



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



Doktora Tezi

**SÖZLÜK TABANLI YAKLAŞIM VE MAKİNE ÖĞRENMESİ İLE
TÜRKÇE DOĞAL DİL İŞLEME: EĞİTİM KURUMLARINDA
DUYGU ANALİZİ**

Harun AKSAYA

Enformatik Anabilim Dalı

Enformatik Programı

**DANIŞMAN
Prof. Dr. Sevinç GÜLSEÇEN**

Aralık, 2021

İSTANBUL

Bu alıřma, 22.12.2021 tarihinde ařağıdaki jüri tarafından Enformatik Anabilim Dalı, Enformatik Programında Doktora tezi olarak kabul edilmiştir.

Tez Jürisi

Prof. Dr. Sevin GÜLSEÇEN (Danışman)
İstanbul Üniversitesi
Fen Fakültesi

Prof. Dr. Zuhâl TANRIKULU
Boğaziçi Üniversitesi
Uygulamalı Bilimler Yüksekokulu

Prof. Dr. Cem Sefa SÜTÇÜ
Marmara Üniversitesi
İletişim Fakültesi

Do. Dr. Çiğdem EROL
İstanbul Üniversitesi
Fen Fakültesi

Dr. Öğr. Üyesi Ahmet Feyzi SATICI
Marmara Üniversitesi
Atatürk Eğitim 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

Bu doktora tez çalışması, yakın tarihte elim bir helikopter kazasında kaybettiğimiz, vatansever ve iyi bir insan olmanın ne demek olduğunu hayatıyla bize öğreten çok değerli abim Şehit Pilot Yüzbaşı Gökhan UYSAL’a atfolunmuştur. Ruhu şad olsun...

Doktora öğrenimim boyunca kendisinden çok şey öğrendiğim, tanımış olmaktan dolayı kendimi şanslı hissettiğim, tez çalışmam boyunca her daim desteğini eksik etmeyen danışmanım, çok kıymetli hocam Sayın Prof. Dr. Sevinç GÜLSEÇEN’e teşekkür eder, sevgi ve saygılarımı sunarım.

Tezimin her aşamasında bilgi birikimini ve tecrübesini paylaşarak benden desteğini esirgemeyen, tez çalışmama ismini veren, çalışmamın bu aşamaya gelmesini sağlayan, çok değerli hocam Sayın Prof. Dr. Zuhâl TANRIKULU’ya teşekkür eder, sevgi ve saygılarımı sunarım.

Gerek yüksek lisans gerekse de doktora öğrenimim boyunca görüş ve önerileri ile her daim ufkumu genişleten, bilgi birikimi ve tecrübesini paylaşarak benden desteğini esirgemeyen, çalışmamın bu aşamaya gelmesini sağlayan, çok değerli hocam Sayın Prof. Dr. Cem Sefa SÜTÇÜ’ye teşekkür eder, sevgi ve saygılarımı sunarım.

Tez çalışmam boyunca danıştığım her konuda destek olan ve tecrübelerini paylaşan Enformatik Bölümünün çok değerli hocaları Araş. Gör. Dr. Serra ÇELİK’e ve Araş. Gör. Dr. Emre AKADAL’a en içten teşekkürlerimi sunarım.

Doktora öğrenimim boyunca aynı sıraları paylaştığım, tez çalışmam boyunca her daim destek aldığım çok değerli arkadaşlarım Öğr. Gör. Dr. Kadir Burak OLGUN’a ve Araş. Gör. Dr. Mehmet ELİBOL’a en içten teşekkürlerimi sunarım.

Beni büyütüp bugünlere getiren, her konuda koşulsuz destekleyen, eğitimim için fedakârlıktan kaçınmayan, haklarını hiçbir zaman ödeyemeyeceğim canım ANNEM’e ve BABAM’a sonsuz teşekkürlerimi sunarım.

Son olarak, bu uzun ve zorlu süreçte dar zamanlarımı genişleten, kasvetli günlerimi aydınlığa yöneltten başlıca desteğim, biricik eşim Elif Hanım’a bir demet çiçektir bu tez çalışması.

Aralık 2021

Harun AKSAYA

İÇİNDEKİLER

Sayfa No

ÖNSÖZ	iv
İÇİNDEKİLER.....	v
ŞEKİL LİSTESİ	vii
TABLO LİSTESİ.....	xi
SİMGE VE KISALTMA LİSTESİ	xiii
ÖZET	xiv
SUMMARY	xvi
• GİRİŞ	1
• GENEL KISIMLAR.....	5
2.1. VERİ	5
2.2. VERİ MADENCİLİĞİ.....	7
2.2.1. Veri Madenciliğinde Kullanılan Yöntem ve Teknikler.....	8
2.2.2. Kullanım Alanları	13
2.3. METİN MADENCİLİĞİ NEDİR?	15
2.3.1. Metin Madenciliğinde Kullanılan Yöntem ve Teknikler	15
2.3.2. Kullanım Alanları	16
2.4. DOĞAL DİL İŞLEME.....	17
2.4.1. Doğal Dil İşleme Çalışmalarında Kullanılan Yöntem ve Teknikler	18
2.4.2. Doğal Dil İşlemenin Kullanım Alanları	18
2.4.3. Doğal Dil İşleme Aşamaları	19
2.4.4. Doğal Dil İşleme Araçları	20
2.5. DUYGU ANALİZİ	27
2.5.1. Duygu Analizi Çalışmalarında Kullanılan Yaklaşımlar.....	28
2.5.1.2. Sözlük Tabanlı Yaklaşım.....	30
2.5.1.3. Makine Öğrenmesi-Yapay Sinir Ağları (YSA) Yaklaşımı	39
2.5.2. Duygu Analizi Çalışmalarında Kullanılan Seviyeler	40
2.5.2.1. Doküman Seviyesinde Duygu Analizi	40
2.5.2.2. Cümle Seviyesinde Duygu Analizi	41
2.5.2.3. Hedef (Aspect) Seviyesinde Sentismental Analiz	41

2.6. BU ALANDA YAPILAN BENZER ÇALIŞMALAR	41
• MALZEME VE YÖNTEM	45
3.1. VERİ TOPLAMA SÜRECİ.....	45
3.1.1. Twitter Üzerinden Veri Toplama	45
3.1.2. Okul.com.tr Platformu Üzerinden Veri Toplama.....	45
3.2. DUYGU SÖZLÜKLERİNİN KULLANIMA HAZIR HALE GETİRİLMESİ	49
3.2.1. SentiTurkNet Sözlüğü'nün Kullanıma Hazır Hale Getirilmesi	49
3.2.2. SentiWordNet Sözlüğü 'nün Kullanıma Hazır Hale Getirilmesi	51
3.3. İTÜ DDİYZ'NİN KULLANIMA HAZIR HALE GETİRİLMESİ.....	56
3.4. VERİ SETİNİN İTÜ DDİYZ SERVİSİ İLE İŞLENMESİ.....	58
3.5. İLİŞKİLİ KELİME TESPİTİ İÇİN YAZILIM GELİŞTİRİLMESİ.....	66
3.6. MAKİNE ÖĞRENMESİ YAKLAŞIMI İÇİN MODELİN GELİŞTİRİLMESİ VE TEST EDİLMESİ.....	69
3.7. SÖZLÜK TABANLI YAKLAŞIM İÇİN MODELİN GELİŞTİRİLMESİ VE TEST EDİLMESİ.....	78
• BULGULAR	83
4.1. MAKİNE ÖĞRENMESİ YAKLAŞIMI İÇİN GELİŞTİRİLEN MODELE AİT BULGULAR.....	83
4.2. SÖZLÜK TABANLI YAKLAŞIM İÇİN GELİŞTİRİLEN MODELE AİT BULGULAR	97
• TARTIŞMA VE SONUÇ	100
5.1. ELDE EDİLEN SONUÇLARIN DEĞERLENDİRİLMESİ	100
5.2. SÖZLÜK TABANLI YAKLAŞIM VE MAKİNE ÖĞRENMESİ TEKNİKLERİNİN KARŞILAŞTIRILMASI	103
5.3. ÖNERİLER	106
KAYNAKLAR	107
EKLER	117
Ek 1: İTÜ DDİYZ Servisi İçin İmzalanan Kullanım Sözleşmesi	117
Ek 2: Makine Öğrenmesi Modeli İçin Geliştirilen Yapıya Ait Kod Blogu.....	118
Ek 3: Makine Öğrenmesi Modeli İçin Elde Edilen Doğruluk Oranlarına Ait Tüm Tablo	120
Ek 4: Sözlük Tabanlı Yaklaşım İçin Geliştirilen Yapıya Ait Kod Blogu	121
ÖZGEÇMİŞ	123

ŞEKİL LİSTESİ

Sayfa No

Şekil 1: Veri Madenciliğinin Diğer Disiplinler ile İlişkisi	8
Şekil 2: Karar Ağacının Yapısı.....	10
Şekil 3: Birliktelik Kurallarının Görselleştirilmesi	11
Şekil 4: Sinir Hücresini Yapısı.	12
Şekil 5: Bir Nöronun Bileşenleri	13
Şekil 6: Metin Madenciliği Aşamaları.....	16
Şekil 7: DDİ Piramidi	19
Şekil 8: İTÜ Türkçe Doğal Dil İşleme Yazılım Zinciri Aşamaları	26
Şekil 9: Duygu Analizi Çalışmalarında Kullanılan Yaklaşımlar.....	29
Şekil 10: Wordnet Kelime Örneği	30
Şekil 11: BalkaNet Görüntüleme ve Düzenleme Editörü.....	31
Şekil 12: KeNet'e Ait Veri Yapısı.....	34
Şekil 13: SenticNet Veri Yapısı.....	36
Şekil 14: HisNet'in Veri Yapısı	39
Şekil 15: Duygu Analizi Çalışmalarında Kullanılan Seviyeler	40
Şekil 16: Okul.com.tr Sitesinin Ana Sayfası	45
Şekil 17: Okul.com.tr Sitesinde Yer Alan Bir Okula Ait Bilgiler	46
Şekil 18: Veri tabanına Aktarılan Yorumlar.....	47
Şekil 19: Yorum Veri Seti İçindeki Hedefler ve Duygu Durumları.....	48
Şekil 20: SentiTurkNet Sözlüğünün MySQL Veri Tabanına Aktarılmış Haline Ait Bir Görünüm.....	50
Şekil 21: SentiWordNet Sözlüğünün Türkçeye Çevrilmesine Ait Adımlar.....	52
Şekil 22: SentiWordNet'in TXT Formatındaki Veri Yapısı	52

Şekil 23: SentiWordNet'in MySQL Veri Tabanına Aktarılmış Haline Ait Bir Görünüm.....	53
Şekil 24: PHP Programlama Dili Kullanılarak Geliştirilen Çeviri Sistemi.....	54
Şekil 25: Türkçeye Çevrilen SentiWordNet'in MySQL Veri Tabanına Aktarılmış Haline Ait Bir Görünüm.....	55
Şekil 26: İTÜ DDİYZ Servisini Test Etmek için Geliştirilen Yazılım Ara Yüzü.....	56
Şekil 27: İTÜ DDİYZ Servisi Tarafından İşlenen Bir Yorum Ait Çıktı	57
Şekil 28: Whole Pipeline Sınıfına Ait Metotlar ve Metnin İşlenme Sırasına Ait Bir Görünüm.....	58
Şekil 29: İTÜ DDİYZ -Tokenizer İşlemine Ait Görünüm	59
Şekil 30: İTÜ DDİYZ -Normalizer İşlemine Ait Görünüm	60
Şekil 31: İTÜ DDİYZ -Morphological Tagger İşlemine Ait Görünüm	61
Şekil 32: İTÜ DDİYZ -Named Entity Recognizer İşlemine Ait Görünüm.....	61
Şekil 33: İTÜ DDİYZ -Dependency Parser İşlemine Ait Görünüm	62
Şekil 34: Yorum Veri Setinin İTÜ DDİYZ ile İşlenmesi için Geliştirilen Yazılım.....	63
Şekil 35: İTÜ DDİYZ ile İşlenmesi için Gönderilen Örnek Bir Yorum Metni	64
Şekil 36: İTÜ DDİYZ ile İşlenmiş Örnek Bir Yorum Metni	64
Şekil 37: İTÜ DDİYZ ile İşlenen Veri Setinin MySQL Veri Tabanına Aktarılmış Haline Ait Bir Görünüm.....	65
Şekil 38: İlişkili Kelimelerin Tespit Edilmesi İçin Geliştirilen Yazılımdan Bir Kesit.....	66
Şekil 39: İlişkili Kelimelerin Tespit Edilmesine Ait Adımlar	67
Şekil 40: Hedef ve İlişkili Kelimelerin Veri Tabanına Aktarılmış Haline Ait Bir Görünüm.....	68
Şekil 41: Makine Öğrenmesi Yaklaşımı İçin Geliştirilen Modelin Yapısı	69
Şekil 42: Modele İşlenmesi İçin Verilen Metin.....	70
Şekil 43: Her Bir Kelimenin Sayısal Değere Dönüştürülmüş Hali	70
Şekil 44: Oluşturulan Kelime Vektörleri.....	70
Şekil 45: Nöron'un Çalışma Şekli	71
Şekil 46: Nöron'un Çalışma Formülü	71

Şekil 47: Aktivasyon Formülü.....	72
Şekil 48: Sigmoid Aktivasyon Fonksiyonuna Ait Grafik.....	72
Şekil 49: Tokenizer Dosyası Oluşturmak İçin Geliştirilen Kod Parçası	73
Şekil 50: Oluşturulan Tokenizer Dosyası	73
Şekil 51: Modelin Oluşturulmasında Kullanılan Kütüphanelere Ait Kod Parçası.....	74
Şekil 52: Veri Setini Eğitim ve Test için Bölen Kod Parçası.....	74
Şekil 53: Her Bir Yorum Metninin Tokenleştirilmesini Sağlayan Kod Parçası.....	75
Şekil 54: Tokenleştirilmiş Bir Yorum Metnine Ait Görünüm	75
Şekil 55: Embeddig ve RNN Yapısını Oluşturan Kod Parçası	76
Şekil 56: Modelin Eğitilmesini Sağlayan Kod Parçası ve Sonuca Ait Görünüm.....	77
Şekil 57: Modelin Test Edilmesini Sağlayan Kod Parçası ve Sonuca Ait Görünüm	77
Şekil 58: Eksiz İlişkili Kelimelerin Veri Tabanına Aktarılması.....	78
Şekil 59: Sözlük Tabanlı Yaklaşım İçin Geliştirilen Model.....	80
Şekil 60: Sözlük Tabanlı Yaklaşım İçin Geliştirilen Modelin Detaylı Hali.....	80
Şekil 61: Duygu Durumuna Karar Veren Fonksiyon	81
Şekil 62: Sözlük Tabanlı Yaklaşım Modelinin Test Edilmesi için Geliştirilen Kod Bloğu	82
Şekil 63: İki Katmanlı Yapıda Modelin Doğruluk Oranlarına Ait Grafik	84
Şekil 64: İki Katmanlı Yapıda Modelin Kayıp Oranlarına Ait Grafik	85
Şekil 65: Üç Katmanlı Yapıda Modelin Doğruluk Oranlarına Ait Grafik	86
Şekil 66: Üç Katmanlı Yapıda Modelin Kayıp Oranlarına Ait Grafik.....	87
Şekil 67: Dört Katmanlı Yapıda Modelin Doğruluk Oranlarına Ait Grafik.....	88
Şekil 68: Dört Katmanlı Yapıda Modelin Kayıp Oranlarına Ait Grafik	89
Şekil 69: Beş Katmanlı Yapıda Modelin Doğruluk Oranlarına Ait Grafik.....	90
Şekil 70: Beş Katmanlı Yapıda Modelin Kayıp Oranlarına Ait Grafik	91
Şekil 71: Modelin Doğruluk Oranlarına Ait Grafik – Tüm Varyasyonlar	92
Şekil 72: Modelin Kayıp Oranlarına Ait Grafik – Tüm Varyasyonlar.....	92

Şekil 73: Karar Matrisi	93
Şekil 74: Sklearn Kütüphanesi Kullanılarak Oluşturulan Karar Matrisi	94
Şekil 75: Kesinlik (<i>Precision</i>) Değerinin Hesaplanması için Kullanılan Formül	95
Şekil 76: Kesinlik (<i>Precision</i>) Değeri	95
Şekil 77: Duyarlılık (<i>Recall</i>) Değerinin Hesaplanması için Kullanılan Formül	96
Şekil 78: Duyarlılık (<i>Recall</i>) Değeri.....	96
Şekil 79: F-score'un Hesaplanması için Kullanılan Formül	96
Şekil 80: F-score Değeri	97
Şekil 81: F-score Sonuçlarına Ait Sklearn Kütüphanesi ile Elde Edilen Sonuçlar	97
Şekil 82: Sözlük Tabanlı Yaklaşımın Doğruluk Oranını (<i>Accuracy</i>) Hesaplamada Kullanılan Formül	98
Şekil 83: Sözlük Tabanlı Yaklaşım ile Elden Edilen Doğruluk Oranlarına Ait Grafik	99
Şekil 84: Birden Fazla Hedef Barındıran Bir Kullanıcı Yorumu	101

TABLO LİSTESİ

Sayfa No

Tablo 1: BalkaNet’de Yer Alan Diller ve Kelime Sayıları	31
Tablo 2: Turkish WordNet’e Ait İstatistikler	32
Tablo 3: Turkish WordNet’e Ait Veri Yapısı.....	32
Tablo 4: KeNet’e Ait İstatistikler	33
Tablo 5: SentiWordNet’in Veri Yapısı.....	35
Tablo 6: SentiWordNet Örnek Verisi.....	36
Tablo 7: SentiTurkNet’in Veri Yapısı	37
Tablo 8: HisNet Sözlüğüne Ait Kutup Kategorileri ve Kelime Sayıları	38
Tablo 9: Veri Setindeki Nitelikler ve Açıklamaları.....	47
Tablo 10: Veri Setindeki Hedefler ve Duygu Durumları	49
Tablo 11: SentiTurkNet Sözlüğünün Veri Yapısı	51
Tablo 12: SentiWordNet Sözlüğünün Veri Yapısı	53
Tablo 13: Türkçeye Çevrilen SentiWordNet Sözlüğünün Veri Yapısı	56
Tablo 14 : Veri Tabanına Aktarılan İşlenmiş Veri Setinin Veri Yapısı	65
Tablo 15: Duygu Durumlarına Göre Hedef Sayıları	67
Tablo 16: Hedef ve İlişkili Kelimelerin Yer Aldığı Veri Tabanının Veri Yapısı.....	68
Tablo 17: Eksiz İlişkili Kelimelerin Yer Aldığı Veri Tabanının Veri Yapısı	79
Tablo 18: Modelin Testi için Kullanılan Farklı Katman, Nöron ve Eğitim Sayısı Değerleri	83
Tablo 19: İki Katmanlı Yapıda Modelin Başarım Oranlarına Ait Tablo	84
Tablo 20: Üç Katmanlı Yapıda Modelin Doğruluk Oranlarına Ait Tablo	86
Tablo 21: Dört Katmanlı Yapıda Modelin Doğruluk Oranlarına Ait Tablo.....	88
Tablo 22: Beş Katmanlı Yapıda Modelin Doğruluk Oranlarına Ait Tablo	90

Tablo 23: Karar Matrisi – Değerlerin Eklenmiş Hali	94
Tablo 24: SentiTurkNet Sözlüğü ile Elden Edilen Doğruluk Oranı.....	98
Tablo 25: SentiWordNet-TR Sözlüğü ile Elden Edilen Doğruluk Oranı.....	98
Tablo 26: SentiWordNet-EN Sözlüğü ile Elden Edilen Doğruluk Oranı.....	99
Tablo 27: Makine Öğrenmesi Yaklaşımına Ait Swot Analizi Tablosu.....	104
Tablo 28: Sözlük Tabanlı Yaklaşımına Ait Swot Analizi Tablosu	105



SİMGE VE KISALTMA LİSTESİ

Kısaltmalar	Açıklama
API	: Application Programming Interface (Uygulama Programlama Arayüzü)
BA	: Bayes Ağı
CNN	: Convolution Neural Network (Konvolüsyonel Sinir Ağı)
DDİ	: Doğal Dil İşleme
DDİYZ	: Doğal Dil İşleme Yazılım Zinciri
DVM	: Destek Vektör Makineleri
GRU	: Gated Recurrent Unit (Geçitli Tekrarlayan Birim)
KTY	: Korpus Tabanlı Yaklaşım
LSTM	: Uzun Kısa Süreli Bellek
LR	: Logistic Regression (Lojistik Regresyon)
ME	: Maksimum Entropi
NBS	: Naive Bayes Sınıflandırıcı
NLTK	: Natural Language Toolkit (Doğal Dil İşleme Araç Seti)
PHP	: Hypertext Preprocessor (Üstünyazı Önışlemcisi)
POS	: Part-Of-Speech
SA	: Sinir Ağları
SQL	: Structured Query Language (Yapılandırılmış Sorgu Dili)
STN:	: SentiTurkNet
STY	: Sözlük Tabanlı Yaklaşım
SWN:	: SentiWordNet
RNN	: Recurrent Neural Network (Tekrarlayan Sinir Ağı)

ÖZET

DOKTORA TEZİ

SÖZLÜK TABANLI YAKLAŞIM VE MAKİNE ÖĞRENMESİ İLE TÜRKÇE DOĞAL DİL İŞLEME: EĞİTİM KURUMLARINDA DUYGU ANALİZİ

Harun AKSAYA

İstanbul Üniversitesi

Fen Bilimleri Enstitüsü

Enformatik Anabilim Dalı

Danışman : Prof. Dr. Sevinç GÜLSEÇEN

Bu tez çalışması kapsamında duygu analizi çalışmalarında yeni bir yaklaşım olan hedef tabanlı duygu analizi çalışması yapılmıştır. Çalışma kapsamında sözlük tabanlı yaklaşım ve makine öğrenmesi tekniklerinden olan yapay sinir ağları kullanılmıştır. Veri setinin oluşturulmasında okul.com.tr sitesi üzerinden elde edilen kullanıcı yorumları kullanılmıştır.

İlk olarak okul.com.tr platformu üzerinden kullanıcı yorumları toplanmıştır. Daha sonra yorumlar içerisinde geçen hedefler tespit edilmiştir. Hedeflerin duygu durumları etiketlenmiştir.

Etiketli veri seti oluşturma süreci tamamlandıktan sonra model oluşturma süreci başlamıştır. Sözlük tabanlı yaklaşım ile oluşturulan model üç farklı sözlük ile test edilmiştir. Bu sözlüklerden ilki Türkçe WordNet tabanlı geliştirilen ve içerisinde 14.795 kelime barındıran SentiTurkNet'dir. İkincisi, İngilizce WordNet tabanlı geliştirilen ve içerisinde 117.659 kelime barındıran SentiWordNet'dir. Üçüncü ise SentiWordNet'in Türkçeye çevrilmesi ile elde edilen ve 73.386 kelime barındıran SentiWordNet-TR'dir. Elde edilen sonuçlar doğrultusunda Türkçe tabanlı geliştirilen SentiTurkNet sözlüğü ile en yüksek doğruluk oranı (%87,7) elde edilmiştir. SentiWordNet sözlüğünün Türkçeye çevrilmesi ile elde edilen SentiWordNet-TR ile %84,1'lik doğruluk oranına ulaşılmıştır. Hedefleri niteleyen ilişkili kelimelerin İngilizceye çevrilmesi ve SentiWordNet sözlüğüne gönderilerek puanlanması senaryosunda ise %2'lik bir artış ile %86,12'lik bir doğruluk oranına erişilmiştir. Bu sonuçlardan da görüldüğü üzere kelime sayısı

bakımından az olmasına rağmen, Türkçe tabanlı olmasından dolayı en yüksek doğruluk oranı SentiTurkNet sözlüğü ile elde edilmiştir.

Makine öğrenmesi yaklaşımı kapsamında çok katmanlı yinelenen yapay sinir ağları (Recurrent Neural Network-RNN) kullanılmıştır. Bu doğrultuda 2, 3, 4, 5 katmanlı ve farklı nöron dizilimlerinden oluşan yapılar oluşturulmuştur. Bu yapıların oluşturulmasında Geçitli Tekrarlayan Birim (Gated Recurrent Unit-GRU) kullanılmıştır. Etiketli veri seti %70'e %30 olacak şekilde bölünmüştür. Bölünen etiketli verilerin %70'i modelin eğitimi için, geriye kalan %30'u ise modeli test etmek için kullanılmıştır. En yüksek doğruluk oranı 3 katmanlı 3-6-12 nöron dizilimli yapıda epoch değeri 10 ile elde edilmiştir. Elde edilen en yüksek doğruluk oranı %96,12' dir. En düşük doğruluk oranı ise 5 katmanlı 96-48-24-12-6 nöron dizilimli yapıda epoch değeri 15 ile elde edilmiştir. Elde edilen en düşük doğruluk oranı ise %92,07 dir. Geliştirilen modelde her bir duygu durumu için f-score değerleri de hesaplanmıştır. Pozitif kutup için precision değeri 0,97, recall değeri 0,96 ve bu doğrultuda f-score değeri 0,96 olarak elde edilmiştir. Negatif kutup için ise precision değeri 0,91, recall değeri 0,93 ve bu doğrultuda f-score değeri 0,92 olarak elde edilmiştir. Bu değerler geliştirilen modelin her bir duygu kutbunu (pozitif-negatif) yüksek oranda doğru sınıflayabildiğini göstermektedir.

Aralık 2021, 141. sayfa.

Anahtar kelimeler: Doğal Dil İşleme, Duygu Analizi, Hedef Tabanlı Duygu Analizi, Makine Öğrenmesi, Sözlük Tabanlı Yaklaşım, Eğitim Kurumları

SUMMARY

Ph.D. THESIS

TURKISH NATURAL LANGUAGE PROCESSING USING DICTIONARY BASED APPROACH AND MACHINE LEARNING: SENTIMENT ANALYSIS IN EDUCATIONAL INSTITUTIONS

Harun AKSAYA

İstanbul University

Institute of Graduate Studies in Sciences

Department of Informatics

Supervisor : Prof. Dr. Sevinç GÜLSEÇEN

Within the scope of this thesis, aspect-based sentiment analysis, which is a new approach in sentiment analysis studies, was conducted. Dictionary-based approach and artificial neural networks were used within the scope of the study. User comments obtained from the okul.com.tr website were used in the creation of the data set.

First of all, user comments were collected on the okul.com.tr platform. Then, the aspects mentioned in the comments were determined. The emotional states of the aspects were labeled.

After the labeled data set creation process was completed, the model creation process started. The model created with the dictionary-based approach has been tested with 3 different dictionaries. The first of these dictionaries is SentiTurkNet, which was developed based on Turkish WordNet and contains 14,795 words. The second is SentiWordNet, which was developed based on English WordNet and contains 117,659 words. The third is SentiWordNet-TR, which was obtained by translating SentiWordNet into Turkish and contains 73,386 words. In line with the results obtained, the highest accuracy rate (87,7%) was obtained with the Turkish-based SentiTurkNet dictionary. With SentiWordNet-TR, which was obtained by translating the SentiWordNet word into Turkish, an accuracy rate of 84,1% was achieved. Finally, an accuracy rate of 86,12% was achieved with an increase of 2% in the scenario translating the related words describing the targets into English and scoring by sending them to

the SentiWordNet dictionary. As can be seen from these results, although the number of words is low, the highest accuracy rate was obtained with the SentiTurkNet dictionary, since it is Turkish-based.

Within the scope of machine learning approach, multi-layer recurrent neural networks (Recurrent Neural Network-RNN) are used. In this direction, structures consisting of 2,3,4,5 layers and different neuron arrays were created. The Gated Recurrent Unit (GRU) was used in the creation of these structures. The labeled dataset was split into 70% as training the model and the remaining 30% as testing the model. The highest accuracy rate was obtained with the epoch value of 10 in the 3-layer structure with 3-6-12 neuron arrays. The highest accuracy rate achieved is 96,12%. The lowest accuracy rate was obtained with the epoch value of 15 in the 5-layer 96-48-24-12-6 neuron array structure. The lowest accuracy rate obtained is 92,07%. In the developed model, f-score values were also calculated for each emotion pole. Precision value for positive pole is 0,97, recall value is 0,96 and f-score value is obtained as 0,96. The precision value for the negative pole is 0,91, the recall value is 0,93, and the f-score value is obtained as 0,92. These values show that the developed model can classify each emotion pole (positive-negative) with a high degree of accuracy.

December 2021, 141. pages.

Keywords: Natural Language Processing, Sentiment Analysis, Aspect-Based Sentiment Analysis, Machine Learning, Dictionary-Based Approach, Educational Institutions

• GİRİŞ

Veri bu yüzyılda yeni nesil petrol olarak tanımlanmaktadır. Veriyi işlemek, içinde yer alan örtük bilgiyi ortaya çıkarmak ve bunu bir katma değere dönüştürmek günümüzün en önemli meziyetleri arasındadır. Veri tek başına bir anlam ifade etmez. Verinin anlamlı hale gelebilmesi ve bir katma değere dönüşebilmesi için yapılandırılması ve düzenli hale getirilmesi gerekmektedir. Böylelikle veri işlenebilir hale gelerek bilgiye dönüşebilir (Gülseçen ve diğ., 2014).

İnternet ve internet teknolojilerinin hayatımıza daha fazla girmesi ile birlikte, eğitimden finansa, eğlenceden iletişime kadar birçok sektörde ciddi dönüşümler yaşanmış, yeni kavram ve metodolojiler ortaya çıkmıştır. Bu dönüşümlerin ülkelerin gelişmişlik düzeyine bakılmaksızın tüm dünya üzerindeki nesiller üzerinde bir etki yarattığı görülmektedir. Aynı zamanda bu etkinin yeni bir toplumsal yapının oluşmasına zemin hazırladığı düşünülmektedir (Aytekin ve Sütçü, 2012).

Bilgi ve iletişim teknolojilerindeki gelişmeler, internet ve internet teknolojilerinin yaygınlaşması ve yeni medya kavramı, bireyleri içerik tüketen konumdan, içerik üreten konuma geçirmiştir. Bu da üretmiş olduğumuz verileri dramatik olarak artırmıştır. Bu gelişmelere bağlı olarak iletişim şekillerimiz de değişmiştir. Örneğin eski zamanlarda insanlar telefonlar üzerinden sesli iletişim kurarken, günümüzde metin tabanlı iletişim mecraları çok daha popüler ve yaygın olarak kullanılmaktadır. Bu da ürettiğimiz metinsel verilerin ciddi oranda artmasına neden olmuştur.

Günümüzde üretmiş olduğumuz verilerin büyük bir kısmının metin tabanlı olması, bu metinlerin düzenlenmesi, yapılandırılarak anlamlı hale getirilebilmesi için bir takım yöntem ve tekniklerin geliştirilmesini zorunlu kılmıştır. Üretilen bu verilerin işlenmesi ve anlamlı bir hale getirilmesi için insanüstü bir güce ihtiyaç vardır. İşte bu noktada makine öğrenmesi ve DDİ yöntem ve teknikleri bize yol göstermektedir (Aytekin ve Sütçü, 2018).

Yapay zekanın alt kümelerinden biri olan Doğal Dil İşleme (DDİ) yöntem ve metodları, makine öğrenmesinin gücünü de arkasına alarak insana özgü olan doğal dili kullanarak kendini ifade

edebilme, bir başkasının ne söylediğini anlayıp kavrayabilme gibi iletişim kurma meziyetlerinin makineler tarafından yapılabilmesini amaçlamaktadır.

