalama	42	42	35	40	43	38	36	39	41	39	37	39

4.3.4. HV-IV ile Çocuk Üçlü Sınıflandırması Sonuçları

- Tekli Yöntem Sonuçları

Bu veri grubunda genel olarak sonuçlar düşüktür. En iyi değer tf için KategoriTekliAğırlıklı yöntemi NB’de %50 olmuştur (Tablo 4.62).

Tablo 4.62: Çocuk DEP x ADHD x SAĞ sınıflandırmasında HV-IV sözlüğü kullanılarak elde edilen Tekli yöntem doğruluk sonuçları.

Çocuk	tf				tf-idf				WLLR			
Özellik	NB	SMO	J48	Ort	NB	SMO	J48	Ort	NB	SMO	J48	Ort
KelimeTekliVarYok	43	41	31	39	43	41	31	39	43	41	31	39
KelimeTekliAğırlıklı	37	40	39	39	42	29	28	33	43	39	33	38
KategoriTekliVarYok	38	23	41	34	38	23	41	34	38	23	41	34
KategoriTekliAğırlıklı	50	37	40	42	37	37	24	33	47	41	36	41

- Liste ve Karma Yöntem Sonuçları

Liste ve karma yöntemlerinde tüm ağırlıklı, ağırlıksız olarak liste ve karma yöntemlerinin de ortalaması alınınca Tablo 4.63 elde edilir. KategoriListe tf-Idf için NB %49 ile en iyi sonucu almıştır.

Tablo 4.63: Çocuk DEP x ADHD x SAĞ Sınıflandırmasında HV-IV kullanılarak Karma, Liste ve Tekli yöntemlerinin kategori ve kelime olarak doğruluk sonuçlarının ortalamaları.

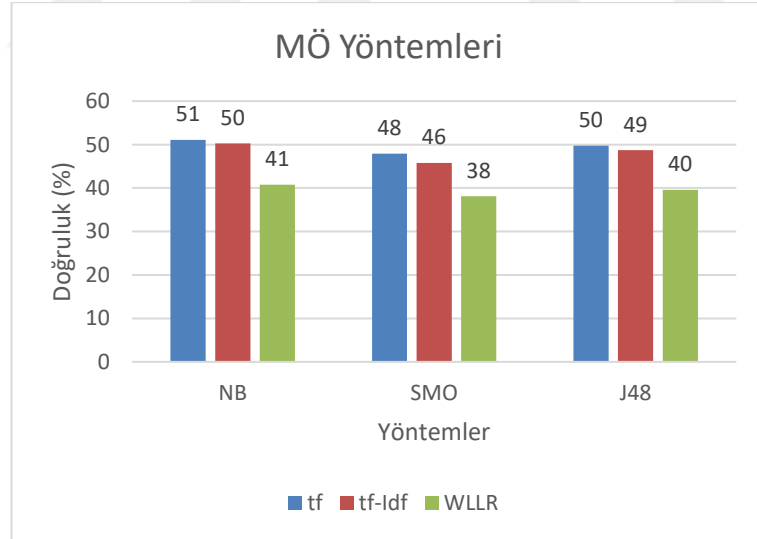
Çocuk	tf				tf-idf				WLLR			
Liste/Karma	NB	SMO	J48	Ort	NB	SMO	J48	Ort	NB	SMO	J48	Ort
Kategori Tüm Liste	44	43	38	42	49	41	43	44	45	45	31	40
Kelime Tüm Liste	48	48	29	42	41	38	38	39	36	42	31	36
Kategori Tüm Karma	30	31	38	33	42	34	34	36	32	35	37	34
Kelime Tüm Karma	40	39	35	38	39	40	29	36	43	36	33	37
Karma, Liste, Tekli Ortalama	41	38	36	38	41	35	34	37	41	38	34	37

Yöntemlere tek tek bakıldığında ise Tablo 4.64’e göre bu veri grubu için tf-Idf’de KategoriListe yöntemi NB’de %49 ile en iyi sonucu vermiştir.

Tablo 4.64: Çocuk DEP x ADHD x SAĞ Sınıflandırmasında HV-IV kullanılarak elde edilen doğruluk sonuçları.

Çocuk	tf				tf-idf				WLLR			
	NB	SMO	J48	Ort	NB	SMO	J48	Ort	NB	SMO	J48	Ort
KategoriListe	44	39	41	41	49	41	43	44	45	45	31	40
KategoriListeAğırlıklı	43	47	36	42								
KategoriKarmaFrek	24	32	42	33	41	38	38	39	28	40	38	36
KategoriKarmaVarYok	35	29	35	33	35	29	35	33	35	29	35	33
KelimeListe	48	48	22	39	39	40	29	36	36	42	31	36
KelimeListeAğırlıklı	47	48	35	44								
KelimeKarmaFrek	34	37	38	36	42	34	34	36	39	30	34	34
KelimeKarmaVarYok	46	41	31	40	46	41	31	40	46	41	31	40
Ortalama	40	40	35	39	42	37	35	38	38	38	33	37

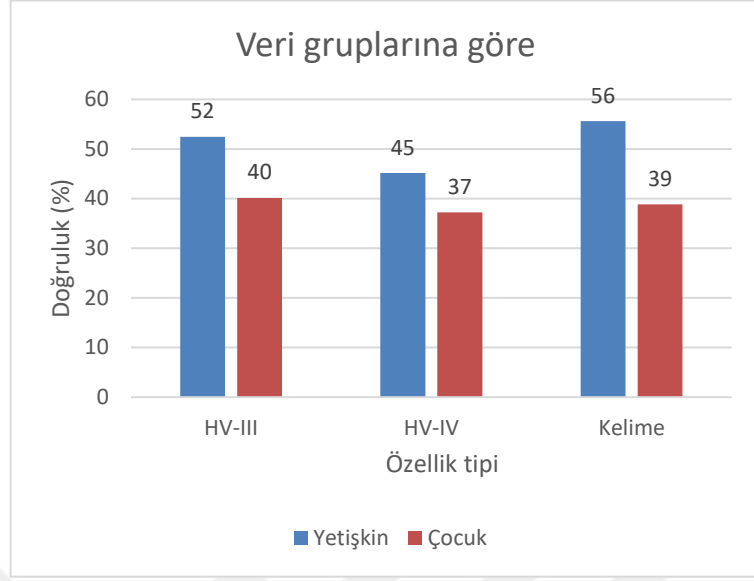
Sınıflandırmalarda Weka programı kullanılmış ve NB, SMO, J48 olmak üzere 3 farklı MÖ algoritması uygulanmıştır. Şekil 4.6 MÖ yöntemlerine göre değişen doğrulukları göstermektedir. Buna göre ikili sınıflandırmada da olduğu gibi NB en iyi sonuçları vermiştir. J48 ise tf, tf-Idf ve WLLR’de NB ile birbirine yakın sonuçlar vermiştir. SMO ise daha düşük değerler almıştır.



Şekil 4.6: MÖ yöntemlerine göre üçlü sınıflandırma doğruluk sonuçları.

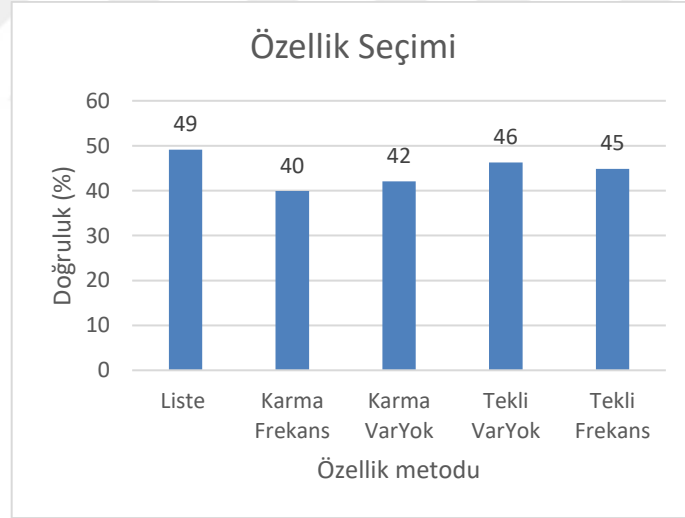
Veri gruplarına bakıldığında tf kullanıldığında doğruluk sonuçları tüm sınıflandırma yöntemlerinde daha yüksek çıkmaktadır. WLLR ise en düşük sonuçları vermiştir (Şekil 4.6).

Yetişkin ve Çocuk verisi üçlü karşılaştırma sonuçlarına göre karşılaştırıldığında yetişkin verisi her özellik tipinde en iyi sonucu vermiştir. En iyi sonuç *Kelime* özellik seçildiğinde alınmıştır. Daha sonra ise HV-III kategorilerinde iyi sonuç alınmıştır. (Şekil 4.3, Şekil 4.7).



Şekil 4.7: Veri tiplerinde özelliklere göre üçlü sınıflandırma doğruluk sonuçları.

Üçlü sınıflandırmada özellik seçim yöntemleri karşılaştırıldığında Şekil 4.8’de görüldüğü gibi liste yöntemi diğer yöntemlere göre daha iyi sonuç vermiştir.



Şekil 4.8: Özellik seçim metotlarının üçlü sınıflandırma sonucu karşılaştırmaları

5. TARTIŞMA VE SONUÇ

Çalışmamızda psikiyatrik rahatsızlığı olan ve olmayan kişilerin konuşma metinleri üzerinde duygu kategorileri de özellik olarak katılarak bir sınıflandırma yapılması hedeflenmiştir. Bu amaçla 14 farklı özellik çıkarım metodu ile otomatik sınıflandırma yapan bir uygulama programı geliştirilmiştir. Bu uygulama programı ilgili klasörlerden alınan eğitim ve test dataları ile bir model hazırlayıp, verilen yeni bir metni sınıflandırabilmektedir. Değişik verilerde yapılan ikili ve üçlü sınıflandırmaların sonuçları verilmiştir.

3 çocuk ve 3 yetişkin olmak üzere 6 tip veri seti üzerinde özellik seçimleri yapılmış ve en iyi metot belirlenmeye çalışılmıştır. Bu bölümde ise elde edilen sonuçlar karşılaştırılacaktır.

Çocuk verileri yönlendirilmiş verilerdir ve psikiyatrik testlerdeki sorulara çocukların verdikleri cevaplardan oluşmaktadır. Yetişkin metinleri ise serbest konuşma metinleridir. 10'ar dakikalık serbest konuşma verileri tüm yetişkin deneklerden uzman psikiyatristler tarafından alınmıştır.

Özellik seçim metotları temelde kelime veya kategorilere dayanmaktadır. Kelimeler morfolojik analiz sonucu elde edilen kelime köklerinden oluşmaktadır. Kategoriler ise kelime köklerinin psikolojik sözlüklerden elde edilen kategorileridir. HV-III ve HV-IV olarak iki farklı psikolojik sözlük kullanılmıştır. Her iki sözlük için ayrı sınıflandırmalar yapıp sonuçları alınmıştır.

Ayırt ediciliği yüksek özelliklerin bulunması için yapılan Idf ve WLLR denemelerinde sonuçlara bakıldığında tf yönteminin dokümanları daha iyi temsil ettiği söylenebilir. İkili sınıflandırma üçlü sınıflandırmaya göre daha başarılıdır.

Psikolojik sözlüklere göre en iyi değerleri veren yöntemlerin bulunması amacıyla doğruluk sonuçlarına bakıldığında genel olarak HV-III kullanımı daha iyi sonuç vermiştir. En iyi sonuç KelimeTekliVarYok metodunda, NB, SMO ve J48 ortalaması olarak tf için %86, idf için %86 ve WLLR için %75 ile yetişkin depresyon verisinde alınmıştır. Ayrıca yine aynı veride KelimeKarmaVarYok_KS sonuçları da benzer şekilde tf için %85, idf için %85 ve WLLR için %73 olarak en iyi sonucu vermiştir. Bu veride sadece NB sınıflandırma değerlerine bakıldığında ise doğruluk tf ve tf-idf'de %94'e kadar çıkmıştır. Cinsiyet verisine bakıldığında ise en iyi değer yetişkin verisindedir. HV-III ile yapılan KelimeKarmaFrek_KS yönteminde SMO sınıflandırmasında %89 sonucu alınmıştır. HV-III sözlüğünün daha iyi sonuç vermesinin

sebebi Türkçeye çevirisinin psikiyatri uzmanı tarafından yapılmış olmasıdır. HV-IV çevirisi ise bizim tarafımızdan birebir kelime karşılığı alınarak yapılmıştır.

Veri tiplerine göre sonuçlar değerlendirildiğine ise yetişkin verisinde çocuk verisine oranla her iki sözlük için de daha iyi sonuç alınmıştır. Çocuklar duygularını gizleme ihtiyacı hissetmedikleri için çocuk verisinde daha iyi sonuçlar beklenmiştir. Ancak verinin yönlendirilmiş olmasından dolayı sonuçlar düşük çıkmıştır. Yetişkin verisi serbest konuşma olduğu için çocuk verisine göre daha iyi sonuç alınmıştır.

Ayrıca ADHD'nin ikili sınıflandırma sonuçlarının düşük olması da çocuk sonuçlarının ortalamasını düşürmüştür. ADHD sonuçlarının düşük olması ise bize ADHD'nin kullanılan dile fazla yansımadığını veya ADHD rahatsızlığı olan kişilerin beraberinde başka rahatsızlıklarının da olması sebebiyle ADHD'nin etkisinin gölgelenmiştir.

Genel olarak ikili sınıflandırma değerleri üçlü sınıflandırmaya göre daha iyi çıkmıştır. Veri sayısının az olması bunda etken olmuştur. Psikiyatrik veri toplanması diğer verilere göre daha zor olduğu için zamanla sayının artırılması ile sonuçlar iyileşecektir. Üçlü ve ikili sınıflandırmada yetişkin verisinde nispeten iyi sonuçlar alınmasının iki sebebi bulunmaktadır. Birincisi serbest konuşma verisinde kişilerin duygularını daha açık etmesi, ikincisi ise çocuk hastalıklarının birbirinden ayrıştırılmasının zor olmasıdır. Çünkü bazı çocuklarda ADHD rahatsızlığı ile depresyon bir arada bulunmaktadır.

Yeni nesil kendisini daha çok elektronik ortamda ifade ettiğinden, bu çalışma önemli bir etkiye sahiptir. Türkçe metin analizinde DDİ yöntemlerini psikolojik metin analizinde kullanan az sayıda çalışma bulunmaktadır. MÖ yöntemlerini kullanarak yapılan zihinsel bozuklukların otomatik sınıflandırılması çalışmasıdır. Bu alandaki yetişkin psikiyatrisinde MÖ yöntemlerinin kullanıldığı ilk sınıflandırma çalışmasıdır. Çocuk psikiyatrisinde ise türkçe üzerine daha önce yapılmış bir çalışma bulunmamaktadır.

Doğruluk sonuçlarını arttırmak için gelecekte MÖ tekniklerinin parametreleri üzerinde değişikliklerin sonuçlara etkisi incelenebilir. Bu çalışmada sunulan *Karma* özellik seçim metodu geliştirilerek daha iyi sonuçlar elde edilebilir. Metin üzerinde ek çalışmalar yapılabilir. Örneğin sadece kelime kökü değil yapım ekleri, zaman bilgisi, cümle öğelerinin bilgisi alınabilir. Denek sayısının artırılmasının da doğruluk sonuçlarını arttıracığı düşünülmektedir. Kategori bazında ise Türkçede yapılmış bir orijinal psikolojik sözlüğe ihtiyaç vardır. Sözlük

hazırlanırken sosyolog, psikolog ve Türkçede uzman kişilerin katkıları gereklidir. Sözlük oluşturulurken kelimelerin anlam olarak toplumda daha sık kullanılan kategorileri belirlenip sözlükte kelimenin karşısındaki ilk sıralara yerleştirilebilir. Bu çalışmamızda kullandığımız HV-III ve HV-IV İngilizceden çeviri sözlüklerdir. Zor bir metin sınıflandırma şekli olan psikolojik metinlerde sınıflandırma sonuçları ümit verici olmuştur.



KAYNAKLAR

- Abadan, N. (1961). Cumhuriyet ve Ulus Gazeteleri Hakkında Muhteva Analizi. *SBF dergisi*, 16.cilt, , No:2, s. 93-188.
- Abbasi, A., ve Chen, H.,. (2007). Affect Intensity Analysis of Dark Web Forums. *Intelligence and Security Informatics* (s. 282-288). New Brunswick: IEEE.
- Abdul-Mageed, M., Diab M., Kübler, S. (2014). SAMAR: Subjectivity and sentiment analysis for Arabic social media. *Computer Speech & Language*, 28(1), 20-37.
- Abelson, R., ve Schank, R. C. (1977). *Scripts, plans, goals and understanding: An inquiry into human knowledge structures* (1 b.). New Jersey: Lawrence Erlbaum Associates.
- Akın, A. A., ve Akın, M. D. (2007). *Zemberek, an open source NLP framework for Turkic languages*. Amazon S3 : https://s3.amazonaws.com/academia.edu.documents/34521696/zemberek_makale.pdf?AWSAccessKeyId=AKIAIWOWYYGZ2Y53UL3A&Expires=1523873817&Signature=fT6o1Vz8h5ws%2FalcRvxD3%2FB2eto%3D&response-content-disposition=inline%3B%20filename%3DZemberek_makale.pdf, [Ziyaret Tarihi: 16 Nisan 2018]
- Albayrak, N. B. (2011). *Opinion and Sentiment Analysis Using Natural Language Processing Techniques* Master Thesis. İstanbul: Fatih University.
- Alşen, S. (2012). *Çocuklar İçin Depresyon Değerlendirme Ölçeği-Gözden Geçirilmiş Formunun Türkçe Uyarlamasının Türk Populasyonundaki Ergenlerde Psikometrik Özellikleri*, Doktora(Uzmanlık) Tezi. İzmir: 9 Eylül Üniversitesi Tıp Fakültesi.
- Aman, S., ve Szpakowicz, S. (2008). Using Roget's Thesaurus for Fine-grained Emotion Recognition. *International Institute of Information Technology*. Prague: IJCNLP.
- Amasyalı M.F., ve Diri B. (2006). Automatic Turkish Text Categorization in Terms of Author, Genre and Gender. *NLDB 2006: Natural Language Processing and Information Systems, Lecture Notes in Computer Science* , 3999(1), 221-226.
- APA. (1994). *Diagnostic and statistical manual of mental disorders (DSM IV)* (4 b.). Washington: American Psychiatric Association.
- APA. (2013). *The definition of attention-deficit/hyperactivity disorder (ADHD) fact sheet*, American Psychiatric Association, American psychiatric publishing. www.dsm5.org/documents/adhdfactsheet.pdf, [Ziyaret Tarihi: 3 Ocak 2016].
- Atkinson, J. W., ve McClelland, D. C.,. (1948). The Effect of Different Intensities of the Hunger Drive on Thematic Apperception. *Journal of Experimental Psychology*, 38(1), 643-658.
- Avanci, J., Assis, S., Oliveira, R. ve Pires, T. (2012). Childhood Depression. Exploring the association between family violence and other psychosocial factors in low-income Brazilian schoolchildren. *Child and Adolescent Psychiatry and Mental Health*, 6, 26.
- Baysal, C. ve Tekarslan, E. (1996). *İşletmeler için davranış bilimleri*. İstanbul: Avcıol Basım-Yayın.
- Beier, H., ve Hanfmann, E. (1956). Emotional attitudes of former Soviet citizens, as studied by the technique of projective questions,. *The Journal of Abnormal and Social Psychology*, 53(2), 143-153.