DDİ yöntem ve teknikleri insan-bilgisayar etkileşiminde de köklü değişiklikler yaratmaktadır. Cep telefonlarımıza giren akıllı asistanlar ya da evlerimizde kullandığımız akıllı hoparlörler sayesinde istediğimiz bilgiye sesli bir komutumuz ile ulaşabiliyor, cihazlarımızı elimize almadan sadece sesimizi kullanarak yemek siparişi bile verebiliyoruz. Diğer taraftan DDİ yöntem ve teknikleri meslekleri de dönüştürmektedir. Örneğin çağrı merkezlerinde insan gücü ile yapılan hizmetler artık yerlerini akıllı chatbotlara bırakmaktadır. Bu tür örnekleri arttırmak mümkündür. Bu da göstermektedir ki DDİ disiplinine olan ilgi ve ihtiyaç her geçen gün artmaktadır (Hark ve diğ., 2017).

Eğitimden finansa, sağlıktan ulaşımaya kadar birçok endüstri ve sektör bilgi ve iletişim teknolojilerinde yaşanan gelişmelere bağlı olarak bir gelişme ve değişim süreci içerisinde. Son yıllarda ortaya çıkan Endüstri 4.0 kavramı da bu değişim ve dönüşümün bir sonucu olarak değerlendirilebilir. Endüstri 4.0 kavramı ile birlikte kurum ve kuruluşlar veri ve bilgi işlem teknolojilerini kullanarak karlılık ve verimliliklerini artırmayı hedeflemektedir (Yılmaz ve Tanrıku, 2017). Bu kapsamda özellikle sosyal medya verileri üzerinden itibar yoklaması yapmak, oluşturdukları kampanyaların, gündemlerin istenilen noktaya gelip gelmediğini analiz etmek için DDİ yöntem ve tekniklerine başvurumaktadırlar.

DDİ çalışmalarının en önemli alt kollarından biri de duygu analizi çalışmalarıdır. Duygu analizi ifadesi ya da diğer bir ifade ile duygu analizi literatüre 2003 yılında “Sentiment Analysis: Capturing Favorability Using Natural Language Processing” çalışması ile girmiştir (Tetsuya, 2003). Duygu analizi, metinsel içeriklerde yer alan markalara, kişilere, ürün ya da bir organizasyona ilişkin duygusal durum bilgilerinin DDİ yöntem ve yaklaşımlarını kullanarak ortaya çıkarılması işidir (Fessini ve diğ., 2014) ve bir sınıflama problemidir.

Duygu analizi için literatürde duygu-fikir madenciliği, duygu sınıflama ifadeleri de kullanılmaktadır (Karaöz ve Gürsoy, 2018; Can ve Alataş, 2017).

Duygu analizi çalışmaları doküman, cümle ve hedef (aspect) olmak üzere üç farklı seviyede yapılmaktadır. Doküman ve cümle seviyesinde yapılan duygu analizi çalışmalarında dokümanın veya cümlenin içerisindeki duygu durumların sayısına bakılmaksızın ortalama duygu durumu tespit edilmektedir. Bundan dolayı duygu durumu kayıpları yaşanabilmektedir.

Hedef seviyesinde duygu analizi çalışmaları ise son yıllarda popüler olmuş bir yaklaşımdır. Bu yaklaşıma göre bir metin içerisinde birden fazla duygu durumu olabilir. Metnin duygu durumuna karar verebilmek için önce metin içerisindeki hedeflerin belirlenmesi gerekir. Her hedefin duygu durumu ayrı ayrı tespit edilmelidir. Bu sayede duygu durumu kaybı minimize edilir.

Bu tez çalışması kapsamında hedef tabanlı duygu analizi tekniği kullanılarak okul.com.tr sitesinden elde edilen kullanıcı yorumlarının pozitif ve negatif olarak yüksek doğruluk oranıyla sınıflandırılması amaçlanmıştır. Sınıflandırma kapsamında sözlük tabanlı yaklaşım ve makine öğrenmesi tekniklerinden yapay sinir ağları (artificial neural networks) kullanılmıştır.

Çalışmanın “Genel Kısımlar” bölümünde “Veri” kavramı ele alınmış ve verinin önemi vurgulanmıştır. Verilerin işlenmesi ve bir katma değere dönüştürülmesi için “Veri Madenciliği” teknikleri ve kullanım alanlarına değinilmiştir. Veri madenciliğinin bir alt kolu olan “Metin Madenciliği” kavramı ele alınmış ve yöntem ile teknikleri irdelenmiştir.

Akabinde DDİ kavramı ele alınmış, yöntem ve teknikleri irdelenmiştir. DDİ’nin kullanım alanları, aşamaları ve araçları konusunda bilgi verilmiştir. Sonrasında DDİ’nin bir alt kolu olan duygu analizi kavramı ele alınmış ve bu aşamada kullanılan yaklaşım ve seviyeler konusunda bilgi verilmiştir. Son olarak alanda yapılan benzer çalışmalardan bahsedilerek “Genel Kısımlar” bölümü tamamlanmıştır.

Çalışmanın “Malzeme ve Yöntem” bölümünde uygulama kısmı ile ilgili bilgiler verilmiştir. İlk olarak veri toplama süreci anlatılmıştır. Akabinde verilerin içerisindeki hedeflerin belirlenmesi ve hedeflerin duygu durumlarının etiketlenmesi süreci anlatılmıştır.

Bir sonraki aşamada sözlük tabanlı yaklaşım kapsamında kullanılan duygu sözlüklerinin kullanıma hazır hale getirilme süreci anlatılmıştır. Bu kapsamda SentiTürkNet ve SentiWordNet sözlükleri için yapılan işlemler anlatılmıştır.

Sözlüklerin hazır hale getirilmesi ile birlikte metin ön işleme sürecinde kullanılan İTÜ DDİYZ servisinin hazır hale getirilmesi süreci anlatılmıştır. Bu servis kullanılarak tüm veri seti işlenmiş, işlenen veriler bir veri tabanına aktarılmıştır.

Hedef tabanlı duygu analizi çalışmalarında hedefi niteleyen kelimelerin tespit edilmesi ve bu kelimeler üzerinden hedefin duygu durumuna karar verilmesi gerekmektedir. Bu doğrultuda ilişkili kelimelerin tespit edilmesine dair bir yazılım geliştirilmiştir. Sonrasında modellerin geliştirilme sürecine geçilmiş ve bu süreç anlatılmıştır. Bu kapsamda makine öğrenmesi ve sözlük tabanlı yaklaşıma göre hazırlanan modeller anlatılmıştır.

Bulgular kısmında her iki yaklaşıma ait doğruluk oranları tablo ve grafikler ile sunulmuştur. Tartışma ve sonuç kısmında ise elde edilen sonuçlar değerlendirilmiş hem çalışma için hem de alan için önerilerde bulunulmuştur.

Tezin son bölümünde ise çalışma boyunca atıfta bulunulan kaynaklar ve ekler yer almaktadır.

• GENEL KISIMLAR

2.1. VERİ

En genel tanımıyla veri, veri tabanlarında yer alan üzerinde herhangi bir işlem yapılmamış ham değerlerdir (Prytherch, 2005). Bu değerler bir ortamdan gözlem ve/veya ölçümler ile elde edilmiş olabilirler (Şeker, 2013). Veriler tek başına bir anlam ifade etmezler. Anlamalı hale gelebilmesi için düzenlenmesi, yapılandırılması, yararlı hale getirilmesi gerekmektedir. Ancak bu şekilde veri bilgiye dönüşür (Gülseçen ve diğ., 2014).

Bir başka tanıma göre ise veri, tek başlarına kullanılamayan, enformasyon ve bilgi için temel oluşturan ham bilgidir (Yılmaz, 2017). Bu ham bilgiler pek çok mecradan elde edilebilirler. Bu mecralar gözlemler, araştırmalar, internet ağları ve her türlü sensörler olabilir (Doğan ve Aytekin, 2016).

Veri 21. yüzyılın petrolü olarak görülmektedir. Veriye sahip olmak ve bu veriyi işleyerek anlamalı hale getirmek ile bir değere dönüştürmek bu çağın en önemli meziyetleri arasındadır.

Veri – Enformasyon – Bilgi

İşlenmemiş veri, sensörlerin, cihazların, yazılım sistemlerinin, çeşitli aletlerin ürettiği, üzerinde herhangi bir işlem yapılmamış çıktılar olarak tanımlanabilir. Bu çıktılar bir konu hakkında bilgi sahibi olma konusunda en az güce sahiptirler (Davenport ve Prusak, 1988). Veriyi işleyerek, düzenleyerek ya da bir problemin çözümünde kullanıp yararlı hale getirerek bilgiye dönüştürebiliriz (Gülseçen ve diğ., 2014).

Bu kapsamda veri, enformasyon ve bilgi kavramlarını ele aldığımızda ise iki farklı dönüşüm karşımıza çıkar. Bunlardan ilki verinin enformasyona dönüşümü, diğeri ise enformasyonun bilgiye dönüşmesidir. Veri belli bir formalizasyon sürecinden geçirilip düzenlendiğinde enformasyona dönüşür. Belli bir amaç ve/veya problemin çözümü konusunda kullanılan enformasyon ise bilgiye dönüşmüş olur (Jensen, 2005).

Aşırı Bilgi Artışı

Eski zamanlarda insanoğlu bilgiye erişebilmek için kütüphanelere, okullara bilginin kaynağına gitmek durumundaydı. Dijitalleşen dünya özellikle internet ve iletişim teknolojilerinin

yaygınlaşması ile birlikte insanoğlunu mekândan bağımsız olarak bilgiye erişebilir hale getirdi. Bu durum şüphesiz çok önemli fırsatlar yarattı. Ancak beraberinde problemler de getirdi. Bilgiye kolay erişmek, kolay üretmek ve paylaşmak hem bilgi çöplüğünün doğmasına hem de aşırı bilgi artışına neden oldu. Bu bilgi bolluğu insanoğlunun istediği bilgiye erişimini zorlaştırdığı gibi fazla bilgiyi işleme konusunda da ciddi bir yük oluşmasına neden oldu (Gülseçen ve diğ., 2014).

Literatüre aşırı bilgilenme olarak da geçen bu durum insanoğlunun işleme kapasitesini aşan bilgiye maruz kalmasından dolayı bilginin yardımcı olmaktan ziyade öğrenme ve bilgiyi kullanma süreçlerine bir engel olduğu kanısını oluşturmaktadır (Bergamaschi ve Leiba, 2010). Bu noktada makinelerin işleme kapasiteleri kullanılarak ortaya çıkan bu durumun insanoğlu lehine dönüştürülmesi çok önem arz etmektedir.

Büyük Veri Kavramı

Son yıllarda dijitalleşen dünyaya paralel olarak ürettiğimiz verilerin boyutları artmış, bu da beraberinde yeni bir kavramı ortaya çıkarmıştır: Büyük veri (Gülseçen ve diğ., 2014).

Büyük veri kavramı ilk kez 2000 yılında gerçekleştirilen 8. Dünya Ekonometri kongresinde araştırmacı Francis X. Diebold tarafından ortaya atılmıştır (Gürsakar, 2020). Bu doğrultuda büyük veri için homojen bir yapısı olmayan çeşitli elektronik içeriklerden oluşan farklı büyüklüklerde tutulan ve geleneksel veri tabanı sistemlerinde saklaması ve işlenmesi çok mümkün olmayan veri kümeleri olduğunu söyleyebiliriz (Gahi ve diğ., 2016).

Büyük veri sayesinde önceleri çok fazla anlam ifade etmeyen, genellikle sistemler için yük sayılacak log kayıtları, sensör verileri vb. verilerden katma değer elde edilebileceği anlaşıldı. Büyük veri teknolojileri sayesinde veriler saklanmaya, analiz edilmeye ve bir amaç ya da bir problemin çözümünde kullanılmaya başlandı (Schönberger ve Cukier, 2013).

İnsanlar internet ve iletişim teknolojilerindeki gelişmelere bağlı olarak hem veri üreten hem de bu verileri tüketen bir konumdadır. Veri bilimi aşırı veri artışına bağlı olarak, bu verilerin makineler tarafından işlenirken ortaya çıkan zorlukları çözmeye çalışan bir disiplin olarak ortaya çıkmıştır (Strawn, 2016).

Bununla birlikte büyük veri bize dünya üzerindeki problemlerin çözümünde geleneksel problem çözme yolları dışında farklı bir yol olabileceğini göstermiştir. Ayrıca büyük veri üzerinden yeni sorunları düşünme imkânı verdiğini de görmekteyiz. İşte bu yeni yaklaşıma “veri güdümlü paradigma” ismi verilmektedir (Kim ve diğ., 2017).

2.2. VERİ MADENCİLİĞİ

Verinin artması ile birlikte, bu veriyi işlemek ve bilgiye dönüştürmek önemli hale gelmiştir. İşletmeler, kurum ve kuruluşlar veri tabanlarında yer alan verileri işleyip bir katma değer elde etmek için gayret ve çaba sarf etmektedir. Verininin büyüklüğünden dolayı bu verilerin insan gücü ile işlenemeyeceği aşikardır. Ham verinin işlenip bilgiye dönüştürülebilmesi için bir takım yöntem ve metotlara ihtiyaç vardır. Veri madenciliği bu ihtiyacı gidermek için ortaya çıkmış bir yöntemdir (Hand, 1998).

Veri madenciliği çok büyük veri yığınları içerisinde, bir amaç veya bir problemin çözümü doğrultusunda anlamlı veri elde etme süreci olarak tanımlanabilir (Ganesh, 2002). Başka bir tanıma göre ise veri madenciliği çok büyük miktardaki ham veri içerisinde gizli kalmış, kullanılabilir faydalı bilgilerin açığa çıkarılma sürecidir (Koyuncugil, 2007). Jacobs’a (1999) göre veri madenciliği, verinin tek başına sağlayamadığı bilgiyi çeşitli yöntem ve analizler ile sağlama işidir. Veri madenciliği çalışmalarındaki asıl amaç fark edilmemiş veri desenlerini tespit etmek ve bunları katma değere dönüştürmektir (Doğan, 2017).

Bu tanımlamalar doğrultusunda veri madenciliğini, veri yığınları içerisinde bir problemin çözümüne katkı sağlayacak, katma değer oluşturabilecek bilgiler elde etme süreci olarak tanımlamak mümkündür.

Verilerden faydalı bilginin çıkarılması sürecinde çok sayıda teknik ve yöntem kullanılır. Bu anlamıyla veri madenciliği multi disiplinler bir çalışma alanıdır. İstatistik, matematik, veri tabanı sistemleri, yapay zeka vb. alanlar veri madenciliği çalışmalarında bir arada kullanılır (Çankırı ve diğ., 2009).



Şekil 1: Veri Madenciliğinin Diğer Disiplinler ile İlişkisi (Çankırı ve diğ., 2009)

Şekil 1’de veri madenciliği çalışmalarının diğer disiplinler ile olan ilişkisi yer almaktadır.

2.2.1. Veri Madenciliğinde Kullanılan Yöntem ve Teknikler

Veri madenciliğinde kullanılan yöntem ve teknikleri denetimli (*supervised*) ve denetimsiz (*unsupervised*) olmak üzere iki kategoriye ayırabiliriz. Veri madenciliği modellerini ise 3 başlık altında incelemek mümkündür.

Bunlar:

1. Sınıflama (*Classification*)
2. Kümeleme (*Clustering*)
3. Birliktelik kurallarıdır (*Association Rules*).

Bu alandan en çok kullanılan modeller ise:

- Regresyon Analizi (*Regression Analysis*)
- Karar Ağaçları (*Decision Trees*)
- Birliktelik Kuralları (*Association Rules*)
- Yapay Sinir Ağları (*Artificial Neural Networks*) (Silahtaroğlu, 2008).

Regresyon Analizi

Regresyon analizi iki ya da daha fazla değişken arasındaki ilişkiyi belirlemek amacıyla yaygın olarak kullanılan bir analiz metotudur. Regresyon analizi sayesinde iki değişken arasında ilişki var olup olmadığı ve bu ilişki ile belirlenebilir (Gürbüz ve Şahin, 2014).

Bir regresyon modeli $Y = \beta_0 + \beta_1 X + \varepsilon$ şeklinde tanımlanabilir (Atkinson, 2000).

Bu bağlamda:

Y: Tahmin edilmeye çalışılan bağımlı değişken

X: Bağımsız değişken

β_1 : Doğrunun eğimi

β_0 : Kesişme noktası

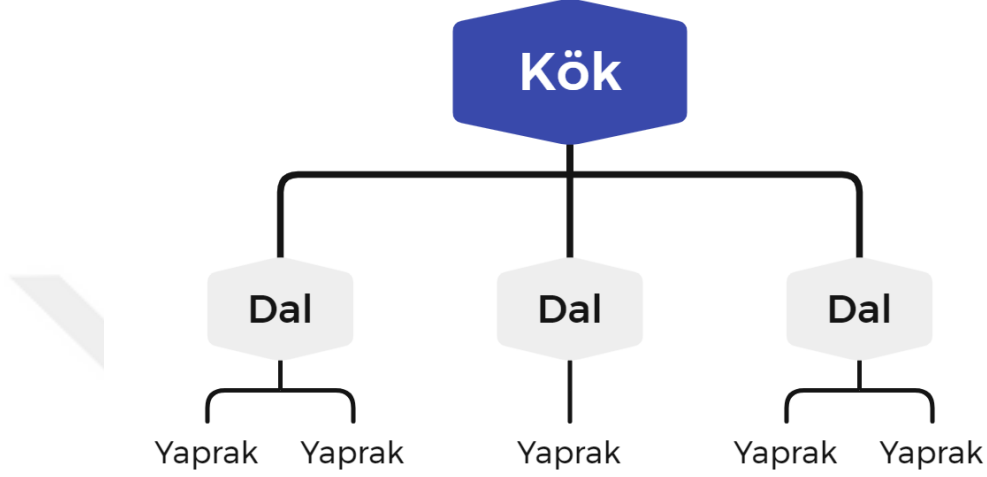
ε : Hata değerini temsil eder.

Karar Ağaçları

Veri madenciliğinde en çok kullanılan yöntemlerden birisi de karar ağaçlarıdır. Karar ağaçlarında en üste bulunan yapıya kök, alttaki elemanlara yaprak ve bunlar arasındaki yapılara ise dal ismi verilmiştir. Karar ağaçları bir kök düğüm ve bu kök düğüme bağlı iç düğümlerden oluşan bir yapıdır.

Bir önceki düğümden veri alan yapılara iç düğüm, çıktıları başka bir düğüm tarafından veri olarak alınmayan yapılara ise yaprak düğüm ismi verilmektedir. Bu teknik kurulumunun nispeten kolay olması, maliyetinin düşük olması, basit yapısı, parametrik olmamasından dolayı bilgi keşfi için uygun olması, veri tabanları ile kolayca entegre olabilmesi ve diğer tekniklere nazaran daha kolay oluşturulması gibi nedenlerden dolayı sınıflandırma ve tahmin etme süreçlerinde sıklıkla kullanılmaktadır (Alpaydın, 2000; Maimon ve Rokach, 2010; Gehrke, 2003).

Bununla birlikte karar ağaçları numerik veri setleri için karmaşık bir ağaç yapısı oluşturması, çok sayıda öznelilik çıktıları konusunda yetersiz olması ve veri setlerindeki küçük değişikliklerde bile stabilizasyon konusunda problemler oluşturması yöntemin negatif yönleri arasındadır (Zhao ve Zhang, 2008).



Şekil 2: Karar Ağacının Yapısı (Kavzoğlu ve Çölkesen, 2010)

Şekil 2’de karar ağacının kök, dal ve yapraklarından oluşan genel yapısı görülmektedir.

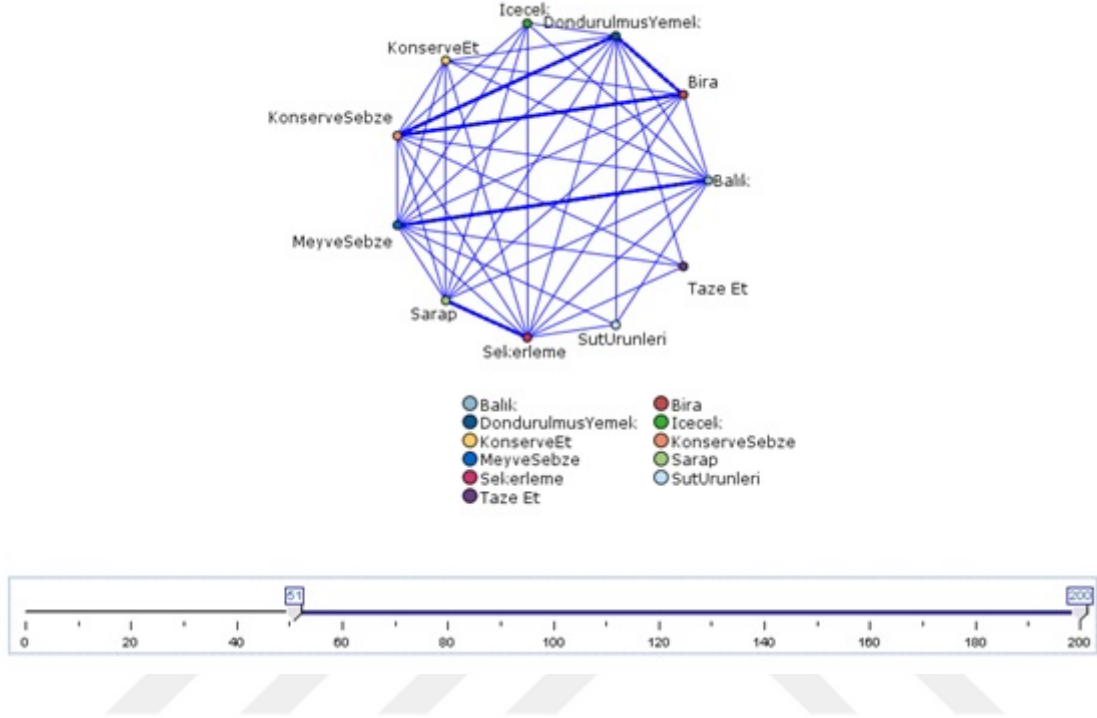
Birliktelik Kuralları

Birliktelik kuralları yapılan etkinliklerin birlikte gerçekleşme durumlarını çözümleyen ve tespit eden veri madenciliği yöntemlerinden biridir. Birliktelik kuralları çok büyük veri yığınları içerisindeki ilişkileri tespit ederler. Ve bu ilişkiler üzerinden yeni kurallar ortaya koyarlar.

Endüstri, veri tabanları içerisinde yer alan veriler arasındaki ilişkilerin tespit edilmesini özellikle satış ve karar süreçlerine katkı sağlamasından dolayı çok önemsemektedir. Bu noktada birliktelik kuralları veriler arasındaki birliktelik ilişkilerinin keşfedilmesi amacıyla çeşitli yöntem ve yaklaşımlar sunar.

Birliktelik kuralları anlatılırken genellikle market sepeti örneği verilir. Sepet içerisinde yer alan ürün verileri analiz edilerek en çok hangi ürün ile ne alındığı tespit edilir. Bu tespitler market-

ürün raf yerleşiminden, pazarlama kampanyalarına kadar birçok süreci doğrudan ve dolaylı olarak etkiler (Agrawal ve Srikant, 1995; Han ve Fu, 1999).



Birliktelik kuralları için yaygın olarak kullanılan algoritmalar şöyledir:

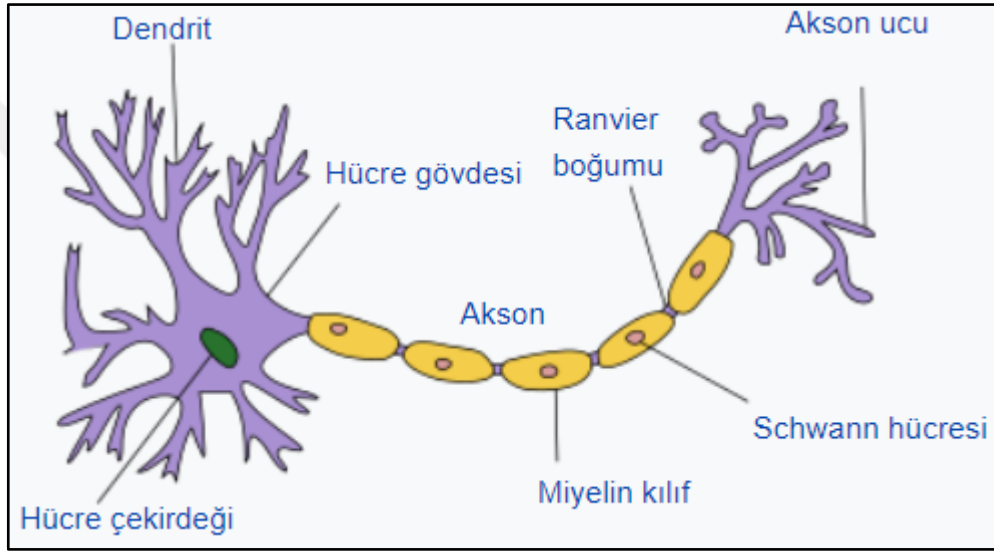
1. Carma
2. Sequence
3. Generalized Rule Induction (GRI)
4. Eclat
5. Apriori

Yapay Sinir Ağları

Yapay sinir ağları, insan beyninin çalışma yapısını referans alarak geliştirilmiş bir makine öğrenmesi modelidir. Yapay sinir ağları ile nöronların yapısı taklit edilir. Farklı bir tanıma göre ise ağırlıklı bağlantılar aracılığı ile birbirlerine bağlanan biyolojik sinir ağlarını referans alan

yazılımlardır. Aynı zamanda sinir ağları gibi davranan sentetik yapılar olduğunu belirten tanımlarda mevcuttur. Yapay sinir ağları, paralel dağıtılmış ağlar, nuroformik ağlar ve bağlantılı ağlar olarak da isimlendirilmektedir (Elmas, 2003; Ergezer ve diğ, 2003).

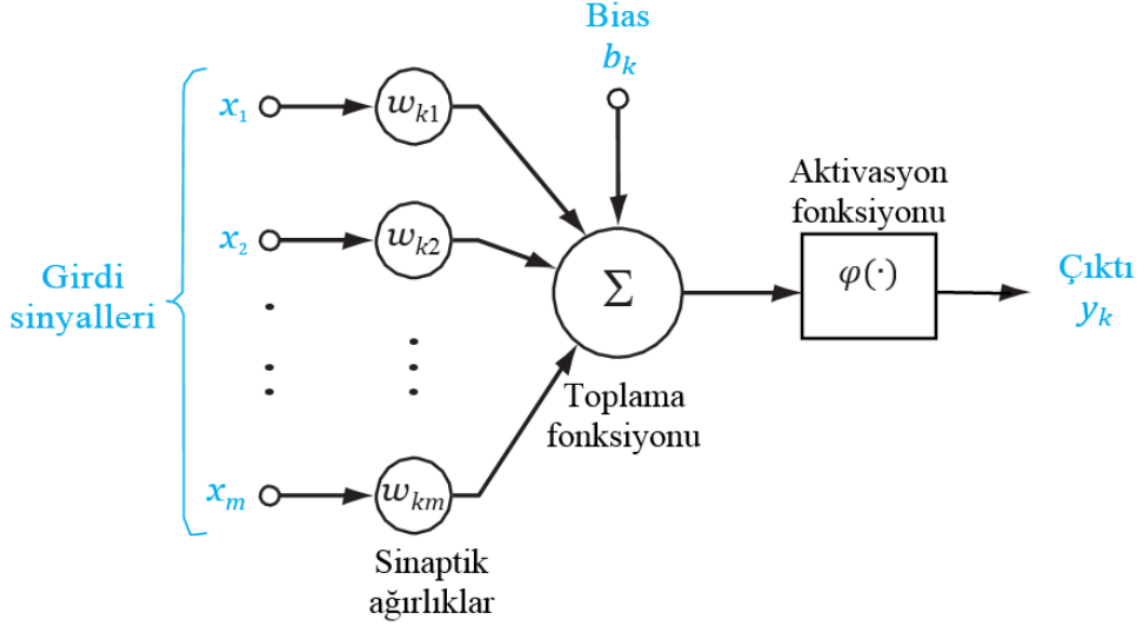
Bir yapay sinir ağı, ona verilen verileri analiz ederek genellemeler ve istatistiksel modeller oluşturur. Daha önce hiç karşılaşmadığı bir veriyi önceki öğrenmelerini kullanarak ne olduğunu tahmin edebilir ya da sınıflayabilir. (Eğrioğlu ve diğ., 2009). Aşağıdaki şekilde bir sinir hücresinin yapısı sunulmuştur.



Şekil 4: Sinir Hücresini Yapısı (Wikipedia, 2021).

Yapay sinir ağları insan sinir sistemini model alarak bu yapıyı bilgisayar sistemlerinde oluşturmaya çalışan, insanlar gibi öğrenme, tahminde bulunma, sınıflandırma gibi işlevleri yerine getiren sistemlerdir. Yapay sinir ağları yapay zeka ve makine öğrenmesinin bir alt disiplini (Kartal ve Özen, 2017; Haykin, 2008).

Yapay sinir ağları, nöron dediğimiz sentetik yapılardan oluşur. Aşağıdaki şekilde bir nörona ait bileşenler sunulmuştur.



Şekil 5: Bir Nöronun Bileşenleri (Kartal ve Özen, 2017)

Bir nöronun çalışma prensibi şu şekildedir (Fyfe, 2000; Kartal ve Özen, 2017).

1. Girdiler ($x_1, x_2, \dots$) her bir nöronda yer alan ağırlık (w) ile çarpılır.
2. Elde edilen çarpımlar ve bias nöronunun ağırlık değeri toplama fonksiyonunda işleme alınır ve toplam bir değer elde edilir.
3. Toplam değer bir sonraki katmana iletilir.
4. Çıktı katmanına gelen toplam değere aktivasyon fonksiyonu uygulanır.
5. Nöronda hesaplanan değer pozitif ise 1 değilse 0 değeri bir sonraki nörona iletilir.

2.2.2. Kullanım Alanları

Çankırı ve diğ.'ne (2017) göre veri madenciliği finans, astronomi, biyoloji, pazarlama, sigorta, bankacılık, tıp ve birçok başka alanda kullanılmaktadır. Bu uygulamaların sektörlere göre dağılımı aşağıda özetlenmiştir.

Bankacılık

- Risk yönetimi,
- Usulsüzlük tespiti

- Satış ve pazar tespiti
- Pazarlama faaliyetleri
- Sadakat programları
- Her türlü tahminleme faaliyetleri

Sigortacılık

- İstismar ve hırsızlık faaliyetleri
- Müşteri bazında poliçe satış tahminleme
- Sorunlu müşterilerin tespiti

Pazarlama

- Satış tahmini
- Kampanya oluşturma ve yönetme süreçleri
- Sadakat programlarının oluşturulması
- Alışveriş alışkanlıkları ve birlikteliklerin tespit edilmesi

Borsa

- Hisse senedi tahminleme
- Piyasa ve sektör analizi

Sağlık ve İlaç

- Teşhis süreçleri
- İlaç ve sağlık ürünleri geliştirme süreçleri
- Tedavi süreçleri

Endüstri

- Kalite kontrol süreçleri
- Optimizasyon ve verimlilik faaliyetleri

Bilim ve Mühendislik

- Bilimsel çalışmalardaki problemlerin çözümü

2.3. METİN MADENCİLİĞİ NEDİR?

Sosyal medya uygulamaları, forumlar, web siteleri ve mesajlaşma uygulamaları gibi internet ve iletişim teknolojileri odaklı gelişmeler daha çok metin tabanlı veriler (yapısal olmayan veriler) üretmemize neden oldu. Bu verilerin işlenmesi ve bir katma değere dönüştürülmesi hem bilim ve teknolojik çalışmalar için hem de endüstri için çok önem arz etmektedir. Metin madenciliği, metni bir veri kaynağı olarak ele alır ve bu metin üzerinden yapılandırılmış veriler elde etmeyi hedefler (Sagayam ve diğ., 2012). Metin Madenciliği ile:

- Metin Sınıflama
- Metin Bölütleme
- Metinden Konu Çıkarma
- Metin Özeti Çıkarma
- Metin Sınıfları Üretme
- Metin Duygu, Görüş Analizi çalışmaları yapılmaktadır (Şeker, 2015).