- Berelson, B. (1952). *Content Analysis in Communication Research*. Illinois: The Free Press.
- Berger, A.L., Pietra, V.J.D., ve Pietra, S.A.D. (1996). A maximum entropy approach to natural language processing. *Computational linguistics*, 22(1), 39-71.
- Boynukalın, Z. (2012). *Emotion Analysis Of Turkish Texts By Using Machine Learning*, Master Thesis. Ankara: Middle East Technical University.
- Carbonell, J. G. (1979). *Subjective Understanding: Computer Models of Belief Systems*, Doctoral Dissertation. Yale: Yale University.
- Chomsky, N. (1957). Logical structures in language. *Association for Information Science and Technology*, 8(9), 141-150.
- Chowdhury, G. G. (2005). Natural language processing. *Information Science and Technologoy*, 51-89.
- Cohen, A. S., Minor, K. S., Baillie, L. E., Dahir, A. M. (2008). Clarifying the Linguistic Signature: Measuring Personality from Natural Speech. *Journal of Personality Assessment*, 90(6), 559-563.
- Colace, F., Santo, M.D., Greco, L., Napoletano, P. (2014). Text classification using a few labeled examples. *Computers in Human Behavior*, 30, 689-697.
- Cortes, C. ve Vapnik, V. (1995). Support-Vector Networks. *Machine Learning, Springer Link*, 20(3), 273-297.
- Eroğul, U. (2009). *Sentiment Analysis in Turkish*. Ankara: Yüksek Lisans Tezi, The Graduate School of Natural and Applied Sciences of Middle East Technical University.
- Erol, A., Mete, L., Doğaner, İ., Sorias, S. (1995). Bilgisayarlı içerik çözümleme yönteminin psikiyatrik tanıda kullanımı. *Türk Psikiyatri Dergisi*, 4(1), 263-268.
- Fox, S., Zickuhr, K., Smith, A. (2009). *19% of Internet Users on Twitter or Another Service*. <http://pewresearch.org/pubs/1385/who-uses-twitter-tweets>, Pew Research Center, [Ziyaret Tarihi: 12 Nisan 2014].
- Fukuhara, T., Nakagawa, H., ve Nishida, T. (2007). Understanding sentiment of people from news articles: Temporal sentiment analysis of social events. *Proceedings of the International Conference on Weblogs and Social Media*. Colorado: ICWSM'2007.
- Fürnkranz, J. (1998). *A Study Using n-gram features for Text Categorization*. Technical Report OEFAI-TR-98-30. Vienna: Austrian Research Institute for Artificial Intelligence.
- Gottschalk, L. A., ve Gleser, G. C. (1969). *The Measurement of Psychological States Through the Content Analysis of Verbal Behavior*. Berkeley: University of California Press.
- Gökğöz, E., Kurt, A., Kulamshayev, K., Kara, M.,. (2011). Two-Level Qazan Tatar Morphology. *1st International Conference on Foreign Language Teaching and Applied Linguistics*. Sarayev: International Burch University.
- Güran, A., Akyokuş, S., Bayazıt, N.G., Gürbüz, M.Z. (2009). Turkish Text Categorization Using N-Gram Words. *International Symposium on Innovations in Intelligent Systems and Applications, INISTA 2009*. Trabzon.
- Helzer, J. E. (2009). DSM categories and dimensions in clinical and research concept. *Dimensional approaches in diagnostic classification: refining the research agenda for DSM-V*, s. 6, Arlington Virginia: American Psychiatric Pub.

- Hurwitz, R. (2002). *Harvard University*. General Inquirer Home Page: <http://www.wjh.harvard.edu/~inquirer/Home.html>, [Ziyaret Tarihi: 6 Ocak 2014].
- Jin, X., Li, Y., Mah, T., Tong, J.,. (2007). Sensitive Webpage Classification for Content Advertising. *Proceedings of the First International Workshop on Data Mining and Audience Intelligence for Advertising, California*, 28-33.
- John, G. H., ve Langley, P. (1996). Estimating Continuous Distributions in Bayesian Classifiers. *Proceedings of the Eleventh Conference on Uncertainty in Artificial Intelligence* (s. 338-345). San Mateo: Morgan Kaufmann Publishers.
- Jurafsky, D. & Martin., J.H. (2009). *Speech and Language Processing: An introduction to natural language processing, computational linguistics, and speech recognition*. (2 b.). New Jersey: Prentice Hall.
- Kaya, M., Fidan, G. ve Toroslu, I.H. (2012). Sentiment Analysis of Turkish Political News. *Proceedings of the 2012 IEEE/WIC/ACM international joint conferences on web intelligence and intelligent agent technology, IEEE Computer Society*, 1(1), 174-180.
- Kessler, R. C., Aguilar-Gaxiola, S., Alonso, J., Chatterji, S., Lee, S., Ormel, J., and Wang, P. S. (2009). The global burden of mental disorders: an update from the WHO World Mental Health (WMH) surveys. *Epidemiologia e psichiatria sociale*, 18-01, 23-33.
- Kessler, R.C., Walters, E.E., ve Forthofer, M.S. (1998). The Social Consequences of Psychiatric Disorders, III: Probability of Marital Stability. *American Journal of Psychiatry*, 155(8), 1092-1096.
- Kim, S. M., Hovy, E. (2007). Crystal: Analyzing Predictive Opinions on the Web. *Proceedings of the 2007 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning*. Prague: USC Information Sciences Institute.
- Kotsiantis, S. B. (2007). Supervised Machine Learning: A Review of Classification. *Informatica*, 31(1), 249-268.
- Kovacs, M. (1981). Rating scales to assess depression in school-aged children. *Acta Paedopsychiatrica: International Journal of Child & Adolescent Psychiatry*, 46(5), 305-315.
- Lewis, D. (1991). *Representation and Learning in Information Retrieval, UMCS-1991-093*. Amherst, MA: Department of Computer Science, University of Massachusetts(UMCS).
- Li, S., Xia, R., Zong, C., Huang, C. (2009). A Framework of Feature Selection Methods for Text Categorization. Singapur: ACL '09 Proceedings of the Joint Conference of the 47th Annual Meeting of the ACL and the 4th International Joint Conference on Natural Language Processing of the AFNLP.
- Liu, J., Cao, Y., Lin, C.Y., Huang, Y., Zhou, M.,. (2007). Low Quality Product Review Detection in Opinion Summarization. *Proceedings of the 2007 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning, Praque*, 334-342.
- Lomax, J. W. (2015). *Encyclopedia Britannica*. Psychiatry: <https://global.britannica.com/topic/psychiatry>, [Ziyaret Tarihi: 12 Nisan 2016].
- McCauley, E., Pavlidis, K., ve Kendall, K. (2001). Developmental precursors of depression: The child and the social environment. *The depressed child and adolescent*, 2 ed., 46-78.

- Mete, L., Schnurr, P.P., Rosenberg, S.D., Oxman, T.E., Doğaner, İ., Sorias, S. (1993). Language Content and Schizophrenia in Acute Phase Turkish Patients. *Social Psychiatry and Psychiatric Epidemiology*, 28(1), 275-280.
- Mladenec, D. ve Grobelnik, M. (1998). Word sequences as features in text learning. *In Proceedings of the 17th Electrotechnical and Computer Science Conference, ERK-98*, Ljubljana, Slovenia.
- Montague, J. C. (1971). *The Effect of Institutionalization on the Social Behavior and Language of Mentally Retarded Children*, Dissertation, p 93. Florida: University of Florida.
- Namenwirth, J.Z. Weberi, R.P. (1987). *Lasswell Dictionary*. Dynamics of Culture, s. 46-53. Boston: Allen and Unwin.
- Ng, V., Dasgupta, S. ve Arifin, S.M.N. . (2006). Examining the Role of Linguistic Knowledge Sources in the Automatic Identification and Classification of Reviews. Dallas: *Proceedings of the COLING, ACL 2006*.
- Nigam, K., McCallum, A., Thrun, S., ve Mitchell, T.,. (2000). Text classification from labeled and unlabeled documents using EM. . *Machine Learning*, 39(2), 103-134.
- Oflazer, K. (1994). Two-level description of Turkish morphology. *Literary and Linguistic Computing*, 9(2), 137-148.
- Orhan, Z., Domnori, E., Giriş, S. F. ve Ceyhan, M. (2014). Customer Satisfaction Measurement Tool by Analysing Turkish Product Reviews. *TURKLANG'2014*, 1(1), 15-23.
- Oxman, T.E., Rosenberg, S.D., Schnurr, P.P., Tucker, G.J. (1988). Diagnostic classification through content analysis of patients' speech. *American Journal of Psychiatry*, 145(4), 464-468.
- Öy, B. (1991). Çocuklar için depresyon ölçeği: Geçerlik ve güvenirlik çalışması. *Türk Psikiyatri Dergisi*, 2(2), 132-136.
- Pang, B., ve Lee, L. (2008). Opinion Mining and Sentiment Analysis. *Foundations and Trends in Information Retrieval*, 2(1-2), 1-135.
- Pang, B., Lee, L. ve Vaithyanathan, S. . (2002). Thumbs up? Sentiment classification using machine learning techniques. (s. 79-86). Philadelphia: In Proceedings of the 2002 Conference on Empirical Methods in Natural Language Processing (EMNLP).
- Parekh, R. (2015a). APS (American Psychiatric Association). Anxiety Disorders: <https://www.psychiatry.org/patients-families/anxiety-disorders/what-are-anxiety-disorders>, [Ziyaret Tarihi: 12 Mayıs 2016].
- Parekh, R. (2015b). APS (American Psychiatric Association), Bipolar Disorders. Bipolar Disorders: <https://www.psychiatry.org/patients-families/bipolar-disorders/what-are-bipolar-disorders>, [Ziyaret Tarihi: 12 Mart 2016].
- Pearl, J. (1998). Graphical Models for Probabilistic and Causal Reasoning. Vol.1f, *Handbook of Defeasible Reasoning and Uncertainty Management Systems* (s. 367-389). Dordrecht: Springer.
- Pennebaker, J. W. (2007). Current issues and new directions in Psychology and Health: Listening to what people say--the value of narrative and computational linguistics in health psychology. *Psychology & Health, American Psychological Association*, 22(6), 631-635.
- Pennebaker, J. W., Francis, M. E., ve Booth, R. C. (2001). *Linguistic Inquiry and Word Count (LIWC): A Computerized Text Analysis Program* (2 b.). NJ: Erlbaum.

- Platt, J. (1999). *Fast training of support vector machines using sequential minimal optimization*, Schölkopf, B., Burges, C., and Smola, A. (Eds.) (1 b.). Advances in Kernel Methods - Support Vector Learning: MIT Press.
- Platt, J. C. (1998). *Sequential Minimal Optimization: A Fast Algorithm for Training Support Vector Machines*. Microsoft Research: Technical Report MSR-TR-98-14.
- Powell, D.L., Powell, R.G. (2016). *Classroom Communication and Diversity* (3 b.). New York: Routledge.
- Purcell, K., Rainie, L., Mitchell, A., Rosenstiel, T., ve Olmstead, K. (2010). *Understanding the Participatory News Consumer*. <http://pewresearch.org/pubs/1508/internet-cell-phone-users-news-social-experience>, Pew Internet and American Life Project, [Ziyaret Tarihi: 14 Mayıs 2016].
- Quinlan, J. R. (1993). *C4.5: Programs for Machine Learning* (1st b.). San Francisco: Morgan Kaufmann Publishers, Inc.
- Rosenberg, S. D., Schnurr, P. P., Oxman, T. E. (1990). Content analysis: A comparison of manual and computerized systems, *Journal of personality assessment*, 54(1-2), 298-310.
- Rosenschein, S. J., ve Shieber, S. M. (1982). Translating English into logical form. In Proceedings of the 20th annual meeting on Association for Computational Linguistics, pp. 1-8. . Menlo Park, CA: Association for Computational Linguistics.
- Saleem, S., Prasad, R., Vitaladevuni, S. N. P., Pacula, M., Crystal, M., Marx, B., Sloan, D., Vasterling, J. ve Speroff, T. (2012). Automatic Detection of Psychological Distress Indicators and Severity Assessment from Online Forum Posts. *Proceedings of COLING 2012: Technical Papers*, 2375-2388.
- Schnurr, P. P., Rosenberg, S. D., ve Oxman, T. E. (1992). Comparison of TAT and free speech techniques for eliciting source material in computerized content analysis. *Journal of Personality Assessment*, 58, 311-325.
- Semin, G. R., Fiedler, K. (1988). The cognitive functions of linguistic categories in describing persons: Social cognition and language. *Journal of personality and Social Psychology*, 54(4), 558.
- Shylov, M. (2008). *DİLMAÇ: Turkish and Turkmen Morphological Analyzer and Machine Translation Program*. Master thesis. İstanbul: Fatih University.
- Sindhu, R., Jandail, R. R. S., Kumar, R. R., (2014). A Novel Approach for Sentiment Analysis and Opinion Mining. *International Journal of Emerging Technology and Advanced Engineering (IJETA)*, www.ijetae.com, 4(4).
- Stone, P. J., Dunphy, D. C., Smith, M. S., ve Ogilvie, D. M. (1966). *The General Inquirer: Computer Approach to Content Analysis*. Cambridge: M.I.T. Press.
- Süyel, Ö. (2002). *Tureng İngilizce-Türkçe Online Sözlük*. <http://tureng.com>, [Ziyaret Tarihi: 9 Aralık 2014].
- Şeker, Ş. E. (2012). *Bilgisayar Kavramları*. <http://bilgisayarkavramlari.sadievrenseker.com/2012/11/13/c4-5-agaci-c4-5-tree/>, [Ziyaret Tarihi: 21 Mayıs 2018].

- Tavşancıl, E., Aslan, E. (2001). *Content Analyse and Application Examples* (1 b.). İstanbul: Epsilon Yayıncılık.
- Tomkins, S. S., ve Tomkins, E. J. (1947). *The Thematic Apperception Test: The theory and technique of interpretation*. New York City, NY, US: Grune & Stratton, xi 297 pp. doi: 10.1037/11186-000.
- Tucker, G.J., ve Rosenberg, S. D. (1975). Computer content analysis of schizophrenic speech: A preliminary report. *The American Journal of Psychiatry*, 132(6), 611-616.
- Turing, A. M. (1950). Computing Machinery and Intelligence. *Oxford University Press*, 59(236), 433-460.
- Vapnik, V. (1982). Estimation of Dependences Based on Empirical Data. *Springer-Verlag*, 40.
- Viney, L. L. (1983). The assessment of psychological states through content analysis of verbal communications. *Psychological Bulletin*, 94(3), 542-563.
- Weathers, F. W., Keane, T. M., Davidson, J. R. T. . (2001). Clinician-Administered PTSD Scale: A review of the first ten years of research. *Depression and Anxiety*, 13(3), 132-156.
- Weizenbaum, J. (1966). ELIZA – A computer program for the study of natural language communication between man and machine. *Communications of the ACM*, 9(1), 36-45.
- WHO (2013). *Premature death among people with severe mental*.
www.who.int/mental_health/management/info_sheet.pdf: World Health Organization,
[Ziyaret Tarihi: 12 Mayıs 2016].
- Wilks, Y. ve Bien, J. (1983). Beliefs, Points of View, and Multiple Environments. *Cognitive Science*, 7(2), 95-119. doi:10.1207/s15516709cog0702_1
- Winograd, T. (1971). *Procedures as a representation for data in a computer program for understanding natural language*. No. MAC-TR-84. CAMBRIDGE: MASSACHUSETTS INST OF TECH PROJECT MAC.
- Wittchen, H. U. (2002). Generalized anxiety disorder: prevalence, burden, and cost to society. *Depression and anxiety*, 16(4), 162-171.
- Witten, I. H. ve Frank, E. (2005). *Data Mining: Practical Machine Learning Tools and Techniques* (2 b.), San Francisco: Morgan Kaufmann.
- World Health Organization. (2014). *Regional Office for Europe. RC63 fact sheet on mental health*.
http://www.euro.who.int/_data/assets/pdf_file/0004/215275/RC63-Fact-sheet-MNH-Eng.pdf,
[Ziyaret Tarihi: 12 Nisan 2016].
- World Health Organization Regional Office for Europe. (2016). *Mental Disorders*.
<http://www.who.int/mediacentre/factsheets/fs396/en/>, [Ziyaret Tarihi: 12 Mart 2016].
- Yüret, D., ve Türe, F. (2006). Learning morphological disambiguation rules for Turkish. New York: Proceedings of the Human Language Technology Conference of the North American Chapter of the ACL. pages 328–334, Association of Computational Linguistics.
- Boynukalın, Z. ve Karagöz, P., (2013). Emotion Analysis on Turkish Texts. *Lecture Notes in Electrical Engineering. Information Sciences and Systems*. 264. Springer, Cham.

Zafer, H.R., Tilki, B., Kurt, A., Kara, M. (2011). Two-level description of kazakh morphology. Sarayevo: *1st International Conference on Foreign Language Teaching and Applied Linguistics*, International Burch University.



EKLER

EK 1. Psikiyatrik hastalıkların DSM-IV'e göre sınıflandırması (APA, 1994)

1. Disosiyatif Bozukluk

- Disosiyatif Bozukluk

2. Anksiyete Bozuklukları

- OKB
- Panik Bozukluk
- PTSS
- Stresli Yaşam
- Anksiyete Bozukluk
- Karma Anksiyete
- Travmatik Bozukluk

3. Duygu Durum Bozuklukları

- Afektif Bozukluk
- Duygulanım Bozukluğu

Bipolar Bozukluklar

- Bipolar Bozukluk
- BAB
- Bipolar Duygulanım Bozukluğu

Depresif Bozukluk

- Distimi
- Distimik Bozukluk
- Depresyon
- Depresif Bozukluk
- Depresif Atak

4. Başka Bir Yerde Sınıflandırılmamış Dürtü Denetim Bozuklukları

- Patolojik Kumar Oynama, Impulsite

5. Genellikle İlk Kez Çocukluk veya Ergenlik Döneminde Tanısı Konan Bozukluklar

Dikkat Eksikliği ve Yıkıcı Davranış Bozuklukları

- Erişkin ADHD
- Davranış Bozukluğu

Tik Bozuklukları

- Tourette

6. Madde Kullanımı ile ilişkili Bozukluklar

- Psikoaktif Madde Kötü Kullanımına Bağlı Zihin Davranış Bozukluğu

Kannabisle İlişkili Bozukluklar

- Esrar Kötüye Kullanımı

Alkol Kullanım Bozuklukları

- Alkol Bağımlılığı
- Alkol Kötüye Kullanımı

7. Kişilik Bozuklukları

- Obsesif Kişilik
- Paranoid Narsistik Kişilik
- Paranoid Obsesif Kişilik
- Narsistik Kişilik

A kümesi

- Antisosyal Kişilik Özellikleri

B kümesi

- B Kümesi Kişilik Özellikleri

C kümesi

- BTA
- Çekingen Kişilik özellikleri
- C Kümesi Kişilik Özelliği

8. Somatoform Bozukluklar

- Hipokondriak Bozukluk
- Konversif Bayılma
- Somatoform Bozukluğu
- Konversiyon

9. Şizofreni Ve Diğer Bozukluklar

- Şizoafektif Bozukluk
- Psikotik Bozukluk

10. Uyum Bozuklukları

- Uyum Bozukluğu

- Depresif Mizaçlı Uyum Bozukluğu
- Anksiyete Mizaçlı Uyum Bozukluğu
- Dürtü Kontrol Bozukluğu

EK 2. Harward-III sözlüğündeki kategoriler

1. Sınıf kategoriler (Varlıklar, Süreçler, Niteleyiciler)

İlk sınıf kategorilere daha yakından bakacak olursak, varlıklar dünyasının 3 ana gruba ayrıldığı görülüyor; sosyal, kültürel ve doğal gruplar. Sosyal evren olarak *kişilere*, *sosyal rollere* ve *topluluklara* işaret ediliyor. SELF kategorisi birinci tekil şahısa, SELVES birinci çoğul şahısa, OTHER ise sen, siz, onlar ve türevlerine işaret eder. Bu sözlüğün amacına uygun olarak o (he, she) ve türevleri erkek rolü (MALE-ROLE) ve kadın rolü (FEMALE-ROLE) altında sınıflandırılmıştır.

Örnek kelime: “Abla” 5, 68 ve 60 nolu kategorilere girmektedir.

Kategori 5: Female role (Kadın rolü)

Kategori 68: Higher-Status (Yüksek statü)

Kategori 60: Family (Aile)

Sosyal Evren

Kişiler

SELF: (Kendi) Tüm birinci tekil şahıs zamirleridir. (ben, beni, bana, kendim)

SELVES: (Kendileri) Tüm birinci çoğul şahıs zamirleri. (biz, bize, bizimki)

OTHER: (Diğer) Tüm ikinci ve üçüncü tekil ve çoğul şahıs zamirleri. (sen, siz, sizi, onlar, onlarınki)

Roller

MALE ROLE: (Eril rol) Tüm erkek rol referansları. (aktör, abi, damat)

FEMALE ROLE: (Dişil rol) Tüm kadın rol referansları. (anne, abla, aktris, teyze, gelin, kız, peri)

NEUTER ROLE: (Nötr rol) Cinsiyet içermeyen rol ve meslek isimleri. (bebek, Türk, kimse, çocuk)

JOB ROLE: (İş rolü) Teorik olarak iki cinsiyete de açık meslek isimleri. (ajan, artist, sanatçı, yazar, kaptan)

Topluluklar

SMALL GROUP: (Küçük grup) Üyelerinin genellikle yüz yüze iletişim yapabildiği gruplar. (ajans, bando, kurul, klüp)

LARGE GROUP: (Büyük grup) Yüz yüze iletişim için çok büyük olan gruplar. (ordu, diyanet, halk)

Kültürel Evren

Kültürel varlıklar

FOOD: (Yiyecek) Yiyecek tipleri ya da yiyecek isimleri. (bakla, çay, şeker, kiraz)

CLOTHING: (Kıyafet) Kıyafet isimleri ya da türleri. (düğme, elbise, kürk)

TOOL: (Araç) Araç-gereç isimleri, el yapımı eşyalar. (otomobil, çanta, ambulans)

Kültürel ortamlar

SOCIAL PLACE: (Sosyal yer) Binalar ve bina bölümleri. Politik sosyal ve ekonomik yerler. (yurtdışı, Türkiye, Amerika, oda, kabin)

Kültürel örüntüler

IDEAL VALUE: (İdeal değerler) Kültürel olarak tanımlanan erdemler, hedefler, değer biçilen koşullar ve aktiviteler. (kabiliyet, yapabilme, imkan, güzellik, cesaret)

DEVIATION VALUES: (Sapma değerleri) Kültürel olarak olumsuz tanımlanan hedefler, koşullar ve aktivite türleri. (anormal, deli, sarhoş)

ACTION-NORM: (Eylem normu) Sosyal davranışların kuralsal örüntüleri. (anlaşma, iş, komisyon, kredi)

MESSAGE-FORM: (Mesaj-formu) İletişim ortamı isimleri, para ve sanat nesnelerini de içerir. (sanat, kitap, nakit)

THOUGHT-FORM: (Düşünce-formu) Mantık türleri ve birimleri. (soyutlama, temel, zıtlık, tahmin)

NONSPESIFIC-OBJECTS: (Belirsiz isimler) Varlıklara soyut referanslar. Entelektüel çağrışımlar. (iş, olay, farklı yönler, sermaye, ayrıntı)

Örnek kelime: “Yunanistan” 20 ve 64 nolu kategorilere girmektedir.

Kategori 20: Social Place (Sosyal alan)

Kategori 64: Political (Politik)

Doğal Evren

BODY PART: (Beden ögesi) Beden öğeleri. (kol, beden, beyin, yanak)

NATURAL OBJECTS: (Doğal varlıklar) İnsanoğlu tarafından yapılmamış objeler. (bitkiler, hayvanlar, mineraller, balık)

NATURALWORLD: (Doğal dünya) Coğrafik yerler, hava durumu ile alakalı ya da evrensel objeler. (hava, sahil, koy, çayır)

Örnek kelime: “Kanat” 80 ve 10 nolu kategorilere girmektedir.

Kategori 80: Ascend Theme (Yükselme ile ilgili tema)

Kategori 10: Body Part (Beden ögesi)

Niteleyiciler

İsim veya fillerin önüne gelen niteleyici sözcükler.

SENSORY-REFERENCE (QUALITY): (Duyusal-referans) Kokular, renkler, tatlar. (siyah, taze, sesli)

QUANTITY-REFERENCE: (Miktar-Referans) Miktar bildiren birim ve ölçüler. (eklemek, arttırmak, büyük, hiç, tam)

SPACE-REFERENCE: (Uzay-Referans) Hacimsel, uzaysal boyutlar. (arka, ön, eğim)

TIME-REFERENCE: (Zaman-Referans) Zaman ifade eden kelimeler. (sonra, öğle, nisan, sonbahar)

Örnek kelime: “Sonsuz” 17 ve 83 nolu kategorilere girmektedir.

Kategori 17: Time-Reference (Zaman ifade eden)

Kategori 83: Death-Theme (Ölüm ile ilgili tema)

Süreçler

Fiillerin sınıflandırılması için kullanılan etiketlerdir. IF, EQUAL, NOT, CAUSE, GOOD, BAD ve OUGHT dâhil değildir. IF, EQUAL, NOT ve CAUSE, düşünce süreçlerini göstermekten ziyade, öncelikle ifade etmek anlamındadır. GOOD, BAD ve OUGHT ise değerlendirici ve ahlaki süreçleri gösterir. Çoğu zaman bu süreçler fiil değildir.

Psikolojik süreçler

Duygular:

AROUSAL: (Uyarılma) Duygusal heyecan durumları. (tutum, uyanık, hissi, alışkanlık)

URGE: (Dürtüsel) Güdüseler. Rüya, hırs, özen, kasıt gibi.

AFFECTION: (Sevgi) Yakınlık ve karşılıklı pozitif ilişki gösteren işaretler. (hayranlık, hoşnutluk, göz alıcı, sevgili, flört)

PLEASURE: (Memnuniyet) Hoşnutluk durumları. (neşe, sevinç, eğlendirici)

DISTRESS: (Sıkıntı) Endişe, üzüntü, korku, suçluluk, ayıplanma, keder, başarısızlık, kararsızlık durumları. (korkmak, panik, kırgınlık, aksilik)

ANGER: (Kızgınlık) Agresif ifadeler. (kızgın, öfke, sinirlilik, nefret)

Düşünce:

SENSE: (Duyu) Algı ve farkındalık. (görünmek, katılım, farkında, okumak)

THINK: (Düşünme) Kavramsal süreçler. (farz etmek, seçim, şüphe, fikir)

IF: (Eğer) Koşullu kelimeler. (neredeyse, şans, değilse, eğer)

EQUAL: (Eşit) Benzerlik ifade eden kelimeler. (benzer, aynı, ibaret)

NOT: (Değil) Olumsuzluk ifade eden kelimeler. (yapamamak, beceriksizlik, değil, fark, hiç, yok).

CAUSE: (Sebeup) Sebeup-sonuç ilişkisi ifade eden kelimeler. (etki, sebeup)

Değerlendirme

GOOD: (İyi) İyinin eşanlamları. (hayran olunan, takdire değer, beğenilen, temiz, adil, uygun)

BAD: (Kötü) Kötünün eşanlamları. (berbat, acı, adi, kaba, ham)

OUGHT: (Zorunluluk) Yapılması doğru olan zorunluluklar. (görev, zorunluluk, münasip)

Örnek kelime: “Vazife” 42 ve 81 nolu kategorilere girmektedir.

Kategori 42: Ought (Zorunluluk)

Kategori 81: Authority-Theme (Otoriteyle ilgili tema)

Davranış Süreçleri

Sosyal duygusal eylemler

COMMUNICATE: (İletişim) Anlam ileten süreçler. (söylev, hitap, kabul, cevap, övünmek)

APPROACH: (Yaklaşım, tutum.) Bir yine doğru hareket ifade eden kelimeler. (ulaşmak, varmak, getirmek)

GUIDE: (Yönlendirme) Yardım ve pozitif yönlendirme. (yardım, izin vermek, faydalanmak).

CONTROL: (Kontrol) Hareketi sınırlandırmak, denetlemek. (tutuklamak, atamak, bağlamak, kısıtlamak)

ATTACK: (Saldırmak) Zarar verici düşmanca hareketleri ifade eden kelimeler. (bunaltmak, saldırı, dövmek, aldatmak)

AVOID: (Kaçınmak) Bir şeyden uzaklaşma hareketlerini ifade eden kelimeler. (terk etmek, mevcut değil, örtmek).

FOLLOW: (Takip etmek) Uyma hareketlerini ifade eden kelimeler. (özür, fikir birliği, anlaşma)

DEFENCE MECHANISM: (Savunma Mekanizması) Savunma ile ilgili kelimeler. (cephe, defans, direnç, müdafaa)

Araçsal eylemler: Bir amaca ulaşmak için yapılan eylemler

ATTEMPT: (Girişim) Hedef doğrultusunda yapılan ve çaba içeren hareketleri ifade eden kelimeler. (amaç, uygulamak, teklif, çaba)

WORK: (İş, çalışma) Yapılacak iş(görev) etkinliği. (ayarlamak, bina etmek, inşa etmek, pişirmek, oluşturmak, yaratmak)

MOVE: (Hareket) Hareket ifade eden kelimeler. (atlamak, çevirmek, dalgalanmak)

GET: (Almak) Sahip olan, tüketen anlamında kelimeler. (sahip, işgal, kilit)

POSSES: (Sahiplenmek) Elde etmek, tüketmek anlamında kelimeler. (erişmek, sağlamak, elde etmek)

EXPEL: (Defetmek, atmak) Kovma, çıkartma ifade eden kelimeler. (atmak, dökmek, tuvalet, düşüş)

Örnek kelime: “Vida” 54, 67, 13 ve 77 nolu kategorilere girmektedir.

Kategori 54: Work (İş)

Kategori 67: Technological (Teknolojik)

Kategori 13: Sign-Reject (Red belirtisi)

Kategori 77: Male-Theme (Erkek ile ilgili tema)

2.Sınıf kategoriler (Kurumsal, Psikolojik, Mevkiler)

Kurumsal alan: Kurallar ve hareketlerin sosyal bağlamının tanımlaması

ACADEMIC: (Akademik) Ödev, doğru, öğretmek, editör.

ARTISTIC: (Artistik) Performans, cazibe, sergilemek, aktris.

COMMUNITY: (Cemaat, halk): Alıştırmak, gelenek, adet, ziyaret.

ECONOMIC: (Ekonomik) İş, maliyet, fiyat, finans.

FAMILY: (Aile) Nişan, evlilik, evlenme, sevgili.

LEGAL: (Yasal) Anlaşma, duruşma, onay, kınamak.

MEDICAL: (Tıbbi) Terapi, tedavi, yaralanma, hastalık.

MILITARY: (Askeri) Savaş, saldırı, dövüş, üniforma, baskın, akın.

POLITICAL: (Politik) Randevu, parti, partizan, kampanya, vergilendirme.

RECREATIONAL: (Eğlence, dinlenmeyle ilgili) Spor, tatil, yüzme, bando, uzanmak.

RELIGIOUS: (Dini) Ayin, sihir, dua, dindarlık.

TECHNOLOGICAL: (Teknolojik) İş, güç, mühendis, yazıcı.

Örnek kelime: “Ahlaki” 22 ve 66 nolu kategorilere girmektedir.

Kategori 22: Ideal Value (İdeal değer)

Kategori 66: Religious (Dini)

Mevki çağrışımları

HIGHER-STATUS: (Yüksek statü) Teyze, analist, doktor, opera.

PEER-STATUS: (Eş-statü) Gelin, koca, eş, aşık, ahabap.

LOWER-STATUS: (Düşük statü) Bebek, çocuk, metres.

Örnek kelime: “Arkadaş” 69, 75 ve 6 nolu kategorilere girmektedir.

Kategori 69: Peer-Status (Eş statü)

Kategori 75: Sign Accept (Kabul belirtisi)

Kategori 6: Neuter Role (Nötr rol)

Psikolojik temalar

OVERSTATE: (Abartmak) Genellikle sıfatlar veya zarflar, bir savunma stili ifade eden etkili, güçlü veya abartılı kelimeler (minnetle, korkunç, berbat, ümitsizce)

UNDERSTATE: (Küçümsemek, tevazu) Genellikle sıfat veya zarflar, şüphe veya kararsızlık ifade eden, bir savunma stili ifade eden, kelimeler (eğilim, tereddüt, adil, kısmen)

SIGN-STRONG: (Güçlü belirtisi) Davranışta güç veya kapasite ifade eden kelimeler (Sihir, teşvik, profesyonel)

SIGN-WEAK: (Zayıf belirtisi) Davranışta zayıflık veya kapasitesizlik ifade eden kelimeler (hüzünlü, erteleme, batmış, utanma, kabul)

SIGN-ACCEPT: (Kabul belirtisi) Kişilerarası kabul ifade eden kelimeler (teşvik, beğeni, takdir)

SIGN-REJECT: (Red belirtisi) Kişilerarası ret ifade eden kelimeler (öfke, kıskançlık, ihanet, küsmek, surat asmak)

MALE-THEME: (Eril tema) Erilliğin psikoanalitik sembolleri (kol-kanat, canavar, kadro, horoz)

FEMALE-THEME: (Dişil tema) Dişilliğin psikoanalitik sembolleri (asalet, filiz, kadife, ağlamak)

SEX-THEME: (Cinsel tema) Doğrudan veya dolaylı cinsel referanslar (nişan, cazibeli, çekici, öpücük)

ASCEND-THEME: (Yükselme teması) Yükselme, düşme, ateş ve su ile ilişkili, Icarus kompleksi ile bağlantı gösteren kelimeler (yanma, canlandırma, uyandırmak, iddia etmek)

AUTHORITY-THEME: (Otorite teması) Otorite ile ilgili kelimeler. (azar, baba, bakan, başvurmak)

DANGER-THEME: (Tehlike teması) Alarm veya tehlike ile ilgili kelimeler (infilak, sapkın, uyarmak, ikaz, germek)

DEATH-THEME: (Ölüm teması) Ölüm ve son ifade eden kelimeler (ezmek, aşağılamak, yanmış)

Örnek kelime: “Aslan” 14, 73 ve 82 nolu kategorilere girmektedir.

Kategori 14: Natural Object (Doğal nesne)

Kategori 73: Sign-Strong (Güç belirtisi)

Kategori 82: Danger Theme (Tehlike ile ilgili tema)

Rollerle ilgili olan kelimeler öncelikle cinsiyete göre veya mesleklere göre etiketlenmeye çalışılıyor. Eğer bu kategorilere girmiyorsa NEUTER-ROLE olarak etiketleniyor. Cinsiyetle ilgili etiketler önemli olduğu için mutlaka etiketlenmelidir. JOB-ROLES etiketleri de aynı şekilde önemli olduğu için yüksek derece ve düşük derece işler olarak etiketlenmelidir. Çünkü sözlüğün yapılış amacı küçük gruplarda bulunan kişilerin cinsiyet ya da statü farkı gibi durumlarda duygusal değişimlerini analiz etmektir. Bu sözlük bu tip duygu kaymalarının dışında diğer çalışmalar için de kullanışlıdır (Montague, 1971).

Topluluklar iki sınıfa ayrılıyor, SMALL-GROUP (küçük çaplı gruplar) ve LARGE GROUP (büyük çaplı gruplar). Küçük gruplar yüz-yüze iletişim yapılabilen gruplar olup, büyük gruplar ise daha genel gruplardır. Bundan dolayı küçük gruplar takım, çete, kulüp vb. Büyük gruplar ise ordu, Türkler, üniversite, toplum vb. Bu şekilde küçük grubun kendisine mi veya başka bir büyük kurumsal topluluğa mı işaret edildiğini anlaşılabılır.

EK 3. Harward-IV sözlüğündeki kategoriler

1. En büyük iki kategori

Pozitif: 1915 pozitif anlam içeren kelimedenden oluşmaktadır. Ancak “evet” (**yes**) kelimesini içermez. Çünkü bu isimde 20 kelimelik ayrı bir kategori vardır. (mümkün, bolluk, kabul)

Negatif: 2291 negatif anlam içeren kelimedenden oluşmaktadır. Ancak “hayır” (**no**) kelimesini içermez. Çünkü bu isimde ret manasını içeren kelimelerin bulunduğu ayrı bir kategori vardır. (terk, çekilmek, kaçmak, dalgın, ani, devamsız)

Diğer kategoriler:

2. Osgood’un üç anlamsal boyutu

Harvard IV kategorileri Charles Osgood’un temel dil anlam ayrımsal bulgularını yansıtır. Osgood’un üç anlamsal boyutu vardır. Harvard sözlüğünün önceki sürümlerinde bu boyutlar her kategori için üç ayrı yoğunluk seviyesi oluşturmaktaydı. Bu sürümde bunlar birleştirildi. Bir kelimenin birden fazla boyutu olabilir. Örneğin “kutlama” (celebration) kelimesi Harvard sözlüğünde *Positiv Pstv Affil Active Ritual* boyutlarına sahiptir.