Metin madenciliği, veri madenciliğinin bir alt kolu olmakla birlikte, metin madenciliği örüntüleri ve istatistiksel modelleri veri tabanlarından ziyade doğal dil metinlerinden çıkarmaktadır (İlhan ve diğ., 2008).

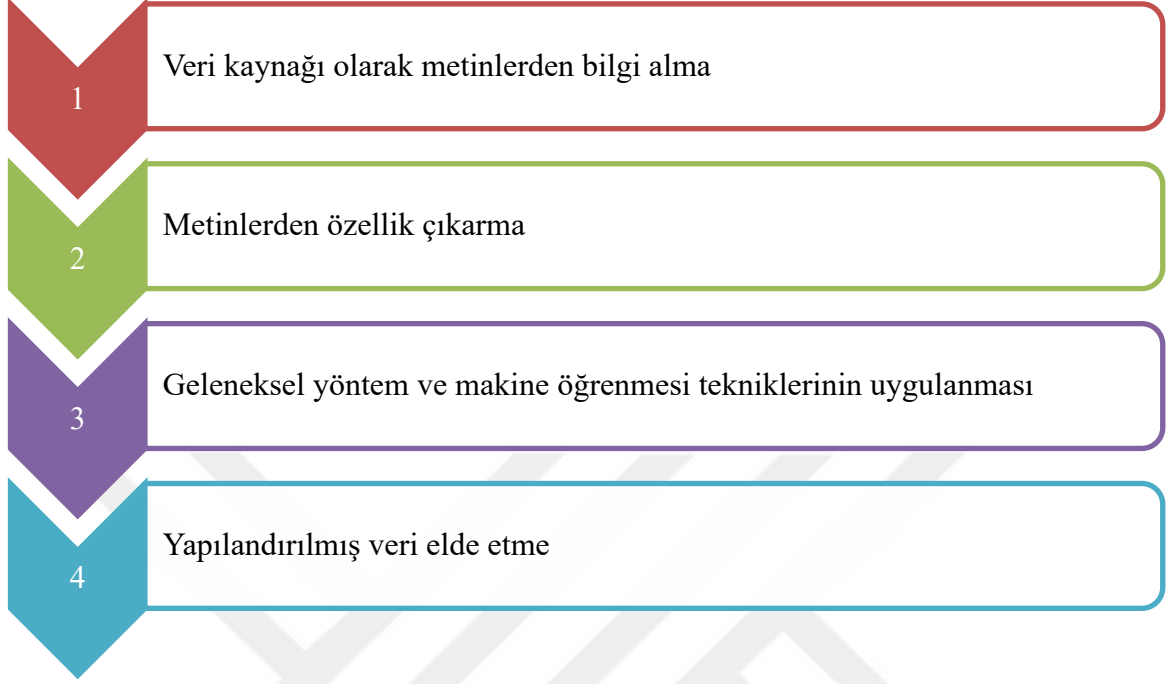
Metin madenciliği disiplinler arası bir çalışma alanıdır. Bu alanda proje ve çalışmalar yapabilmek için enformatik, makine öğrenmesi, veri madenciliği, istatistik ve dil bilim uzmanlıklarına gerek duyulmaktadır. Aynı zamanda doğal dillere bağlı olarak farklı yaklaşım ve metotlar da geliştirmek gerekebilir (Weiss ve diğ., 2010).

2.3.1. Metin Madenciliğinde Kullanılan Yöntem ve Teknikler

Veri madenciliği için kullanılan yöntem ve tekniklerin metin madenciliği içinde kullanıldığını söylenebilir. Bunları; sınıflandırma, birliktelik analizi, bilgi çıkarım ve kümeleme olarak sıralayabiliriz (Tunalı, 2011). Metin madenciliği çalışmaları aşağıdaki şekilde de görüldüğü üzere dört aşamadan oluşmaktadır. Bunlar sırasıyla (Şeker, 2015):

1. Veri kaynağı olarak metinlerden bilgi alma
2. Metinlerden özellik çıkarma
3. Geleneksel yöntem ve makine öğrenmesi tekniklerinin uygulanması

4. Yapılandırılmış (yapısal) veri elde etme.



Şekil 6: Metin Madenciliği Aşamaları (Şeker, 2015)

2.3.2. Kullanım Alanları

Metin tabanlı iletişim ve üretimlerimiz hemen hemen her sektörde son yıllarda dramatik olarak artış göstermiştir. Hal böyleyken metin madenciliği birçok sektör için rekabet ve katma değer oluşturma açısından ciddi fırsatlar barındırmaktadır (Çelik, 2020; Şeker, 2015).

- Finans
- Yayıncılık
- Veri Keşif
- Arama ve Bilgiye Erişme
- Sosyal Medya Ölçümleme
- Ulusal Güvenlik ve İstihbarat
- Bilimsel Çalışmalar
- Pazarlama

Metin madenciliğinin kullanıldığı alanlardan bazılarıdır (Fan ve diğ., 2006; Şeker, 2015).

2.4. DOĞAL DİL İŞLEME

Doğal dil insanların birbirleri ile iletişime geçmek için kullandıkları ve kendilerini ifade etmelerine imkan sağlayan bir araçtır (Chomsky, 1986). İnsanlığın sahip olduğu kültürel, sanatsal ve bilimsel birikiminin arkasında kullanmış oldukları dil vardır.

Diller yaşayan organizmalar gibidir. Yeni sözcükler eklenir, kullanılmayan sözcükler unutulurlar. Ancak dillerin yapıları ve kuralları uzun süre değişmeden kalırlar. Doğal Dil İşleme (DDİ) adı verilen bilim alanı bilişim teknolojilerinin gücünü arkasına alarak insan dillerinin makineler tarafından işlenebilmesini hedefler (Adalı, 2012).

En basit tanımıyla DDİ insanlara özgü dillerin bilgisayar sistemleri içerisinde işlenerek kullanılmasını amaç edinen bir yapay zeka alt disiplini. Dillerin matematiksel modellerini oluşturarak bilgisayarlar tarafından anlaşılmasını sağlar ve yeniden üretilmesi için çözümler geliştirir.

Yine başka bir tanıma göre DDİ, doğal dilleri matematiksel olarak temsil ve analiz etmek için kullanılan bir disiplindir (Khurana ve diğ., 2017). Ayrıca literatürde DDİ için istatistiksel dil modelleme olarak tanımlamalar da bulunmaktadır (Ofłazer, 2016). DDİ metinsel verilerin işlenebilmesi için teknik ve yöntemler içerir (Saquete ve diğ., 2020).

Bununla birlikte DDİ, yapay zeka, dilbilim ve bilgisayar destekli dilbilim gibi çok farklı disiplinlerdeki kuram, yöntem ve teknolojileri bir araya getiren bir bilim dalıdır (Delibaş, 2008).

Cambria ve White'a (2014) göre ise DDİ, insan dilinin otomatik işleme ve analizi için geliştirilmiş bir takım hesaplama teknikleridir. Bu alandaki DDİ çalışmalarındaki amaç doğal dilleri aynı bir insan gibi anlama ve oluşturabilme süreçleri için yöntem ve modeller geliştirmektir (Young ve diğ., 2018).

DDİ çalışmalarında doğal dil ifadelerine dilbilgisi kuralları uygulamak yerine dilin istatistiksel modelleri üzerinde durulur. Bu alan dil bilimi ve DDİ yöntemlerinin birlikte kullanılması ile mümkün olmaktadır (Schütze ve Manning, 1999).

İnternet ve web teknolojilerinin gelişmesi ile birlikte ürettiğimiz metinsel verilerde dramatik bir artış gözlemlenmektedir. Bu verilerin anlamlı hale getirilmesi endüstriler ve pazarlama sektörü için çok önemlidir. Bununla birlikte yapılandırılmamış doküman, belge vb. arşiv yığınları da anlamlandırılmaya muhtaç durumdadır. Bu durum DDİ çalışmalarına olan ilgiliyi her geçen gün artırmaktadır (Hark ve diğ., 2017).

Metin madenciliği çalışmaları çoğu zaman DDİ çalışmaları ile birlikte yürütülmektedir. Bununla birlikte ikisi arasında şöyle bir fark vardır. DDİ çalışmalarında makine öğrenmesi temelli yaklaşımlar kullanılırken, metin madenciliği çalışmalarında istatistiksel ve matematiksel yöntemler kullanılır (Göker, 2017).

2.4.1. Doğal Dil İşleme Çalışmalarında Kullanılan Yöntem ve Teknikler

Metin ve veri madenciliğinde kullanılan yöntem ve teknikler DDİ çalışmalarında da kullanılmakta ve bunlar iki farklı yaklaşım ile yapılmaktadır. Bunlar:

1. Kural Tabanlı Yaklaşımlar.
2. Makine Öğrenmesi Tabanlı Yaklaşımlar.

Literatüre göre makine öğrenmesi temelli yaklaşımlar, kural tabanlı yaklaşımlara göre daha iyi sonuçlar vermektedir (Bill, 1992; Gimenez ve Marquez, 2004; Song ve Lee, 2013).

2.4.2. Doğal Dil İşlemenin Kullanım Alanları

Adalı'ya (2012) göre DDİ kapsamında şu çalışmalar yapılmaktadır:

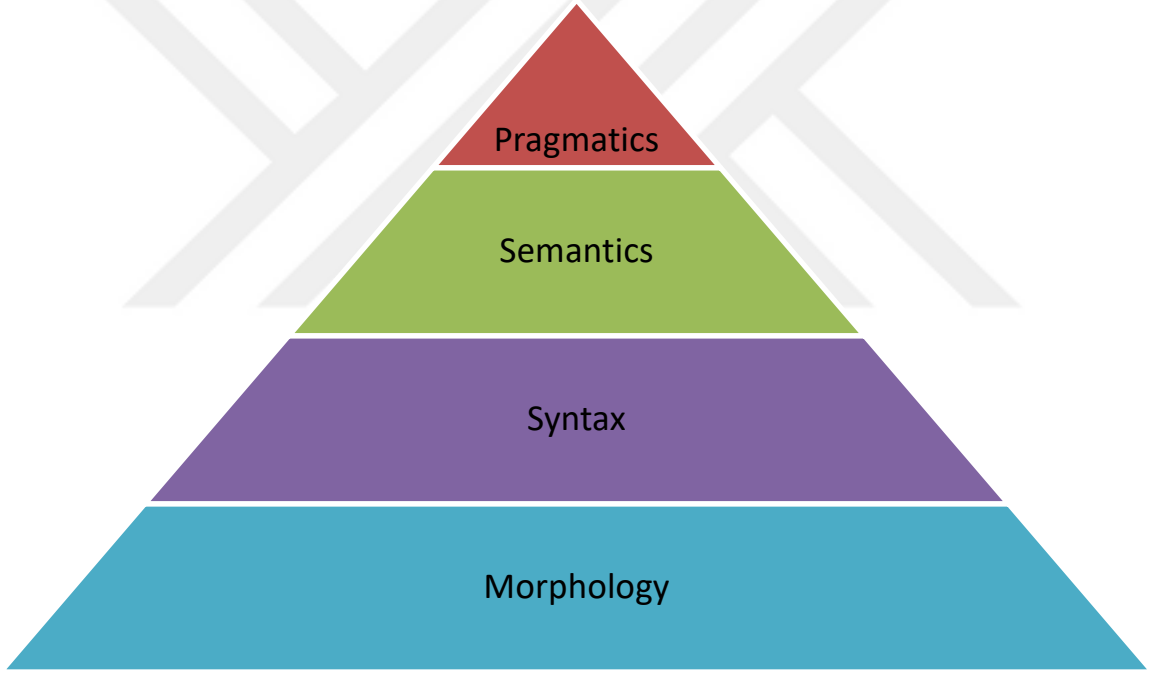
- Basılı bir metni okuma (optik olarak metin okuma) ve okuma yanlışlarını düzeltme
- Metinden özet çıkarma
- Metinden bilgi çıkarma
- Bilgiye erişim sağlama
- Bir metni anlama
- Bilgisayarla sesli etkileşim
- Bir metni seslendirme

- Ses dalgalarını metne dönüştürme
- Soru yanıt dizgeleri (Chatbotlar)
- Okuma yardımcı araçları
- Yazım yardımcı araçları
- Diller arası çeviri

2.4.3. Doğal Dil İşleme Aşamaları

Doğal Dil İşleme Piramidi

Towards Data Science'a (2017) göre DDİ çalışmaları dört fazlı bir piramitten oluşmaktadır.



Şekil 7: DDİ Piramidi (Toward Data Science, 2017).

1. Morphology-Biçim Bilimsel Analiz Çalışmaları

- Yazım ve imla hataları
- Cümle içindeki kelimelerin tespiti
- Kelimelerin türleri ve aldığı ekler

2. Syntax-Söz Dizimsel Analiz Çalışmaları

- a. Cümle kurulum yapısı
- b. Gramer yapısı

3. Semantic-Anlamsal Analiz Çalışmaları

- a. Varlık tanıma (Şehir, ülke, özel isim)
- b. Kelimeler arası ilişkilerin tespit edilmesi

4. Pragmatics-Pragmatik Analiz Çalışmaları

- a. Metinden konu tespiti
- b. Özet çıkarma
- c. Yazar tespit etme
- d. Duygu analizi
- e. Soru cevap sistemleri
- f. Çeviri sistemleri

2.4.4. Doğal Dil İşleme Araçları

DDİ çalışmalarında piramidinin ilk üç aşaması olan biçim bilimsel, söz dizimsel ve anlamsal analiz çalışmaları seviyesindeki süreçlerde yer alan işlemleri gerçekleştiren birtakım araçlar vardır. Bu araçlar DDİ piramidin ilk 3 aşamasında yer alan süreçleri tamamlayarak, piramidin dördüncü aşamasında yer alan pragmatik analiz çalışmalarına odaklanmamızı sağlarlar.

DDİ çalışmaları için hem Türkçe hem de diğer diller için birçok araç ve kütüphaneler geliştirilmiştir. Bunlar:

1. NLTK (Natural Language Toolkit)

NLTK çeşitli DDİ araçları sunan Python programlama dili kullanılarak geliştirilmiş bir DDİ kütüphanesidir. Sahip olduğu özellikler şunlardır (Bird, 2006):

- Tokenizasyon işlemleri (Tokenization)
- Stopwords kaldırma işlemleri (Stopwords)

- Normalizasyon işlemleri (Normalization)
- Kelime tür tespiti işlemleri (Part-of-speech (POS) Tagging)
- Duygu analizi işlemleri (Sentiment Analysis)
- Metin sınıflama işlemleri (Text Classification)
- Erişim Adresi: <https://www.nltk.org/>

2. SpaCy

Spacy, Python programlama dili kullanılarak geliştirilmiş, açık kaynak kodlu bir DDİ kütüphanesidir. Sahip olduğu özellikler şunlardır (Schmitt ve diğ., 2019):

- Tokenizasyon işlemleri (Tokenization)
- Kelime tür tespiti işlemleri (Part-of-speech (POS) Tagging)
- Kelimeler arası bağılıkların tespiti işlemleri (Dependency Parsing)
- Lemmatization (Kelimelerin köküne erişme işlemleri)
- Tümcelerin sonunu belirleme işlemleri (Sentence Boundary Detection)
- Varlık tanıma işlemleri (Named Entity Recognition)
- Varlık bağlantıları işlemleri (Entity Linking)
- Benzerlik işlemleri (Similarity)
- Metin sınıflama işlemleri (Text Classification)
- Kural tabanlı eşleştirme işlemleri (Rule-based Matching)
- Eğitim işlemleri (Training)
- Serileştirme işlemleri (Serialization)
- Erişim Adresi: <https://spacy.io/>

3. TextBlob

TextBlob, metin verilerini işlemek için Python programlama dili ile geliştirilmiş bir DDİ kütüphanesidir. Sağladığı özelliklere erişmek için bir Application Programming Interface (Uygulama Programlama Arayüzü) API arayüzü sunar. Sahip olduğu özellikler şunlardır (Bose ve diğ., 2020):

- İsim çıkarma işlemleri (Noun Phrase Extraction)
- Kelime tür tespiti işlemleri (Part-of-speech (POS) Tagging)
- Duygu analizi işlemleri (Sentiment Analysis)
- Sınıflandırma işlemleri (Classification)
- Tokenizasyon işlemleri (Tokenization)
- Kelime ve cümle sıklıkları işlemleri (Word and Phrase Frequencies)
- Ayırıştırma işlemleri (Parsing)
- Çoğullaştırma ve tekilleştirme işlemleri (Word Inflection)
- N-Grams işlemleri (N-Grams)
- Yazım düzeltme işlemleri (Spelling Correction)
- Model ve dil ekleme işlemleri (Models or Languages Through Extensions)
- WordNet entegrasyonu (WordNet Integration)
- Erişim Adresi: <https://textblob.readthedocs.io/en/dev/>

4. Stanford CoreNLP

Standford Üniversitesi DDİ çalışma grubu tarafından Java programlama dili kullanılarak geliştirilen açık kaynak kodlu bir DDİ kütüphanesidir. Sahip olduğu özellikler şunlardır (Manning ve diğ., 2014):

- Tokenizasyon işlemleri (Tokenization)
- Cümle bölme işlemleri (Sentence Splitting)
- Kelime tür tespiti işlemleri (Part-of-speech (POS) Tagging)
- Kelimelerin köküne erişme işlemleri (Lemmatization)
- Varlık tanıma işlemleri (Named Entity Recognition)
- Kelimeler arası bağılıkların tespiti işlemleri (Dependency Parsing)
- Erişim Adresi: <https://stanfordnlp.github.io/CoreNLP/>

5. Scikit-Learn

DDİ çalışmalarında en çok kullanılan makine öğrenmesi yöntem ve tekniklerini içeren Python programlama dili kullanılarak geliştirilen bir kütüphanedir. Sahip olduğu özellikler şunlardır (Pedregosa ve diğ., 2011):

- Sınıflandırma işlemleri (Classification)
- Regresyon işlemleri (Regression)
- Kümeleme işlemleri (Clustering)
- Boyut indirgeme işlemleri (Dimensionality Reduction)
- Özellik çıkarma ve normalleştirme işlemleri (Preprocessing)
- Erişim Adresi: <https://scikit-learn.org/>

6. Polyglot

Polyglot çok dilli uygulamaları destekleyen java tabanlı bir DDİ kütüphanesidir. Sahip olduğu özellikler şunlardır (Nystrom ve diğ., 2003):

- Tokenizasyon işlemleri – 165 dil (Tokenization)
- Dil algılama işlemleri – 196 dil (Language Detection)
- Varlık tanıma işlemleri – 40 dil (Named Entity Recognition)
- Kelime tür tespiti işlemleri – 16 dil (Part-of-speech (POS) Tagging)
- Duygu analizi işlemleri – 136 dil (Sentiment Analysis)
- Kelime gösterim işlemleri – 137 dil (Word Embeddings)
- Morfolojik analiz işlemleri – 135 dil (Morphological Analysis)
- Harf çevirisi işlemleri – 69 dil (Transliteration)
- Erişim Adresi: <https://polyglot.readthedocs.io/en/latest/index.html>

7. Torchtext

Torchtext, pytorch kütüphanesi kullanılarak geliştirilmiş içinde DDİ araçları ve popüler veri setleri barındıran açık kaynak kodlu DDİ platformudur. Sahip olduğu özellikler şunlardır (Puri ve Catanzaro, 2019):

- Duygu analizi işlemleri (Sentiment Analysis)
- Soru Sınıflandırma işlemleri (Question Classification)
- Dil modelleme işlemleri (Language Modeling)
- Makine çevirisi işlemleri (Machine Translation)
- Soru cevap işlemleri (Question Answering)
- Erişim Adresi: <https://pytorch.org/text/stable/index.html>

8. AllenNLP

AllenNLP, Pytorch kütüphanesi kullanılarak geliştirilmiş içinde derin öğrenme modelleri barındıran açık kaynak kodlu DDİ platformudur. Sahip olduğu özellikler şunlardır (Gardner ve diğ., 2018):

- Sınıflandırma işlemleri (Classification)
- Varlık tanıma işlemleri (Named Entity Recognition)
- Duygu analizi işlemleri (Sentiment Analysis)
- Kelime tür tespiti işlemleri (Part-of-speech (POS) Tagging)
- Erişim Adresi: <https://allennlp.org/>

9. Zemberek

Zemberek, Türkçe dili için geliştirilen java tabanlı açık kaynak kodlu bir DDİ kütüphanesidir. Sahip olduğu özellikler şunlardır (Akın ve Akın, 2007):

- Morfolojik analiz işlemleri (Morphological Analysis)
- Tokenizasyon işlemleri (Tokenization)
- Normalizasyon işlemleri (Normalization)
- Varlık tanıma işlemleri (Named Entity Recognition)
- Sınıflandırma işlemleri (Classification)
- Dil tanıma işlemleri (Language Identification)
- Dil modelleme işlemleri (Language Modeling)
- Erişim adresi: <https://github.com/ahmetaa/zemberek-nlp#readme>

10. TurkishNLP

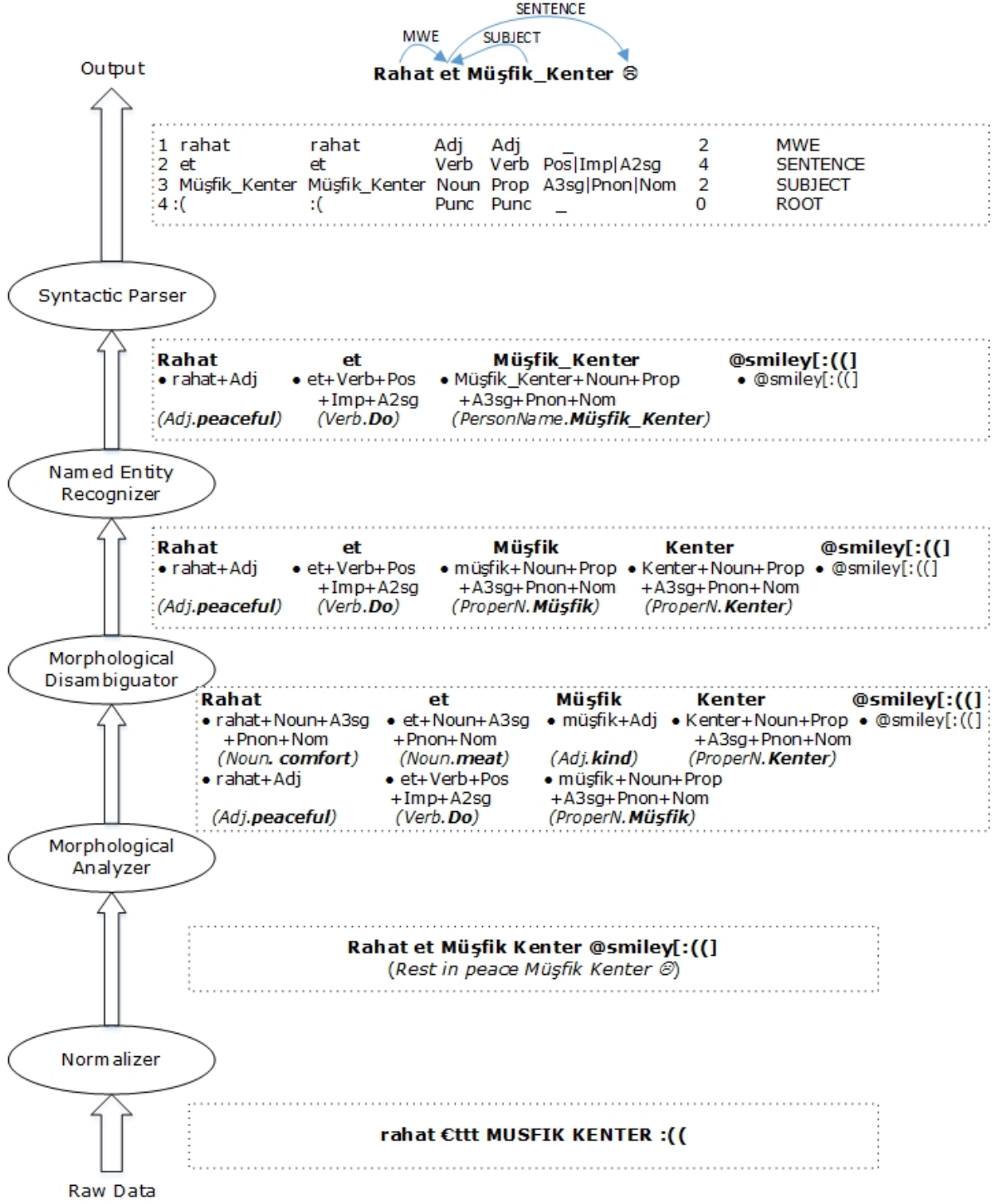
TurkishNLP, Python kullanılarak geliştirilmiş henüz beta aşamasında olan açık kaynak kodlu DDİ kütüphanesidir. Sahip olduğu özellikler şunlardır:

- Türkçe dil tespit işlemleri (Language Detection)
- Yazım hatası düzeltme işlemleri (Typo Correction)
- Hecelere ayırma işlemleri (Syllabication)
- Büyük ünlü uyumu işlemleri (Vowel Harmony)
- Türkçe köken kontrolü işlemleri (Is Turkish Origin)
- Boşluksuz yazılan yazıyı düzeltme işlemleri (Correct Text Without WhiteSpace)
- Erişim Adresi: <https://github.com/MeteHanc/turkishnlp>

11. İTÜ Doğal Dil İşleme Yazılım Zinciri (İTÜ DDİYZ)

İTÜ doğal dile işleme yazılım zinciri, İTÜ DDİ çalışma gurubu tarafından geliştirilmiş bir Türkçe DDİ platformudur. Bu platform şu özelliklere sahiptir (Eryiğit, 2014).

- Tokenizasyon işlemleri (Tokenizer)
- Ünlü ve aksanlı kelime İşlemleri (Deasciifier)
- Yazım düzeltme işlemleri (Spelling Corrector)
- Normalizasyon işlemleri (Normalizer)
- Türkçe kontrol işlemleri (isTurkish)
- Morfolojik analiz işlemleri (Morphological Analyzer)
- Varlık tanıma işlemleri (Named Entity Recognizer)
- Bağımlılık ayrıştırma işlemleri (Dependency Parser)



Şekil 8: İTÜ Türkçe Doğal Dil İşleme Yazılım Zinciri Aşamaları (Eryiğit, 2014)

Şekil 8’de İTÜ DDİ yazılım zincirinin beş aşamadan oluşan yapısı görülmektedir. Sisteme “Raw Data” olarak giren bir metin, yazılım zincirinin sonunda işlenmiş ve bir dizi haline gelmiş şekilde çıkmaktadır.

2.5. DUYGU ANALİZİ

Duygu analizi ifadesi literatüre 2003 yılında “Sentiment Analysis: Capturing Favorability Using Natural Language Processing” çalışması ile girmiştir (Tetsuya ve Jeonghee, 2003).

Duygu analizi, metinsel içerikler içerisinde yer alan markalara, kişilere, ürün ya da bir organizasyona ilişkin duygusal durum bilgilerinin DDİ yöntem ve yaklaşımlarını kullanarak ortaya çıkarılması işidir (Fersini ve diğ., 2014). Duygu analizi çalışmaları ile metinsel ifadelerden görüş, tutum, yaklaşım ve duygu durumları çıkarılmaya çalışılır (Pang ve Lee, 2005). Duygu analizi aslında bir sınıflama işlemidir ve üzerinde çalışılan metinsel veriler gruplara ayrılarak sınıflandırılır (Onan ve Korukoğlu, 2017).

İnternet ve sosyal mecraların gelişmesi ile birlikte iletişim biçimimizde değişiklikler oldu. Sosyal medya paylaşımlarımız, mesajlaşma uygulamalarındaki mesaj verilerimiz ile büyük metinsel veriler oluşturmuş durumdayız. Bu metinsel verilerin işlenmesi ve anlamlı bilgiler çıkarılması hem bireysel kullanıcılar için hem de kurum ve firmalar için çok önem arz etmektedir (Xia ve diğ., 2011; Çetin ve Eryiğit, 2018). Bu verilerin elle işlenmesi ve sınıflandırılması verinin büyüklüğünden dolayı mümkün değildir. Bunun için çeşitli yöntem ve yaklaşımlara ihtiyaç bulunmaktadır (Ayvaz ve Shiha, 2018).

Web, yapısal olmayan metinsel verilerden oluşan büyük bir ambar durumundadır. Sosyal mecralar, forumlar, kullanıcı görüşlerini içeren siteler, duygu ve görüş barındırmaktadır. Duygu analizi ile bu verileri kullanarak, anlamlı bilgiler ortaya çıkarabilir. Bu bilgilerde yöneticilerin ya da bireylerin karar verme süreçlerine destek olabilir (Fessini ve diğ., 2016). Duygu analizi çalışmaları ile metinsel verilerin pozitif, negatif ve nötr durumlarını tespit etmek ile birlikte o metinden mutlu, hüznü, neşeli, heyecanlı gibi duygu durumları da çıkarılabilmektedir (Burcu ve Şimşek, 2018).

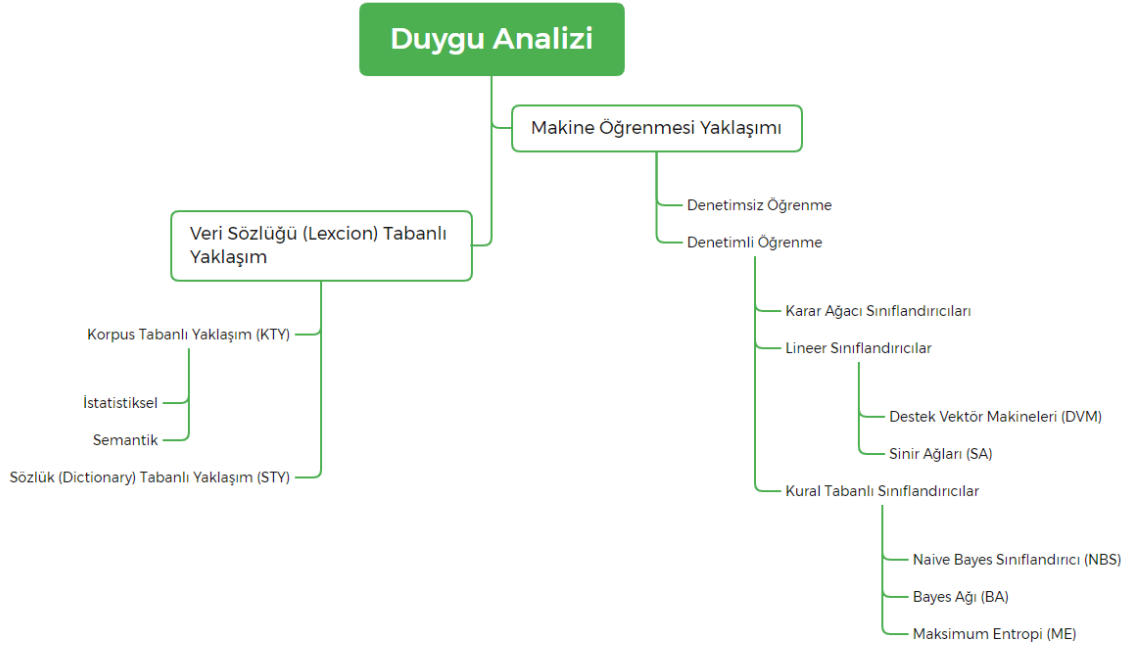
Literatürde duygu analizi, görüş madenciliği, fikir madenciliği ya da duygu durumu sınıflandırma olarak da geçmektedir (Karaöz ve Gürsoy, 2018; Can ve Alataş, 2017).

Duygu analizi çalışmaları ile şu işlemler gerçekleştirilebilir (Liu, 2012):

1. **Duygu Sınıflama:** Metinsel bir veride duygu durumu tespiti ve sınıflama işlemleri yapılabilir. Bu sınıflama pozitif, negatif şeklinde olabileceği gibi mutlu, mutsuz, heyecanlı, karamsar şeklinde de olabilir (Yu ve diğ., 2013; Saleh ve diğ., 2011).
2. **Öznellik Tespit Etme:** Metinsel bir ifadenin öznel mi yoksa nesnel mi olup olmadığının tespit edilmesine dair çalışmalar yapılır (Barbosa ve Feng, 2010).
3. **Görüş Sağlama:** Genellikle doküman seviyesinde yapılan çalışmalardır. Belli konular hakkında görüş bildiren metinlerin toplanmasına dair çalışmalar yapılır (Guo ve Wan, 2012).
4. **Görüş Özetleme:** Büyük bir metinsel veri içinden, belgelerden konu özünü tespit etmeye dair çalışmalar yapılır (Wang ve diğ., 2013).
5. **Alay ve İroni Tespit Etme:** Alay ve ironik ifadelerin tespit edilmesine dair çalışmalar yapılır. Alay ve ironiler dil ve kültüre göre farklılık gösterdiğinden duygu analizi çalışmalarının en zor çalışmaları olarak görülmektedir (Reyes ve Rosso, 2012).
6. **Diğer İşlemler:** Spam mesaj tespiti, yazar tespiti, sahte hesaplardan atılan manipülatif yorumların tespit edilmesine dair çalışmalar yapılır (Finn ve Kushmerick, 2006; Ott ve diğ., 2012).

2.5.1. Duygu Analizi Çalışmalarında Kullanılan Yaklaşımlar

Duygu analizi çalışmalarında kullanılan yöntemleri, makine öğrenmesine dayalı yöntemler ve sözlüğe dayalı yöntemler olmak üzere iki sınıfta toplayabiliriz (Medhat ve diğ., 2014).



Şekil 9: Duygu Analizi Çalışmalarında Kullanılan Yaklaşımlar (Medhat ve diğ., 2014)

Yukarıdaki şekilde de görüldüğü üzere duygu analizi çalışmalarında veri sözlüğü tabanlı yaklaşım ve makine öğrenmesi yaklaşımları kullanılmaktadır. Makine öğrenmesi yaklaşımı kendi içerisinde denetimli ve denetimsiz öğrenme olarak ikiye ayrılmaktadır.

Denetimli öğrenme altında ise karar ağacı sınıflandırıcıları, lineer sınıflandırıcılar ve kural tabanlı sınıflandırıcılar yer almaktadır. Sözlük tabanlı yaklaşım ise kendi içerisinde korpus tabanlı yaklaşım ve sözlük tabanlı yaklaşım olmak üzere ikiye ayrılmaktadır.

Makine öğrenmesi yaklaşımlarından gözetimli öğrenme modelleri yüksek başarımları ile ön plana çıkmak ile birlikte veri etiketleme süreci ve model oluşturma süreçlerinin zorluklarını da barındırmaktadır.

Sözlük tabanlı yaklaşım ise teknik anlamda makine öğrenmesi modellerine göre daha kolay olmak ile birlikte nispeten daha düşük başarımlarına sahiptir (Catal ve Nangir, 2017).

2.5.1.2. Sözlük Tabanlı Yaklaşım

Sözlük tabanlı yaklaşımda, bir metinde geçen duygu durumlarının tespiti için her bir kelimenin duygusal durumu, kullanılan sözlük yardımı ile hesaplanır (Taboada, 2011). Duygu durumunun tespit edilmesinde DDİ araç ve yöntemleri ile duygu terimleri sözlükleri kullanılır. Bu tür çalışmalarda herhangi bir etiketli veriye ya da eğitilmiş bir modele ihtiyaç yoktur (Özyurt ve Akcayol, 2018). Bununla birlikte kullanılan duygu sözlüğü ne kadar kapsamlı ise alınan sonuçlarda bir o kadar iyi olacaktır. Duygu sözlükleri, ilgili senaryo için elle yazılarak oluşturulabileceği gibi bu kapsamda önceden hazırlanmışta olabilirler (Liu, 2010).

Sözlük Nedir?

DDİ süreçlerinde kelimelerin anlamlarını bilmek çok önemlidir. Bu kapsamda sözlük geliştirilmiştir.

WordNet

Wordnet sözlüğü Princeton Üniversitesi'nde Bilişsel Bilimler çalışma grubu tarafından hazırlanan İngilizce dili için hazırlanmış bir sözlüktür (Miller ve diğ., 1990). Bu sözlüğü diğer sözlüklerden ayıran özelliklerinden biri kelimeler arasındaki ilişkiler ve eş anlamlı sözlüklere yer vermesidir.

A-Labels	Examples
EMOTION	noun anger#1, verb fear#1
MOOD	noun animosity#1, adjective amiable#1
TRAIT	noun aggressiveness#1, adjective competitive#1
COGNITIVE STATE	noun confusion#2, adjective dazed#2
PHYSICAL STATE	noun illness#1, adjective all_in#1
EDONIC SIGNAL	noun hurt#3, noun suffering#4
EMOTION-ELICITING SITUATION	noun awkwardness#3, adjective out_of_danger#1
EMOTIONAL RESPONSE	noun cold_sweat#1, verb tremble#2
BEHAVIOUR	noun offense#1, adjective inhibited#1
ATTITUDE	noun intolerance#1, noun defensive#1
SENSATION	noun coldness#1, verb feel#3

Şekil 10: Wordnet Kelime Örneği (Miller, 1995)

Wordnet veri tabanında 207.106 kelime barındırmaktadır. Wordnet'in oluşturulma amacı İngilizce kelimeler arası ilişkilerin ortaya konulmasıdır (Horák ve diğ., 2006).

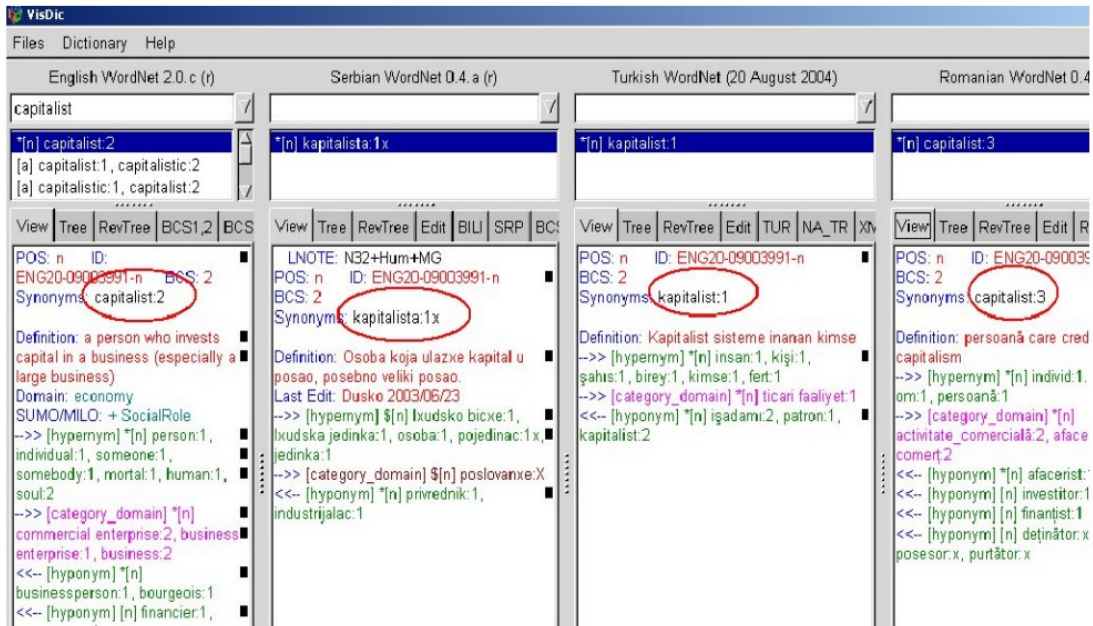
BalkaNet

BalkaNet, Balkan diller için ayrı WordNet'lerden oluşan çok dilli bir yapıda sözlük veri tabanının geliştirilmesi ile oluşturulmuştur. Bu proje kapsamında Yunanca, Bulgarca, Sırpça, Romanca, Çekçe ve Türkçe dilleri için WordNet veri tabanları oluşturulmuştur (Bilgin ve diğ., 2004).

<i>WordNet</i>	<i>Eş Anlamlı Kelime Sayısı</i>
<i>Türkçe</i>	10280
<i>Bulgarca</i>	15007
<i>Çekçe</i>	26525
<i>Yunanca</i>	15781
<i>Rumence</i>	14707
<i>Sırpça</i>	4772

Tablo 1: BalkaNet’de Yer Alan Diller ve Kelime Sayıları (Stamou ve diğ., 2004)

Tablo 1’de görüldüğü üzere en fazla kelime sayısı 26.525 kelime ile Çek dilinde en az ise 4772 kelime sayısı ile Sırp dilinde olduğu görülmektedir.



Şekil 11: BalkaNet Görüntüleme ve Düzenleme Editörü (Antonic ve Krstev, 2008)

Şekil 11’de VisDic (Horák ve diğ., 2004) editöründe de görüldüğü üzere kelimelerin Balkan dillerindeki karşılıkları ve eş anlamlıları görülmektedir.

Turkish WordNet

BalkaNet projesi kapsamında Türkçe dili için geliştirilen bir sözlüktür. Sözlüğe dair istatistikler aşağıda yer almaktadır.

<i>Turkish WordNet</i>	<i>Sayılar</i>
<i>Eş Anlamlı Kelimeler</i>	<i>11628</i>
<i>Eş Anlamlı Kelime Üyeleri</i>	<i>16095</i>
<i>Ortalama Eş Anlam Kelime Büyüklüğü</i>	<i>1.38</i>
<i>Açıklamalar</i>	<i>4913</i>

Tablo 2: Turkish WordNet’e Ait İstatistikler (Çetinoğlu ve diğ., 2018)

Tablo 2’de görüldüğü üzere sözlük içerisinde 11628 kelime, 16095 eş anlamlı kelime üyesi ve 4913 kelimeye ait açıklamalar bulunmaktadır.

<i>Turkish WordNet</i>	<i>Değer</i>
<i>Synonyms</i>	<i>güzelleştirmek, süslemek</i>
<i>Gloss</i>	<i>daha güzel hale getirmek</i>
<i>Pos tag</i>	<i>Verb</i>
<i>Synset Label</i>	<i>Pos</i>
<i>Hypernym synset label</i>	<i>Pos</i>
<i>Near-antonym synset label</i>	<i>Neg</i>
<i>Equivalent English synset</i>	<i>Ameliorate, improve, better, amend...</i>

Tablo 3: Turkish WordNet’e Ait Veri Yapısı (Dehkharghani ve diğ., 2016)

Tablo 3’te BalkaNet projesi kapsamında geliştirilen Turkish WordNet’e ait veri yapısı görüntülenmektedir.

Kenet

Türkçe için literatürde iki wordnet vardır. Bunlardan ilki BalkaNet projesi kapsamında geliştirilen Turkish WordNet diğeri ise KeNet’dir. KeNet, Turkish WordNet’e göre daha kapsamlıdır. 76.757 eş anlamlı kelime barındırmaktadır (Bakay ve diğ., 2021).

<i>KeNet</i>	<i>Sayılar</i>
<i>İsimler</i>	<i>44074</i>
<i>Fiiller</i>	<i>17791</i>
<i>Sıfatlar</i>	<i>12416</i>
<i>Zarflar</i>	<i>2550</i>
<i>Ünlemler</i>	<i>3342</i>
<i>Zamirler</i>	<i>68</i>
<i>Bağlaçlar</i>	<i>60</i>
<i>Edatlar</i>	<i>29</i>
<i>Toplam</i>	<i>77330</i>

Tablo 4: KeNet’e Ait İstatistikler (Bakay ve diğ., 2021)

Tablo 4’te görüldüğü üzere sözlük içerisinde toplamda 77330 kelime bulunmaktadır. Bu kelimelerden 44074’ü isim, 17791’i fiil, 12416’sı sıfat, 2550’si zarf, 3342’si ünlem, 68’i zamir, 60’ı bağlaç ve 29’u edattır.

```

<SYNSET>
  <ID>TUR10-0038510 </ID>
  <synset member>anne<SENSE>2</SENSE>
  </synset member>
  <POS>n</POS>
  <DEF>... </DEF>
  <EXAMPLE>... </EXAMPLE>
</SYNSET>

```

Şekil 12: KeNet'e Ait Veri Yapısı (Bakay ve diğ., 2021)

Yukarıdaki şekilde KeNet'e ait veri yapısı görüntülenmektedir.

Duygu Sözlükleri Nelerdir?

Duygu analizi çalışmaları için geliştirilen ve barındırdığı her kelimenin sayısal olarak duygu durumunu içeren sözlüklere duygu sözlüğü (sentiment dictionary) denmektedir (Uçan, 2014). Duygu sözlükleri duygu ile ilgili bilinen kelimelerden oluşan kelime listelerinin bir araya getirilmesi ile oluşmuş, duygu analizi yapmak için geliştirilmiş sözlüklerdir (Agarwal ve Mittal, 2015).

Duygu sözlükleri her bir kelime için olumlu, olumsuz ve nötr olarak kategori edilmiş puanlamalar içermektedir (Suat ve Çınar, 2019). İngilizce ve Türkçe dilleri için geliştirilen duygu sözlükleri aşağıda listelenmiştir.

Duygu Sözlükleri (İngilizce)

1. SentiWordNet
2. SenticNet

Duygu Sözlükleri (Türkçe)

1. SentiTurkNet
2. HisNet

SentiWordNet

SentiWordNet İngilizce dilinde hazırlanmış bir duygu sözlüğüdür (Esuli ve Sebastiani, 2006). Sözlük içindeki kelimelerin negatif ve pozitif skorları yer almaktadır. Bu skorlar kullanılarak ilgili kelimenin negatif, pozitif ya da nötr olma durumuna karar verilebilir (Denecke, 2008).

POS	<i>Sözlükteki kelimelerin türleri. İsim (n), Fiil (v), Sıfat (a), Zarf (r)</i>
ID	<i>Her bir kelimeye ait numara.</i>
PosScore (Pozitiflik Skoru)	<i>İlgili kelimenin pozitiflik değeridir. Bu değer 0 ile 1 arasındadır. Nötr ise değer 0'dır. Değer 1'e ne kadar yakınsa ilgili kelimenin pozitif kutbu o kadar yüksektir.</i>
NegScore (Negatiflik Skoru)	<i>İlgili kelimenin negatiflik değeridir. Bu değer 0 ile 1 arasındadır. Nötr ise değer 0'dır. Değer 1'e ne kadar yakınsa ilgili kelimenin negatif kutbu o kadar yüksektir.</i>
SynsetTerms (Eş Anlamlı Kelimeler)	<i>İlgili kelimeye ait eş anlamlı kelimeleri içerir.</i>
Glos (Anlam)	<i>Eş anlamlı kelimelerin anlamlarını içerir.</i>

Tablo 5: SentiWordNet'in Veri Yapısı (Sevindi, 2013)

Tablo 5'te görüldüğü üzere SentiWordNet sözlüğünün veri yapısı altı farklı yapıdan oluşmaktadır. Bunlar POS, ID, PosScore, NegScore, SynsetTerms ve Glos. Sözlük 117,659 kelime barındırmaktadır.

POS	<i>a</i>
ID	00682419
PosScore	0,125
NegScore	0
SynsetTerms	hearing-impaired#1 hard-of-hearing#1
Glos	having a hearing loss

Tablo 6: SentiWordNet Örnek Verisi (Sevindi, 2013)

Tablo 6’da SentiWordNet sözlüğünde yer alan bir veriye ait detaylar sunulmuştur. Kelimenin pozitif skorunun 0,125 olduğu gözükmemektedir.

SenticNet

SenticNet 2009 yılında MIT Medya laboratuvarında geliştirilmiş bir duygu sözlüğüdür. İçeriğinde 200.001 kelime barındırmaktadır (Cambria ve diğ., 2020).

```

<rdf:Description rdf:about="http://sentic.net/api/en/concept/capital_benefit">
  <rdf:type rdf:resource="http://sentic.net/api/concept"/>
  <text xmlns="http://sentic.net">capital benefit</text>
  <semantics xmlns="http://sentic.net">
    <concept xmlns="http://sentic.net" rdf:resource="http://sentic.net/api/en/concept/target"/>
    <concept xmlns="http://sentic.net" rdf:resource="http://sentic.net/api/en/concept/intention"/>
    <concept xmlns="http://sentic.net" rdf:resource="http://sentic.net/api/en/concept/aerobics"/>
    <concept xmlns="http://sentic.net" rdf:resource="http://sentic.net/api/en/concept/aquarobics"/>
    <concept xmlns="http://sentic.net" rdf:resource="http://sentic.net/api/en/concept/circuit_training"/>
  </semantics>
  <sentic xmlns="http://sentic.net">
    <introversion xmlns="http://sentic.net" rdf:datatype="http://www.w3.org/2001/XMLSchema#float">0</introversion>
    <temper xmlns="http://sentic.net" rdf:datatype="http://www.w3.org/2001/XMLSchema#float">0.823</temper>
    <attitude xmlns="http://sentic.net" rdf:datatype="http://www.w3.org/2001/XMLSchema#float">0</attitude>
    <sensitivity xmlns="http://sentic.net" rdf:datatype="http://www.w3.org/2001/XMLSchema#float">0</sensitivity>
  </sentic>
  <moodtags xmlns="http://sentic.net">
    <concept xmlns="http://sentic.net" rdf:resource="http://sentic.net/api/en/concept/calmness"/>
    <concept xmlns="http://sentic.net" rdf:resource="http://sentic.net/api/en/concept/calmness"/>
  </moodtags>
  <polarity xmlns="http://sentic.net">
    <value xmlns="http://sentic.net">positive</value>
    <intensity xmlns="http://sentic.net" rdf:datatype="http://www.w3.org/2001/XMLSchema#float">0.823</intensity>
  </polarity>
</rdf:Description>

```

Şekil 13: SenticNet 6 Veri Yapısı (Cambria ve diğ., 2020)

SenticNet veri yapısında da görüldüğü üzere sözlük xml formatında sunulmaktadır. Xml yapısı içerisinde yer alan polarite alanında kelimenin kutup değeri (polarity value) değeri ve yoğunluk (intensity) değerleri gözükmemektedir.

SentiTurkNet

SentiTurkNet duygu sözlüğü Sabancı Üniversitesi araştırmacıları tarafından geliştirilmiştir. Sözlüğün yapısal geliştirilmesinde WordNet, SentiWordNet ve SenticNet sözlüklerinden faydalanmışlardır. Çalışmanın kaynağı olarak ise sadece BalkaNet projesi kapsamında geliştirilen Turkish WordNet kullanılmıştır (Dehkharghani ve diğ., 2016).

<i>Synonyms</i>	<i>Turkish Gloss</i>	<i>Polarity Label</i>	<i>POS Tag</i>	<i>Neg Value</i>	<i>Obj Value</i>	<i>Pos Value</i>
ayva reçeli	Meyveleri şekerle kaynatarak hazırlanan tatlı	<i>o</i>	<i>n</i>	0,06	0,872	0,068
gül reçeli	Meyveleri şekerle kaynatarak hazırlanan tatlı	<i>o</i>	<i>n</i>	0,06	0,872	0,068
incir reçeli	Meyveleri şekerle kaynatarak hazırlanan tatlı	<i>o</i>	<i>n</i>	0,06	0,872	0,068
iştah	Yemek yeme isteği.	<i>o</i>	<i>n</i>	0,06	0,872	0,068
iştahlı	İştahlı olan, boğazlı	<i>p</i>	<i>a</i>	0,06	0,462	0,478
iştahsız	Yemek yeme isteği olmayan, boğazsız	<i>n</i>	<i>a</i>	0,48	0,452	0,068
iştahsızlık	İştahsız olma durumu	<i>n</i>	<i>n</i>	0,512	0,42	0,068
kayısı reçeli	Meyveleri şekerle kaynatarak hazırlanan tatlı	<i>o</i>	<i>n</i>	0,06	0,872	0,068
kayısı suyu	Meyve suyu	<i>o</i>	<i>n</i>	0,06	0,872	0,068
kısayol	Bilgisayarda bir dosyadan diğer dosyaya ulaşmak	<i>o</i>	<i>n</i>	0,06	0,872	0,068
mikrobiyolojik	Mikroblar ile uğraşan bilim	<i>o</i>	<i>a</i>	0,06	0,872	0,068
minibüs	10, 12 kişilik küçük otobüs	<i>o</i>	<i>n</i>	0,06	0,872	0,068
müşteşarlık	Müşteşarın görevi veya makamı	<i>o</i>	<i>n</i>	0,06	0,872	0,068
nevresim	Torba biçiminde dikilmiş yorgan çarşafı	<i>o</i>	<i>n</i>	0,06	0,872	0,068
neyzen	Ney üfleyici, ney çalan kimse	<i>o</i>	<i>n</i>	0,06	0,872	0,068

Tablo 7: SentiTurkNet'in Veri Yapısı (Dehkharghani ve diğ., 2016)

Tablo 7’de görüldüğü üzere SentiTurkNet sözlüğünde yer alan her bir kelimenin pozitif, negatif ve objektif değerleri vardır. Bu değerlerden hangisi yüksek ise kelimenin duygu kutbu da o olmaktadır. Sözlük 14795 kelime barındırmaktadır.

HisNet

Bir grup araştırmacı SentiTurkNet sözlüğünün sınırlı kapsamından dolayı alternatif bir duygu sözlüğü geliştirmişlerdir. HisNet adı verilen bu sözlük KeNet sözlüğü üzerinde geliştirilmiştir. Sözlük 76,825 kelime içermektedir (Özçelik ve diğ., 2021).

<i>Kutup Seviyesi</i>	<i>Kelime Sayısı</i>
<i>Kesinlikle Olumlu</i>	1038
<i>Çok Olumlu</i>	451
<i>Olumlu</i>	456
<i>Az Olumlu</i>	1234
<i>Objektif</i>	63534
<i>Şiddetli Olumsuz</i>	4430
<i>Çok Olumsuz</i>	1465
<i>Olumsuz</i>	1238
<i>Az Olumsuz</i>	3360

Tablo 8: HisNet Sözlüğüne Ait Kutup Kategorileri ve Kelime Sayıları (Özçelik ve diğ., 2021)

Tabloda 8’de görüldüğü üzere HisNet sözlüğünde dokuz farklı kategoride duygu kutbu bulunmaktadır. En çok kelimenin “Objektif” kutup kategorisinde en az kelimenin ise “Çok Olumlu” kutup kategorisinde olduğu görülmektedir.

```

<WORD><NAME>avurtlamak</NAME><PScore>0.0</PScore><NScore>0.375</NScore></WORD>
<WORD><NAME>buğday güvesi</NAME><PScore>0.0</PScore><NScore>0.0</NScore></WORD>
<WORD><NAME>korakor</NAME><PScore>0.0</PScore><NScore>0.0</NScore></WORD>
<WORD><NAME>fesat karıştırmak</NAME><PScore>0.0</PScore><NScore>1.0</NScore></WORD>
<WORD><NAME>mayışmak</NAME><PScore>0.0</PScore><NScore>0.0</NScore></WORD>
<WORD><NAME>faş etmek</NAME><PScore>0.0</PScore><NScore>0.0</NScore></WORD>
<WORD><NAME>sülükçülük</NAME><PScore>0.0</PScore><NScore>0.0</NScore></WORD>
<WORD><NAME>kanasta</NAME><PScore>0.0</PScore><NScore>0.0</NScore></WORD>
<WORD><NAME>işgüzarlık etmek</NAME><PScore>0.0</PScore><NScore>0.25</NScore></WORD>
<WORD><NAME>avane</NAME><PScore>0.0</PScore><NScore>0.375</NScore></WORD>
<WORD><NAME>gözünden kaçırmak</NAME><PScore>0.0</PScore><NScore>0.25</NScore></WORD>
<WORD><NAME>içi vık vık etmek</NAME><PScore>0.0</PScore><NScore>0.25</NScore></WORD>
<WORD><NAME>inakçı</NAME><PScore>0.0</PScore><NScore>0.0</NScore></WORD>
<WORD><NAME>kümes</NAME><PScore>0.0</PScore><NScore>0.0</NScore></WORD>
<WORD><NAME>son kartını oynamak</NAME><PScore>0.0</PScore><NScore>0.0</NScore></WORD>
<WORD><NAME>canı bayılmak</NAME><PScore>0.0</PScore><NScore>0.25</NScore></WORD>
<WORD><NAME>herkes</NAME><PScore>0.0</PScore><NScore>0.0</NScore></WORD>
<WORD><NAME>karşılaştırmalı dil bilgisi</NAME><PScore>0.0</PScore><NScore>0.0</NScore></WORD>
<WORD><NAME>peyrev</NAME><PScore>0.0</PScore><NScore>0.0</NScore></WORD>
<WORD><NAME>katkı maddesi</NAME><PScore>0.0</PScore><NScore>0.0</NScore></WORD>

```

Şekil 14: HisNet'in Veri Yapısı (Özçelik ve diğ., 2021)

Yukarıda yer alan HisNet'in veri yapısında da görüldüğü üzere, sözlükte her bir kelimenin pozitif ve negatif skorları yer almaktadır.

2.5.1.3. Makine Öğrenmesi-Yapay Sinir Ağları (YSA) Yaklaşımı

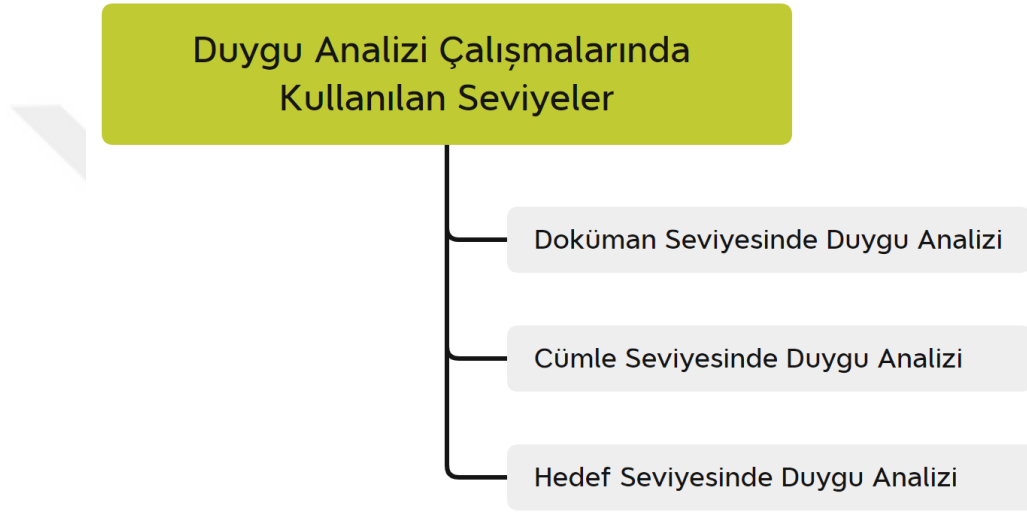
Duygu analizi çalışmalarında kullanılan yaklaşımlarından biri de makine öğrenmesi algoritmalarıdır. Bu yaklaşımda daha önce etiketli veriler kullanılarak eğitilen bir model kullanılarak sınıflama işlemleri yapılabilmektedir. Bu yaklaşımda duygu sözlüklerine ihtiyaç yoktur. Ancak modelin eğitilmesi için gerekli olan etiketli veri oluşturma süreci zahmetli ve zaman alan bir süreçtir (Kearney ve Liu, 2013).

Makine öğrenmesine dayalı duygu analizi çalışmalarında oluşturulan model etiketli bir eğitim verisi ile önceden eğitilmeli, test verisi ile de başarı oranı test edilmelidir (Özyurt ve Akçayol, 2018). Literatürdeki son çalışmalara göre duygu analizi çalışmalarında makine öğrenmesinin alt disiplini olan derin öğrenme yöntemlerinin kullanılması başarı oranını ciddi oranda artırmıştır (Collobert ve diğ., 2011). YSA yapısı büyük verilerin işlenmesi ve sınıflandırılmasında sözlük tabanlı yaklaşımlara göre çok daha hızlı ve başarılı sonuçlar ortaya

koyduğu tespit edilmiştir (Song ve Lee, 2013). Ancak bu yaklaşımda geliştirilen yapay zeka modelinin daha iyi sonuçlar verebilmesi için çok miktarda etiketli veriye ihtiyaç vardır.

2.5.2. Duygu Analizi Çalışmalarında Kullanılan Seviyeler

Duygu analizi çalışmaları, çalışmanın kapsamı ve veri setlerinin yapısı doğrultusunda üç farklı çalışma seviyesinde ele alınmaktadır. Bunlar sırasıyla doküman, cümle ve hedef (özellik, kategori, hedef) seviyesinde yapılan çalışmalardır (Özyurt ve Akçayol, 2018).



Şekil 15: Duygu Analizi Çalışmalarında Kullanılan Seviyeler (Ozyurt ve Akcayol, 2018)

Bu üç seviyeden en detaylı olanı hedef seviyesinde yapılan çalışmalardır. En genel olanı ise doküman seviyesinde yapılan çalışmalardır. Bununla birlikte hedef seviyesinde yapılan çalışmalar, doküman ve cümle seviyesinde yapılan çalışmalara göre çok daha fazla zaman ve teknik detay barındırmaktadır (Ghag ve Shah, 2013).

2.5.2.1. Doküman Seviyesinde Duygu Analizi

Bu seviyede yapılan çalışmalarda bir belge, makale ya da birden fazla cümleden oluşan kullanıcı görüşünün tamamı bir bütün olarak değerlendirilir. Duygu durumu cümle ya da özellik bazında detaya inilmeden doküman seviyesinde pozitif, negatif ya da nötr olarak tespit edilir. Bu seviyede yapılan çalışmalar diğer seviyelere göre daha kolay olmak ile birlikte duygu

durumu genellemesi yaptığından duygu kayıpları söz konusu olmaktadır (Wilson ve diğ., 2005).

2.5.2.2. Cümle Seviyesinde Duygu Analizi

Bu seviyede yapılan çalışmalarda bir doküman içerisinde yer alan her bir cümlelerin duygu analizi ayrı ayrı yapılır. Bu sayede doküman seviyesinde yaşanabilecek duygu durumu kayıpları engellenmiş olur. Doküman seviyesinde yapılan duygu durumu analizlerine göre daha detaylı ve hassas sonuçlar elde edilir. Ancak bu seviyede yapılan analizlerde bir cümle içerisinde birden fazla duygu durumu olduğunda duygu kayıpları yaşanabilmektedir (Meena ve Prabhakar, 2007).

2.5.2.3. Hedef (Aspect) Seviyesinde Sentismental Analiz

Hedef tabanlı duygu analizi çalışmaları sözcük ve doküman seviyesinde yapılan çalışmalara göre çok daha yeni ve yakın döneme ait çalışmalardır (Özyurt ve Akçayol, 2018).