Pstv: 1045 olumlu kelime içerir. Ancak bu, eski sürümdeki şeklidir. Yeni sürümde kategoriye farklı kelimeler eklenerek *Positiv* ismi verilmiştir. Eskisi de korunmuştur. (başarılı, gerçek, yeterli, güncel)

Bu kategoride olan ve aynı zamanda bağlılık ya da dayanışma gösteren 557 kelime *Affil* isimli kategoriye de girmektedir.

Ngstv: 1160 olumsuz kelime içerir. Ancak bu, eski sürümdeki şeklidir. Yeni sürümde eskisi de korunmuş ve kategoriye farklı kelimeler eklenerek *Negativ* ismi verilmiştir. (istismar, saçma, kaza, suçlamak)

Bu kategoride olan ve aynı zamanda saldırganlık veya düşmanlık tutumu ya da endişesi gösteren 833 kelime *Hostile* isimli kategoriye de girmektedir.

Strong: Güç ima eden 1902 kelime bu kategoriye girer. (mümkün, yetenek, feshetmek, mutlak)

Bu kategorinin otorite, güç veya kontrol endişesi gösteren 689 kelimesi *Power* olarak da etiketlenir.

Weak: Zayıflık veya güçsüzlük ima eden 755 kelime bu kategoriye girer. (sefil, dalgın, amatör, yalnız)

Bu kategorinin otoriteye veya güce boyun eğmek, başkalarına bağımlılık veya başkalarına korunmasızlık veya vazgeçme gösteren 284 kelimesi *Submit* olarak da etiketlenir.

Active: Etkin bir yönelim ifade eden 2045 kelimedenden oluşmaktadır. (aklamak, vurgulamak, ivme, başarılı, eylem)

Passive: Edilgen bir yönelim ifade eden 911 kelimedenden oluşmaktadır. (çekilmek, nefret etmek, kabul, dalgın)

3. Zevk, acı, erdem ve ahlak sözcükleri.

Bu kategoriler genellikle güç gösteren, erdem ve zayıflık gösteren kelimelerdir. Aynı zamanda pozitif ve negatif olarak da etiketlenmiştir.

Pleasure (Zevk): Memnuniyet gösteren 168 kelime içerir, bunların içinde güven, ilgi ve vaat gösteren kelimeler de vardır. (hayranlık, kültür, takdir, mutlu)

Pain (Acı): Acı, güven eksikliği veya bağıllık gösteren 254 kelime içerir. (korkmak, acı, ağrı, hastalık, kaygı)

Feel (Hissetme): Acı ve memnuniyet dışındaki minnettarlık, ilgisizlik veya iyimserlik gibi duyguları tarif eden 49 kelime içeriyor. (sevimsiz, duygusuz, imrenmek, huysuz)

Arousal (Uyarılma): Acı ve memnuniyet dışındaki heyecan gösteren 166 kelimeyi içerir, bunların içinde yakınlık ya da düşmanlıktan gelen uyarılma dâhildir. (nefret, hırçın, uyarı, öfke, kışkırtmak, iştah, beklenti)

EMOT (Heyecan): 311 kelimedenden oluşan his veya duyguyla ilişkili kelimeleri içeren, daha öncelikle belirsizliği kaldırmak için kullanılan kategoridir. (korkmak, saldırgan, sıkıntı, kaygı, utanma, mutlu, dingin, neşeli)

Virtue (Erdem): Özellikle toplumdaki orta sınıfın bakış açısı, ahlaki onay ya da iyi talihin değerlendirmesini ifade eden 719 kelime. (doğru, hassasiyet, titizlik, gerçek, uyum, yeterli)

Vice (Kusur): Ahlakî olarak kabul edilmeyen veya talihsizlik değerlendirmesi gösteren 685 kelimedenden oluşmaktadır. (anormal, hırçın, istismar, naz, sıkıntı)

4. Abartı veya hafife almayı gösteren kelimelerdir. Sıklıkla duygu varlığını veya eksikliğini yansıtırlar.

Overst (Abartı) : Hız, frekans, nedensellik, kapsayıcılık, miktar, doğruluk, geçerlilik, kapsam, boyut, netlik, yoğunluk, benzerlik, kesinlik ve aşırılık kavramlarını vurgulayan 696 kelime. (mutlak, bolluk, teşekkür, titiz, aslında, tüm, yalnız, süper, anarşi)

Undrst (Küçümsemek): Ovrst kategorisindeki kavramların vurgusuz ve kısıtlı olanlarını ifade eden 319 kelimeden oluşmaktadır. (yaklaşık, kaza, belirsiz, sevimsiz, kısa, biraz)

5. Belirli bir kurum dilini yansıtan kelimeler. Bu kategoriler sosyolojik bakış açısını yansıtır. Bu kategorinin yüksek sayıda çıkması metnin bir kuruma ait olduğunu, örneğin bir avukat, bir profesör ya da bir subaya ait olduğunu gösterebilir.

Academ (Akademik): Akademik, entelektüel veya eğitimsel konuları içeren 153 kelimeden oluşur. Bu kelimeler içinde alanların ana başlık isimleri de yer almaktadır. (biyolojik, astronomi, kurul, yerleşke, danışman, bilgi)

Doctrin (Öğreti): İnanç veya bilgi gibi organize sistemlerle ilgili olan 217 kelimeden oluşur. Uygulamalı bilgi, mistik inanç ve akademik sanatı da içerir. (yönetim, idari, tarım, mimarlık, sanatsal)

Econ@ (Ekonomik): Ekonomik, ticari, endüstriyel ve iş yönelimini alanındaki roller, eylemler, topluluklar, soyut fikirler ve semboller (para, mal) içeren 510 kelimedir. (servet, hesap, muhasebe, ücret, iflas, müzayede)

Exch (Alım-Satım): Ticaretle, alış ve satışla ilgili 60 kelimeden oluşmaktadır.

Ayrıca 502 kelimelik **ECON** kategorisi bulunmaktadır. Bunlardan 269 kelimesi Econ@ kategorisi ile ortaktır. Bu kategori GI tarafından belirsizliği (muğlâklık) kaldırmak için kullanılır. (ödünç, satın almak, alım, tüketim, maliyet)

Exprsv (İfade): Sanat, spor ve kendini ifade ile ilgili 205 kelime içerir. Son çalışmalarda spor ile ilgili 186 kelime içeren **Sports** kategorisi ve 138 kelime içeren sanat ile ilgili **Arts** kategorisi oluşturulmuştur. Henüz eklenmemiştir. (estetik, aktör, sanatsal, küresel, yazar)

Legal (Yasal): Hukuki, adli veya polisiye konularla ilgili 192 kelimeden oluşmaktadır. (aklamak, iddia, af örgütü, geçerli, temyiz, tutuklama)

Milit (Askeri): Askeri konularla ilgili 88 kelimeden oluşmaktadır. (pusu, mühimmat, silahlı, ordu, savaş, bomba)

Polit@ (Politika): Net bir siyasi karaktere sahip 263 kelimeden oluşur. Siyasi rolleri, toplulukları, eylemleri, fikirleri, ideolojileri ve sembolleri içerir. (büyükelçi, kampanya, ittifak, kapitalizm, aday). Ayrıca 507 kelimelik **POLIT** kategorisi bulunmaktadır. Bu Polit@ kategorisinden daha geniştir ve GI tarafından belirsizliği kaldırmak için kullanılır.

Relig (Din): Dini metafizik, doğaüstü veya ilgili felsefi konulara ilişkin 103 kelimedenden oluşmaktadır. (inanç, mukaddes, melek, Kuran, kilise, Allah)

6. Roller, toplumlar, ritüeller, insanlar arası ilişki çeşitleri gibi kurumsal bağlamlardan herhangi birini ifade eden kelimelerdir.

Rol (Rol): Sosyologlar tarafından kullanılan, bireysel insan davranış örneklerini gösteren 569 kelimeyi içerir. (yetişkin, aktör, yönetici, danışman, yazar, gelin, müşteri, albay)

COLL (Toplumlar): Tüm insan topluluklarını gösteren 191 kelimeyi içerir. Belirsizliği kaldırmak için kullanılır. (hava kuvvetleri, kitle, kurul, sivil)

Work (İş): Sosyolojik olarak tanımlanmış iş yapma yollarını gösteren kelimelerdir. (işlemek, yük, satın alma, geçim, telafi, sözleşme)

Ritual (Ritüel, Ayin): İş ile ilgisi olmayan sosyal ritüelleri gösteren 134 kelimeyi içerir. (ziyafet, doğum günü, kutlama, tören, kamp)

SocRel (Sosyal ilişkiler): Sosyal olarak tanımlanmış kişilerarası süreçleri gösteren 577 kelimedir. (suçlama, hasım, eşlik, kabul, istismar, tembih)

7. Atfedici toplumsal kategoriler yanı sıra, insan ve hayvanı işaret eden genel ifadeler.

Race (İrk): Irksal veya etnik karakterleri gösteren 15 kelimedenden oluşur. (kızılderili, zenci, önyargı)

Kin@ (Akrabalık): Akrabalığı gösteren 50 kelimedenden oluşur. (gelin, kardeş, abi, ata, teyze, baba, boşanma, aile, halk, genetik)

MALE (Erkek): Erkekler ve sosyal rolleri ile ilişkili 56 kelimedenden oluşur. Belirsizlik kaldırılmasında işaret olarak kullanılır. (işadamı, papaz, Mesih, şoför, baba, koca, bay)

Female (Bayan): Bayanlar ve sosyal rolleri ile ilişkili 43 kelimedenden oluşur. (teyze, gelin, kız, prenses, anne, kadın)

NonAdlt (Yetişkin olmayan): Bebeklikten ergenliğe kadar kullanılan 25 kelimedenden oluşur. (kız, çocuk, bebek, buluş)

HU (İnsan): İnsanı ve rollerini gösteren 795 kelimedenden oluşur. (tanıdık, ergen, ajan, danışman, amatör, büyükelçi, ittifak)

ANI (Hayvan): Hayvan ve hayvan topluluklarını gösteren 72 kelimedenden oluşur. (hayvan, sinek, kuş, deve,)

8. Yerleri ve aralarındaki güzergâhları ifade eden kelimeler.

Place (Yer): 318 kelimedenden oluşmaktadır. Alt kategorilere ayrılır: *Social, Region, Route, Aquatic, Land, Sky*

Social (Sosyal): Yerler için oluşturulmuş 111 kelimedenden oluşur. Tipik olarak sosyal etkileşim sağlar ve sınırlı yer kaplarlar. (havaalanı, doğum yeri, bina, kabin, yerleşke)

Region (Bölge): 61 kelimedenden oluşur. (şehir, çevre, ülke, doğu)

Route (Rota): 23 kelimedenden oluşur. (meydan, köprü, mahalle, kavşak)

Aquatic (Su ile ilgili): 20 kelimedenden oluşur. (körfez, kanal, sel, okyanus, göl)

Land (Kara): Doğada oluşmuş yerleri gösteren 63 kelimedenden oluşur. (Çöl, sahil)

Sky (Gökyüzü): Hava şartları ile ilgili 34 kelimedenden oluşur. (atmosfer, sis, ufuk, meteor, gezegen yağmur)

9. Varlıklar için kullanılan kelimeler.

Object (Nesne): 661 kelimedenden oluşur. Alt kategorilere ayrılır: *Tool, Food, Vehicle, BldgPt, ComnObj, NatObj, BodyPt*.

Tool (Araç): 318 kelimedir. (klima, çapa, silah, ok)

Food (Yiyecek): 80 kelimedir. (kahvaltı, kek, şeker, ekmek, yağ, fasulye, limon)

Vehicle (Araç): 39 kelimedir. (uçak, otomobil, tren, paraşüt, gemi, roket)

BldgPt (Bina öğeleri): Binalar, odaları ve diğer parçaları ifade eden 46 kelimedir. (daire, asansör, zemin, giriş, salon, laboratuvar)

ComnObj (İletişim nesneleri): İletişim araçlarını belirten 104 kelimedir. (albüm, sanat, kitap, sertifika, film, dolar, bayrak, doküman)

NatObj (Doğal nesneler): Bitkiler, mineraller gibi insan ve hayvanların dışındaki nesneleri gösteren 61 kelimedir. (Kül, atom, kalsiyum, kömür, petrol, çekirdek, ağaç)

BodyPt (Beden öğeleri): Beden bölümlerini ifade eden 80 kelimedenden oluşur. (kol, bilek, sırt, ayak, kan, yüz, hemoglobin, diş, damak)

10. İletişim süreçleri.

ComForm (İletişim şekli): İletişim ortamı, biçimi ve şekli ile ilgili 895 kelimedir. (tavsiye, fikir birliği, alarm, ilan, pazarlık). Belirsizliğin kaldırılması için kullanılan ve 412 kelimedenden oluşan **COM (İletişim)** kategorisi de vardır.

Say (Söylemek): Söyleme ve anlatma ifadelerinden oluşan 4 kelimedir. (söyle, de, anlat, konuşmak)

11. Teşvik ile ilişkili kelimeler.

Need (İhtiyaç): Niyet veya ihtiyaç ifadesi için kullanılan 76 kelimedir. (imrenmek, merak, arzu, umutlu, sabırsız, yalvarmak, niyet, teşvik)

Goal (Hedef): Fiziksel ya da zihinsel uğraşların sonunda varmak istenilen hedefi gösteren 53 kelimedir. (amaç, başarı, fetih, fayda, seçim, kader)

Try (Dene): Bir hedefe ulaşmak için gerçekleştirilen faaliyetleri gösteren 70 kelimedir. Fakat hedeflere ulaşıldığı bilgisini içermeyen kelimelerdir. (amaçlamak, uygulamak, girişim, yaklaşmak)

Means (Usûl): Nesneleri, eylemleri veya hedeflere ulaşmada kullanılan yöntemleri gösteren 244 kelimedir. (erişim, boyunca). Lasswell sözlüğündeki **Means** 77 kelime içeriyor ve bu iki kategorinin sadece 16 kelimesi örtüşüyor.

Persist (Israr): Dayanıklılık ve ısrar etmek ile ilgili 64 kelimedir. (daima, devamlı, çıkmaz, gecikme, tahammül, sürdürmek, azmetmek)

Compleat (Tamamlama): Süreç devam etse bile hedefe ulaşıldığını gösteren 81 kelimedir. (başarmak, elde etmek, yenmek, ortaya çıkarmak, tedavi, üstesinden gelmek, mezuniyet)

Finish (Bitirme): Sürecin bittiğini gösteren 87 kelimedir. (imha, patlama, yok etmek, iptal, bitiş)

Fail (Başarısızlık): Hedeflere ulaşamadığını gösteren 137 kelimedir. (terk, iflas, gaf, afet)

12. Diğer süreç ya da değişim kelimeleri.

NatrPro (Doğal süreçler): Doğada bulunan süreçleri (doğumdan ölüme) gösteren 217 kelimedir. (yaş, uyanık, doğum, bayılmak, ısırma)

Değişim süreçlerini gösteren kategoriler aşağıdaki gibidir:

Begin (Başla) : Süreç başlangıcı gösteren 56 kelimeden oluşan kelimedir. (başlamak, girişimci, açmak, başta, köken).

Vary (Değişim): Artma, azalma, başlama ve bitiş anlamlarının dışındaki değişim süreçlerini gösteren 98 kelimedir. (köpürmek, dönüşüm, evrim)

Increas (Artma): 111 kelimedir. (ivme, artırmak, yardımcı, fazlasıyla, gelişme, güçlenme)

Decreas (Azalma): 82 kelimedir. (hafiflemek, ucuzlatmak, aşındırmak, kesmek, çöküş)

Finish (Bitiş): 87 kelimedir. (ölüm, tufan, imha, patlama)

Hareket süreçlerini gösteren kategoriler aşağıdaki gibidir:

Stay (Kalmak): 125 kelimeden oluşur. (tutuklama, çapa, kamp, beklemek)

Rise (Yükselmek): 25 kelimeden oluşur. (uçmak, atlamak, çıkmak)

Exert (Sarf): 194 kelimeden oluşur. (faaliyet, iliştiirmek, bozmak, öldürmek, tutuklamak, yağlamak)

Fetch (Almak): Taşımayı da içeren 129 kelimeden oluşur. (satın almak, getirmek, hırsız, hayırsever, sevk)

Travel (Seyahat): Yatay düzlemde bir yerden başka yere fiziksel hareketi belirten 209 kelimedir. (varış, macera, ulaşım)

Fall (Düşüş): 42 kelimeden oluşur. (damla, düşüş, alçalmak, çökmek)

13. Akli yönelim (bilmek, değerlendirmek ve problem çözmek)

Think (Düşünmek): Rasyonel düşünce süreçlerinin varlığına ya da yokluğuna atıfta bulunan 81 kelimedir. (çıkarım, çağırışım, uyanıklık, zeki, idrak, unutkanlık)

Know (Bilmek): Farkındalık veya gafletin, kesinlik veya belirsizlik, benzerlik ya da farklılık, genelliği veya özgünlük, önem veya önemsizlik, varlığını veya yokluğunu gösteren aynı zamanda zihinsel, kavramsal veya fikir bileşenlerini gösteren 348 kelimedir. (özet, soyut, benzer, hesaplama, sınıflandırma, ipucu, karşılaştırma, eşdeğer)

Causal (Nedensel): Bir olayın olmasının başka bir olayın olmasına neden olduğunu gösteren 112 kelimedir. (ihmal, çünkü, şans eseri, sonuç, kasıt)

Ought (Gereklik): Toplumsal olarak yapılması iyi olan ve gereklik ifade eden 26 kelimedir. (saygılı, kibar, temkin, hürmet, formalite, zorunlu, vicdanlı)

Perciev (Algılamak): Duyular vasıtasıyla bir şey algılamayı ifade eden 192 kelimedir. (belirsizlik, görünüm, farkındalık, keşif, güvenilmez, şüphe)

Compare (Karşılaştırmak): Karşılaştırma ifade eden 21 kelimedir. (en iyi, daha iyi, az, çok, yüksek, alçak, daha)

Eval@ (Değerlendirmek): Olumlu veya olumsuz yargı ya da değerlendirme ifade eden 205 kelimedir. (tavsiye, doğrulama, alkışlama, güzelleştirmek, suçsuz)

EVAL (Değerlendirmek): Belirsizliği kaldırmak için kullanılan 314 kelimelik Eval@’den daha geniş bir kategoridir.