Cümle seviyesinde yapılan duygu analizleri her ne kadar doküman seviyesinde yapılan analiz çalışmalarına göre daha hassas ve detaylı sonuçlar verse de bazı durumlarda bir cümlelerin gerçek duygu durumunu yansıtamayabilirler. Örneğin bir cümle içerisinde hem pozitif hem de negatif duygu durumu yer alabilir. Hedef seviyesinde yapılan çalışmalar bu problemin giderilmesi için geliştirilmiştir. Hedef tabanlı duygu analizi çalışmalarında önce cümle içerisinde geçen hedefler (aspect) belirlenir. Daha sonra bu hedefleri niteleyen kelimeler üzerinden sözlük tabanlı ya da makine öğrenmesi yaklaşımları kullanılarak her bir hedefin duygu durumuna karar verilir (Liu, 2012; Çetin ve Eryiğit, 2018).

2.6. BU ALANDA YAPILAN BENZER ÇALIŞMALAR

Duygu analizi çalışmalarında yapılan ilk çalışmalardan biri Pang, Lee ve Vaithyanatham tarafından yapılan çalışmadır. Bu çalışmada film yorumları verilerini kullanarak yorumların duygu durumlarını çeşitli sınıflandırma algoritmaları ile test etmişlerdir. En iyi sonucu ise destek vektör makineleri ile almışlardır. %82,9'luk bir başarı oranına ulaşmışlardır (Pang ve diğ., 2002).

Pennacchiatti ve Popescu yapmış oldukları çalışmada twitter verileri üzerinde çalışmışlardır. Twitter verileri üzerinden kişilerin etnik ve siyasi yönelimlerini gruplayarak tespit etmeye çalışmışlardır. Bu sınıflama işlemi yaparken makine öğrenmesi tekniklerini kullanmışlardır. Sonuçların çeşitlilik gösterdiğini belirtmişlerdir (Pennacchiatti ve Popescu, 2011).

Go ve arkadaşları yapmış olduğu çalışmada Twitter verileri üzerinde denetimli öğrenme yaklaşımı ile duygu analizi çalışması yapmışlardır. Çalışma kapsamında yaklaşık 1.600.000 tweet kullanmışlardır. Çalışma sonucunda destek vektör makinelerinde %82,2 Navie Bayes’de ise %83’lük doğruluk oranı elde etmişlerdir (Go ve diğ., 2009).

Matsumoto ve arkadaşları yapmış olduğu çalışmada film yorumları veri seti kullanarak duygu sınıflama yapmışlardır. N-gram yöntemine alternatif olarak geliştirdikleri sözcükler arası ilişki yöntemini kullanmışlar ve Destek Vektör Makine algoritmasında %92,9 oranında başarı elde etmişlerdir (Matsumoto ve diğ., 2005).

Nikfarjam ve arkadaşları ise sağlık verileri üzerinde duygu analizi çalışması yürütmüşlerdir. Çeşitli sağlık forumlarından ve Twitter üzerinden elde ettikleri yorumları ADRMine ve Destek Vektör Makineleri algoritmaları test etmişlerdir. En yüksek skoru ADRMine ile %82,1 olarak elde etmişlerdir (Nikfarjam ve diğ., 2015).

Katz ve arkadaşları ise ortaya koymuş oldukları ConSent içerik tabanlı duygu analizi modeli ile bir çalışma yapmışlardır. TripAdvisor ve IMDB’den elde ettikleri toplamda 32.000 yorumu sınıflamışlardır. Sınıflama sonucunda önerdikleri ConSent modeli en yüksek sonuç aldıkları Destek Vektör Makineleri modeline çok yakın sonuç vermiştir. ConSent modeli ise yorumlar içerisinde anahtar kelimelerin belirlenip özniteliklerinin çıkarılması esasına dayanmaktadır (Katz ve diğ., 2015).

Nizam ve Akın yapmış olduğu çalışmada toplamda 2824 twetten oluşan veri setini Naive Bayes, destek vektör makineleri, karar ağacı ve k-en yakın komşu algoritmaları kullanılarak duygu analizi sınıflama işlemine tabi tutmuşlardır. Sınıflamada pozitif, negatif ve nötr olmak üzere 3 kutup için çalışma yapılmıştır. En iyi performansı %72,33 ile destek vektör makineleri ile elde etmişlerdir (Nizam ve diğ., 2014).

Sevindi yapmış olduğu çalışmada film yorumlarını pozitif ve negatif duygu kutuplarına göre sınıflamıştır. Sınıflama için makine öğrenmesi tekniklerinden Destek Vektör Makinelerini,

Karar Ağaçlarını ve k-En Yakın Komşu ve Naive Bayes modellerini test etmiştir. Bununla birlikte sözlük tabanlı yaklaşımı kullanarak test etmiştir. En yüksek skoru Destek Vektör Makineleri ile elde etmiştir. 0,82 f-skoru elde etmişlerdir (Sevindi, 2013).

Yang ve arkadaşları e-ticaret sitesi üzerinde elde ettikleri kullanıcı görüşlerini kendi geliştirdikleri derin öğrenme temelli SLCABG modelini kullanarak sınıflamışlardır. %93,5 oranında bir doğruluk oranı elde etmişlerdir (Yang ve diğ., 2020).

Bir başka çalışmada ise Twitter ve film yorumlarından elde edilen yorumlardan oluşan veri seti duygu analizi yapılarak sınıflandırılmaya çalışılmıştır. Sınıflama işleminde sözlük tabanlı yaklaşım ve makine öğrenmesi yaklaşımları test edilmiştir. Elde edilen bulgular doğrultusunda makine öğrenmesi yaklaşımının (%89,5 doğruluk oranı) sözlük tabanlı yaklaşıma (%79 doğruluk oranı) göre daha iyi sonuç verdiği tespit edilmiştir. En iyi sonuç ise Vektör Destek Makineleri ile elde edilmiştir (Türkmenoğlu ve Tantuğ, 2014).

Tang ve arkadaşları Twitter verileri üzerinde duygu analizi çalışması yapmışlardır. Sınıflama sürecinde Uzun Kısa Süreli Bellek (LSTM) ve hedefe bağlı LSTM modellerini test etmişler. Elde ettikleri bulgular doğrultusunda hedefe bağlı LSTM modelinin normal LSTM modeline göre daha başarılı sonuçlar verdiğini tespit etmişlerdir (Tang ve diğ., 2015).

Ayata ve arkadaşları Twitter'dan elde ettikleri veriler üzerinden duygu analizi çalışması yapmışlardır. Çalışmada Destek Vektör Makineleri ve Rastgele Orman yöntemlerini kullanmışlardır. Elde ettikleri sonuçlar doğrultusunda Destek Vektör Makineleri yönteminin Rastgele Orman yöntemine göre daha başarılı sonuçlar verdiğini gözlemlemişlerdir. Bununla birlikte veri setinin domain (alan) bazlı olmasının başarı oranını artırdığını tespit etmişlerdir (Ayata ve diğ., 2017).

Karakuş ve arkadaşları film yorumları üzerinde duygu analizi çalışması yapmışlardır. Sınıflama aşamasında Konvansiyonel Sinir Ağları (CNN) ve LSTM kullanmışlardır. Ayrıca nu iki modeli birleştirerek yeni bir model oluşturmuşlar ve başarı oranlarını test etmişlerdir (Karakuş ve diğ., 2018).

Bir başka çalışmada ise e-ticaret sitelerinden elde edilen kullanıcı yorumları kullanılmıştır. Yapılan çalışmada Rastgele Orman ve LSTM ve CNN yöntemleri kullanılmıştır. En yüksek sonuç %94,21 ile LSTM'de elde edilmiştir (Pervan ve Keleş, 2019).

Bir başka çalışmada Twitter'dan elde edilen veriler üzerinden duygu analizi çalışması yapılmıştır. Çalışmada yorumlardan öznitelikler çıkartılmış ve bu kapsamda yorumlar sınıflandırılmıştır. Unigram için %77,15, big-gram için ise %76,88 lik başarı oranları elde etmişlerdir. Kelime sayısının azlığı sınıflandırma işlemini zorlaştırarak başarı oranını düşürdüğünü gözlemlemişlerdir (Al-Bnd ve Salman, 2015).

Nassar ve Sezer doktora tez çalışmalarında çeşitli merkezlerden elde ettikleri 1000 film yorumunu sözlük tabanlı yaklaşım ve makine öğrenmesi teknikleri sınıflamaya çalışmıştır. Bulgular neticesinde en yüksek başarı oranını yapay sinir ağları (0,92 f-score) ile elde etmişlerdir. Makine öğrenmesi yaklaşımının sözlük tabanlı yaklaşıma göre daha iyi sonuçlar verdiğini gözlemlemişlerdir (Nassar ve Sezer, 2019).

Rahim Dehkharghani, Berrin Yanıkoğlu ve Yücel Sayın ve Kemal Oflazer'in "Sentiment analysis in Turkish at different granularity levels" isimli çalışmalarında ise Türkçe film yorumları üzerinde çeşitli seviyelerde (cümle, doküman ve hedef) duygu analizi çalışması yapmışlardır. Elde ettikleri sonuçlar doğrultusunda en yüksek başarı oranını %79,56 ile hedef seviyesinde yapılan çalışmada elde etmişlerdir. Çalışmada model olarak sözlük tabanlı yaklaşım tercih edilmiştir (Dehkharghani ve diğ., 2017).

• MALZEME VE YÖNTEM

3.1. VERİ TOPLAMA SÜRECİ

3.1.1. Twitter Üzerinden Veri Toplama

Veri toplama sürecinde ilk olarak Twitter üzerinden veriler toplanmaya çalışılmıştır. Belli konu etiketlerinde veriler toplanmış ve değerlendirilmiştir. Elde edilen verilerin kutupsal anlamda nötr duygu durumu taşıdığı daha çok tavsiye niteliği taşıyan “meli, malı” şeklinde biten ifadeler olduğu görülmüştür.

3.1.2. Okul.com.tr Platformu Üzerinden Veri Toplama

Daha sonra “okul.com.tr” platformu incelenmiş ve buradaki kullanıcı görüşlerinin kullanılabileceği tespit edilmiştir. Okul.com.tr içinde anaokulu, ilkokul, ortaokul ve lise düzeyinde özel ve devlet okullarına dair bilgilerin yer aldığı bir sitedir. Bu siteyi kullanıcılar ziyaret edip okullar ve sahip olduğu imkanlar hakkında bilgi alabilmektedirler.



Şekil 16: Okul.com.tr Sitesinin Ana Sayfası

2 Eylül 2021 tarihi itibari ile sitede kayıtlı İstanbul, Ankara, İzmir, Bursa, Antalya, Kocaeli ve Adana illerinde yer alan 14.523 okul bulunmaktadır.

Kalamış Şehit Murat Özyalçın İlkokulu

0 216 348 11**

Numarayı Göster

Galeri

Okulun İmkanları

Yorumlar

İletişim Bilgileri

Fiziksel İmkanlar

✓ Yemekhane

✓ Laboratuvar

✓ Kantin

Hizmetler

✓ Güvenlik

✓ Rehberlik

✓ Servis

Dil Olanakları

✓ İngilizce

Yorumlar (3)

Hepsi Öğrenci Veli

S

Butik yapısı ile küçük yaş grubundaki çocuklar için çok ideal bir okul. Tecrübeli kadrosu ile devletin özel okulu gibi.

★★★★★

+1 Beğen

FA

Butik bir okul gibi öğrencilerle birebir ilgili bir okul. öğretmenleri sevecen.okulun eksiklikleri sosyal aktiviteler düzenlenmiyor.fiziki imkanları yetersiz. öğretmenler yaşlı olduğu için çocukların enerjisine yetişemiyor,rehberlik servisi işlevini yerine getiremiyor maalesef.

★★★★☆

+1 Beğen

MÖ

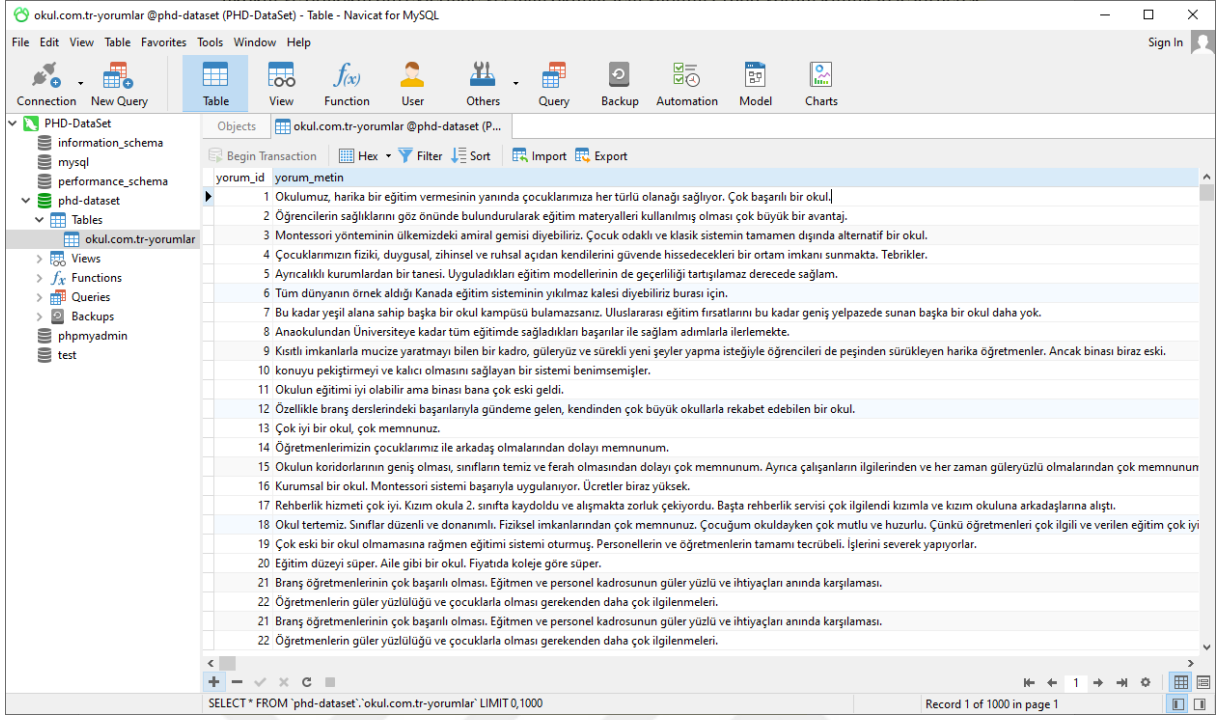
Güvenlik sorumlusu 1 kişi bence yetersiz bir sayı. Öğretmenimizi seviyoruz. Müzelere ve akvaryumlara gezilere gidiyorlar.

★★★★★

+0 Beğen

Şekil 17: Okul.com.tr Sitesinde Yer Alan Bir Okula Ait Bilgiler

Şekil 17’de görüldüğü üzere sitede bir okula ait bilgiler yer almaktadır. Okula ait görseller, imkanları, yorumlar ve iletişim bilgileri yer almaktadır. Yorum bilgilerini kullanmak üzere site yöneticileri ile iletişime geçilmiş ve izin alınmıştır. İzin kapsamında İstanbul’da yer alan ilkokul düzeylerinde bulunan okullar için yapılmış 1000 yorum kimliksizleştirilerek toplanmış ve işlenmek amacıyla bu araştırma kapsamında oluşturulan bir veri tabanına aktarılmıştır.



Şekil 18: Veri Tabanına Aktarılan Yorumlar

Şekil 18’de Okul.com.tr sitesinden MySQL veri tabanına aktarılan yorumların görünümü mevcuttur.

Sütun Adı	Açıklaması
<i>Yorum_id</i>	<i>Yoruma verilen benzersiz numara</i>
<i>Yorum_metin</i>	<i>Kullanıcı yorumu</i>

Tablo 9: Veri Setindeki Nitelikler ve Açıklamaları

Tablo 9’da görüldüğü gibi veri setinde “*Yorum_id*” alanı yoruma verilen benzersiz numarayı içermektedir. “*Yorum_metin*” alanı ise kimliksizleştirilmiş kullanıcı yorumunu içermektedir. Toplamda 1000 farklı yorum veri tabanına kullanılmak üzere aktarılmıştır.

Verilerin Etiketlenmesi Süreci

Yorum metni içindeki hedeflerin ve her bir hedefin duygu durumunun tespit edilmesi için üç farklı eğitim uzmanından destek alınmıştır. Her bir eğitim uzmanı yorumların ayrı ayrı hedeflerini tespit etmiş ve duygu durumlarını etiketlemiştir. Üç uzmandan alınan geri bildirimler doğrultusunda yorumların nihai hedef ve duygu durumlarına karar verilmiştir.

yorum_id	yorum_metin	aspect
1	Okulumuz, harika bir eğitim vermesinin yanında çocuklarınıza her türlü olanağı sağlıyor. Çok başarılı bir okul.	Okul-p, Eğitim-p, Olanak-p
2	Öğrencilerin sağlığını göz önünde bulundurularak eğitim materyalleri kullanılmış olması çok büyük bir avantaj	Eğitim Materyali-p
3	Montessori yönteminin ülkemizdeki amiral gemisi diyebiliriz. Çocuk odaklı ve klasik sistemin tamamen dışında a	Yöntem-p, Sistem-p, Okul-p
4	Çocuklarımızın fiziki, duygusal, zihinsel ve ruhsal açıdan kendilerini güvende hissedebilecekleri bir ortam imkanı sun	Ortam-p
5	Ayrıcalıklı kurumlardan bir tanesi. Uyguladıkları eğitim modellerinin de geçerliliği tartışılmaz derecede sağlam.	Kurum-p, Eğitim Modeli-p
6	Tüm dünyanın örnek aldığı Kanada eğitim sisteminin yıkılmaz kalesi diyebiliriz burası için.	Eğitim Sistemi-p
7	Bu kadar yeşil alana sahip başka bir okul kampüsü bulamazsınız. Uluslararası eğitim fırsatlarını bu kadar geniş ye	Kampüs-p, Eğitim Fırsatları-p
10	Anaokulundan Üniversiteye kadar tüm eğitimde sağladıkları başarılar ile sağlam adımlarla ilerlemekte.	Eğitim-p
11	Kısıtlı imkanlarla mucize yaratmayı bilen bir kadro, güler yüz ve sürekli yeni şeyler yapma isteğiyle öğrencileri de	Kadro-p, Öğretmenler-p, Bina-n
12	konuyu pekiştirmeyi ve kalıcı olmasını sağlayan bir sistemi benimsemişler.	Sistem-p
13	Okulun eğitimi iyi olabilir ama binası bana çok eski geldi.	Eğitim-p, Bina-n
14	Özellikle branş derslerindeki başarılarıyla gündeme gelen, kendinden çok büyük okullarla rekabet edebilen bir ok	Branş Dersleri-p, Okul-p
15	Çok iyi bir okul, çok memnunuz.	Okul-p
16	Öğretmenlerimizin çocuklarımız ile arkadaş olmalarından dolayı memnunuz.	Öğretmen-p
17	Okulun koridorlarının geniş olması, sınıfların temiz ve ferah olmasından dolayı çok memnunuz. Ayrıca çalışanla	Okul-p, Sınıf-p, Çalışan-p
18	Kurumsal bir okul. Montessori sistemi başarıyla uygulanıyor. Ücretler biraz yüksek.	Okul-p, Sistem-p, Ücret-n
19	Rehberlik hizmeti çok iyi. Kızım okula 2. sınıfta kaydoldu ve alışmakta zorluk çekiyordu. Başta rehberlik servisi çok	Rehber Hizmeti-p
20	Okul tertemiz. Sınıflar düzenli ve donanımlı. Fiziksel imkanlarından çok memnunuz. Çocuğum okuldayken çok rı	Okul-p, Sınıf-p, Fiziksel İmkan-p, Öğretmen-p, Eğitim
21	Çok eski bir okul olmamasına rağmen eğitimi sistemi oturmuş. Personellerin ve öğretmenlerin tamamı tecrübeli.	Okul-n, Eğitim Sistemi-p, Personel-p, Öğretmen-p
22	Eğitim düzeyi süper. Aile gibi bir okul. Fiyatıda koleje göre süper.	Eğitim-p, Okul-p, Fiyat-p
23	Branş öğretmenlerinin çok başarılı olması. Eğitimci ve personel kadrosunun güler yüzü ve ihtiyaçları anında kar	Branş Öğretmen-p, Eğitimci-p, Personel-p
24	Öğretmenlerin güler yüzü ve çocuklarla olması gerekenden daha çok ilgilenmeleri.	Öğretmen-p
25	Öğretmenlerin ilgisinden çok memnun kaldık. Oyun içerikli ders programları planlamaları çok hoşuma gitti.	Öğretmen-p, Ders Programı-p
26	Özellikle matematik derslerindeki başarı çok iyi.	Matematik Dersi-p

Şekil 19: Yorum Veri Seti İçindeki Hedefler ve Duygu Durumları

Şekil 19'da uzmanlar tarafından hedef ve duygu durumları tespit edilmiş her bir yoruma ait verilerin görünümü mevcuttur. Hedeflerin duygu durumları pozitif ve negatif olarak etiketlenmiştir. Bir yorumda birden fazla hedef olması durumunda hedefler aralarına virgül konularak eklenmiştir.

Şekil 19'da görüldüğü üzere bir yorum genellikle birden fazla duygu durumu içermektedir. Bu durum hedef tabanlı yaklaşımın önemini de göstermektedir. Hedef tabanlı yaklaşım sayesinde duygu kaybı yaşamadan bir yorum hakkında duygu analizi yapılabilir (Pontiki ve diğ., 2016).

<i>Sütun Adı</i>	<i>Açıklaması</i>
<i>Yorum_id</i>	<i>Yoruma verilen benzersiz numara</i>
<i>Yorum_metin</i>	<i>Kullanıcı yorumu</i>
<i>Hedef</i>	<i>Yorum içinde geçen hedefler (aspect) ve duygu durumları</i>

Tablo 10: Veri Setindeki Hedefler ve Duygu Durumları

Tablo 10’da görüldüğü üzere veri setinde “*Yorum_id*” alanı yoruma verilen benzersiz numarayı içermektedir. “*Yorum_metin*” alanı ise kimliksizleştirilmiş kullanıcı yorumunu içermektedir. Hedef alanı ise yorum içinde geçen hedefler ve her bir hedefin duygu durumunu içermektedir.

3.2. DUYGU SÖZLÜKLERİNİN KULLANIMA HAZIR HALE GETİRİLMESİ

Çalışma kapsamında SentiWordNet ve SentiTurkNet sözlükleri kullanılmıştır. SentiWordNet İngilizce dilinde hazırlanmış bir sözlüktür. SentiWordNet 117.659 kelime barındırmaktadır (Esuli ve Sebastiani, 2006). SentiTurkNet sözlüğü ise Türkçe dilinde hazırlanmış bir sözlüktür. SentiTurkNet ise 14.795 kelime barındırmaktadır (Dehkharghani ve diğ., 2016).

3.2.1. SentiTurkNet Sözlüğü'nün Kullanıma Hazır Hale Getirilmesi

CVS formatındaki sözlük dosyası çalışma grubunun sitesinden indirilmiş ve sql scripti ile bir MySQL veri tabanına aktarılmıştır. Şekil 20’de görüldüğü üzere CVS formatındaki sözlük dosyası daha kolay erişilebilmesi ve işlem yapılabilmesi için bir MySQL veri tabanına aktarılmıştır. Sözlük 12 sütundan oluşmaktadır.

sentitur_id	synonyms	TurkishGloss	PolarityLabel	Postag	neg_value	obj_value	pos_value	Eng_Synonyms	EnglishGloss	Swnpas	Swnneg
1	ayva reçeli	Meyveleri şekerle o	n	n	0,06	0,872	0,068	jam#1	preserve of crushed fruit	0	0
2	gül reçeli	Meyveleri şekerle o	n	n	0,06	0,872	0,068	jam#1	preserve of crushed fruit	0	0
3	incir reçeli	Meyveleri şekerle o	n	n	0,06	0,872	0,068	jam#1	preserve of crushed fruit	0	0
4	iştah	Yemek yeme isteğ o	n	n	0,06	0,872	0,068	(Null)	(Null)	(Null)	(Null)
5	iştahlı	İştahı olan, boğaz p	a	a	0,06	0,462	0,478	(Null)	(Null)	(Null)	(Null)
6	iştahsız	Yemek yeme isteğ n	a	a	0,48	0,452	0,068	(Null)	(Null)	(Null)	(Null)
7	iştahsızlık	İştahsız olma dur. n	n	n	0,512	0,42	0,068	(Null)	(Null)	(Null)	(Null)
8	kayısı reçeli	Meyveleri şekerle o	n	n	0,06	0,872	0,068	jam#1	preserve of crushed fruit	0	0
9	kayısı suyu	meyve suyu o	n	n	0,06	0,872	0,068	fruit_juice#1 fruit_crn drink produced by squeez	(Null)	(Null)	(Null)
10	kısayol	bilgisayarda bir dı o	n	n	0,06	0,872	0,068	(Null)	(Null)	(Null)	(Null)
11	mikrobiyolojik	mikroplar la uğraş o	a	a	0,06	0,872	0,068	(Null)	(Null)	(Null)	(Null)
12	minibüs	10, 12 kişilik küçül o	n	n	0,06	0,872	0,068	(Null)	(Null)	(Null)	(Null)
13	müşterilik	Müşterinin görevi o	n	n	0,06	0,872	0,068	(Null)	(Null)	(Null)	(Null)
14	nevresim	Torba biçiminde c o	n	n	0,06	0,872	0,068	(Null)	(Null)	(Null)	(Null)
15	neyzen	Ney üfleyici, ney ç o	n	n	0,06	0,872	0,068	(Null)	(Null)	(Null)	(Null)
16	ofsayt	Futbolde bir oyun n	n	n	0,502	0,43	0,068	(Null)	(Null)	(Null)	(Null)
17	portakal reçeli	Meyveleri şekerle o	n	n	0,06	0,872	0,068	jam#1	preserve of crushed fruit	0	0
18	postane, postahane	Posta ile gönderik o	n	n	0,06	0,872	0,068	(Null)	(Null)	(Null)	(Null)
19	rekortmen	Rekor kıran kimse p	n	n	0,06	0,06	0,88	(Null)	(Null)	(Null)	(Null)
20	repo	Bankalar arası işle o	n	n	0,06	0,872	0,068	(Null)	(Null)	(Null)	(Null)
21	sadrazam	Osmanlı İmparator p	n	n	0,06	0,452	0,488	(Null)	(Null)	(Null)	(Null)
22	vişne reçeli	Meyveleri şekerle o	n	n	0,06	0,872	0,068	jam#1	preserve of crushed fruit	0	0
23	vişne suyu	meyve suyu o	n	n	0,06	0,872	0,068	fruit_juice#1 fruit_crn drink produced by squeez	(Null)	(Null)	(Null)
24	şeftali reçeli	Meyveleri şekerle o	n	n	0,06	0,872	0,068	(Null)	(Null)	(Null)	(Null)
25	şeftali suyu	meyve suyu o	n	n	0,06	0,872	0,068	fruit_juice#1 fruit_crn drink produced by squeez	(Null)	(Null)	(Null)

Şekil 20: SentiTurkNet Sözlüğünün MySQL Veri Tabanına Aktarılmış Haline Ait Bir Görünüm

Tablo 11’de görüldüğü üzere sözlük 12 sütundan oluşmaktadır. *Sentitur_ID* her kelimeye verilen benzersiz numarayı içermektedir.

Synonyms alanı kelimenin kendisini içermektedir. *TurkishGloss* alanı kelimenin Türkçe açıklamasını içermektedir. *PolarityLabel* alanı kelimenin duygu durumunu içermektedir.

Duygu durumu pozitif ise “p”, negatif ise “n”, objektif ise “o” harfî yer almaktadır. *Postag* alanı kelimenin tür bilgisini içermektedir. *Postag* alanı “a” ise kelime sıfat, “n” ise isim, “v” ise fiil anlamına gelmektedir. *Neg_Value* alanı kelimenin negatif skorunu, *Pos_Value* alanı kelimenin pozitif skorunu, *Obj_Value* alanı ise kelimenin objektif skorunu içermektedir.

Bu üç skurun toplamı 1 yapmaktadır. Her bir kelime için bu üç skordan hangisi 1’e daha yakın ise kelimenin duygu durumu ona daha yakındır.

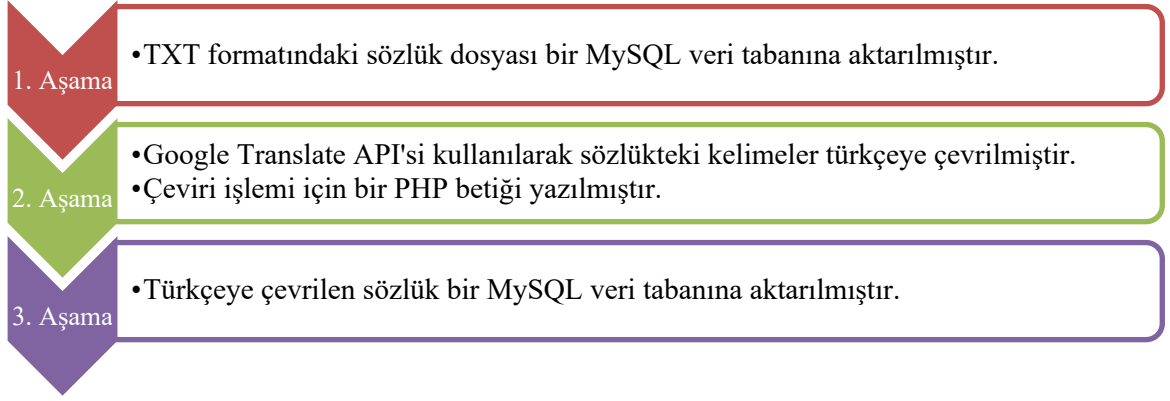
<i>Sütun Adı</i>	<i>Açıklaması</i>
<i>Sentitur_Id</i>	<i>Kelimeye verilen benzersiz numara</i>
<i>Synonyms</i>	<i>Sözlükte yer alan kelime</i>
<i>TurkishGloss</i>	<i>Türkçe açıklaması</i>
<i>PolarityLabel</i>	<i>Polarite etiketi</i>
<i>Postag</i>	<i>Kelime tür etiketi</i>
<i>Neg_Value</i>	<i>Kelimenin negatif puanı</i>
<i>Obj_Value</i>	<i>Kelimenin objektif puanı</i>
<i>Pos_Value</i>	<i>Kelimenin pozitif puanı</i>
<i>Eng_Synonyms</i>	<i>Kelimenin İngilizcedeki karşılığı</i>
<i>EnglishGloss</i>	<i>İngilizce açıklaması</i>
<i>Swnpos</i>	<i>Kelimenin SentiWordNet pozitif puanı</i>
<i>Swnneg</i>	<i>Kelimenin SentiWordNet negatif puanı</i>

Tablo 11: SentiTurkNet Sözlüğünün Veri Yapısı

Eng_Synonyms alanı kelimenin SentiWordNet (SWN) sözlüğündeki karşılık olan kelime ya da kelime grubunu içermektedir. *EnglishGloss* alanı kelimenin SWN'deki İngilizce açıklamasını içermektedir. *Swnpos* alanı kelimenin SWN'deki pozitif duygu skorunu, *Swnneg* alanı ise kelimenin SWN'deki negatif duygu skorunu içermektedir.

3.2.2. SentiWordNet Sözlüğü'nün Kullanıma Hazır Hale Getirilmesi

SentiWordNet sözlüğünün kullanıma hazır hale getirilmesi için 3 aşamadan oluşan bir yol izlenmiştir.