Solve (Çözmek): Çoğunlukla fiillerden oluşan, problem çözme ile ilgili 189 kelimedir. (çözmek, açmak)

Abs@ (Soyut): Soyut anlatım eğilimi gösteren 189 kelimedir. Ayrıca bulunan ABC kategorisi ise belirsizliği kaldırmak için kullanılan 276 kelimedir. (eylem, ideoloji, kapitalizm, inanç, uygarlık, kavram, vicdan, ölüm)

Quality (Kalite): İnsan duyuları ile ölçülebilen, nitelik ya da nitelik derecelerini ifade eden 344 kelimedir. İyi ve kötü huyları ifade eden kelimeler ayrı bir kategoride sınıflandırılmıştır. (şık, incelik, yoğun, olağanüstü)

Quantity (Miktar): Miktar belirten kelimelerdir. Sayılar da dâhildir. Sayılar için ayrıca NUMB, ORD ve CARD kategorileri de vardır. (bolluk, miktar, yaklaşık, büyük, milyar, sınırsız, koleksiyon, küme, hiç, ikinci, bir)

NUMB (Sayılar) : 51 kelimedir. ORD ve CARD kategorilerinin birleşmesinden oluşur. (bir, beş, on beş, ikinci, son)

ORD (Kaçıncı): Kaçıncı olduğunu ifade eden 15 kelimedir. (ikinci, önceki, son, onuncu, ilk)

CARD (Sayılar): Sayılara ait 36 kelimedir. (sıfır, bir, düzine, milyon)

FREQ (Sıklık): Sıklık ifade eden 46 kelimedir. Belirsizlik kaldırmak için de kullanılır. (nadir, tekrar, yıllık, ortak, sık, asla, kez, kere)

DIST (Mesafe): Mesafe ve ölçüleri ifade eden 19 kelimedir. (metre, mil, adım, genişlik)

Time@ (Zaman): Zaman ifade eden veya olayların sürelerini ifade eden 273 kelimedir. Hız kavramını ifade eden kelimeleri de içerir. (sonra, öğle, nisan, sonbahar, yıldönümü, yaz, yüzyıl)

TIME (Zaman): Time@’dan daha az kelimesi vardır. Belirsizliği kaldırmak için işaret olarak kullanılır.

Space (Mekânsal) : Uzay ve mekân ifade eden 273 kelimedir. Belirsizliği kaldırmak için ise **POS (Konum)** (35 kelime) ve **DIM (Boyut)** (49 kelime) kategorileri vardır. (yurtdışı, kavşak, yan, açı, aşağıda, köprü, santimetre, merkez)

Rel (İlişki): İnsanlar, yerler, nesneler ve fikirler arasındaki soyut ilişkileri ifade eden 136 kelimedir. Uzay ve zaman ilişkileri içinde değildir. (uyum, beton, orantılı, birlikte, eşit, on beşinci)

COLOR (Renk): Belirsizlik için kullanılan 21 renk kelimesidir. (bronz, kahverengi, turkuaz, ak, kara, al, sarımtırak, masmavi)

14. “Ben”, “Sen”, ”Biz”, “Siz” yönelimli zamirler ve isimler.

Self (Kendi): Birinci tekil şahıs zamirleridir. 7 tanedir. (ben, benim, kendim, benim)

Our (Bizim): Birinci çoğul şahıs zamirleridir. 6 tanedir. (biz, bizim, bizimki, kendimiz)

You (Siz, Sen): İkinci tekil ve çoğul şahıs zamirleridir. 9 kelimedir. (sen, siz, seninki)

Name (İsim): Özel isimlerdir. 86 kelimedir. (Türkiye, Amerika, Afrika, İngilizce, Avrupa)

Nation (Millet): Lasswell sözlüğünden alınan özel isimlerdir. (Arnavutluk, Çek, Libya, Mali, Suriye)

15. “Evet”, “Hayır”, inkâr ve ünlem ifade eden kelimeler.

Yes (Evet): 20 onaylama kelimesidir. (tabii ki, peki, mutlak, kesinlikle, tamam, öyle)

No (Hayır): 7 onaylamama kelimesidir. İçinde olan hayır (no) kelimesi yokluk ve olumsuzluğun belirsizliğini de giderir. (inkâr, hiç, asla, ret, yanlış)

Negate (Olumsuz): Olumsuz anlam veren 217 kelimedir. (değil, asla, tiksime, dezavantaj, aykırı, nahoş, yalanlama, ret, vazgeç)

Intrj (Argo): Nida, sıradan konuşma ve argo 42 kelimedir. (ulan, harbi, oley, koçum, fıstık, tosun, inek, ya, he, ha, be, tıkmak, kanka, çaktım, trollemek, moruk)

Sosyal idrak temelli yeni kategoriler:

16. Fiil türleri:

IAV (Yorum fiilleri): Bir eylemi yorumlayan fiiller. 1947 kelimedir. (feshetmek, terk etmek, sil, anlat, temizle, yaz, al, taşı, kaldır, tak, iç, kullan, ısıt, topla, çal, karıştır, tara, aç, kapat araştır, çiz, öde, besle, büyüt, öldür, yut, incit, dengele, dinle, sür, gör, bul, kes, sev, kaz)

DAV (Tarif fiilleri): Bir eylemin kendisinin tarifi 540 kelimedir. (sarar, karar, morar, küflen, kızar, bayatla, paslan, uza, büyü, sol, acık, ağar, zayıfla, eri, olgunlaş, çürü, yeşer, hastalan, körel, piş, soğuk)

SV (Hal fiilleri): Akli veya hissi hal gösteren 102 fiildir. Genelde gözlemlenebilen olaylardan çıkarılır. (sevmek, güvenmek, nefret etmek)

17. Sıfat türleri:

IPadj (İlişki sıfatları): İnsanlar arası ilişkileri ifade eden 117 kelimedir. (mesafeli, resmi, kaba, anlayışlı)

IndAdj (Bireysel sıfatlar): İnsanları birbirinden ayıran özellikler, ilişkilerinden bağımsız sıfatları ifade eden 637 kelimedir. (tutumlu, dalgın, atılgan, sorumlu, maceracı)

Lasswell sözlüğü kategorileri.

Bu sözlük kelimeleri saygı anlayışına göre dört grubuna ayırır: güç, dürüstlük, saygı, yakınlık. Aynı zamanda refah durumuna göre dört gruba ayırır: zenginlik, refah, aydınlanma ve beceri. Her alan içinde, kazanç, kayıp, katılımcı, hedef ve bölge gibi alt kategorileri olabilir. Ana kategorilerde ve alt kategorilerde sözcükler bulunur. Sözlük yazarları bir kelime birden fazla ana kategoriye veya alt kategoriye girse bile, sadece birine koymaya çalışmışlardır. (General Inquirer Sitesi, 2002)

18. Güç:

PowGain (Güç kazanımı): Güç artmasını gösteren 65 kelimedir. (işbirliği, teşvik, kuvvetlendirmek, takip)

PowLoss (Güç kaybı): Güç kaybını gösteren 109 kelimedir. (çıkarılma, saldırı, zorlayacak)

PowEnds (Güç hedefi): Güç kullanmanın hedefini gösteren 30 kelimedir. (hırs, yarışma, iddia)

PowAren (Güç bölgesi): Ulus-devletler dışındaki siyasi bölgesel ayrımları gösteren 53 kelimedir. (Afrika, imparatorluk, ilçe, kıta, ülke)

PowCon (Güç çatışması): Güç çatışmasının farklı şekillerini ifade eden 228 kelimedir. (ihale, çalkalama, pusu, çatışma)

PowCoop (Güç işbirliği): Güçte işbirliğinin farklı şekillerini ifade eden 118 kelimedir. (yerleştirme, ortak, ittifak)

PowAuPt (Yetkili güç katılımcıları): Güçte yetki sahipleri, güç kullanımının bireysel ve toplumsal aktörlerini ifade eden 134 kelimedir. (yönetim, büyükelçi, yetki, patron)

PowPt (Güç katılımcıları): Sıradan güç sahibi, yetkisi olmayan aktörleri ifade eden 81 kelimedir. (aktör, sivil, aday)

PowDoct (Güç ilkesi): Güç ilkeleri, güç ilişkileri ve uygulamaları hakkında bilinen fikirleri ifade eden 42 kelimedir. (burjuvazi, kapitalizm, rakip, muhafazakâr)

PowAuth (Güç yetkisi): Yetkiye bağlı güç, resmi gücü içeren araçları ve şekilleri ifade eden 79 kelimedir. (idari, yönetmek, evlatlık)

PowOth (Güçle ilgili diğer kavramlar): Artık kategori, güçle alakalı diğer alt kategorilere ait olmayan 332 kelimedir. (başarısız, feshetmek, tahsis, özerklik)

PowTot (Güçle ilgili tüm kavramlar): Tüm güç kategorilerine giren 1266 kelimedir. (feshetmek, başarısız, ani, razı olmak)

19. Dürüstlük ahlaki değerlerle ilgilidir. .

RcEthic (Dürüstlük etiği): Etik, toplumsal düzene ilişkin değerleri ifade eden 151 kelimedir. (ihanet, iyi, karakter, sansür)

RcRelig (Dini dürüstlük): Metafizik, mistik veya doğaüstü temellere dayanan dürüstlikle ilgili 83 kelimedir. (inançlı, mukaddes, Müslüman)

RcGain (Dürüstlük kazanımı): Dürüstlük kazanımı, af ve ibadet gibi 30 kelimedir. (iman, şeffaf, bağışlamak, af)

RcLoss (Dürüstlük kaybı): Dürüstlük kaybı, günah ve iftira gibi 12 kelimedir. (suçlamak, hükümlü, kınamak, şerefsizlik)

RcEnds (Dürüstlük hedefleri): Dürüstlük hedefleri, cennet veya yapılması doğru olan 33 kelimedir. (bağışlanmak, yaratan, geleneksel, cennet)

RcTot (Dürüstlükle ilgili tüm kelimeler): Dürüstlükle ilgili tüm kategorilerde yer alan 310 kelimedir. (iman, inançlı, suçlamak, ihanet)

20. Konum, onur, tanınma ve itibardan gelen saygı.

RspGain (Saygı kazanımı): Saygı kazanmayla ilgili 26 kelimedir. (tebrik)

RspLoss (Saygı kaybı): Saygı kaybetmeyle ilgili 38 kelimedir. (ayıp)

RspOth (Diğer): Saygı kazanma veya kaybetme dışındaki diğer saygıyla ilgili 182 kelimedir. (takdir, hayranlık, özür)

RspTot (Toplam): Saygı ile ilgili tüm kategorilerde yer alan 245 kelimedir. (istismar, teşekkür, kabul)

21. Sevgi, merhamet ve dostlukla ilgili yakınlıktır.

AffGain (Sevgi kazanımı): Sevgi kazanımı ile ilgili 35 kelimedir. (kalp, aşk, evlenmek)

AffLoss (Sevgi kaybı): Sevgi kaybı ve umursamazlığı ile ilgili 11 kelimedir. (öfke, boşanma, bırakma, yalnız)

AffPt (Sevgi katılımcısı): Arkadaş ve aile ile ilgili 55 kelimedir. (teyze, gelin, baba, kuzen)

AffOth (Diğer): Diğer alt kategorilere girmeyen 96 kelimedir. (tanıdık, kardeşlik, yoldaşlık)

AffTot (Toplam): Tüm sevgi kelimelerinin yer aldığı 196 kelimedir. (bırak, sevgi, tanıdık)

22. Servet, sahip olduğunun değerini bilmektir.

WltPt (Servet katılımcısı): Servet katılımcıları, iş ve ticarete çeşitli roller için kullanılan 52 kelimedir. (denetçi, işadamı, şirket, iştirakçi)

WltTran (Servet transferi): Alış, satış gibi servetin yer değiştirmesi ile ilgili 53 kelimedir. (uygun, müzayede, ödünç)

WltOth (Diğer): Diğer alt kategorilere girmeyen 271 kelimedir. (bolluk, hesap, servet)

WltTot (Toplam): Servet alanındaki tüm kelimelerin yer aldığı 378 kelimedir.

23. Canlının sağlık ve güvenliği refahtır.

WlbGain (Refah kazanımı): Refah kazanımıyla ilgili 37 kelimedir. (küresel, kür, kolaylık, rahatlık)

WlbLoss (Refah kaybı): Refah kaybıyla ilgili 60 kelimedir. Üzgün olmayı da içerir. (musibet, depresyon, sakatlık, ölüm)

WlbPhys (Refahın fiziksel yönü): Refahın fiziksel taraflarını çağrıştıran 226 kelimedir. Refahın yokluğu ile ilgili olanlar da dâhildir. (hayat dolu, doğum, nefes)

WlbPsyc (Refahın psikolojik yönü): Refahın psikolojik taraflarını çağrıştıran 226 kelimedir. Refahın yokluğu ile ilgili olanlar da dâhildir. 139 kelimedir. (sakin, neşeli, öfkeli, keyifli)

WlbPt (Refah katılımcısı): Refahı çağrıştıran rollerle ilgili 27 kelimedir. (Bebekler, doktorlar, tatilciler)

WlbTot (Toplam): Refah ile ilgili tüm kavramları içeren 487 kelimedir.

24. Aydınlanma, bilgi, iç görü ve kişisel ve kültürel ilişkiler ile ilgili bilgilerdir.

EnlGain (Bilgi kazanımı): Aydınlanma ile ilgili 146 kelimedir. (analiz, duyuru, anlamak, açıkça, cevap)

EnlLoss (Bilgi kaybı): Bilgi kaybetme ile ilgili 27 kelimedir. (yanıltıcı, aldatmak, şaşkınlık)

EnlEnds (Bilgi edinme hedefi): İçsel aydınlanma fikirlerinin peşinde olduğunu gösteren 18 kelimedir. (Düşünmek, keşfetmek, incelemek, büyülemek, genellemek, ispat, arama)

EnIPt (Aydınlanma katılımcısı): Aydınlanma katılımcıları, laik aydınlanmadaki rolleri ifade eden 61 kelimedir. (analist, danışman, yazar)

EnIOth (Diğer): Diğer alt kategorilere girmeyen 585 kelimedir. (özet, akademi, soyutlama, hassasiyet)

EnITot (Toplam): Tüm aydınlanma kavramlarını içeren 835 kelimedir.

25. Sanat, estetik gibi alanlarda becerileri yansıtır.

SklAsth (Estetik becerisi): Estetik becerisini, çoğunlukla sanatsal becerileri ifade eden 35 kelimedir. (estetik, güzel, edebiyat, opera, şiir)

SklPt (Beceri katılımcısı): Beceri katılımcıları, esnaf ve meslekler hakkında 64 kelimedir. (sanatçı, marangoz, memur, doktor, dalgıç)

SklOth (Diğer): Yukarıdakilerin dışında beceri ile ilgili 158 kelimedir. (uzman, tasarımcı, estetik, amatör)

SklTot (Toplam): Tüm beceri kavramlarının yer aldığı 257 kelimedir.

26. Değer alanlarından herhangi birine özgü olmayan sözlük kategorileri.

TrnGain (İşlem kazancı): Başarı ile tamamlanmış bir işlemin kazancını ifade eden 129 kelimedir. (iyileştirmek, güvence, hafiflemek)

TrnLoss (İşlem kaybı): Sorunlardan dolayı tamamlanamamış bir işlemin kaybını ifade eden 113 kelimedir. (üzmek, rahatsız, şaşırtmak, hasar, düşme)

TranLw (İşlem): Genel manada bir işlemin ya da değiş-tokuş ile ilgili 334 kelimedir. (tadilat, söndürmek)

MeansLw (Araçlar): Araç veya yararları veya bunların eksikliğini ifade eden 78 kelimedir. (kapasite, maliyet, cihaz, ekipman, tesis). Aynı zamanda daha sınırlı bir kategori olan **Means** ile örtüşür.

EndsLw (Hedef): İsteddiğimiz ya da istemediğimiz hedef ya da sonuçlarını ifade eden 270 kelimedir. (başarı, aktif, faaliyet, seçim, tutarlı)

ArenaLw (Bölge): Güç bölgelerinin (**PowAren**) dışındaki ortamlar için kullanılan 34 kelimedir. (şehir, uygarlık, ülke)

PtLw (Katılımcı): Yukarıdaki kategorilerin hiç birine girmeyen oyuncularla (aktörler) ilgili 68 kelimedir. (karakter, rehber, mirasçı)

Nation (Ulus): Milletlerden oluşan 169 kelimedir. (Afganistan, Arnavutluk)

Anomie (Tercih inkarı): İyi bir değer tercihi inkârını gösteren 30 kelimedir. (hayal kırıklığı, anarşi, kaos, umutsuz)

NegAff (Olumsuz duygular): Olumsuz duygular ve duygusal ret gösteren 193 kelimedir. (anormal, olumsuz, karşı)

PosAff (Olumlu duygular): Olumlu duygular, kabul, takdir, duygusal destek ifade eden 126 kelimedir. (kabul, maceracı, coşkulu)

SureLw (Kesinlik): Kesinlik, belirlilik, istikrarlı olmakla ilgili 175 kelimedir. (mutlak, gerçek, daima)

If (Eğer): Belirsizlik ve şüphe ifade eden 132 kelimedir. (yaklaşık, kararsız, belirsizlik)

NotLw (Ret): Bir şeyi ret etmeyi ifade eden 25 kelimedir (değil, asla, yok)

TimeSpc (Zaman-Mekân): Genel zaman-mekân kavramlarını içeren 428 kelimedir. (sonra, önce, yanında, önünde)

FormLw (Biçim): Biçimleri, standartları, araçlar ve iletişim kurallarını ifade eden 368 kelimedir (ihale, danışma, kabul). 895 kelimelik *ConForm* kategorisinin hemen hemen bir alt kümesidir.

EK 4. Metot Kısaltmaları ve Açıklamaları.