Şekil 21: SentiWordNet Sözlüğünün Türkçeye Çevrilmesine Ait Adımlar

İlk aşamada .txt uzantılı sözlük dosyası bir MySQL veri tabanına aktarılmıştır. Aktarım sırasında .txt uzantılı dosyası içerisindeki her bir satır bir SQL sorgusu haline getirilmiş ve bu şekilde veri tabanına aktarılmıştır.

```

SentiWordNet_3.0.0.txt
1 # SentiWordNet v3.0.0 (1 June 2010)
2 # Andrea Esuli
3 # SentiWordNet is distributed under the Attribution-ShareAlike 4.0 Unported (CC BY-SA 4.0) license.
4 # http://creativecommons.org/licenses/by-sa/4.0/
5 # For any information about SentiWordNet:
6 # Web: http://sentiwordnet.isti.cnr.it
7
8 # Data format.
9 # SentiWordNet v3.0 is based on WordNet version 3.0.
10 # WordNet website: http://wordnet.princeton.edu/
11 #
12 # The pair (POS,ID) uniquely identifies a WordNet (3.0) synset.
13 # The values PosScore and NegScore are the positivity and negativity
14 # score assigned by SentiWordNet to the synset.
15 # The objectivity score can be calculated as:
16 # ObjScore = 1 - (PosScore + NegScore)
17 # SynsetTerms column reports the terms, with sense number, belonging
18 # to the synset (separated by spaces).
19 #
20 # POS ID PosScore NegScore SynsetTerms Gloss
21 a 00001740 0.125 0 able#1 (usually followed by `to') having the necessary means or skill or know-how or authority to do something; "able to swim"; "sh
22 a 00002098 0 0.75 unable#1 (usually followed by `to') not having the necessary means or skill or know-how; "unable to get to town without a car"; "un
23 a 00002312 0 0 dorsal#2 abaxial#1 facing away from the axis of an organ or organism; "the abaxial surface of a leaf is the underside or side facing aw
24 a 00002527 0 0 ventral#2 adaxial#1 nearest to or facing toward the axis of an organ or organism; "the upper side of a leaf is known as the adaxial surf
25 a 00002730 0 0 acroscopic#1 facing or on the side toward the apex
26 a 00002843 0 0 basiscopic#1 facing or on the side toward the base
27 a 00002956 0 0 abducting#1 abducent#1 especially of muscles; drawing away from the midline of the body or from an adjacent part
28 a 00003131 0 0 adductive#1 adducting#1 adducent#1 especially of muscles; bringing together or drawing toward the midline of the body or toward an adja
29 a 00003356 0 0 nascent#1 being born or beginning; "the nascent chicks"; "a nascent insurgency"
30 a 00003553 0 0 emerging#2 emergent#2 coming into existence; "an emergent republic"
31 a 00003700 0.25 0 dissilient#1 bursting open with force, as do some ripe seed vessels
32 a 00003829 0.25 0 parturient#2 giving birth; "a parturient heifer"
33 a 00003939 0 0 dying#1 in or associated with the process of passing from life or ceasing to be; "a dying man"; "his dying wish"; "a dying fire"; "a dyi
34 a 00004171 0 0 moribund#2 being on the point of death; breathing your last; "a moribund patient"

```

Şekil 22: SentiWordNet'in TXT Formatındaki Veri Yapısı

Şekil 22'de SentiWordnet sözlüğünün bir .txt uzantılı dosya içindeki veri yapısı gözükmemektedir. Her bir kelime ayrı bir satırda yer alacak şekilde ve her bir sütun için bir tab boşluk bırakılmış şekilde hazırlanmıştır.

id	type	word_id	pstv_score	ngtv_score	word	word_explain
1	a	2098	0	0,75	unable#1	(usually followed by 'to') not having the necessary means or skill or know-how; "unable"
2	a	2312	0	0	dorsal#2 abaxial#1	facing away from the axis of an organ or organism; "the abaxial surface of a leaf is the u"
3	a	2527	0	0	ventral#2 adaxial#1	nearest to or facing toward the axis of an organ or organism; "the upper side of a leaf is t"
4	a	2730	0	0	acrescopic#1	facing or on the side toward the apex
5	a	2843	0	0	basiscopic#1	facing or on the side toward the base
6	a	2956	0	0	abducting#1 abducent#1	especially of muscles; drawing away from the midline of the body or from an adjacent
7	a	3131	0	0	adductive#1 adducting#1 adducent#1	especially of muscles; bringing together or drawing toward the midline of the body or t
8	a	3356	0	0	nascent#1	being born or beginning; "the nascent chicks"; "a nascent insurgency"
9	a	3553	0	0	emerging#2 emergent#2	coming into existence; "an emergent republic"
10	a	3700	0,25	0	dissilient#1	bursting open with force, as do some ripe seed vessels
11	a	3829	0,25	0	parturient#2	giving birth; "a parturient heifer"
12	a	3939	0	0	dying#1	in or associated with the process of passing from life or ceasing to be; "a dying man"; "I
13	a	4171	0	0	moribund#2	being on the point of death; breathing your last; "a moribund patient"
14	a	4296	0	0	last#5	occurring at the time of death; "his last words"; "the last rites"
15	a	4413	0	0	abridged#1	(used of texts) shortened by condensing or rewriting; "an abridged version"
16	a	4615	0	0	shortened#4 cut#3	with parts removed; "the drastically cut film"
17	a	4723	0	0	half-length#2	abridged to half its original length
18	a	4817	0	0	potted#3	(British informal) summarized or abridged; "a potted version of a novel"
19	a	4980	0	0	unabridged#1	(used of texts) not shortened; "an unabridged novel"
20	a	5107	0,5	0	uncut#7 full-length#2	complete; "the full-length play"
21	a	5205	0,5	0	absolute#1	perfect or complete or pure; "absolute loyalty"; "absolute silence"; "absolute truth"; "ab
22	a	5473	0,75	0	direct#10	lacking compromising or mitigating elements; exact; "the direct opposite"
23	a	5599	0,5	0,5	unquestioning#2 implicit#2	being without doubt or reserve; "implicit trust"
24	a	5718	0,125	0	infinite#4	total and all-embracing; "God's infinite wisdom"
25	a	5839	0,5	0,125	living#3	(informal) absolute; "she is a living doll"; "scared the living daylights out of them"; "bea

Şekil 23: SentiWordNet'in MySQL Veri tabanına Aktarılmış Haline Ait Bir Görünüm

Şekil 23'te görüldüğü üzere TXT formatındaki sözlük dosyası daha kolay erişilebilmesi ve işlem yapılabilmesi için bir MySQL veri tabanına aktarılmıştır. Sözlük altı sütundan oluşmaktadır.

Sütun Adı	Açıklaması
Word_Id	Kelimeye verilen benzersiz numara
Type	Kelime türü
NgTv_Score	Kelimenin negatif puanı
PstV_Score	Kelimenin pozitif puanı
Word	Sözlükte yer alan kelime
Word_Explain	Kelimeye ait açıklama

Tablo 12: SentiWordNet Sözlüğünün Veri Yapısı

Tablo 12'de görüldüğü üzere sözlük altı sütundan oluşmaktadır. *Word_Id* alanı her kelimeye verilen benzersiz numarayı içermektedir. *Type* alanı kelime tür bilgisini içermektedir. *Type*

alanı “a” ise kelime sıfat, “n” ise isim, “v” ise fiil anlamına gelmektedir. *Ngty_Score* alanı kelimenin negatif skorunu, *Pstv_Score* alanı kelimenin pozitif skorunu içermektedir. *Word* alanı kelimenin kendisini içermektedir. *Word_Explain* alanı ise kelimenin Türkçe açıklamasını içermektedir.

İkinci aşamada Google’ın Translate API’si kullanılarak bir yazılım geliştirilmiştir. Bu yazılım ile SentiWordNet sözlüğü Türkçeye çevrilmiştir. Yazılım PHP programlama dili kullanılarak geliştirilmiştir. Yazılım ile çevirisi yapılan her bir kelime otomatik olarak yeni bir veri tabanına aktarılmıştır.

Google Translate (<https://translate.google.com>), hem web sitesi üzerinden hem de API üzerinden hizmet veren bir Google ürünüdür. Google Translate uygulaması ile 100’ün üzerinde dilde çeviri yapmak mümkündür. Çeviri algoritmalarında makine öğrenmesi tekniklerinden faydalanmaktadırlar. Bu hizmeti kullanarak metin çevirisi yanında yüklediğiniz bir dokümanı da (.doc, .docx, .odf, .pdf, .ppt, .pptx, .ps, .rtf, .txt, .xls veya .xlsx dosyası) otomatik olarak çevirebilirsiniz. Bu servisi web ara yüzünden ya da API üzerinden kullanmak ücretsizdir (Aiken ve Balan, 2011).

```

10     $tr = new TranslateClient('en', 'tr');
11     $cevrimeyenKelimeSayisi=0;
12     $cevriyenKelimeSayisi=0;
13     $eklemeSayac1=0;
14     $eklemeSayac2=0;
15     $i=0;
16     $query = $db->query("SELECT * FROM en_source", PDO::FETCH_ASSOC);
17     if ( $query->rowCount() ){
18         foreach( $query as $row ){
19             $dizi = explode("#",$row['word']);
20             $ingilizce = $dizi[0];
21             $turkce = $tr->translate($ingilizce);
22             echo $i++;
23             echo "İngilizce: ".$ingilizce." ";
24             echo "Türkçe: ".$turkce;
25             if ($ingilizce==$turkce) {
26                 echo "çevrilemedi.";
27                 $cevrimeyenKelimeSayisi++;
28                 $query1 = $db->prepare("INSERT INTO untranslate SET type = ?, word_id = ?, pstv_score = ?, ngty_score = ?, word ");
29                 $insert1 = $query1->execute(array($row['type'], $row['word_id'], $row['pstv_score'], $row['ngty_score'], $row['w
30             if ($insert1) {
31                 $eklemeSayac1 ++;
32             }
33         }

```

Şekil 24: PHP Programlama Dili Kullanılarak Geliştirilen Çeviri Sistemi

Üçüncü aşamada çevrilen kelimeler otomatik olarak bir MySQL veri tabanına aktarılmıştır.

word_id	word_tr	pstv_score	ngtv_score	type	word_en	word_explain
2098	acı	0	0,75	a	unable	(usually followed by 'to') not having the necessary means or skill or know-how; "i
2312	sırt	0	0	a	dorsal	facing away from the axis of an organ or organism; "the abaxial surface of a leaf i
2527	karnı	0	0	a	ventral	nearest to or facing toward the axis of an organ or organism; "the upper side of a
2956	kaçarak	0	0	a	abducting	especially of muscles; drawing away from the midline of the body or from an adju
3356	doğan	0	0	a	nascent	being born or beginning; "the nascent chicks"; "a nascent insurgency"
3553	gelişmekte olan	0	0	a	emerging	coming into existence; "an emergent republic"
3829	doğuran	0,25	0	a	parturient	giving birth; "a parturient heifer"
3939	ölen	0	0	a	dying	in or associated with the process of passing from life or ceasing to be; "a dying m
4171	can çekişen	0	0	a	moribund	being on the point of death; breathing your last; "a moribund patient"
4296	son	0	0	a	last	occurring at the time of death; "his last words"; "the last rites"
4413	kısaltılmış	0	0	a	abridged	(used of texts) shortened by condensing or rewriting; "an abridged version"
4615	kısaltılmış	0	0	a	shortened	with parts removed; "the drastically cut film"
4723	yarı uzunluk	0	0	a	half-length	abridged to half its original length
4817	saksılı	0	0	a	potted	(British informal) summarized or abridged; "a potted version of a novel"
4980	kısaltılmamış	0	0	a	unabridged	(used of texts) not shortened; "an unabridged novel"
5107	kesilmemiş	0,5	0	a	uncut	complete; "the full-length play"
5205	kesin	0,5	0	a	absolute	perfect or complete or pure; "absolute loyalty"; "absolute silence"; "absolute truth
5473	direkt	0,75	0	a	direct	lacking compromising or mitigating elements; exact; "the direct opposite"
5599	kayıtsız şartsız	0,5	0,5	a	unquestioning	being without doubt or reserve; "implicit trust"
5718	sonsuz	0,125	0	a	infinite	total and all-embracing; "God's infinite wisdom"
5839	yaşam	0,5	0,125	a	living	(informal) absolute; "she is a living doll"; "scared the living daylight out of them"
6032	bağil	0,25	0,5	a	relative	estimated by comparison; not absolute or complete; "a relative stranger"
6245	ilişkisel	0	0	a	relational	having a relation or being related
6336	emici	0	0	a	absorptive	having power or capacity or tendency to absorb or soak up something (liquids or

Şekil 25: Türkçeye Çevrilen SentiWordNet'in MySQL Veri tabanına Aktarılmış Haline Ait Bir Görünüm

Çeviri sırasında sadece kelime alanının çevirisi yapılmıştır. Şekil 25'te görüldüğü üzere kelimenin Türkçe karşılığı "*Word_Tr*" alanında tutulmaktadır.

Sütun Adı	Açıklaması
<i>Word_Id</i>	Kelimeye verilen benzersiz numara
<i>Type</i>	Kelime türü
<i>NgTv_Score</i>	Kelimenin negatif puanı
<i>PstV_Score</i>	Kelimenin pozitif puanı
<i>Word_Tr</i>	Türkçeye çevrilen sözlükte yer alan kelime
<i>Word_En</i>	Sözlükte yer alan kelime (İngilizce)
<i>Word_Explain</i>	Kelimeye ait açıklama

Tablo 13: Türkçeye Çevrilen SentiWordNet Sözlüğünün Veri Yapısı

Tablo 13’de görüldüğü üzere sözlük yedi sütundan oluşmaktadır. *Word_Id* alanı her kelimeye verilen benzersiz numarayı içermektedir. *Type* alanı kelime tür bilgisini içermektedir. *Type* alanı “a” ise kelime sıfat, “n” ise isim, “v” ise fiil anlamına gelmektedir. *Ngtv_Score* alanı kelimenin negatif skorunu, *Pstv_Score* alanı kelimenin pozitif skorunu içermektedir. *Word* alanı kelimenin kendisini içermektedir. *Word_Explain* alanı kelimenin Türkçe açıklamasını içermektedir. *Word_Tr* alanı ise kelimenin Türkçeye çevrilmiş halini içermektedir.

3.3. İTÜ DDİYZ’NİN KULLANIMA HAZIR HALE GETİRİLMESİ

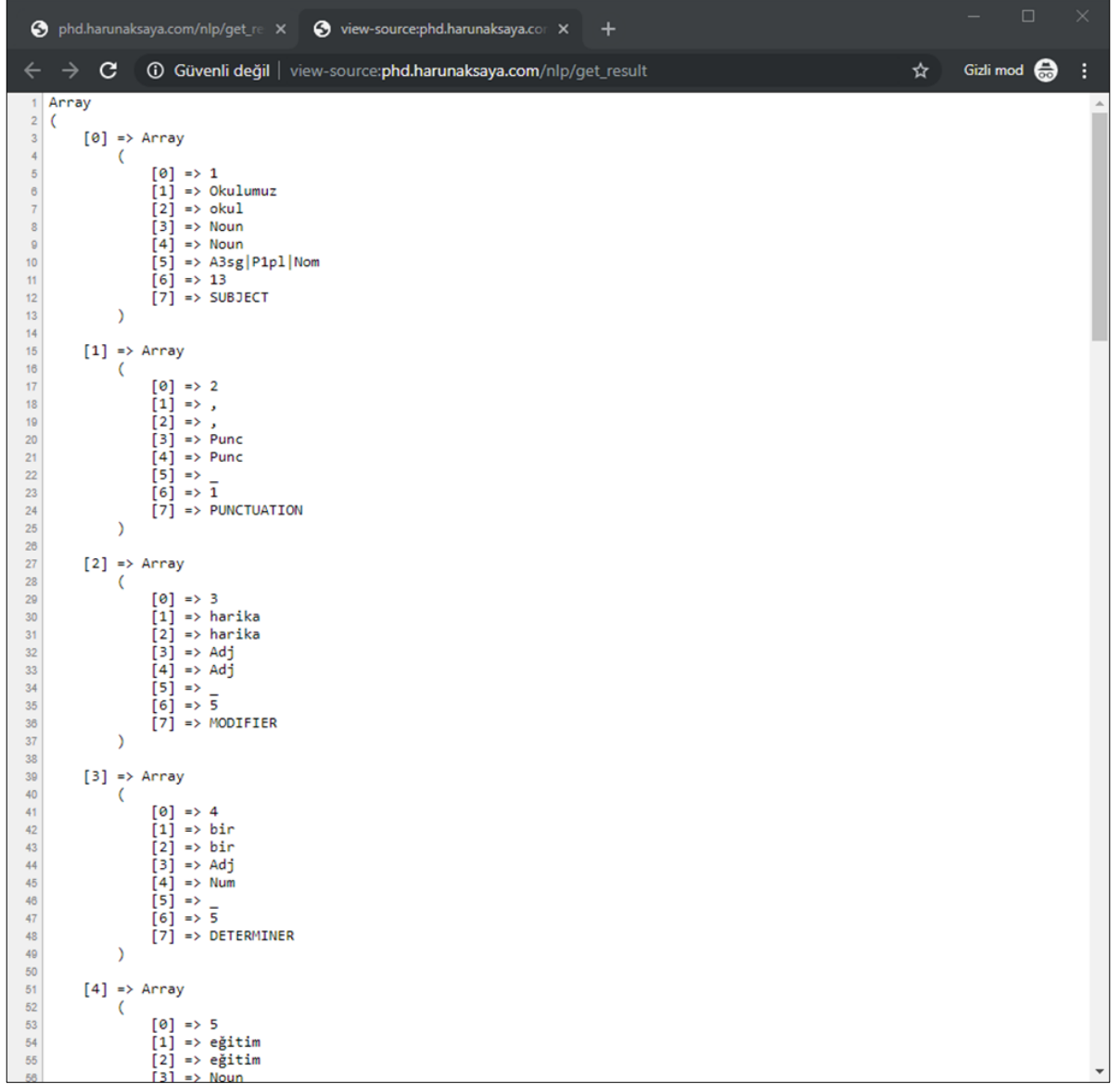
İTÜ tarafından geliştirilen Türkçe Doğal Dil İşleme Yazılım Zinciri’nin kullanımı için birim yetkilisi Doç. Dr. Gülşen Cebiroğlu Eryiğit ile iletişime geçilmiş ve talepte bulunulmuştur. Tarafımıza gönderilen kullanım sözleşmesi EK 1’de yer almaktadır.

Göndermiş oldukları kullanıcı adı ve şifre bilgilerini kullanarak servis erişimi sağlandı. Servis entegrasyon çalışmalarını tamamlayarak bir yazılım ara yüzü geliştirildi ve servis test edildi. Servis, Normalizer, Morphological Analyzer, Morphological Disambiguator, Named Entity Recognizer ve Syntactic Parser fonksiyonlarına sahiptir (Eryiğit, 2014; Eryiğit ve diğ., 2008).

Şekil 26: İTÜ DDİYZ Servisini Test Etmek için Geliştirilen Yazılım Ara Yüzü

Şekil 26’da İTÜ DDİYZ servisini test etmek için geliştirilen yazılımın ara yüzü görülmektedir.

Yorum metni alanına yazılan bir metin “Gönder” butonuna tıklandığında servis aracılığı ile İTÜ DDİYZ’ne gönderilip işlenmektedir. İşlenen metin, bir dizi formatında geri dönmektedir.



```

1 Array
2 (
3   [0] => Array
4   (
5     [0] => 1
6     [1] => Okulumuz
7     [2] => okul
8     [3] => Noun
9     [4] => Noun
10    [5] => A3sg|Pl|Nom
11    [6] => 13
12    [7] => SUBJECT
13  )
14
15  [1] => Array
16  (
17    [0] => 2
18    [1] => ,
19    [2] => ,
20    [3] => Punc
21    [4] => Punc
22    [5] => _
23    [6] => 1
24    [7] => PUNCTUATION
25  )
26
27  [2] => Array
28  (
29    [0] => 3
30    [1] => harika
31    [2] => harika
32    [3] => Adj
33    [4] => Adj
34    [5] => _
35    [6] => 5
36    [7] => MODIFIER
37  )
38
39  [3] => Array
40  (
41    [0] => 4
42    [1] => bir
43    [2] => bir
44    [3] => Adj
45    [4] => Num
46    [5] => _
47    [6] => 5
48    [7] => DETERMINER
49  )
50
51  [4] => Array
52  (
53    [0] => 5
54    [1] => eğitim
55    [2] => eğitim
56    [3] => Noun

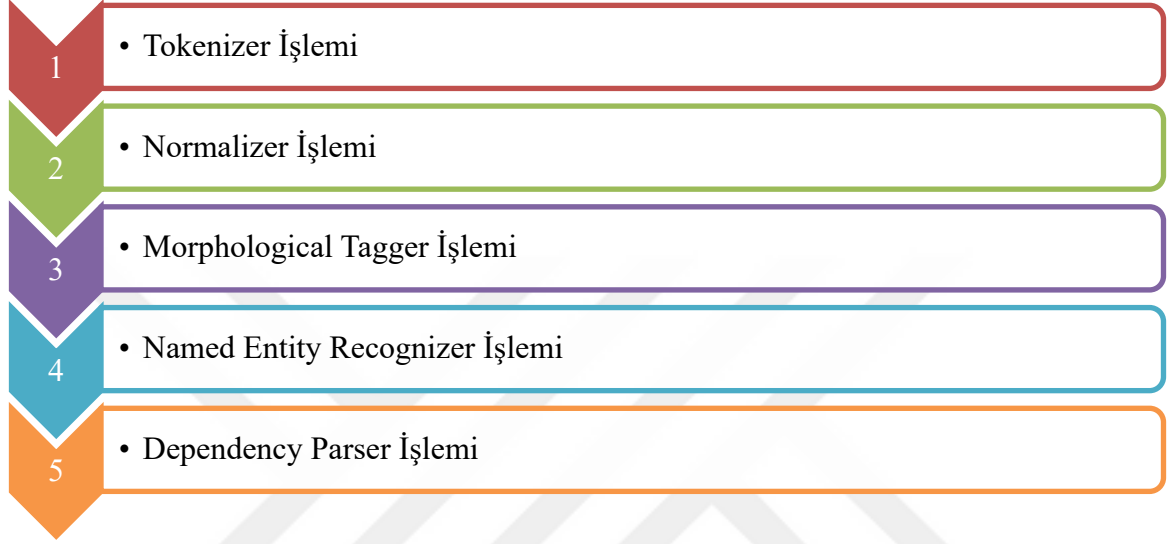
```

Şekil 27: İTÜ DDİYZ Servisi Tarafından İşlenen Bir Yorum Ait Çıktı

Şekil 27’de İTÜ DDİYZ servisine gönderilen ve işlenip geri gönderilen bir yorum metnine ait çıktı gözükmemektedir. Metinde yer alan her bir kelimenin eksiz hali, türü, almış olduğu ekin türü ve ilişkili olduğu kelimenin satır sayısı bilgisi yer almaktadır.

3.4. VERİ SETİNİN İTÜ DDİYZ SERVİSİ İLE İŞLENMESİ

Okul.com.tr sitesinden elde edilen yorumlar İTÜ DDİYZ “Whole Pipeline” sınıfı kullanılarak işlenmiştir. Bu sınıf, Tokenizer, Normalizer, Morphological Tagger, Named Entity Recognizer, Dependency Parser metodlarına sahiptir.



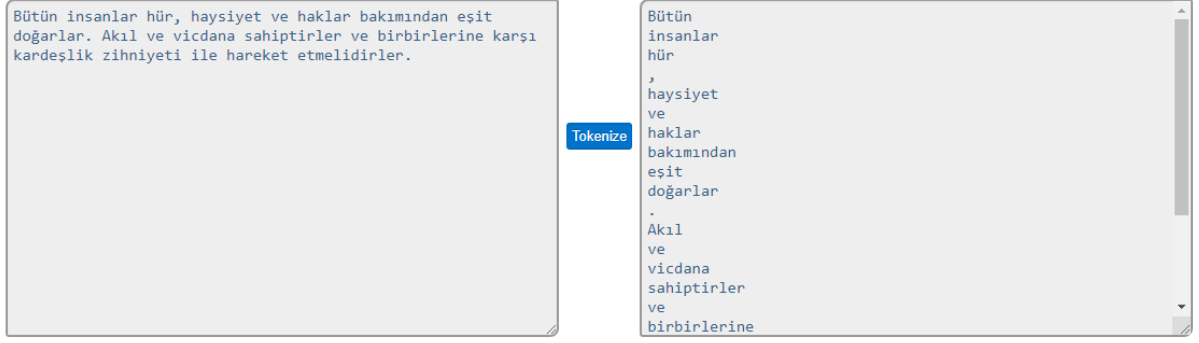
Şekil 28: Whole Pipeline Sınıfına Ait Metotlar ve Metnin İşlenme Sırasına Ait Bir Görünüm

Şekil 28’de görüldüğü üzere işlenmek üzere servise gönderilen bir metin önce “Tokenizer” işlemine tabi tutulmaktadır. Sonrasında sırasıyla “Normalizer”, “Morphological Tagger”, “Named Entity Recognizer”, “Dependency Parser” ve son olarak “UD Mapper” işlemine tabi tutulmaktadır.

Tokenizer İşlemi

Bu adımda İTÜ DDİYZ servisine verilen bir metin cümlesi, kelime düzeyinde parçalarına ayrılır. Tokenizer aşaması normalizasyon sürecinin ilk aşamasıdır. Bir metni tokenizer işlemine tabi tutmadan sağlıklı bir şekilde diğer ön işleme süreçlerini gerçekleştiremezsiniz (Torunoğlu ve diğ., 2014).

Tokenizer



Şekil 29: İTÜ DDİYZ -Tokenizer İşlemine Ait Görünüm (İTÜ DDİYZ, 2021)

Şekil 29’da İTÜ DDİYZ –Tokenizer işlemine ait bir görsel bulunmaktadır. Servise verilen bir yorum metni, kelime ve noktalama işareti düzeyinde parçalanmıştır.

Normalizer İşlemi

Normalizer işlemi DDİ sürecinin olmaz ise olmaz bir adımıdır. DDİ çalışmalarında toplanan metinler genellikle sorunlu ve düzeltilmeye muhtaçtır. Bu süreçte yedi farklı işlem yapılır:

1. Letter Case Transformation,
2. Replacement Rules and Lexicon Lookup,
3. Proper Noun Detection,
4. Deasciification,
5. Vowel Restoration,
6. Accent Normalization,
7. Spelling Correction (Eryiğit ve diğ., 2017).

Dependency Parser

İTÜ DDiYZ’de yer alan Dependency Parser aracı ile kelimeler arası ilişkilerin tespit edilebilmektedir. Kelimeler arası ilişkilerin tespit edilebilmesi DDİ çalışmalarında ve de özellikle hedef tabanlı duygu analizi çalışmalarında çok önem arz etmektedir (Eryiğit ve diğ., 2008).

Dependency Parser

☒ Formal language ☐ Noisy language

1	Bütün	bütün	Adj	Adj	—			
2	insanlar	insan	Noun	Noun	Noun	A3p1 Pnon Nom		
3	hür	hür	Adj	Adj	—			
4	,	,	Punc	Punc	—			
5	haysiyet	haysiyet	Noun	Noun	A3sg Pnon Nom			
6	ve	ve	Conj	Conj	—			
7	haklar	hak	Noun	Noun	A3p1 Pnon Nom			
8	bakımından	bakım	Noun	Noun	A3sg P3sg Abl			
9	eşit	eşit	Adj	Adj	—			
10	doğarlar	doğ	Verb	Verb	Pos Aor A3p1			
11	.	.	Punc	Punc	—			

Parse

1	Bütün	bütün	Adj	Adj	—	2	MODIFIER	
2	insanlar	insan	Noun	Noun	Noun	A3p1 Pnon Nom	10	SUBJECT
3	hür	hür	Adj	Adj	—	10	MODIFIER	
4	,	,	Punc	Punc	—	3	PUNCTUATION	
5	haysiyet	haysiyet	Noun	Noun	A3sg Pnon Nom	7	COORDINATION	
6	ve	ve	Conj	Conj	—	5	CONJUNCTION	
7	haklar	hak	Noun	Noun	A3p1 Pnon Nom	8	POSSESSOR	
8	bakımından	bakım	Noun	Noun	A3sg P3sg Abl	10	MODIFIER	
9	eşit	eşit	Adj	Adj	—	10	MODIFIER	
10	doğarlar	doğ	Verb	Verb	Pos Aor A3p1	10	PREDICATE	
11	.	.	Punc	Punc	—	10	PUNCTUATION	

Şekil 33: İTÜ DDiYZ -Dependency Parser İşlemine Ait Görünüm (İTÜ DDiYZ, 2021)

Şekil 33’te görüldüğü üzere İTÜ DDiYZ Dependency Parser işlemi uygulanmış bir metin cümlesi görülmektedir. Görsel üzerindeki oklar kelimeler arası bağılıkları göstermek amacıyla kullanılmıştır. Birinci satırda yer alan “Bütün” kelimesi 2. satırda yer alan “insanlar” kelimesine, “insanlar” kelimesi ise 10. satırda yer alan “doğarlar” kelimesine bağlıdır.

Yorum Veri Setinin İTÜ DDİYZ ile İşlenmesi İçin Geliştirilen Yazılım

Okul.com.tr sitesinden alınan 1000 yorumun İTÜ DDİYZ ile işlenmesi için bir yazılım geliştirilmiştir.

```

1  <?php
2      ini_set('MAX_EXECUTION_TIME', '-1');
3      $db = new PDO("mysql:host=localhost;dbname=phd2;charset=utf8", "root", "");
4      $url = 'http://tools.nlp.itu.edu.tr/SimpleApi';
5      $query = $db->query("SELECT * FROM okul_yorum_v3 where yorum_id = 826", PDO::FETCH_ASSOC);
6      if ( $query->rowCount() ){
7          foreach( $query as $row ){
8              $yorum_metin = $row['yorum_metin'];
9              $yorum_id = $row['yorum_id'];
10             $context = stream_context_create(array(
11                 'http' => array(
12                     'method' => 'POST',
13                     'header' => 'Content-type: application/x-www-form-urlencoded',
14                     'content' => http_build_query(
15                         array(
16                             'token' => '4wwdzhoGJbT1KhPB1wLu8PHp4JD0zLu8',
17                             'tool' => 'pipelineFormal',
18                             'input' => $yorum_metin,
19                             'type' => 'text',
20                             'data' => utf8_encode('Hello, world! (æøâ&øÅ)')
21                         )
22                     ),
23                     'timeout' => 60
24                 )
25             ));
26             $resp = file_get_contents($url, TRUE, $context);
27             $arr = explode("\n", $resp);
28             for ($i=0; $i < count($arr); $i++) {
29                 $temp[$i] = explode("\t", $arr[$i]);
30             }
31             unset($arr);
32             $test2 = serialize($temp);
33             //echo $test2;die;
34             unset($temp);
35             $query = $db->prepare("INSERT INTO yorum_parsed SET parsed_content =
36             :parsed_content, yorum_id = :id, content = :yorum_metin");
37             $insert = $query->execute(array(
38                 "parsed_content" => $test2,
39                 "id" => $yorum_id,
40                 "yorum_metin" => $yorum_metin
41             ));
42             if ( $insert ){
43                 $last_id = $db->lastInsertId();
44                 print "insert işlemi başarılı!";
45                 unset($test2);
46             }
47         }
48     }
49     ?>

```

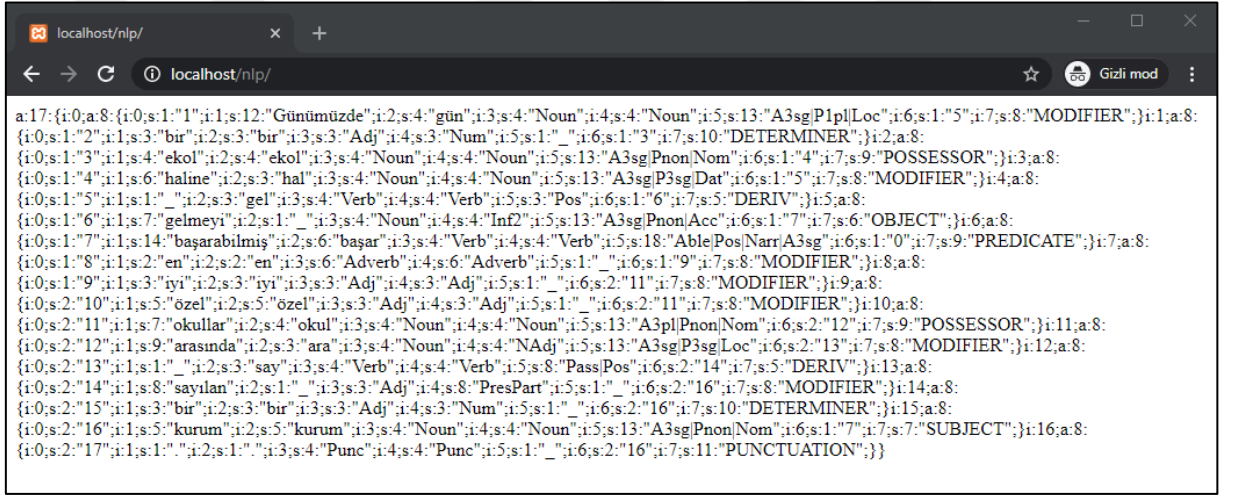
Şekil 34: Yorum Veri Setinin İTÜ DDİYZ ile İşlenmesi için Geliştirilen Yazılım

Geliştirilen bu yazılım İTÜ DDIYZ'ne bağlanıp her bir yorum metninin işlenmesi ve servisten dönen sonucun bir MySQL veri tabanına aktarılması işlemlerini yerine getirmektedir.