Metot	Açıklama
1 KategoriListe	Kategori Liste Frekansı
2 KategoriListe_HS	Kategori Liste Frekansı Herikisini-Sil
3 KategoriListe_KS	Kategori Liste Frekansı Küçüğü-Sil
4 KategoriListeAğırlıklı	Kategori Liste Ağırlıklı Frekansı
5 KategoriListeAğırlıklı_HS	Kategori Liste Ağırlıklı Frekansı Herikisini-Sil
6 KategoriListeAğırlıklı_KS	Kategori Liste Ağırlıklı Frekansı Küçüğü-Sil
7 KategoriTekliVarYok	Kategori Tekli Var-Yok Bilgisi
8 KategoriTekliAğırlıklı	Kategori Tekli Frekans Ağırlıklı
9 KategoriKarmaVarYok	Kategori Karma Var-Yok Bilgisi
10 KategoriKarmaVarYok_HS	Kategori Karma Var-Yok Bilgisi Herikisini-Sil
11 KategoriKarmaVarYok_KS	Kategori Karma Var-Yok Bilgisi Küçüğü-Sil
12 KategoriKarmaFrekans	Kategori Karma Frekans Ağırlıklı
13 KategoriKarmaFrekans_HS	Kategori Karma Frekans Ağırlıklı Herikisini-Sil
14 KategoriKarmaFrekans_KS	Kategori Karma Frekans Ağırlıklı Küçüğü-Sil
15 KelimeKarmaVarYok	Kelime Karma Var-Yok Bilgisi
16 KelimeKarmaVarYok_HS	Kelime Karma Var-Yok Bilgisi Herikisini-Sil
17 KelimeKarmaVarYok_KS	Kelime Karma Var-Yok Bilgisi Küçüğü-Sil
18 KelimeKarmaFrekans	Kelime Karma Frekans Ağırlıklı
19 KelimeKarmaFrekans_HS	Kelime Karma Frekans Ağırlıklı Herikisini-Sil
20 KelimeKarmaFrekans_KS	Kelime Karma Frekans Ağırlıklı Küçüğü-Sil
21 KelimeListe	Kelime Liste Frekansı
22 KelimeListe_HS	Kelime Liste Frekansı Herikisini-Sil
23 KelimeListe_KS	Kelime Liste Frekansı Küçüğü-Sil
24 KelimeListeAğırlıklı	Kelime Liste Ağırlıklı Frekansı
25 KelimeListeAğırlıklı_HS	Kelime Liste Ağırlıklı Frekansı Herikisini-Sil
26 KelimeListeAğırlıklı_KS	Kelime Liste Ağırlıklı Frekansı Küçüğü-Sil
27 KelimeTekliVarYok	Kelime Tekli Var-Yok Bilgisi
28 KelimeTekliAğırlıklı	Kelime Tekli Frekans
29 WLLRKategoriListe_HS	WLLR Ağırlıklı Kategori Liste Frekansı Herikisini-Sil
30 WLLRKategoriListe_KS	WLLR Ağırlıklı Kategori Liste Frekansı Küçüğü-Sil
31 WLLRKategoriTekliVarYok	WLLR Var-Yok Kategori Tekli
32 WLLRKategoriTekliAğırlıklı	WLLR Ağırlıklı Kategori Tekli Frekans
33 WLLRKategoriKarmaAğırlıklı_HS	WLLR Ağırlıklı Kategori Karma Frekans_ Herikisini-Sil
34 WLLRKategoriKarmaAğırlıklı_KS	WLLR Ağırlıklı Kategori Karma Frekans_ Küçüğü-Sil
35 WLLRKategoriKarmaVarYok_HS	WLLR Var-Yok Kategori Karma_ Herikisini-Sil
36 WLLRKategoriKarmaVarYok_KS	WLLR Var-Yok Kategori Karma_ Küçüğü-Sil
37 WLLRKelimeKarmaAğırlıklı_HS	WLLR Ağırlıklı Kelime Karma_ Herikisini-Sil
38 WLLRKelimeKarmaAğırlıklı_KS	WLLR Ağırlıklı Kelime Karma_ Küçüğü-Sil
39 WLLRKelimeKarmaVarYok_HS	WLLR Var-Yok Kelime Karma_ Herikisini-Sil
40 WLLRKelimeKarmaVarYok_KS	WLLR Var-Yok Kelime Karma_ Küçüğü-Sil
41 WLLRKelimeListe_HS	WLLR Ağırlıklı Kelime Liste Frekansı Herikisini-Sil
42 WLLRKelimeListe_KS	WLLR Ağırlıklı Kelime Liste Frekansı Küçüğü-Sil
43 WLLRKelimeTekliVarYok	WLLR Var-Yok Kelime Tekli
44 WLLRKelimeTekliAğırlıklı	WLLR Ağırlıklı Kelime Tekli Frekans
45 IDFKategoriListe_HS	IDF Ağırlıklı Kategori Liste Frekansı Herikisini-Sil
46 IDFKategoriListe_KS	IDF Ağırlıklı Kategori Liste Frekansı Küçüğü-Sil
47 IDFKategoriTekliAğırlıklı	IDF Ağırlıklı Kategori Tekli Frekans
48 IDFKategoriKarmaAğırlıklı_HS	IDF Ağırlıklı Kategori Karma Frekans_ Herikisini-Sil
49 IDFKategoriKarmaAğırlıklı_KS	IDF Ağırlıklı Kategori Karma Frekans_ Küçüğü-Sil
50 IDFKelimeKarmaAğırlıklı_HS	IDF Ağırlıklı Kelime Karma_ Herikisini-Sil
51 IDFKelimeKarmaAğırlıklı_KS	IDF Ağırlıklı Kelime Karma_ Küçüğü-Sil
52 IDFKelimeListe_HS	IDF Ağırlıklı Kelime Liste Frekansı Herikisini-Sil
53 IDFKelimeListe_KS	IDF Ağırlıklı Kelime Liste Frekansı Küçüğü-Sil
54 IDFKelimeTekliAğırlıklı	IDF Ağırlıklı Kelime Tekli Frekans

EK 5. Özellik seçimi yöntemlerine ve eşik değerlerine göre HV-III ile yapılan çocuk DEP x SAĞ sınıflandırmasının NB, SMO, J48 teknikleri uygulanarak elde edilen Liste ve Karma yöntem doğruluk sonuçları.

DEP_IDF												
	Eşik=0			Eşik=5			Eşik=25			Eşik=50		
Özellik	NB	SMO	J48	NB	SMO	J48	NB	SMO	J48	NB	SMO	J48
KategoriListe_HS	69	69	80	69	69	80	71	73	76	73	73	67
KategoriListe_KS	69	69	80	69	69	77	71	71	77	71	71	76
KategoriKarmaFrek_HS	73	69	67	74	76	67	74	67	67	73	69	64
KategoriKarmaFrek_KS	73	69	67	73	69	67	73	69	67	73	67	67
KelimeListe_HS	69	67	56	69	67	51	67	69	53	71	69	46
KelimeListe_KS	69	67	56	69	67	55	69	67	45	69	67	51
KelimeKarmaFrek_HS	67	58	69	67	60	69	64	60	67	66	60	75
KelimeKarmaFrek_KS	67	58	69	67	58	69	67	58	69	67	58	69
DEP_WLLR												
	Eşik=0			Eşik=5			Eşik=25			Eşik=50		
Özellik	NB	SMO	J48	NB	SMO	J48	NB	SMO	J48	NB	SMO	J48
KategoriListe_HS	75	77	69	76	78	73	74	76	73	67	67	62
KategoriListe_KS	75	77	69	75	73	68	75	73	69	71	73	71
KategoriKarmaFrek_HS	71	73	60	74	71	60	69	69	64	73	67	62
KategoriKarmaFrek_KS	71	73	60	71	73	60	71	73	60	71	73	60
KategoriKarmaVarYok_HS	82	60	75	82	60	75	80	68	69	66	66	64
KategoriKarmaVarYok_KS	82	60	75	82	60	75	82	60	75	82	60	75
KelimeListe_HS	75	77	64	75	77	64	77	78	67	76	76	64
KelimeListe_KS	75	77	64	75	77	64	75	77	62	75	77	64
KelimeKarmaFrek_HS	71	64	62	73	62	62	64	60	66	60	55	60
KelimeKarmaFrek_KS	71	64	62	71	64	62	71	64	62	71	64	62
KelimeKarmaVarYok_HS	78	67	71	78	65	71	76	64	71	54	60	75
KelimeKarmaVarYok_KS	78	69	71	78	69	71	78	69	71	78	69	71
DEP_FREKANS												
	Eşik=0			Eşik=5			Eşik=25			Eşik=50		
Özellik	NB	SMO	J48	NB	SMO	J48	NB	SMO	J48	NB	SMO	J48
KategoriListe_HS	69	67	71	73	71	71	73	73	69	62	69	69
KategoriListe_KS	69	67	71	71	69	71	69	73	69	75	76	75
KategoriListeAğırlıklı_HS	75	75	71	75	75	71	75	73	73	58	65	64
KategoriListeAğırlıklı_KS	75	75	71	75	75	71	82	74	66	75	75	71
KategoriKarmaFrek_HS	49	69	64	76	62	62	73	68	49	62	71	58
KategoriKarmaFrek_KS	49	69	64	49	69	64	49	69	64	49	69	64
KategoriKarmaVarYok_HS	78	52	75	80	58	75	80	65	71	66	66	64
KategoriKarmaVarYok_KS	78	52	75	78	52	75	78	52	75	78	52	75
KelimeListe_HS	65	66	54	71	70	52	78	76	64	75	75	67
KelimeListe_KS	67	66	54	64	68	66	74	69	54	76	70	60
KelimeListeAğırlıklı_HS	73	75	67	73	75	65	73	71	67	71	70	73
KelimeListeAğırlıklı_KS	73	75	67	73	75	65	73	75	67	73	73	69
KelimeKarmaFrek_HS	67	62	55	66	63	60	64	62	51	69	64	54
KelimeKarmaFrek_KS	65	62	55	58	64	58	67	64	58	64	64	51
KelimeKarmaVarYok_HS	69	65	75	67	65	75	67	65	76	67	58	73
KelimeKarmaVarYok_KS	69	65	75	69	65	75	69	65	75	69	65	75

EK 6. Özellik seçimi yöntemlerine ve eşik değerlerine göre HV-III ile yapılan çocuk ADHD x SAĞ sınıflandırmasının NB, SMO, J48 yöntemleri uygulanarak elde edilen doğruluk sonuçları.

ADHD_TF-IDF												
Eşik=0				Eşik=5			Eşik=25			Eşik_50		
Özellik	NB	SMO	J48	NB	SMO	J48	NB	SMO	J48	NB	SMO	J48
KategoriListe_HS	48	59	57	48	50	52	61	57	54	52	57	52
KategoriListe_KS	48	61	57	48	57	54	48	57	54	48	54	54
KategoriKarmaFrek_HS	67	63	44	63	65	46	72	69	50	61	59	59
KategoriKarmaFrek_KS	67	65	44	67	63	44	67	61	44	67	63	44
KelimeListe_HS	48	50	58	48	52	59	48	52	58	46	52	57
KelimeListe_KS	48	52	58	48	52	54	48	54	59	46	50	42
KelimeKarmaFrek_HS	56	55	46	67	63	44	57	52	50	63	63	57
KelimeKarmaFrek_KS	56	55	46	56	52	54	56	55	46	56	55	46
ADHD_WLLR												
Eşik =0				Eşik =5			Eşik =25			Eşik =50		
Özellik	NB	SMO	J48	NB	SMO	J48	NB	SMO	J48	NB	SMO	J48
KategoriListe_HS	61	54	70	65	63	65	55	63	64	50	49	45
KategoriListe_KS	61	56	70	65	63	71	63	65	72	61	65	71
KategoriKarmaFrek_HS	46	48	55	51	58	46	70	66	54	61	57	52
KategoriKarmaFrek_KS	46	48	55	46	48	55	46	48	55	46	48	55
KategoriKarmaVarYok_HS	64	61	46	73	55	44	57	55	53	50	50	52
KategoriKarmaVarYok_KS	64	61	46	64	61	46	64	59	46	64	59	46
KelimeListe_HS	59	59	53	59	59	53	56	57	59	65	55	60
KelimeListe_KS	59	61	53	59	59	53	56	59	51	54	61	55
KelimeKarmaFrek_HS	55	47	48	52	54	49	52	65	60	59	50	47
KelimeKarmaFrek_KS	55	50	48	55	48	48	55	47	48	55	47	48
KelimeKarmaVarYok_HS	65	53	66	65	56	61	65	55	59	70	58	63
KelimeKarmaVarYok_KS	65	53	66	65	53	66	65	53	66	65	53	66
ADHD_TF												
Eşik =0				Eşik =5			Eşik =25			Eşik =50		
Özellik	NB	SMO	J48	NB	SMO	J48	NB	SMO	J48	NB	SMO	J48
KategoriListe_HS	65	54	50	48	50	48	44	50	50	48	54	50
KategoriListe_KS	65	54	50	56	55	48	56	56	48	59	63	56
KategoriListeAğırlıklı_HS	43	44	50	57	59	54	55	56	55	48	55	46
KategoriListeAğırlıklı_KS	43	50	50	57	59	54	70	67	59	56	58	55
KategoriKarmaFrek_HS	58	52	54	54	52	52	65	74	56	65	61	54
KategoriKarmaFrek_KS	58	52	54	58	52	54	58	52	54	58	52	54
KategoriKarmaVarYok_HS	69	61	50	61	56	50	57	52	56	52	48	52
KategoriKarmaVarYok_KS	69	61	50	69	61	50	69	61	50	69	61	50
KelimeListe_HS	63	63	50	58	69	59	58	56	52	54	54	39
KelimeListe_KS	63	63	50	59	70	58	69	65	59	63	63	54
KelimeListeAğırlıklı_HS	65	67	57	65	69	57	52	56	60	65	65	57
KelimeListeAğırlıklı_KS	65	67	57	65	69	57	63	60	67	54	59	55
KelimeKarmaFrek_HS	56	58	61	48	56	54	59	46	54	56	54	50
KelimeKarmaFrek_KS	56	58	61	56	58	61	56	58	61	56	58	61
KelimeKarmaVarYok_HS	58	54	61	59	61	67	61	54	61	61	61	67
KelimeKarmaVarYok_KS	58	54	61	58	54	61	58	54	61	58	54	61

EK 7. Özellik seçimi yöntemlerine ve eşik değerlerine göre HV-III ile yapılan çocuk K x E sınıflandırmasının NB, SMO, J48 yöntemleri uygulanarak elde edilen doğruluk sonuçları.

Cinsiyet_TF-IDF												
Eşik=0				Eşik=5			Eşik=25			Eşik_50		
Özellik	NB	SMO	J48	NB	SMO	J48	NB	SMO	J48	NB	SMO	J48
KategoriListe_HS	71	73	73	71	73	73	69	73	73	69	73	73
KategoriListe_KS	71	73	73	71	73	73	71	73	73	69	73	76
KategoriKarmaFrek_HS	56	67	53	64	73	76	60	73	82	64	73	73
KategoriKarmaFrek_KS	56	67	53	56	64	53	56	67	53	56	64	53
KelimeListe_HS	71	73	80	71	73	69	73	73	62	73	73	64
KelimeListe_KS	71	73	80	71	73	80	69	73	71	76	73	76
KelimeKarmaFrek_HS	69	73	73	69	76	76	69	76	69	73	78	76
KelimeKarmaFrek_KS	69	73	73	69	73	73	69	73	73	69	73	73
Cinsiyet_WLLR												
Eşik =0				Eşik =5			Eşik =25			Eşik =50		
Özellik	NB	SMO	J48	NB	SMO	J48	NB	SMO	J48	NB	SMO	J48
KategoriListe_HS	49	50	50	53	50	60	50	50	54	58	50	50
KategoriListe_KS	49	50	50	54	50	56	57	50	51	56	50	51
KategoriKarmaFrek_HS	57	51	57	57	53	58	60	50	57	60	50	50
KategoriKarmaFrek_KS	57	51	57	57	51	57	57	51	57	57	51	57
KategoriKarmaVarYok_HS	55	51	54	57	61	53	55	54	50	51	50	50
KategoriKarmaVarYok_KS	55	51	54	55	51	54	55	51	54	55	51	54
KelimeListe_HS	50	50	51	51	50	50	54	50	58	57	51	57
KelimeListe_KS	50	50	51	51	50	50	50	50	50	50	50	53
KelimeKarmaFrek_HS	53	51	65	61	50	65	54	50	65	53	50	65
KelimeKarmaFrek_KS	53	53	65	53	50	65	53	51	65	53	50	65
KelimeKarmaVarYok_HS	50	57	63	50	56	64	50	51	64	50	56	64
KelimeKarmaVarYok_KS	50	55	63	50	57	63	50	57	63	50	57	63
Cinsiyet_TF												
Eşik =0				Eşik =5			Eşik =25			Eşik =50		
Özellik	NB	SMO	J48	NB	SMO	J48	NB	SMO	J48	NB	SMO	J48
KategoriListe_HS	69	73	73	71	73	73	80	73	73	71	73	73
KategoriListe_KS	69	73	73	73	73	73	78	73	69	78	73	73
KategoriListeAğırlıklı_HS	71	73	73	69	73	73	64	73	71	69	73	73
KategoriListeAğırlıklı_KS	71	73	73	69	73	73	80	73	69	82	73	76
KategoriKarmaFrek_HS	64	73	58	73	73	67	62	71	69	64	73	73
KategoriKarmaFrek_KS	64	73	58	64	73	58	64	73	58	64	73	58
KategoriKarmaVarYok_HS	71	78	67	73	71	67	73	73	71	73	73	73
KategoriKarmaVarYok_KS	71	76	67	71	76	67	71	76	67	71	76	67
KelimeListe_HS	60	60	29	76	73	69	73	73	73	76	73	67
KelimeListe_KS	60	60	29	73	73	71	71	73	76	71	73	71
KelimeListeAğırlıklı_HS	76	73	73	76	73	73	76	73	80	82	80	78
KelimeListeAğırlıklı_KS	76	73	73	76	73	73	71	73	76	76	73	82
KelimeKarmaFrek_HS	78	73	53	76	24	53	78	80	56	80	82	53
KelimeKarmaFrek_KS	78	73	53	78	73	53	78	73	53	78	73	53
KelimeKarmaVarYok_HS	76	87	53	78	82	56	76	76	62	78	80	53
KelimeKarmaVarYok_KS	76	87	53	76	87	53	76	87	53	76	87	53

EK 8. Özellik seçimi yöntemlerine ve eşik değerlerine göre yapılan çocuk DEP x SAĞ sınıflandırmasının HV-IV sözlüğü kullanılarak ve NB, SMO, J48 teknikleri uygulanarak elde edilen Liste ve Karma yöntem doğruluk sonuçları.

DEP_IDF												
Eşik=0				Eşik=5			Eşik=25			Eşik_50		
Özellik	NB	SMO	J48	NB	SMO	J48	NB	SMO	J48	NB	SMO	J48
KategoriListe_HS	71	69	69	71	69	73	71	69	73	62	67	69
KategoriListe_KS	71	69	69	71	69	73	71	69	71	71	69	69
KategoriKarmaFrek_HS	67	65	76	63	57	66	62	69	66	54	58	62
KategoriKarmaFrek_KS	67	64	76	67	64	76	67	64	76	67	64	76
KelimeListe_HS	69	67	56	69	67	51	67	69	53	71	69	46
KelimeListe_KS	69	67	56	69	67	55	69	67	45	69	67	51
KelimeKarmaFrek_HS	67	58	69	67	60	69	64	60	67	66	60	75
KelimeKarmaFrek_KS	67	58	69	67	58	69	67	58	69	67	58	69
DEP_WLLR												
Eşik =0				Eşik =5			Eşik =25			Eşik =50		
Özellik	NB	SMO	J48	NB	SMO	J48	NB	SMO	J48	NB	SMO	J48
KategoriListe_HS	80	80	65	75	73	71	77	71	71	73	71	71
KategoriListe_KS	80	80	65	75	73	71	75	75	71	75	75	71
KategoriKarmaFrek_HS	69	71	64	65	62	62	65	55	62	51	51	51
KategoriKarmaFrek_KS	69	71	64	69	71	64	69	71	64	69	71	64
KategoriKarmaVarYok_HS	51	53	53	49	51	51	51	51	51	51	51	51
KategoriKarmaVarYok_KS	51	53	53	51	53	53	51	53	53	51	53	53
KelimeListe_HS	75	77	64	75	77	64	77	78	67	76	76	64
KelimeListe_KS	75	77	64	75	77	64	75	77	62	75	77	64
KelimeKarmaFrek_HS	71	64	62	73	62	62	64	60	66	60	55	60
KelimeKarmaFrek_KS	71	64	62	71	64	62	71	64	62	71	64	62
KelimeKarmaVarYok_HS	78	67	71	78	65	71	76	64	71	54	60	75
KelimeKarmaVarYok_KS	78	69	71	78	69	71	78	69	71	78	69	71
DEP_FREKANS												
Eşik =0				Eşik =5			Eşik =25			Eşik =50		
Özellik	NB	SMO	J48	NB	SMO	J48	NB	SMO	J48	NB	SMO	J48
KategoriListe_HS	73	69	71	75	75	71	65	67	71	66	71	69
KategoriListe_KS	73	69	71	71	69	67	71	71	69	73	73	73
KategoriListeAğırlıklı_HS	69	69	71	73	75	67	71	75	71	66	71	69
KategoriListeAğırlıklı_KS	69	69	71	75	75	67	75	75	60	75	75	58
KategoriKarmaFrek_HS	57	49	64	64	51	65	67	67	75	62	58	64
KategoriKarmaFrek_KS	57	49	64	57	49	64	57	49	64	57	49	64
KategoriKarmaVarYok_HS	68	53	62	71	65	62	74	73	73	67	67	65
KategoriKarmaVarYok_KS	68	53	62	68	53	62	68	53	62	68	53	62
KelimeListe_HS	65	71	49	69	75	52	78	76	57	77	77	63
KelimeListe_KS	67	71	49	57	75	64	76	76	53	75	77	51
KelimeListeAğırlıklı_HS	75	77	67	75	78	65	75	76	67	77	77	78
KelimeListeAğırlıklı_KS	75	78	67	75	78	65	75	78	67	75	77	69
KelimeKarmaFrek_HS	66	60	62	66	62	62	62	57	60	64	57	51
KelimeKarmaFrek_KS	64	60	62	64	60	62	64	60	62	64	60	62
KelimeKarmaVarYok_HS	69	58	73	67	58	73	67	63	75	65	56	75
KelimeKarmaVarYok_KS	69	58	73	69	58	73	69	58	73	69	58	73

EK 9. Özellik seçimi yöntemlerine ve eşik değerlerine göre yapılan çocuk ADHD x SAĞ sınıflandırmasının NB, SMO, J48 yöntemleri uygulanarak ve HV-IV kullanılarak elde edilen doğruluk sonuçları.

ADHD_TF-IDF												
Eşik=0				Eşik=5			Eşik=25			Eşik_50		
Özellik	NB	SMO	J48	NB	SMO	J48	NB	SMO	J48	NB	SMO	J48
KategoriListe_HS	48	58	50	50	50	50	52	48	54	52	54	54
KategoriListe_KS	48	52	50	48	50	50	48	50	57	50	52	50
KategoriKarmaFrek_HS	59	61	54	59	63	55	56	56	56	48	52	50
KategoriKarmaFrek_KS	59	61	54	59	61	54	59	60	54	59	56	54
KelimeListe_HS	48	52	58	48	52	59	48	52	58	46	52	57
KelimeListe_KS	48	52	58	48	54	54	48	54	59	46	52	42
KelimeKarmaFrek_HS	56	55	46	56	52	54	57	52	50	63	63	57
KelimeKarmaFrek_KS	56	55	46	56	55	46	56	55	46	56	55	46
ADHD_WLLR												
Eşik =0				Eşik =5			Eşik =25			Eşik =50		
Özellik	NB	SMO	J48	NB	SMO	J48	NB	SMO	J48	NB	SMO	J48
KategoriListe_HS	66	63	57	56	56	65	48	56	52	48	54	52
KategoriListe_KS	66	61	57	56	56	59	57	61	65	57	61	65
KategoriKarmaFrek_HS	62	59	48	51	44	48	50	50	50	50	50	50
KategoriKarmaFrek_KS	62	59	48	62	57	48	62	57	48	62	57	48
KategoriKarmaVarYok_HS	50	50	50	50	50	50	50	50	50	50	50	50
KategoriKarmaVarYok_KS	50	50	50	50	50	50	50	50	50	50	50	50
KelimeListe_HS	59	62	53	59	57	53	56	57	59	65	57	60
KelimeListe_KS	59	61	53	59	61	53	56	59	51	54	61	55
KelimeKarmaFrek_HS	55	47	48	52	54	49	52	65	60	59	50	47
KelimeKarmaFrek_KS	55	47	48	55	48	48	55	47	48	55	47	48
KelimeKarmaVarYok_HS	65	53	66	65	56	61	65	55	59	70	58	63
KelimeKarmaVarYok_KS	65	53	66	65	53	66	65	53	66	65	53	66
ADHD_TF												
Eşik =0				Eşik =5			Eşik =25			Eşik =50		
Özellik	NB	SMO	J48	NB	SMO	J48	NB	SMO	J48	NB	SMO	J48
KategoriListe_HS	54	61	48	61	56	54	48	61	48	44	50	48
KategoriListe_KS	54	59	48	61	63	54	54	55	54	54	59	54
KategoriListeAğırlıklı_HS	61	54	44	56	57	56	57	59	50	46	52	50
KategoriListeAğırlıklı_KS	61	54	44	59	68	56	60	55	54	54	54	54
KategoriKarmaFrek_HS	44	48	46	54	50	39	59	58	50	50	50	48
KategoriKarmaFrek_KS	44	48	46	44	48	46	44	48	46	44	50	46
KategoriKarmaVarYok_HS	50	43	52	59	50	56	59	52	53	52	45	50
KategoriKarmaVarYok_KS	50	43	52	50	39	52	50	43	52	50	43	52
KelimeListe_HS	63	63	50	58	69	59	58	56	52	54	54	39
KelimeListe_KS	63	63	50	59	70	58	69	65	59	63	63	54
KelimeListeAğırlıklı_HS	65	65	57	65	69	57	52	56	60	65	65	57
KelimeListeAğırlıklı_KS	65	69	57	65	69	57	63	61	67	54	59	55
KelimeKarmaFrek_HS	56	58	61	48	56	54	59	46	54	56	54	50
KelimeKarmaFrek_KS	56	58	61	56	58	61	56	58	61	56	58	61
KelimeKarmaVarYok_HS	58	54	61	59	61	67	61	54	61	61	61	67
KelimeKarmaVarYok_KS	58	54	61	58	54	61	58	54	61	58	54	61

EK 10. Özellik seçimi yöntemlerine ve eşik değerlerine göre HV-IV ile yapılan çocuk K x E sınıflandırmasının MÖ yöntemleri uygulanarak elde edilen doğruluk sonuçları.

Cinsiyet_TF-IDF												
Eşik=0				Eşik=5			Eşik=25			Eşik_50		
Özellik	NB	SMO	J48	NB	SMO	J48	NB	SMO	J48	NB	SMO	J48
KategoriListe_HS	53	50	50	50	50	50	51	50	50	51	50	50
KategoriListe_KS	53	50	50	51	50	50	54	50	50	51	50	50
KategoriKarmaFrek_HS	54	51	49	57	51	53	55	50	53	50	50	50
KategoriKarmaFrek_KS	54	51	49	54	51	49	54	51	49	54	51	49
KelimeListe_HS	53	50	53	53	50	54	53	50	60	50	50	54
KelimeListe_KS	53	50	53	53	50	54	53	50	63	53	50	58
KelimeKarmaFrek_HS	57	54	67	61	65	65	59	64	64	61	58	62
KelimeKarmaFrek_KS	57	54	67	57	54	67	57	54	67	57	54	67
Cinsiyet_WLLR												
Eşik =0				Eşik =5			Eşik =25			Eşik =50		
Özellik	NB	SMO	J48	NB	SMO	J48	NB	SMO	J48	NB	SMO	J48
KategoriListe_HS	50	50	50	50	50	50	51	50	50	51	50	50
KategoriListe_KS	50	50	50	50	50	50	50	50	50	50	50	50
KategoriKarmaFrek_HS	51	50	50	50	50	49	50	50	50	50	50	50
KategoriKarmaFrek_KS	51	50	50	51	50	50	51	50	50	51	50	50
KategoriKarmaVarYok_HS	50	50	50	50	50	50	50	50	50	50	50	50
KategoriKarmaVarYok_KS	50	50	50	50	50	50	50	50	50	50	50	50
KelimeListe_HS	50	50	51	51	50	50	54	50	58	57	51	57
KelimeListe_KS	50	50	51	51	50	50	50	50	50	50	50	53
KelimeKarmaFrek_HS	53	50	65	61	50	65	54	50	65	53	51	65
KelimeKarmaFrek_KS	53	51	65	53	51	65	53	50	65	53	50	65
KelimeKarmaVarYok_HS	50	57	63	50	56	64	50	51	64	50	56	64
KelimeKarmaVarYok_KS	50	59	63	50	57	63	50	54	63	50	56	63
Cinsiyet_TF												
Eşik =0				Eşik =5			Eşik =25			Eşik =50		
Özellik	NB	SMO	J48	NB	SMO	J48	NB	SMO	J48	NB	SMO	J48
KategoriListe_HS	50	50	50	50	50	50	50	50	50	50	50	50
KategoriListe_KS	50	50	50	50	50	50	50	50	50	50	50	50
KategoriListeAğırlıklı_HS	50	50	50	50	50	50	50	50	50	50	50	50
KategoriListeAğırlıklı_KS	50	50	50	50	50	50	50	50	50	50	50	50
KategoriKarmaFrek_HS	54	50	51	57	49	52	54	50	53	51	50	50
KategoriKarmaFrek_KS	54	50	51	54	50	51	54	50	51	54	51	51
KategoriKarmaVarYok_HS	49	60	57	53	53	50	50	50	51	50	50	50
KategoriKarmaVarYok_KS	49	60	57	49	60	57	49	60	57	49	60	57
KelimeListe_HS	50	53	49	53	50	53	51	50	54	50	50	50
KelimeListe_KS	50	53	49	47	49	49	53	50	56	55	51	56
KelimeListeAğırlıklı_HS	50	50	50	50	50	50	55	50	54	54	54	57
KelimeListeAğırlıklı_KS	50	50	50	50	50	50	51	50	50	50	50	57
KelimeKarmaFrek_HS	57	44	57	56	49	63	57	53	60	57	46	67
KelimeKarmaFrek_KS	57	44	57	57	44	57	57	44	57	57	44	57
KelimeKarmaVarYok_HS	53	58	62	53	56	60	53	51	57	50	58	58
KelimeKarmaVarYok_KS	53	58	62	53	58	62	53	58	62	53	58	62

EK 11. Özellik seçimi yöntemlerine ve eşik değerlerine göre HV-III ile yapılan yetişkin DEP x SAĞ sınıflandırmasının NB, SMO, J48 teknikleri uygulanarak elde edilen Liste ve Karma yöntem doğruluk sonuçları.

DEP_IDF												
Eşik=0				Eşik=5			Eşik=25			Eşik=50		
Özellik	NB	SMO	J48	NB	SMO	J48	NB	SMO	J48	NB	SMO	J48
KategoriListe_HS	54	63	56	54	57	52	61	55	72	59	61	75
KategoriListe_KS	54	59	56	52	57	54	61	59	61	59	67	65
KategoriKarmaFrek_HS	67	78	69	72	85	65	70	74	65	57	56	65
KategoriKarmaFrek_KS	67	78	69	67	78	69	67	78	69	67	78	69
KelimeListe_HS	65	63	52	63	65	48	74	72	63	76	67	58
KelimeListe_KS	59	67	54	54	65	54	63	70	71	78	83	61
KelimeKarmaFrek_HS	87	89	65	85	93	67	81	93	76	85	91	74
KelimeKarmaFrek_KS	81	91	61	81	91	61	81	91	61	81	91	61
DEP_WLLR												
Eşik =0				Eşik =5			Eşik =25			Eşik =50		
Özellik	NB	SMO	J48	NB	SMO	J48	NB	SMO	J48	NB	SMO	J48
KategoriListe_HS	63	62	62	70	72	67	72	77	60	57	55	52
KategoriListe_KS	63	58	62	78	75	73	78	77	75	77	75	73
KategoriKarmaFrek_HS	73	73	57	73	73	60	68	77	65	55	53	58
KategoriKarmaFrek_KS	73	73	57	73	73	57	73	73	57	73	73	57
KategoriKarmaVarYok_HS	58	57	50	68	62	63	63	55	55	55	55	55
KategoriKarmaVarYok_KS	58	57	50	58	57	50	58	57	50	58	55	50
KelimeListe_HS	67	70	58	70	68	62	77	82	70	75	78	72
KelimeListe_KS	67	70	58	70	70	60	75	70	75	78	75	70
KelimeKarmaFrek_HS	82	87	58	85	85	58	80	83	65	78	72	63
KelimeKarmaFrek_KS	83	83	58	83	83	58	83	83	58	83	83	58
KelimeKarmaVarYok_HS	80	73	67	83	73	67	82	73	67	78	77	72
KelimeKarmaVarYok_KS	78	75	67	78	75	67	78	75	67	78	75	67
DEP_FREKANS												
Eşik =0				Eşik =5			Eşik =25			Eşik =50		
Özellik	NB	SMO	J48	NB	SMO	J48	NB	SMO	J48	NB	SMO	J48
KategoriListe_HS	57	59	63	59	69	52	71	69	72	58	58	67
KategoriListe_KS	57	59	63	67	67	61	78	74	76	83	82	85
KategoriListeAğırlıklı_HS	41	39	50	57	65	59	69	74	69	59	59	67
KategoriListeAğırlıklı_KS	41	42	50	57	67	46	85	85	82	85	89	76
KategoriKarmaFrek_HS	74	76	68	71	59	52	61	78	46	56	54	61
KategoriKarmaFrek_KS	74	76	68	74	76	68	74	76	68	74	76	68
KategoriKarmaVarYok_HS	69	59	59	70	63	70	61	61	63	56	56	59
KategoriKarmaVarYok_KS	69	59	59	69	59	59	69	59	59	69	59	59
KelimeListe_HS	59	78	56	69	80	65	63	83	42	80	78	63
KelimeListe_KS	50	78	56	56	72	50	76	79	65	78	87	67
KelimeListeAğırlıklı_HS	54	59	52	57	76	72	81	76	69	85	72	76
KelimeListeAğırlıklı_KS	54	57	52	59	76	66	81	82	69	82	81	72
KelimeKarmaFrek_HS	89	52	65	80	71	65	82	91	65	85	41	67
KelimeKarmaFrek_KS	79	55	65	79	55	65	79	55	65	79	55	65
KelimeKarmaVarYok_HS	94	92	63	94	92	70	94	91	72	87	89	70
KelimeKarmaVarYok_KS	94	90	70	94	90	70	94	90	70	94	90	70

EK 12. Özellik seçimi yöntemlerine ve eşik değerlerine göre HV-III ile yapılan yetişkin MAN x SAĞ sınıflandırmasının NB, SMO, J48 teknikleri uygulanarak elde edilen Liste ve Karma yöntem doğruluk sonuçları.

MAN_IDF												
Eşik=0				Eşik=5			Eşik=25			Eşik=50		
Özellik	NB	SMO	J48	NB	SMO	J48	NB	SMO	J48	NB	SMO	J48
KategoriListe_HS	65	62	57	57	57	43	65	63	68	67	60	53
KategoriListe_KS	65	62	57	58	58	57	68	60	57	70	65	65
KategoriKarmaFrek_HS	67	70	65	57	60	57	68	73	63	65	57	65
KategoriKarmaFrek_KS	67	70	65	67	70	65	67	70	65	67	70	65
KelimeListe_HS	62	58	62	62	58	57	60	58	73	62	57	63
KelimeListe_KS	58	65	70	60	60	67	65	63	60	62	58	58
KelimeKarmaFrek_HS	65	78	65	63	75	62	65	78	62	65	73	55
KelimeKarmaFrek_KS	65	82	63	65	82	63	65	82	63	65	82	63
MAN_WLLR												
Eşik =0				Eşik =5			Eşik =25			Eşik =50		
Özellik	NB	SMO	J48	NB	SMO	J48	NB	SMO	J48	NB	SMO	J48
KategoriListe_HS	78	82	77	80	82	78	78	82	73	57	63	50
KategoriListe_KS	78	82	77	80	82	78	77	80	67	77	82	72
KategoriKarmaFrek_HS	58	73	67	68	70	60	68	63	58	55	57	55
KategoriKarmaFrek_KS	58	73	67	58	73	67	58	73	67	58	73	67
KategoriKarmaVarYok_HS	53	57	58	57	53	52	63	60	58	52	52	52
KategoriKarmaVarYok_KS	53	57	58	53	57	58	53	57	58	53	57	58
KelimeListe_HS	80	75	70	80	78	73	80	73	73	75	73	75
KelimeListe_KS	80	78	70	80	78	73	83	80	82	78	77	73
KelimeKarmaFrek_HS	75	68	55	75	77	63	72	67	63	70	73	67
KelimeKarmaFrek_KS	78	75	60	78	75	60	78	75	60	78	75	60
KelimeKarmaVarYok_HS	73	72	55	72	68	57	73	70	73	65	63	67
KelimeKarmaVarYok_KS	70	67	58	70	67	58	70	67	58	70	67	58
MAN_FREKANS												
Eşik =0				Eşik =5			Eşik =25			Eşik =50		
Özellik	NB	SMO	J48	NB	SMO	J48	NB	SMO	J48	NB	SMO	J48
KategoriListe_HS	65	60	47	60	58	48	75	77	63	55	52	53
KategoriListe_KS	65	62	47	70	62	50	82	78	75	82	83	78
KategoriListeAğırlıklı_HS	72	73	62	67	68	68	72	65	58	55	58	50
KategoriListeAğırlıklı_KS	72	73	62	68	68	68	82	82	73	85	80	80
KategoriKarmaFrek_HS	55	55	60	75	57	57	68	63	67	55	57	55
KategoriKarmaFrek_KS	55	55	60	55	55	60	55	55	60	55	55	60
KategoriKarmaVarYok_HS	67	55	60	65	63	60	63	67	65	65	65	65
KategoriKarmaVarYok_KS	67	55	60	67	55	60	67	55	60	67	55	60
KelimeListe_HS	60	53	55	55	68	60	73	72	65	63	67	58
KelimeListe_KS	62	52	48	58	60	55	60	67	75	73	78	78
KelimeListeAğırlıklı_HS	65	77	62	68	78	68	72	75	68	80	78	82
KelimeListeAğırlıklı_KS	65	77	62	68	78	68	72	78	67	87	83	73
KelimeKarmaFrek_HS	65	53	73	67	53	73	65	50	73	70	73	52
KelimeKarmaFrek_KS	65	60	73	65	60	73	65	60	73	65	60	73
KelimeKarmaVarYok_HS	70	75	50	67	75	53	68	77	52	65	78	60
KelimeKarmaVarYok_KS	75	75	55	75	75	55	75	75	55	75	75	55

EK 13. Özellik seçimi yöntemlerine ve eşik değerlerine göre HV-III ile yapılan yetişkin K x E sınıflandırmasının NB, SMO, J48 teknikleri uygulanarak elde edilen Liste ve Karma yöntem doğruluk sonuçları.