“Günümüzde bir ekol haline gelmeyi başaramış en iyi özel okullar arasında sayılan bir kurum.”

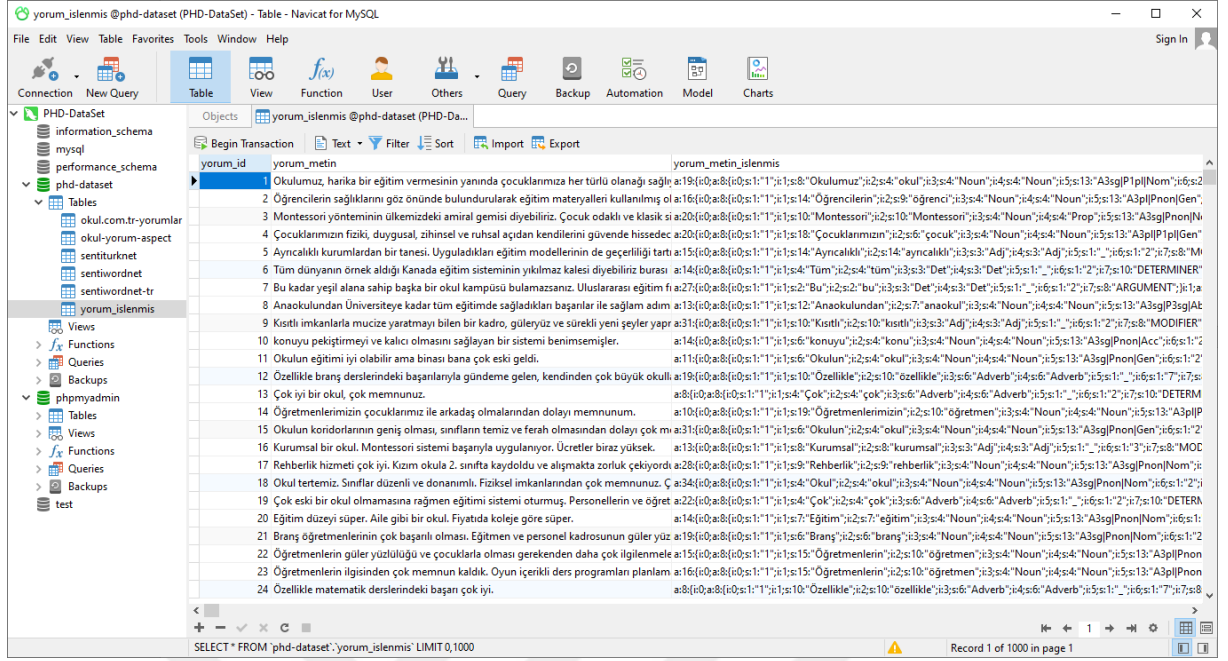
Şekil 35: İTÜ DDIYZ'ne İşlenmesi için Gönderilen Örnek Bir Yorum Metni

Şekil 35'te işlenmesi için İTÜ DDIYZ'ne gönderilecek örnek bir yorum metni görülmektedir.



Şekil 36: İTÜ DDIYZ ile İşlenmiş Örnek Bir Yorum Metni

Şekil 36'da İTÜ DDIYZ ile işlenmiş yorum metninin çıktısı görüntülenmektedir. Veri seti içerisinde yer alan her bir yorum için bu işlem yapılmış, işlenen yorum metni ilişkisel bir yapıda MySQL veri tabanına aktarılmıştır.



Şekil 37: İTÜ DDİYZ ile İşlenen Veri Setinin MySQL Veri tabanına Aktarılmış Haline Ait Bir Görünüm

Şekil 37’de İTÜ DDİYZ ile işlenip bir MySQL veri tabanına aktarılmış yorum veri setine ait bir görünüm yer almaktadır. Bu veri seti 1000 satırdan oluşmaktadır.

Tablo 14 : Veri tabanına Aktarılan İşlenmiş Veri Setinin Veri Yapısı

Sütun Adı	Açıklaması
<i>Yorum_Id</i>	<i>Her bir yoruma verilen benzersiz numara</i>
<i>Yorum_Metin</i>	<i>İşlenmemiş yorum metni.</i>
<i>Yorum_Metin_Islenmis</i>	<i>İTÜ DDİYZ ile işlenmiş yorum metni</i>

Tablo 14’te İTÜ DDİYZ’de işlenip aktarılmış yorumları içeren veri tabanına ait veri yapısı gözükmemektedir. *Yorum_Id* alanı her bir yoruma verilen benzersiz numarayı içermektedir. *Yorum_Metin* alanı işlenmemiş yorum metnini, *Yorum_Metin_Islenmis* alanı ise İTÜ DDİYZ ile işlenmiş yorum metnine ait verileri içermektedir.

3.5. İLİŞKİLİ KELİME TESPİTİ İÇİN YAZILIM GELİŞTİRİLMESİ

Hedef tabanlı duygu analizi çalışması için hedefi niteleyen kelimelerin tespit edilmesi çok önem arz etmektedir. Hedefin duygu durumuna karar vermek için hedefi niteleyen kelimelerin duygu durumuna bakılır. İlişkili kelimelerin tespiti için önce İTÜ DDİYZ’nde yer alan Dependency Parser kullanılmıştır. Ancak burada yeterince performans alınamadığı gözlenmiş ve ilişkili kelimelerin tespiti için bir yazılım geliştirilmiştir.

```

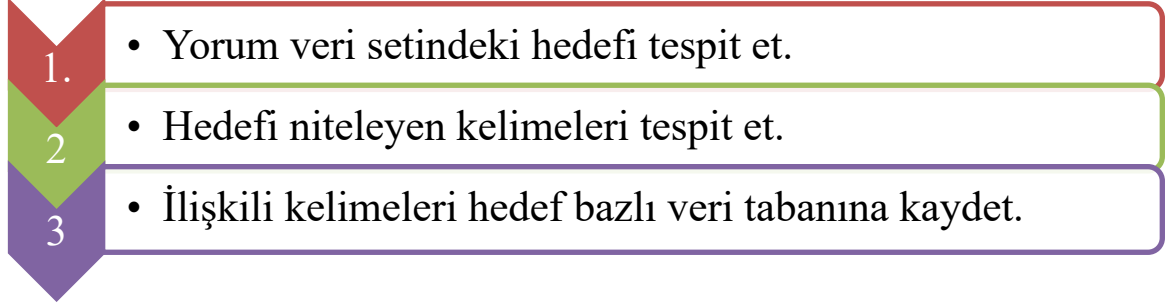
54 //Serilaze edilmiş yorum cümlesi işlenebilmesi için unserialize ediliyor...
55 $yorum_parsed = unserialize($row['yorum_parsed']);
56
57 //Yorum cümlesinin boyutu alınıyor.
58 $dizi_boyutu = count($yorum_parsed);
59 echo "Dizi boyutu: ". $dizi_boyutu;
60 echo "<br>";
61
62 //Aspect kaç kelimeden oluşuyor.
63 $baslangicAspectDizi=explode(" ", $baslangic_aspect); // cümlemiz boşluklardan bölünecek
64
65 //İlgili aspect tek bir kelimeden oluşuyorsa
66 if (count($baslangicAspectDizi)==1) {
67 //Cümle içinde aspectler tespit ediliyor.
68 for ($i=0; $i < $dizi_boyutu; $i++) {
69
70 //Türkçe karakter probleminin giderilmesi
71 $bul = array("ö");
72 $degistir = array("o");
73 $yorum_parsed[$i][1] = str_replace($bul, $degistir, $yorum_parsed[$i][1]);
74 $yorum_parsed[$i][2] = str_replace($bul, $degistir, $yorum_parsed[$i][2]);
75
76 $bul = array("ç");
77 $degistir = array("c");
78 $yorum_parsed[$i][1] = str_replace($bul, $degistir, $yorum_parsed[$i][1]);
79 $yorum_parsed[$i][2] = str_replace($bul, $degistir, $yorum_parsed[$i][2]);
80
81
82 $bul = array("ı");
83 $degistir = array("i");
84 $yorum_parsed[$i][1] = str_replace($bul, $degistir, $yorum_parsed[$i][1]);
85 $yorum_parsed[$i][2] = str_replace($bul, $degistir, $yorum_parsed[$i][2]);
86
87 if (strstr($yorum_parsed[$i][2], $baslangicAspectDizi[0]) or strstr($yorum_parsed[$i][1], $baslangicAspectDizi[0])){
88
89 echo "Aspect: ".$baslangicAspectDizi[0];
90 echo "<br>";
91 echo "Yorum İçindeki Karşılığı: ".$yorum_parsed[$i][2];
92 echo "<br>";
93 echo "Yorum İçindeki İndeksi: ".$i;
94 $aspectler[] = $i;
95 echo "<br>";
96 print_r($yorum_parsed);
97 echo "<br>";
98 echo "Tespit edilen aspectler:";
99 echo "<br>";
100 print_r($aspectler);
101 echo "<br>";
102 }

```

Şekil 38: İlişkili Kelimelerin Tespit Edilmesi İçin Geliştirilen Yazılımdan Bir Kesit

Şekil 38’de hedefleri niteleyen ilişkili kelimelerin tespiti için geliştirilen yazılımdan bir kesit görülmektedir. Script ilişkili kelimeleri tespit edip bu kelimeleri ilgili yorum verisi ile ilişkili bir formatta MySQL veri tabanına aktarmaktadır.

İlişkili kelimelerin tespiti için geliştirilen yazılımın işlem aşamaları aşağıdaki gibidir:



Şekil 39: İlişkili Kelimelerin Tespit Edilmesine Ait Adımlar

İlişkili kelimelerin tespit edilmesinde şu hususlara dikkat edilmektedir.

1. Hedef cümle içerisinde kaç kez geçiyor?
2. Hedef tek kelimedenden mi yoksa çok kelimedenden mi oluşuyor?
3. Hedefi niteleyen kelime hedef ile aynı cümle içerisinde mi?

Script 1000 yorum üzerinde çalıştırılmış, ilişkili kelimeler MySQL veri tabanına aktarılmıştır. İlişkili kelimeler hedef bazlı olarak tespit edildiği için cümle içerisinde hedef sayısı kadar veri tabanında yeni satır oluşturulmuştur.

Hedeflere dair ilişkili kelime tespiti çalışması sonucunda bir yorum metni içinde birden fazla hedef olmasından ötürü 1000 yoruma ait 1925 hedef oluşmuştur. Her bir hedefin bağlı olduğu yorum, hedefin duygu durumu ve ilişkili kelimeleri ilişkisel formatta geliştirilen yazılım ile kaydedilmiştir.

Tablo 15: Duygu Durumlarına Göre Hedef Sayıları

<i>Duygu Durumu</i>	<i>Hedef Sayısı</i>
<i>Negatif</i>	452
<i>Pozitif</i>	1473
<i>Toplam:</i>	1925

Tablo 15'te görüldüğü üzere 1925 hedefin 452'sinin duygu durumu negatif 1473'ü ise pozitifdir.

yorum_id	yorum_metin	hedef	duygu_durumu	yorum_metin_islenmis	iliskili_kelimeler
1	Okulunuz, harika bir eğitim vermesinin yanında çok Olanak	p	a:19(i0:a8(i0:s1:1":i1:s8: eğitim vermesinin yanında çocuklarınıza her türlü olanağı sağlıyor		
1	Okulunuz, harika bir eğitim vermesinin yanında çok Eğitim	p	a:19(i0:a8(i0:s1:1":i1:s8: Okulunuz harika bir eğitim vermesinin yanında çocuklarınıza her türlü olanağı		
1	Okulunuz, harika bir vermesinin yanında çocukların Okul	p	a:19(i0:a8(i0:s1:1":i1:s8: Okulunuz harika bir eğitim vermesinin yanında Çok başarılı bir okul		
2	Öğrencilerin sağlıklarını göz önünde bulundurularak Eğitim Materyali	p	a:16(i0:a8(i0:s1:1":i1:s8: Öğrencilerin sağlıklarını göz önünde bulundurularak eğitim materyalleri kullanılmış olması çok b		
3	Montessori yönteminin ülkemizdeki amiral gemisi d Okul	p	a:20(i0:a8(i0:s1:1":i1:s8: ve klasik sistemin tamamen dışında alternatif bir okul		
3	Montessori yönteminin ülkemizdeki amiral gemisi d Sistem	p	a:20(i0:a8(i0:s1:1":i1:s8: Çocuk odaklı ve klasik sistemin tamamen dışında alternatif bir okul		
3	Montessori yönteminin ülkemizdeki amiral gemisi d Yöntem	p	a:20(i0:a8(i0:s1:1":i1:s8: Montessori yönteminin ülkemizdeki amiral gemisi diyebiliriz		
4	Aynacılıkları kurumlarından bir tanesi. Uyguladıkları eğiti Eğitim Modeli	p	a:20(i0:a8(i0:s1:1":i1:s8: ruhsal açıdan kendilerini güvende hissedecekleri bir ortam imkanı sunmakta		
5	Aynacılıkları kurumlarından bir tanesi. Uyguladıkları eğiti Eğitim Modeli	p	a:15(i0:a8(i0:s1:1":i1:s8: Uyguladıkları eğitim modellerinin de geçerliliği tartışılmaz derecede sağlam		
6	Tom dünyanın örnek aldığı Kanada eğitim sisteminin Eğitim Sistemi	p	a:15(i0:a8(i0:s1:1":i1:s8: Aynacılıkları kurumlarından bir tanesi		
7	Bu kadar yeşil alana sahip başka bir okul kampüsü b Eğitim Fırsatları	p	a:14(i0:a8(i0:s1:1":i1:s8: Tom dünyanın örnek aldığı Kanada eğitim sisteminin yıkılmaz kalesi diyebiliriz burası için		
7	Bu kadar yeşil alana sahip başka bir okul kampüsü b Kampüs	p	a:27(i0:a8(i0:s1:1":i1:s8: Uluslararası eğitim fırsatlarını bu kadar geniş yelpazede sunan başka		
8	Anaokulundan Üniversiteye kadar tüm eğitimde sağ Eğitim	p	a:27(i0:a8(i0:s1:1":i1:s8: kadar yeşil alana sahip başka bir okul kampüsü bulamazsınız		
9	Kıstlı imkanlarla mucize yaratmayı bilen bir kadro, g Bina	n	a:13(i0:a8(i0:s1:1":i1:s8: Anaokulundan Üniversiteye kadar tüm eğitimde sağladıkları başarılar ile sağlam adımlarla ilerlen		
9	Kıstlı imkanlarla mucize yaratmayı bilen bir kadro, g Öğretmenler	p	a:31(i0:a8(i0:s1:1":i1:s8: Ancak binası biraz eski		
9	Kıstlı imkanlarla mucize yaratmayı bilen bir kadro, g Kadro	p	a:31(i0:a8(i0:s1:1":i1:s8: 11 imkânlarla öğrencileri de peşinden sürükleyen harika öğretmenler		
10	konuyu pekiştirmeyi ve kalıcı olmasını sağlayan bir s Sistem	p	a:14(i0:a8(i0:s1:1":i1:s8: 6 kalıcı olmasını sağlayan bir sistemi benimsemişler		
11	Okulun eğitimi iyi olabilir ama binası bana çok eski ; Bina	p	a:11(i0:a8(i0:s1:1":i1:s8: Okulun eğitimi iyi olabilir ama binası bana çok eski geldi		
11	Okulun eğitimi iyi olabilir ama binası bana çok eski ; Eğitim	p	a:11(i0:a8(i0:s1:1":i1:s8: Okulun eğitimi iyi olabilir ama binası bana çok eski		
12	Özellikle branş derslerindeki başarılarıyla gündeme g Okul	p	a:19(i0:a8(i0:s1:1":i1:s8: 11 gündeme gelen kendinden çok büyük okullara rekabet edebilen bir okul		
12	Özellikle branş derslerindeki başarılarıyla gündeme g Branş Dersleri	p	a:19(i0:a8(i0:s1:1":i1:s8: 11 Özellikle branş derslerindeki başarılarıyla gündeme gelen kendinden		
13	Çok iyi bir okul, çok memnunuz. Okul	p	a:8(i0:a8(i0:s1:1":i1:s8: 4 " Çok iyi bir okul çok memnunuz		
14	Öğretmenlerimizin çocuklarımız ile arkadaş olmaları Öğretmen	p	a:10(i0:a8(i0:s1:1":i1:s8: 11 Öğretmenlerimizin çocuklarımız ile arkadaş olmalarından dolayı		
15	Okulun koridorlarının geniş olması, sınıfların temiz v Çalışan	p	a:31(i0:a8(i0:s1:1":i1:s8: Ayrıca çalışanların ilgisinden ve her zaman güler yüzlü		
15	Okulun koridorlarının geniş olması, sınıfların temiz v Sınıf	p	a:31(i0:a8(i0:s1:1":i1:s8: Okulun koridorlarının geniş olması sınıfların temiz ve ferah olmasından dolayı çok		
15	Okulun koridorlarının geniş olması, sınıfların temiz v Okul	p	a:31(i0:a8(i0:s1:1":i1:s8: Okulun koridorlarının geniş olması sınıfların temiz		
16	Kurumsal bir okul. Montessori sistemi başarıyla uygı Ücret	n	a:13(i0:a8(i0:s1:1":i1:s8: Ücretler biraz yüksek		
16	Kurumsal bir okul. Montessori sistemi başarıyla uygı Sistem	p	a:13(i0:a8(i0:s1:1":i1:s8: Montessori sistemi başarıyla uygulanıyor		

Şekil 40: Hedef ve İlişkili Kelimelerin Veri Tabanına Aktarılmış Haline Ait Bir Görünüm

Şekil 40'ta hedef ve ilişkili kelimelerin veri tabanına aktarılmış yorum veri setine ait bir görünüm yer almaktadır. Bu veri seti 1925 satırdan oluşmaktadır.

Tablo 16: Hedef ve İlişkili Kelimelerin Yer Aldığı Veri Tabanının Veri Yapısı

Sütun Adı	Açıklaması
<i>Yorum_Id</i>	<i>Her bir yoruma verilen benzersiz numara</i>
<i>Yorum_Metin</i>	<i>İşlenmemiş yorum metni</i>
<i>Hedef</i>	<i>Yorum metni içinde geçen hedef</i>
<i>Duygu_Durumu</i>	<i>Hedefin duygu durumu</i>
<i>Yorum_Metin_Islenmis</i>	<i>İTÜ DDİYZ ile işlenmiş yorum metni</i>
<i>İlişkili_Kelimeler</i>	<i>Hedefi niteleyen kelimeler</i>

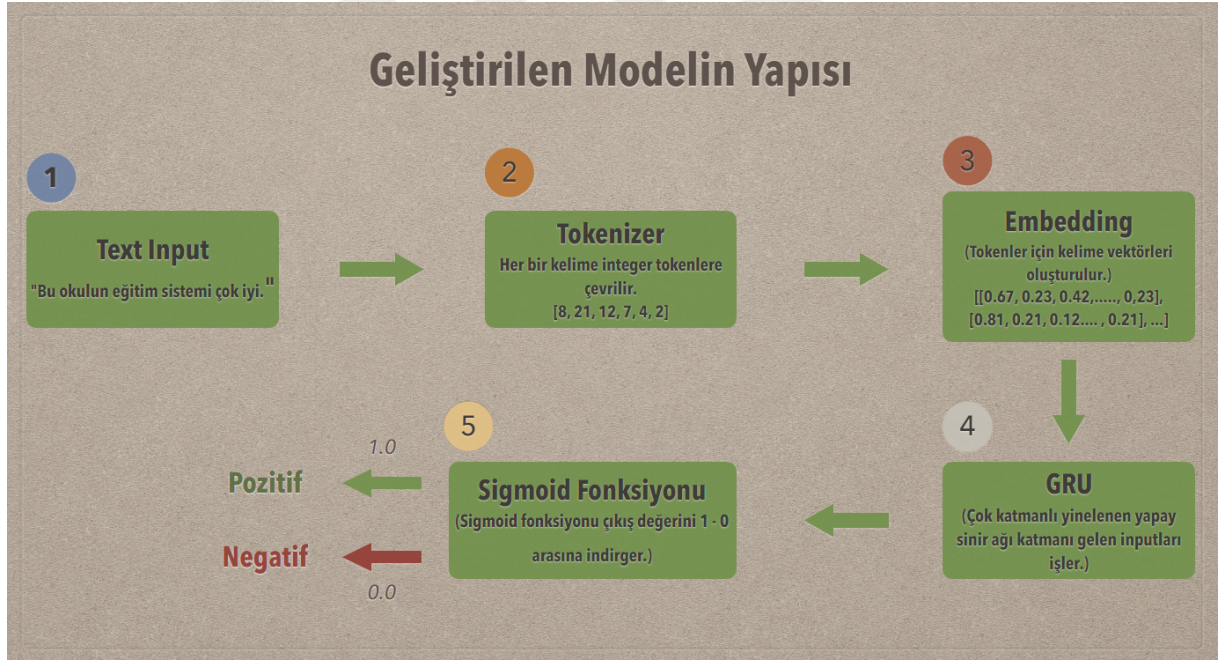
Tablo 16'da hedef ve ilişkili kelimeleri içeren veri tabanına ait veri yapısı gözükmemektedir. *Yorum_Id* alanı her bir yoruma verilen benzersiz numarayı içermektedir. *Yorum_Metin* alanı

işlenmemiş yorum metnini, *Yorum_Metin_Islenmis_alanı* İTÜ DDİYZ ile işlenmiş yorum metnine ait verileri içermektedir. *Hedef* alanı metin içerisinde yer alan hedefi içermektedir. *İliskili_Kelimeler* alanı ise hedefi niteleyen kelimeleri içermektedir.

3.6. MAKİNE ÖĞRENMESİ YAKLAŞIMI İÇİN MODELİN GELİŞTİRİLMESİ VE TEST EDİLMESİ

Çalışma kapsamında makine öğrenmesi yaklaşımı içinde yer alan çok katmanlı yinelenen yapay sinir ağları (Recurrent Neural Network-RNN) kullanılmıştır.

Modelin yapısı:



Şekil 41: Makine Öğrenmesi Yaklaşımı İçin Geliştirilen Modelin Yapısı

Şekil 41’de görüldüğü üzere çok katmanlı yinelenen yapay sinir ağları ile geliştirilen model beş aşamadan oluşmaktadır.

1. **Aşama:** Her bir hedef için tespit edilen ilişkili kelimeler modele input olarak dahil olur.

{bu okulun eğitim sistemi çok iyi}

Şekil 42: Modele İşlenmesi İçin Verilen Metin

Şekil 42’de modele işlenmesi için verilen ilişkili kelimelerin yer aldığı metin verisi yer almaktadır.

2. **Aşama (Tokenizer):** Her bir kelime tüm veri seti kullanılarak oluşturulan tokenizer dosyası kullanılarak tokenler, sayısal değere dönüştürülür.

[8, 21, 12, 7, 4, 2]

Şekil 43: Her Bir Kelimenin Sayısal Değere Dönüştürülmüş Hali

Şekil 43’te her bir kelimenin tokenizer dosyası kullanılarak sayısal değerlere dönüştürülmüş hali yer almaktadır.

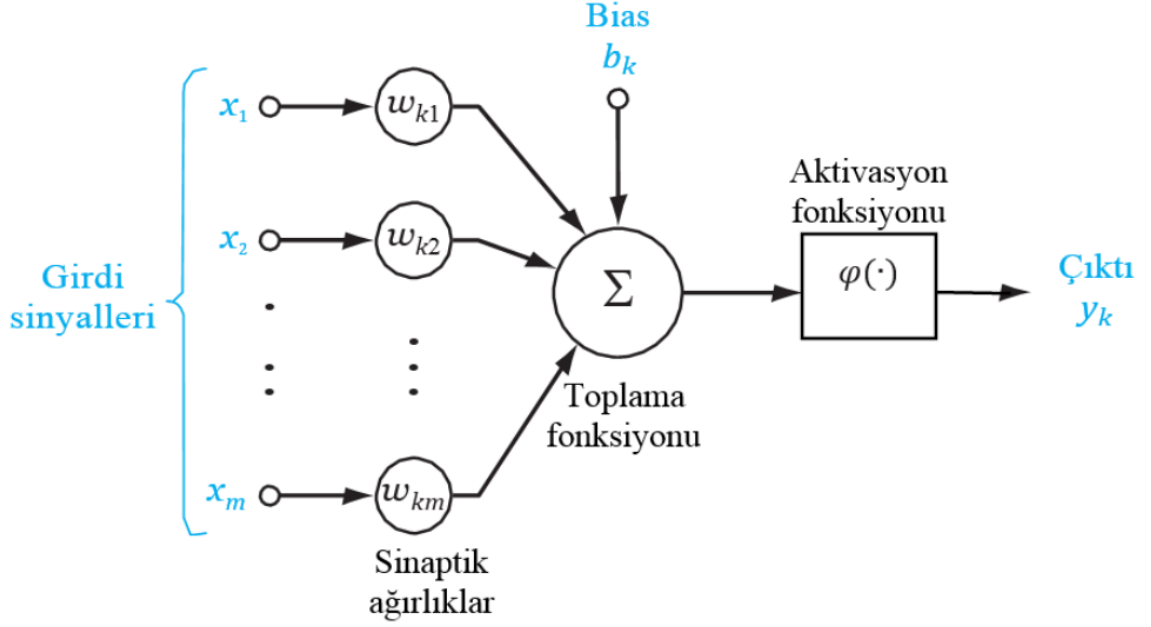
3. **Aşama (Embedding):** Tokenler için kelime vektörleri oluşturulur.

[[0.67, 0.23, 0.42,....., 0.23],
[0.81, 0.21, 0.12,....., 0.21], ...]

Şekil 44: Oluşturulan Kelime Vektörleri

Şekil 44’te tokenler için oluşturulan kelime vektörleri yer almaktadır. Kelime vektörleri kelimelerin sayısal temsillerini güçlendirmek için kullanılmaktadır.

4. **Aşama (RNN):** Çok katmanlı yinelenen yapay sinir ağı gelen inputları işler.



Şekil 45: Nöron'un Çalışma Şekli (Kartal ve Özen, 2017)

Şekil 45'te yapay sinir ağlarının temel taşı olan nöronların çalışma şekli yer almaktadır.

$$u_k = \sum_{j=1}^m w_{kj} x_j$$

Şekil 46: Nöron'un Çalışma Formülü (Haykin, 2008)

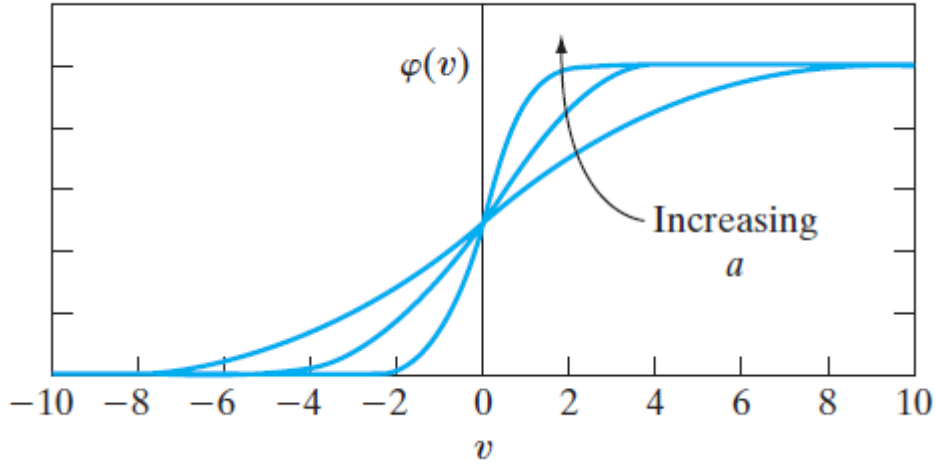
Yapay sinir ağlarının temel yapı taşı olan nöronlarda her bir x girdisi w ağırlık değeri ile çarpılır. Bu çarpımların toplamı alınır ve aktivasyon fonksiyonundan geçirilir. Bu değer çıktı bilgisidir.

5. **Aşama (Sigmoid Fonksiyonu):** Sigmoid fonksiyonu çıkış değerini 1-0 arasına sıkıştırır. Eğer değer bire yakınsa pozitif, sıfıra yakınsa negatif olarak sınıflanır.

$$y_k = \varphi(u_k + b_k)$$

Şekil 47: Aktivasyon Formülü (Haykin, 2008)

Şekil 47’de aktivasyon fonksiyonuna ait formül yer almaktadır.



Şekil 48: Sigmoid Aktivasyon Fonksiyonuna Ait Grafik (Haykin, 2008)

Şekil 48’de yer alan grafikte görüldüğü üzere Sigmoid fonksiyonu 0-1 arasında değer üreten doğrusal olmayan bir fonksiyondur.

Modelin Geliştirilme Süreci

Bu bölümde oluşturulan modelin uygulama aşaması anlatılacaktır. Uygulama için Python programlama dili ve Anaconda geliştirme ortamı kullanılmıştır.