Cinsiyet_IDF												
Özellik	Eşik =0			Eşik =5			Eşik =25			Eşik =50		
	NB	SMO	J48	NB	SMO	J48	NB	SMO	J48	NB	SMO	J48
KategoriListe_HS	41	59	59	38	52	59	45	59	59	45	62	41
KategoriListe_KS	41	59	59	38	59	59	38	52	62	38	52	52
KategoriKarmaFrek_HS	38	52	38	38	55	41	62	66	72	52	55	34
KategoriKarmaFrek_KS	38	52	38	38	52	38	38	52	38	38	52	38
KelimeListe_HS	45	55	52	52	59	48	55	62	41	55	62	62
KelimeListe_KS	45	55	52	45	59	52	55	59	55	66	62	55
KelimeKarmaFrek_HS	59	59	76	59	62	76	52	62	76	52	59	66
KelimeKarmaFrek_KS	59	59	76	59	59	76	59	59	76	59	59	76
Cinsiyet_WLLR												
Özellik	Eşik =0			Eşik =5			Eşik =25			Eşik =50		
	NB	SMO	J48	NB	SMO	J48	NB	SMO	J48	NB	SMO	J48
KategoriListe_HS	67	65	69	79	75	67	53	50	57	48	54	54
KategoriListe_KS	67	65	69	79	75	83	79	78	83	78	78	86
KategoriKarmaFrek_HS	82	69	44	78	90	43	86	85	61	56	51	54
KategoriKarmaFrek_KS	82	69	44	82	69	44	82	69	44	82	69	44
KategoriKarmaVarYok_HS	50	50	51	49	50	51	51	50	50	50	50	50
KategoriKarmaVarYok_KS	50	50	51	50	50	51	50	50	51	50	50	51
KelimeListe_HS	64	57	68	79	72	83	76	71	81	72	51	72
KelimeListe_KS	64	57	68	81	74	82	81	79	81	67	76	67
KelimeKarmaFrek_HS	86	72	79	88	89	79	85	76	79	59	60	75
KelimeKarmaFrek_KS	85	89	79	85	89	79	85	89	79	85	89	79
KelimeKarmaVarYok_HS	50	38	49	50	53	47	50	40	51	50	68	67
KelimeKarmaVarYok_KS	50	57	49	50	57	49	50	57	49	50	57	49
Cinsiyet_FREKANS												
Özellik	Eşik =0			Eşik =5			Eşik =25			Eşik =50		
	NB	SMO	J48	NB	SMO	J48	NB	SMO	J48	NB	SMO	J48
KategoriListe_HS	55	59	59	69	59	59	41	59	59	45	59	55
KategoriListe_KS	55	59	59	62	62	59	59	59	62	66	59	62
KategoriListeAğırlıklı_HS	52	38	48	72	79	59	45	59	79	45	59	59
KategoriListeAğırlıklı_KS	52	38	48	69	79	52	69	72	69	69	66	66
KategoriKarmaFrek_HS	66	55	41	41	41	41	55	66	66	55	55	62
KategoriKarmaFrek_KS	66	55	41	66	55	41	66	55	41	66	55	41
KategoriKarmaVarYok_HS	55	41	59	62	62	66	55	55	59	41	41	41
KategoriKarmaVarYok_KS	55	38	59	55	38	59	55	38	59	55	38	59
KelimeListe_HS	41	76	45	55	72	55	48	66	59	66	69	48
KelimeListe_KS	38	76	45	55	62	62	72	69	62	76	72	59
KelimeListeAğırlıklı_HS	59	72	72	69	72	69	62	66	79	62	62	45
KelimeListeAğırlıklı_KS	59	72	72	69	72	72	66	76	76	62	66	69
KelimeKarmaFrek_HS	66	69	41	59	41	41	59	38	41	66	59	41
KelimeKarmaFrek_KS	59	59	41	59	59	41	59	59	41	59	59	41
KelimeKarmaVarYok_HS	62	79	48	62	86	48	69	83	48	59	76	38
KelimeKarmaVarYok_KS	59	79	48	59	79	48	59	79	48	59	79	48

EK 14. Özellik seçimi yöntemlerine ve eşik değerlerine göre yapılan yetişkin DEP x SAĞ sınıflandırmasının HV-IV kullanılarak elde edilen Liste ve Karma yöntem doğruluk sonuçları.

DEP_IDF												
Eşik=0				Eşik=5			Eşik=25			Eşik_50		
Özellik	NB	SMO	J48	NB	SMO	J48	NB	SMO	J48	NB	SMO	J48
KategoriListe_HS	43	42	50	43	40	50	37	47	50			
KategoriListe_KS	43	42	50	43	40	50	38	38	50	43	48	50
KategoriKarmaFrek_HS	63	52	58	62	58	58	57	52	75	50	50	50
KategoriKarmaFrek_KS	63	53	58	63	53	58	63	53	58	63	53	58
KelimeListe_HS	55	62	63	60	60	62	63	62	60	65	67	65
KelimeListe_KS	58	60	65	60	62	62	65	68	52	72	73	72
KelimeKarmaFrek_HS	72	88	60	72	92	60	75	87	58	75	83	58
KelimeKarmaFrek_KS	73	92	60	73	92	60	73	92	60	73	92	60
DEP_WLLR												
Eşik =0				Eşik =5			Eşik =25			Eşik =50		
Özellik	NB	SMO	J48	NB	SMO	J48	NB	SMO	J48	NB	SMO	J48
KategoriListe_HS	80	75	78	78	72	75	82	80	77	57	55	52
KategoriListe_KS	80	75	78	78	75	78	78	78	78	78	78	78
KategoriKarmaFrek_HS	58	73	78	70	73	78	75	73	73	50	50	50
KategoriKarmaFrek_KS	58	72	78	58	77	78	58	73	78	58	73	78
KategoriKarmaVarYok_HS	50	50	50	50	50	50	50	50	50	50	50	50
KategoriKarmaVarYok_KS	50	50	50	50	50	50	50	50	50	50	50	50
KelimeListe_HS	67	70	58	70	68	62	77	77	70	75	78	72
KelimeListe_KS	67	70	58	70	72	60	75	70	75	78	77	70
KelimeKarmaFrek_HS	82	87	58	85	85	58	80	83	65	78	72	63
KelimeKarmaFrek_KS	83	83	58	83	83	58	83	83	58	83	83	58
KelimeKarmaVarYok_HS	80	73	67	83	73	67	82	73	67	80	77	72
KelimeKarmaVarYok_KS	78	75	67	78	75	67	78	75	67	78	75	67
DEP_FREKANS												
Eşik =0				Eşik =5			Eşik =25			Eşik =50		
Özellik	NB	SMO	J48	NB	SMO	J48	NB	SMO	J48	NB	SMO	J48
KategoriListe_HS	55	60	50	58	60	48	78	77	63			
KategoriListe_KS	55	65	50	62	68	57	77	78	67	78	78	78
KategoriListeAğırlıklı_HS	57	62	50	65	63	60	82	82	70			
KategoriListeAğırlıklı_KS	57	62	50	65	63	60	80	80	87	82	82	87
KategoriKarmaFrek_HS	60	47	72	70	63	72	72	73	70	50	50	50
KategoriKarmaFrek_KS	60	47	72	60	47	72	60	47	72	60	47	72
KategoriKarmaVarYok_HS	62	53	47	67	48	57	65	62	62	50	50	50
KategoriKarmaVarYok_KS	62	53	47	62	53	47	62	53	47	62	53	47
KelimeListe_HS	57	75	52	65	75	53	63	72	55	77	73	63
KelimeListe_KS	55	72	52	53	55	48	70	70	62	78	83	72
KelimeListeAğırlıklı_HS	53	53	70	70	75	78	78	80	68	83	72	75
KelimeListeAğırlıklı_KS	52	55	68	68	77	73	83	83	77	85	78	75
KelimeKarmaFrek_HS	82	68	62	82	55	62	77	78	62	77	48	62
KelimeKarmaFrek_KS	88	63	62	88	63	62	88	63	62	88	63	62
KelimeKarmaVarYok_HS	92	92	58	92	92	58	88	92	58	85	88	57
KelimeKarmaVarYok_KS	92	90	58	92	90	58	92	90	58	92	90	58

EK 15. Özellik seçimi yöntemlerine ve eşik değerlerine göre yapılan yetişkin MAN x SAĞ sınıflandırmasının HV-IV kullanarak elde edilen Liste ve Karma yöntem doğruluk sonuçları.

MAN_IDF												
Eşik =0				Eşik =5				Eşik =25			Eşik =50	
Özellik	NB	SMO	J48	NB	SMO	J48	NB	SMO	J48	NB	SMO	J48
KategoriListe_HS	47	50	47	48	55	50	47	52	47			
KategoriListe_KS	57	50	52	47	53	50	53	53	50	68	68	62
KategoriKarmaFrek_HS	63	67	50	65	63	52	63	72	65	55	55	53
KategoriKarmaFrek_KS	72	68	52	70	70	53	70	70	53	70	70	53
KelimeListe_HS	65	72	70	62	63	55	62	65	67	63	67	70
KelimeListe_KS	65	68	67	63	65	60	67	68	57	68	67	60
KelimeKarmaFrek_HS	68	70	60	62	75	62	62	80	62	67	73	63
KelimeKarmaFrek_KS	65	78	55	65	82	63	65	82	63	65	82	63
MAN_WLLR												
Eşik =0				Eşik =5				Eşik =25			Eşik =50	
Özellik	NB	SMO	J48	NB	SMO	J48	NB	SMO	J48	NB	SMO	J48
KategoriListe_HS	82	82	78	77	78	80	75	75	77	55	62	52
KategoriListe_KS	82	82	78	80	80	82	83	85	80	83	85	80
KategoriKarmaFrek_HS	50	57	50	55	58	58	73	73	72	50	50	50
KategoriKarmaFrek_KS	50	55	50	50	58	50	50	57	50	50	57	50
KategoriKarmaVarYok_HS	50	50	50	50	50	50	50	50	50	50	50	50
KategoriKarmaVarYok_KS	50	50	50	50	50	50	50	50	50	50	50	50
KelimeListe_HS	80	78	70	80	78	73	80	73	73	75	75	75
KelimeListe_KS	80	78	70	80	78	73	83	80	82	78	77	73
KelimeKarmaFrek_HS	75	68	55	75	77	63	72	67	63	70	73	67
KelimeKarmaFrek_KS	78	75	60	78	75	60	78	75	60	78	75	60
KelimeKarmaVarYok_HS	73	72	55	72	68	57	73	70	73	65	63	67
KelimeKarmaVarYok_KS	70	67	58	70	67	58	70	67	58	70	67	58
MAN_FREKANS												
Eşik =0				Eşik =5				Eşik =25			Eşik =50	
Özellik	NB	SMO	J48	NB	SMO	J48	NB	SMO	J48	NB	SMO	J48
KategoriListe_HS	77	82	68	73	77	68	75	75	65			
KategoriListe_KS	75	73	67	70	67	62	77	77	77	78	77	77
KategoriListeAğırlıklı_HS	72	68	60	77	78	77	77	77	77			
KategoriListeAğırlıklı_KS	72	68	57	73	75	73	80	77	78	83	82	75
KategoriKarmaFrek_HS	67	55	60	55	65	48	65	65	62	50	50	55
KategoriKarmaFrek_KS	67	60	63	67	52	60	58	50	53	58	50	53
KategoriKarmaVarYok_HS	68	65	63	63	65	63	67	70	68	57	58	58
KategoriKarmaVarYok_KS	77	73	72	77	73	72	72	72	70	72	72	70
KelimeListe_HS	63	62	62	70	73	62	67	63	57	55	68	58
KelimeListe_KS	70	62	55	67	65	57	60	65	67	63	70	68
KelimeListeAğırlıklı_HS	70	75	53	70	78	60	68	75	67	77	78	75
KelimeListeAğırlıklı_KS	68	77	63	72	80	75	68	77	65	80	77	75
KelimeKarmaFrek_HS	63	58	72	65	60	72	65	52	73	68	65	55
KelimeKarmaFrek_KS	63	50	73	65	58	70	65	60	73	65	60	73
KelimeKarmaVarYok_HS	70	77	62	67	77	65	70	77	52	68	75	55
KelimeKarmaVarYok_KS	70	75	55	68	78	60	75	75	55	75	75	55

EK 16. Özellik seçimi yöntemlerine ve eşik değerlerine göre yapılan yetişkin K x E sınıflandırmasının HV-IV sözlüğü kullanılarak elde edilen Liste ve Karma yöntem doğruluk sonuçları.

Cinsiyet_IDF												
Özellik	Eşik =0			Eşik =5			Eşik =25			Eşik =50		
	NB	SMO	J48	NB	SMO	J48	NB	SMO	J48	NB	SMO	J48
KategoriListe_HS	25	50	50	40	50	33	50	50	56	63	50	38
KategoriListe_KS	25	50	50	25	50	33	28	47	51	29	47	53
KategoriKarmaFrek_HS	38	64	71	67	61	69	54	56	50	51	54	51
KategoriKarmaFrek_KS	38	64	71	38	63	71	38	64	71	38	64	71
KelimeListe_HS	44	50	28	43	50	45	49	50	62	70	50	70
KelimeListe_KS	44	50	29	44	50	27	69	67	45	71	69	47
KelimeKarmaFrek_HS	67	72	72	71	72	71	74	71	71	70	69	67
KelimeKarmaFrek_KS	72	71	72	72	71	72	72	71	72	72	71	72
Cinsiyet_WLLR												
Özellik	Eşik =0			Eşik =5			Eşik =25			Eşik =50		
	NB	SMO	J48	NB	SMO	J48	NB	SMO	J48	NB	SMO	J48
KategoriListe_HS	61	63	65	63	54	65	44	50	50	48	52	54
KategoriListe_KS	61	59	65	60	57	61	61	57	67	61	54	67
KategoriKarmaFrek_HS	53	53	51	43	38	57	50	50	50	50	50	50
KategoriKarmaFrek_KS	53	53	51	53	53	51	53	53	51	53	53	51
KategoriKarmaVarYok_HS	50	50	50	50	50	50	50	50	50	50	50	50
KategoriKarmaVarYok_KS	50	50	50	50	50	50	50	50	50	50	50	50
KelimeListe_HS	64	57	68	79	71	83	76	72	81	72	51	72
KelimeListe_KS	64	57	68	81	74	82	81	79	81	67	79	67
KelimeKarmaFrek_HS	80	60	81	82	78	81	82	69	81	61	56	75
KelimeKarmaFrek_KS	80	79	81	80	79	81	80	79	81	80	79	81
KelimeKarmaVarYok_HS	50	39	49	50	54	49	50	43	54	50	68	64
KelimeKarmaVarYok_KS	50	57	49	50	57	49	50	57	49	50	57	49
Cinsiyet_FREKANS												
Özellik	Eşik =0			Eşik =5			Eşik =25			Eşik =50		
	NB	SMO	J48	NB	SMO	J48	NB	SMO	J48	NB	SMO	J48
KategoriListe_HS	28	50	50	61	51	50	59	53	57	58	50	50
KategoriListe_KS	28	50	50	51	33	51	53	67	59	57	74	53
KategoriListeAğırlıklı_HS	44	50	50	62	58	79	55	50	56	54	50	50
KategoriListeAğırlıklı_KS	44	50	50	58	60	65	54	54	64	54	54	67
KategoriKarmaFrek_HS	46	56	75	56	50	74	54	56	56	50	54	47
KategoriKarmaFrek_KS	46	56	75	46	56	75	46	56	75	46	56	75
KategoriKarmaVarYok_HS	53	54	53	53	53	50	57	58	53	50	54	49
KategoriKarmaVarYok_KS	53	54	53	53	54	53	53	54	53	53	54	53
KelimeListe_HS	42	56	79	50	50	38	50	50	60	71	51	64
KelimeListe_KS	43	56	76	74	69	53	72	72	83	75	77	77
KelimeListeAğırlıklı_HS	74	67	58	61	68	75	77	67	78	76	68	71
KelimeListeAğırlıklı_KS	74	67	58	60	67	63	85	81	86	83	83	83
KelimeKarmaFrek_HS	72	63	71	72	68	71	72	68	71	68	65	75
KelimeKarmaFrek_KS	71	65	71	71	65	71	71	65	71	71	65	71
KelimeKarmaVarYok_HS	67	74	51	67	75	50	67	77	70	50	72	71
KelimeKarmaVarYok_KS	67	72	49	67	72	49	67	72	49	67	72	49

ÖZGEÇMİŞ

Kişisel Bilgiler	
Adı Soyadı	Mine MERCAN
Doğum Yeri	Uşak
Doğum Tarihi	09.09.1965
Uyruğu	☒ T.C. ☐ Diğer:
Telefon	05333692165
E-Posta Adresi	minemercan@gmail.com

Eğitim Bilgileri	
Lisans	
Üniversite	İstanbul Teknik Üniversitesi
Fakülte	Elektrik-Elektronik Fakültesi
Bölümü	Kontrol ve Bilgisayar Mühendisliği
Mezuniyet Yılı	13.02.1989

Yüksek Lisans	
Üniversite	Fatih Üniversitesi
Enstitü Adı	Fen Bilimleri Enstitüsü
Anabilim Dalı	Bilgisayar Mühendisliği Ana Bilim Dalı
Programı	Bilgisayar Mühendisliği Programı

Doktora	
Üniversite	İstanbul Üniversitesi
Enstitü Adı	Fen Bilimleri Enstitüsü
Anabilim Dalı	Bilgisayar Mühendisliği Anabilim Dalı
Programı	Bilgisayar Mühendisliği Programı

Makale ve Bildiriler	
Z.Şevkli, M.Mercan, A.Kurt, "A Middleware Approach to Storing and Querying XML Documents in Relational Databases", <i>Lecture Notes in Computer Science</i> (Springer ISSN: 0302-9743), 3261, 223-233 (2004)	
M. Mercan, Z. Orhan, A. Sertbaş S. Nuhoglu, "Computer-Aided Analysis of Written Texts for Detecting The Symptoms of Psychological Disorder", International Conference on Electronics, Computer and Computation (ICECCO 2012), Turgut Ozal University, Ankara.	
Z. Orhan, M.Mercan, M. Ceyhan, S. F. Giriş, "Text Analysis System for School-Age Children to Detect Psychiatric Disorders", International Conference on E-Education, (ICEE 2014), September 26-27, 2014, Bosnia and Herzegovina.	

- Z. Orhan, M. Mercan, A. Sertbaş, “Turkish Text Analysis System For Automatic Detection Of Psychiatric Disorders”, International Conference on Turkic Language Processing, 6-7 November 2014, (TURKLANG’14), İstanbul.
- Z. Orhan, M. Ceyhan, M. Mercan, S. F. Giriş, “Yazılı metinlerdeki dil kullanım özellikleri ile psikolojik rahatsızlıkların teşhisi”, TBD Ulusal Bilişim Kurultayı 6-9 Kasım 2014 (Bilişim 2014).
- M. Mercan, Z. Orhan, A. Sertbaş, “Computer-Aided Diagnosis of Warning Signs of Mental Disorders in the Early Stages by Text Analysis of School-Age Children”, *Health Informatics Journal*, 2018, (revizyon aşaması).