Tokenizer Dosyası Oluşturma Süreci

Kelimelerin sayısal ifadelere dönüştürülebilmesi için mevcut veri seti kullanılarak ilgili domaini temsil edecek bir tokenizer dosyası oluşturulmuştur. Tokenizer dosyası oluşturmak için Python programlama dili ve Keras kütüphanesi kullanılmıştır.

```

tokenizer = Tokenizer(
    num_words=None,
    filters='!#$%&()*+,-./:;<=>?@[\]^_`{|}~\t\n',
    lower=True,
    split=" ",
    char_level=False,
    oov_token=None,
    document_count=0
)

tokenizer.fit_on_texts(data)

#tokenleri bir dosyaya kaydediyor.
import pickle
with open('tokenizer.pickle','wb') as handle:
    pickle.dump(tokenizer, handle, protocol=pickle.HIGHEST_PROTOCOL)

```

Şekil 49: Tokenizer Dosyası Oluşturmak İçin Geliştirilen Kod Parçası

Şekil 49’da Tokenizer dosyası oluşturmak için geliştirilen kod parçası görülmektedir. Kod parçasının üst bölümünde bir tokenizer sınıfı oluşturulmuş ve parametreleri tanımlanmıştır. Tokenizer sınıfı içinde “*num_words*”, “*filters*”, “*lower*”, “*split*”, “*char_level*”, “*split*”, “*char_level*”, “*oov_token*” ve “*document_count*” parametreleri bulunmaktadır. Kod parçasının 2. bölümünde oluşturulan parametreler dikkate alınarak ve veri seti kullanılarak tokenizer dosyası oluşturulmuştur. “*tokenizer.pickle*” ismi ile kaydedilmiştir

tokenizer.json

```

+ seçilmiş": 965, "alıp": 966, "buraya": 967, "dönemde": 968, "binasıyla": 969, "anlayışıyla": 970, "imkanlarıyla": 971, "sayılmaz": 972, "olmasıyla": 973, "kulübü": 974, "elit": 975, "ar
+ 94, "görevlileri": 995, "bilgiler": 996, "gidin": 997, "nasıl": 998, "görün": 999, "insanlardan": 1000, "ona": 1001, "alınmış": 1002, "yanındaki": 1003, "tabi": 1004, "çeşitliliği": 100
+ şarısının": 1024, "sebebi": 1025, "yoktur": 1026, "tekrar": 1027, "2": 1028, "uzaklar": 1029, "çeşitlilik": 1030, "ettiği": 1031, "ağırılık": 1032, "işine": 1033, "almak": 1034, "sunulan
+ eve": 1052, "değer": 1053, "hiçbir": 1054, "misyon": 1055, "yönetimdeki": 1056, "sonra": 1057, "diye": 1058, "hassasiyetle": 1059, "yaratılması": 1060, "rehber": 1061, "gerektiği": 1062
+ ri": 1079, "hesaplarına": 1080, "planlı": 1081, "imkankarlarına": 1082, "ekip": 1083, "ayrı": 1084, "tecrübe": 1085, "sıcacık": 1086, "profilinden": 1087, "diyemem": 1088, "imkanları":
+ iyetlerinin": 1105, "niteliğine": 1106, "bakınca": 1107, "yıllardır": 1108, "çizgisini": 1109, "ilim": 1110, "bl": 1111, "kullanmaları": 1112, "içli": 1113, "olmadığını": 1114, "burdan":
+ 1131, "hassas": 1132, "tavri": 1133, "noktaya": 1134, "gelmiş": 1135, "aralığı": 1136, "farketmeksizin": 1137, "tutulmuş": 1138, "deneyin": 1139, "yapılabildiği": 1140, "başarabilmiş":
+ yon": 1157, "yönetiyor": 1158, "sacma": 1159, "ciddi": 1160, "geliştirmesi": 1161, "oturmamış": 1162, "tavsiye": 1163, "etmem": 1164, "gecirebilmeyi": 1165, "atölyeleri": 1166, "kullan
+ n": 1183, "vermek": 1184, "amacıyla": 1185, "kadrosuna": 1186, "tartışmaya": 1187, "hocalarından": 1188, "anlayışından": 1189, "çeşitliliğinin": 1190, "sıkmadan": 1191, "çocuklara": 119
+ 1209, "okulunda": 1210, "asla": 1211, "sanslıyız": 1212, "ücret": 1213, "eğitimcilere": 1214, "yerde": 1215, "ihtiyac": 1216, "varlığı": 1217, "gelişimini": 1218, "yöneticisinden": 121
+ 1235, "sistemine": 1236, "yemekleri": 1237, "dili": 1238, "hizmetten": 1239, "gerçekleştirebilecekleri": 1240, "faydasını": 1241, "şehirin": 1242, "orta": 1243, "birde": 1244, "boşlukt
+ yabilecekleri": 1262, "parlayan": 1263, "yıldızı": 1264, "üstü": 1265, "düzenlenmiş": 1266, "içerikleriyle": 1267, "düşündüğüm": 1268, "bekleneni": 1269, "okudukları": 1270, "faydalandı
+ ın": 1288, "becerileri": 1289, "alsın": 1290, "diyorsanız": 1291, "kaldı": 1292, "altıyapıya": 1293, "sağlıklarını": 1294, "önünde": 1295, "bulundurulur": 1296, "yönetiminin": 1297, "am
+ larını": 1314, "yelpazede": 1315, "bulamazsanız": 1316, "anaokulundan": 1317, "üniversiteye": 1318, "sağladıkları": 1319, "adınlarla": 1320, "ilerlenekte": 1321, "istegiyle": 1322, "pes
+ 1338, "olmasından": 1339, "ücretler": 1340, "başarıyla": 1341, "uygulanıyor": 1342, "fiyatıda": 1343, "koleje": 1344, "düzeyi": 1345, "anında": 1346, "yüzölçüğü": 1347, "içerikli": 1348
+ in": 1367, "takipçileri": 1368, "bahçede": 1369, "ilkokul": 1370, "değiller": 1371, "kişiliği": 1372, "çözme": 1373, "kurucularının": 1374, "sirketini": 1375, "sevmiyorum": 1376, "sevd
+ 1393, "kapasitesinin": 1394, "cocukla": 1395, "çözüm": 1396, "verdiğiniz": 1397, "tuzla": 1398, "arayışı": 1399, "babalara": 1400, "ilgili": 1401, "şisli": 1402, "tarafındakiler": 1403, "sağd
+ 18, "özelliklerini": 1419, "taşınan": 1420, "olduğundan": 1421, "gözüm": 1422, "ifade": 1423, "edebilmeye": 1424, "onu": 1425, "bilmem": 1426, "benden": 1427, "yıldız": 1428, "alın": 14
+ 1445, "ülkemiz": 1446, "kurumların": 1447, "yetistirdiğini": 1448, "bakmamızı": 1449, "etmeden": 1450, "etrafımızdan": 1451, "deneyimleyerek": 1452, "öğrenmelerini": 1453, "programının"
+ ": 1470, "fazlaca": 1471, "yoğun": 1472, "bende": 1473, "bölümü": 1474, "soforleri": 1475, "eve": 1476, "yaklaştığında": 1477, "arayacak": 1478, "telefonuna": 1479, "mesaj": 1480, "kodl
+ : 1498, "çalıştıran": 1499, "beklerseniz": 1500, "hepsini": 1501, "karşılayabilecek": 1502, "kapasitede": 1503, "içerisi": 1504, "mı": 1505, "kuvvetli": 1506, "gereğinden": 1507, "gelis
+ 1527, "koşullarda": 1580, "kütüphaneleri": 1581, "bellirlenen": 1582, "zamanlarda": 1583, "işlerinde": 1584, "özgüvenleri": 1585, "edilebilir": 1586, "tecrübeyi": 1587, "peşinde": 1
+ 1587, "yandan": 1606, "alleyiz": 1607, "çizgide": 1608, "ilerliyorlar": 1609, "çağında": 1610, "cocuğu": 1611, "bakın": 1612, "sağlığını": 1613, "olumsuz": 1614, "yönde": 161
+ teblen": 1632, "gayesi": 1633, "içerinde": 1634, "çalışmalarını": 1635, "beğendiğimiz": 1636, "kulvarda": 1637, "değerlendirilebilecek": 1638, "dışa": 1639, "kaydımızı": 1640, "yaptırd
+ "ınanıyorum": 1657, "düzeyinde": 1658, "genişliğe": 1659, "birleştirerek": 1660, "kurdıkları": 1661, "vermiştim": 1662, "sindi": 1663, "yavrum": 1664, "sevdiriyorlar": 1665, "yarım": 16
+ ": 1684, "tereddütsüz": 1685, "edeceğim": 1686, "cevremizde": 1687, "cocuğunu": 1688, "yollayanlardan": 1689, "yorumlar": 1690, "almıştık": 1691, "pişman": 1692, "olmadık": 1693, "defa
+ ca": 1711, "alakayı": 1712, "hocalardan": 1713, "görüyorsunuz": 1714, "tuzla'da": 1715, "arayanların": 1716, "mutlaka": 1717, "uğraması": 1718, "cocuğunuzla": 1719, "günden": 1720, "sev
+ ilyon": 1737, "hazırlamak": 1738, "gelmesi": 1739, "uyguluyorlar": 1740, "dikkat": 1741, "olmaz": 1742, "gururlandırdı": 1743, "konuşmuşuzda": 1744, "hoşnut": 1745, "bişeyse": 1746, "b

```

C:\Users\user\Desktop\tokenizer.json 131754

CR LF UTF-8 JSON GitHub Git (0)

Şekil 50: Oluşturulan Tokenizer Dosyası

Şekil 50’de yorum veri seti kullanılarak oluşturulan tokenizer dosyası görülmektedir. Tokenizer dosyası ne kadar kapsamlı olursa ilgili domaini o kadar iyi temsil eder. Bu temsiliyet durumu modelin başarımlarını etkiler.

Modelin Oluşturulmasında Kullanılan Kütüphaneler

Modelin oluşturulmasında numpy, pandas, json ve tensorflow keras kütüphaneleri kullanılmıştır.

```
import numpy as np
import pandas as pd
import json

from tensorflow.python.keras.models import Sequential
from tensorflow.python.keras.layers import Dense, GRU, Embedding, CuDNNGRU
from tensorflow.keras.optimizers import Adam
from sklearn.utils import shuffle
```

Şekil 51: Modelin Oluşturulmasında Kullanılan Kütüphanelere Ait Kod Parçası

Şekil 51’de modelin oluşturulması için kullanılan kütüphanelere ait kod parçası yer almaktadır. Model oluşturma sürecinde yukarıda belirtilen kütüphanelere ait Sequential, Dense, GRU, Embedding, Adam ve Sklearn kütüphaneleri de kullanılmıştır.

Veri Setinin Eğitim ve Test İçin Bölünmesi

Veri seti modelin eğitilmesi ve sonrasında test edilebilmesi amacıyla ikiye bölünmüştür.

```
cutoff = int(len(data) * 0.70)
x_train, x_test = data[:cutoff], data[cutoff:]
y_train, y_test = target[:cutoff], target[cutoff:]
```

Şekil 52: Veri Setini Eğitim ve Test için Bölen Kod Parçası

Şekil 52’de veri setinin eğitim ve test için bölen kod parçası görülmektedir. Veri setinin %70’i eğitim için %30’u ise test için kullanılmıştır.

Her Bir Yorumun Tokenizer Dosyası Kullanılarak Tokenleştirilmesi

Bu aşamada her bir yorumu kelime bazında tokenleştirilmiştir. Tokenleştirilerek oluşturulacak dizi içerisinde aynı sayıda sayı ile temsil edilebilmesi için yorumların başına 0 koyarak 20 kelimeye tamamlanmıştır.

```
with open('tokenizer.json') as json_dosyasi:
    json_tokenizer = json.load(json_dosyasi)

def tokenlestir(yorumListesi):
    y_yorumlar = []
    for yorum in yorumListesi:
        y_yorum = []
        for kelime in yorum.split():
            if (len(y_yorum) < 20 and kelime in json_tokenizer):
                y_yorum.append(json_tokenizer[kelime])

        if (len(y_yorum) < 20):
            sifirlar = list(np.zeros(20 - len(y_yorum), dtype = int))
            y_yorum = sifirlar + y_yorum

        y_yorumlar.append(y_yorum)
    return np.array(y_yorumlar, dtype = np.dtype(np.int32))
```

Şekil 53: Her Bir Yorum Metninin Tokenleştirilmesini Sağlayan Kod Parçası

Şekil 53'te her bir yorum metninin kelime bazında tokenleştirilmesini sağlayan kod parçası yer almaktadır. Eğitim ve test kümeleri bu kod bloğu yardımı ile tokenleştirilmiştir.

Tokenleştirilmiş Bir Yorum Metni

Tokenleştirilen her bir yorum içerisindeki kelimeler sayısal ifadelere dönüşür. Sayısal ifadeler bir dizi haline getirilir.

```
test_kume[10]
array([ 0,  0,  0,  0,  0,  0,  0,  0,  0,  0,  0,  0,  1,
        61, 26, 693, 63, 23, 25, 4])
```

Şekil 54: Tokenleştirilmiş Bir Yorum Metnine Ait Görünüm

Şekil 54'te tokenleştirilmiş ve bir dizi haline getirilmiş bir yorum metnine ait görünüm yer almaktadır. Dizilerin aynı boyutta temsili için başlarına eksik olduğu sayı kadar 0 eklenmiştir.

Modelin Oluşturulması

Modelin oluşturulmasında üç katmanlı yinelenen yapay sinir ağı kullanılmıştır. Model üç ana bölümden oluşmaktadır:

1. Embedding Katmanı (Kelime Vektörleri)
2. GRU Katmanı (Çok katmanlı Yinelenen Yapay Sinir Ağı)
3. Sigmoid Aktivasyon Fonksiyonu Katmanı

```
model = Sequential()
embedding_size = 20

model.add(Embedding(input_dim=2000,
                    output_dim=embedding_size,
                    input_length=20,
                    name='embedding_layer'))

model.add(GRU(units=12, return_sequences=True))
model.add(GRU(units=6, return_sequences=True))
model.add(GRU(units=3))
model.add(Dense(1, activation='sigmoid'))

optimizer = Adam(lr=1e-3)

model.compile(loss='binary_crossentropy',
              optimizer=optimizer,
              metrics=['accuracy'])

model.summary()
```

Şekil 55: Embeddig ve RNN Yapısını Oluşturan Kod Parçası

Şekil 55'te Embedding ve RNN yapısını oluşturan kod parçası görülmektedir. RNN 3 katmanlı yapıdadır. İlk katman 16, ikinci katman sekiz, üçüncü katman dört nörondan oluşmaktadır. RNN'den çıkan değer sigmoid aktivasyon fonksiyonundan işlenerek 1-0 aralığına indirgenir. Değer sıfıra yakınsa negatif, bire yakınsa pozitif olarak sınıflandırılır.

Modelin Eğitilmesi

Model, eğitim sayısı (epoch) 5 ve *batch_size* = 10 olacak şekilde eğitim veri seti kullanılarak eğitilmiştir.

```

model.fit(egitim_kume, y_train, epochs = 5, batch_size = 10)

Epoch 1/5
135/135 [=====] - 3s 21ms/step - loss: 0.3022 - accuracy: 0.9354
Epoch 2/5
135/135 [=====] - 3s 23ms/step - loss: 0.2296 - accuracy: 0.9391
Epoch 3/5
135/135 [=====] - 3s 24ms/step - loss: 0.2272 - accuracy: 0.9391
Epoch 4/5
135/135 [=====] - 3s 23ms/step - loss: 0.1957 - accuracy: 0.9413
Epoch 5/5
135/135 [=====] - 3s 23ms/step - loss: 0.1345 - accuracy: 0.9562
<tensorflow.python.keras.callbacks.History at 0x1ca50501a88>

```

Şekil 56: Modelin Eğitilmesini Sağlayan Kod Parçası ve Sonuca Ait Görünüm

Şekil 56’da modelin eğitilmesini sağlayan kod parçası ve sonuca ait görünüm yer almaktadır. Şekilde de görüldüğü üzere eğitim sayısı (epoch) 1/5’ de *loss* değeri 0,3022 ve *accuracy* (doğruluk oranı) değeri 0,9354 olarak tespit edilmiştir. Eğitim sayısı (epoch) 5/5’de ise *loss* değeri 0,1345 ve *accuracy* değeri 0,9562 olarak tespit edilmiştir. Değerlerden de görüldüğü üzere eğitim sayısı arttıkça kayıp (loss) değeri azalmış doğruluk oranı (accuracy) değeri artmıştır. Buda modelin tasarımsal olarak doğru olduğunu gösterir. Şayet geliştirilen modelde tasarımsal bir sorun olsaydı eğitim sayısı arttıkça istikrarsız bir kayıp (loss) ve doğruluk oranı (accuracy) değerleri tespit edilebilirdi.

Modelin Test Edilmesi

Bu aşamada toplam veri setinin %30’luk kısmı ile modelin başarı oranı test edilmiştir. %30’luk bu kısım modelin daha önce hiç karşılaşmadığı verilerden oluşmaktadır.

```

model.evaluate(test_kume, y_test)

19/19 [=====] - 0s 4ms/step - loss: 0.2137 - accuracy: 0.9325
[0.2137160748243332, 0.9325259327888489]

```

Şekil 57: Modelin Test Edilmesini Sağlayan Kod Parçası ve Sonuca Ait Görünüm

Şekil 57’de modelin test edilmesini sağlayan kod parçası ve modelin başarı sonuca ait görünüm yer almaktadır. Model ilk kez karşılaştığı verileri sınıflandırmada 0,9325 *accuracy* değeri ve 0,2137 *loss* değeri elde etmiştir.

3.7. SÖZLÜK TABANLI YAKLAŞIM İÇİN MODELİN GELİŞTİRİLMESİ VE TEST EDİLMESİ

Sözlük tabanlı yaklaşım için üç aşamadan oluşan bir yol izlenmiştir:

1. Yorum veri setinin işlenmeye hazır hale getirilmesi,
2. Modelin geliştirilmesi,
3. Farklı sözlükler ile test edilmesi ve başarı oranlarının tespit edilmesi.

1. Yorum Veri Setinin İşlenmeye Hazır Hale Getirilmesi

Makine öğrenmesi modeli için kullanılan ilişkili kelimeler, İTÜ DDIYZ aracı kullanılarak eklerinden kurtarılmış, ilişkili kelimelerin eksiz halleri veri tabanına aktarılmıştır.

yorum_id	yorum_metin	hedef	duygu_durumu	yorum_metin_islenmis	iliskili_kelimeler	eksiz_iliskili_kelimeler	eksiz_iliskili_kelimeler_en
1	Okulumuz, harika ! Olanak	p	a:19:(0;a:8:(0;s:1:"1";i:s:8: eğitim vermesinin y			edim ver yan çocuk her ti education,give,side,child,each,variety,possibility,provide	
1	Okulumuz, harika ! Eğitim	p	a:19:(0;a:8:(0;s:1:"1";i:s:8: Okulumuz harika b			okul harika bir eğitim ver y school,great,one,education,give,side,child,each,variety,possibility	
1	Okulumuz, harika ! Okul	p	a:19:(0;a:8:(0;s:1:"1";i:s:8: Okulumuz harika b			okul harika bir eğitim ver y school,great,one,education,give,side,very,successful,one,school	
2	Öğrencilerin sağlık Eğitim Mat	p	a:16:(0;a:8:(0;s:1:"1";i:s:11: Öğrencilerin sağl			öğrenci sağlık göz ön bulu student,health,eye,front,find,education,materiel,use,be,very,big,one,advantage	
3	Montessori yöntemi Okul	p	a:20:(0;a:8:(0;s:1:"1";i:s:11: Çocuk odaklı ve klas			ve klasik sistem tamamen and, classical,system,completely,tooth,alternative,one,school	
3	Montessori yöntemi Sistem	p	a:20:(0;a:8:(0;s:1:"1";i:s:11: Çocuk odaklı ve klas			çocuk odak ve klasik sisten child, focus, and, classical, system, completely, tooth, alternative, one, school	
3	Montessori yöntemi Yöntem	p	a:20:(0;a:8:(0;s:1:"1";i:s:11: Montessori yöntem			Montessori yöntem ülke ar Montessori,method,country,admiral,ship,also	
4	Çocuklarımızın fizik Ortam	p	a:20:(0;a:8:(0;s:1:"1";i:s:11: ruhsal açıdan kendil			ruhsal açı kendi güven hiss spiritual,pain,own,confidence,feel,one,environment,facility,sun	
5	Aynca kurumu Eğitim Mo	p	a:15:(0;a:8:(0;s:1:"1";i:s:11: Uyguladıkları eğitim			uygul eğitim model de ge apply,education,model,also,validity,Discuss,degree,sturdy	
5	Aynca kurumu Kurum	p	a:15:(0;a:8:(0;s:1:"1";i:s:11: Aynca kurumu			aynca kurumu bir tane privileged,institution,one,piece	
6	Tüm dünyanın öm Eğitim Sıst	p	a:14:(0;a:8:(0;s:1:"1";i:s:4: Tüm dünyanın örne			tüm dünya örnek al Kanad all,world,sample,get,Canada,education,system,break down,Castle,also,here,for	
7	Bu kadar yeşil alan Eğitim Fırsı	p	a:27:(0;a:8:(0;s:1:"1";i:s:2: Uluslararası eğitim f			ulus eğitim fırsat bu kadar nation,education,opportunity,this,until, Large,fan,sun,another	
7	Bu kadar yeşil alan Kampüs	p	a:27:(0;a:8:(0;s:1:"1";i:s:2: kadar yeşil alana sah			kadar yeşil alan sahip başk until,green,domain,owner,another,one,school,campus,find	
8	Anaokulundan Üni Eğitim	p	a:13:(0;a:8:(0;s:1:"1";i:s:1: Anaokulundan Üniv			anaokul üniversite kadar ti pre school,university,until,all,education,provide,succes,with,sturdy,step,go ahead	
9	Kısıtlı imkanlarla m Bina	n	a:31:(0;a:8:(0;s:1:"1";i:s:11: Ancak binası biraz e			ancak bina biraz eski but,building,some,old	
9	Kısıtlı imkanlarla m Öğretmen	p	a:31:(0;a:8:(0;s:1:"1";i:s:11: isteğiyle öğrenciler			istek öğrenci de peş sürükl claim,student,also,advance,drag,great,teacher	
9	Kısıtlı imkanlarla m Kadro	p	a:31:(0;a:8:(0;s:1:"1";i:s:11: imkanlarla mucize y			imkan mucize yarat bir bir facility,miracle,create,know,one,staff,smiling face,and,continuous,new,thing	
10	konuyu pekiştirme Sistem	p	a:14:(0;a:8:(0;s:1:"1";i:s:6: kalıcı olmasını sağla			kal ol sağla bir sistem beni remain,be,provide,one,system,if mine	
11	Okulun eğitimi iyi i Bina	n	a:11:(0;a:8:(0;s:1:"1";i:s:6: Okulun eğitimi iyi i			okul eğitim iyi ol ama bina school,education,good,be,but,building,l,very,old,come	
11	Okulun eğitimi iyi i Eğitim	p	a:11:(0;a:8:(0;s:1:"1";i:s:6: Okulun eğitimi iyi i			okul eğitim iyi ol ama bina school,education,good,be,but,building,l,very,old	
12	Özellikle branş deri Okul	p	a:19:(0;a:8:(0;s:1:"1";i:s:11: gündeme gelen ken			gündem gel kendi çok büyük agenda,come,own,very,big,school,competition,meat,one,school	
12	Özellikle branş deri Branş Dersl	p	a:19:(0;a:8:(0;s:1:"1";i:s:11: Özellikle branş dersl			özellikle branş ders başan especially,branch,lesson,succes,agenda,come,own	
13	Çok iyi bir okul, ço Okul	p	a:8:(0;a:8:(0;s:1:"1";i:s:4: Çok iyi bir okul çok			çok iyi bir okul çok memni very,good,one,school,very,satisfied	
14	Öğretmenlerimizin Öğretmen	p	a:10:(0;a:8:(0;s:1:"1";i:s:11: Öğretmenlerimizin			öğretmen çocuk ile arkada teacher,child,with,friend,be,because	
15	Okulun koridorları Çalışan	p	a:31:(0;a:8:(0;s:1:"1";i:s:6: Ayrıca çalışanların			aynca çalışan ilgi ve her zai also,working,interest,and,each,time,smiling face	
15	Okulun koridorları Sınıf	p	a:31:(0;a:8:(0;s:1:"1";i:s:6: Okulun koridorların			okul koridor geniş ol sınıf t school,corridor, Large,be,class, clean, and, spacious, be, because, very	
15	Okulun koridorları Okul	p	a:31:(0;a:8:(0;s:1:"1";i:s:6: Okulun koridorların			okul koridor geniş ol sınıf t school,corridor, Large,be,class, clean	

Şekil 58: Eksiz İlişkili Kelimelerin Veri tabanına Aktarılması

Şekil 58’de sözlük tabanlı yaklaşım sürecinde kullanılmak üzere ilişkili kelimelerin eksiz halleri her bir hedef ile ilişkili bir formatta veri tabanına kaydedilmiştir. İTÜ DDIYZ aracı ile bir yorumu işlediğinizde her bir kelimenin eksiz halini oluşturabilmektedir. Servisin bu özelliği kullanılarak eksiz ilişkili kelimeler elde edilmiştir.

Tablo 17: Eksiz İlişkili Kelimelerin Yer Aldığı Veri tabanının Veri Yapısı

<i>Sütun Adı</i>	<i>Açıklaması</i>
<i>Yorum_Id</i>	<i>Her bir yoruma verilen benzersiz numara</i>
<i>Yorum_Metin</i>	<i>İşlenmemiş yorum metni</i>
<i>Hedef</i>	<i>Yorum metni içinde geçen hedef</i>
<i>Duygu_Durumu</i>	<i>Hedefin duygu durumu</i>
<i>Yorum_Metin_İslenmis</i>	<i>İTÜ DDİYZ ile işlenmiş yorum metni</i>
<i>İliskili_Kelimeler</i>	<i>Hedefi niteleyen kelimeler</i>
<i>Eksiz_Iliskili_Kelimeler</i>	<i>Hedefi niteleyen kelimelerin kök hali</i>
<i>Eksiz_Iliskili_Kelimeler_En</i>	<i>Hedefi niteleyen kelimelerin kök hali (İngilizce)</i>

Tablo 17’de eksiz ilişkili kelimelerin yer aldığı veri tabanının veri yapısı yer almaktadır. *Yorum_Id* alanı her bir yoruma verilen benzersiz numarayı içermektedir. *Yorum_Metin* alanı işlenmemiş yorum metnini, *Yorum_Metin_Islenmis alanı* İTÜ DDİYZ ile işlenmiş yorum metnine ait verileri içermektedir. *Hedef* alanı metin içerisinde yer alan hedefi içermektedir. *İliskili_Kelimeler* alanı ise hedefi niteleyen kelimeleri içermektedir. *Eksiz_Iliskili_Kelimeler* alanında eklerinden arındırılmış ilişkili kelimelerin yer almaktadır. *Eksiz_Iliskili_Kelimeler_En* alanında ise bu verinin İngilizceye çevrilmiş hali yer almaktadır.

2. Modelin Yapısı

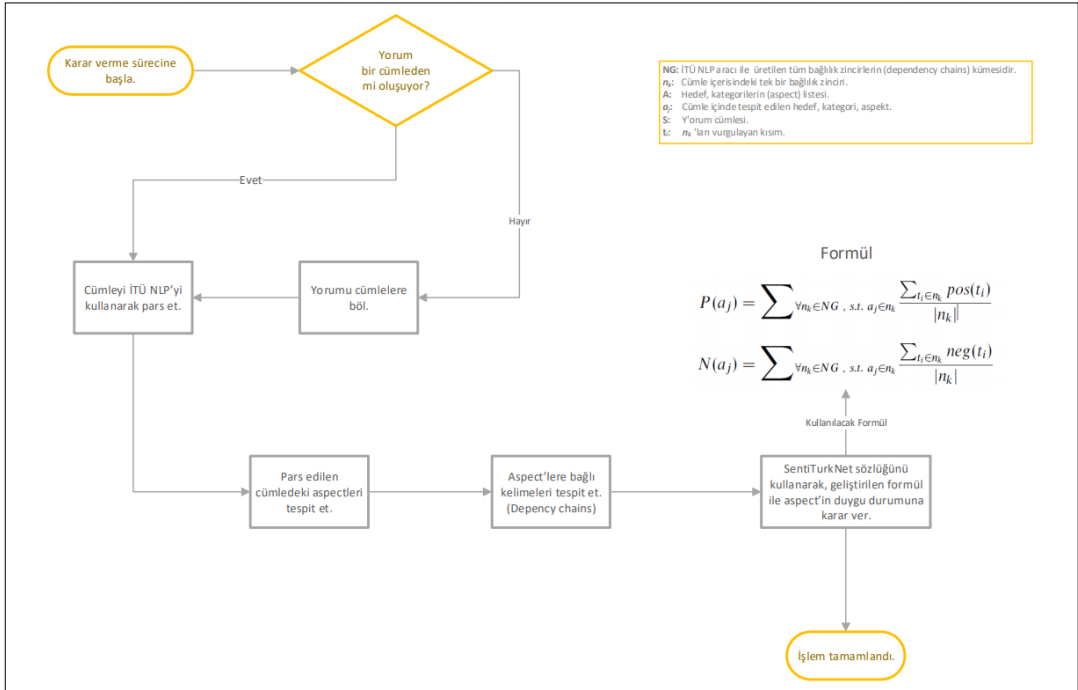
Modelin hazırlanmasında beş adımdan oluşan bir yapı oluşturulmuştur. Bu adımlar sırayla:

1. Yorum cümlesi işleme alınır.
2. Yorum cümlesi içindeki hedefler tespit edilir.
3. Hedeflere bağlı kelime (eksiz) zincirleri oluşturulur.
4. Her bir hedefe ait kelime zincirleri puanlama için sözlüğe gönderilir.
5. Hedefin duygu durumuna karar verilir.



Şekil 59: Sözlük Tabanlı Yaklaşım İçin Geliştirilen Model

Şekil 59’da sözlük tabanlı yaklaşım için geliştirilen model gözükmemektedir. Model beş aşamadan oluşmaktadır. Modelin dördüncü aşamasında yer alan her sözlük süreci için SentiWordNet ve SeniTurknet olmak üzere iki farklı sözlük kullanılmıştır.



Şekil 60: Sözlük Tabanlı Yaklaşım İçin Geliştirilen Modelin Detaylı Hali

Şekil 60'ta sözlük tabanlı yaklaşım için geliştirilen modelin detayı hali görülmektedir.

$P(a_i) = \frac{\sum_{k_j \in t_y} pos(k_j)}{ t_y }$ $N(a_i) = \frac{\sum_{k_j \in t_y} neg(k_j)}{ t_y }$	$P(a_i) > N(a_i) \Rightarrow \text{Positive}$ $P(a_i) < N(a_i) \Rightarrow \text{Negative}$
a_i : Her bir hedef k_j : ilişkili her bir kelime t_y : Hedef ile ilişkili tüm kelimeler $ t_y $: İlişkili kelimelerin sayısı	

Şekil 61: Duygu Durumuna Karar Veren Fonksiyon (Dehkharghani ve diğ., 2017)

Şekil 61'de duygu durumuna karar veren fonksiyon yer almaktadır. Her bir hedefin pozitif puanının tespit edilebilmesi için:

1. İlişkili olan her bir kelime sözlüğe gönderilir.
2. Sözlükten aldığı pozitif puanlar toplanır.
3. Toplam puan ilişkili kelime sayısına bölünür.
4. Ortalama pozitif puanı tespit edilir.

Bir hedefin negatif puanının tespit edilebilmesi için:

1. İlişkili olan her bir kelime sözlüğe gönderilir.
2. Sözlükten aldığı negatif puanlar toplanır.
3. Toplam puan ilişkili kelime sayısına bölünür.
4. Ortalama negatif puanı tespit edilir.

Ortalama pozitif puan, ortalama negatif puandan yüksek ise hedefin duygu durumu pozitif
Ortalama negatif puan, ortalama pozitif puandan yüksek ise hedefin duygu durumu negatif kabul edilir.

```

1  <?php
2
3  //Veritabanı bağlantısı yapılıyor.
4  ini_set('MAX_EXECUTION_TIME', '-1');
5  $db = new PDO("mysql:host=localhost;dbname=phd4;charset=utf8", "root", "");
6
7  //Kullanılacak değişkenler tanımlanıyor.
8  $sayac = 0;
9
10 //İlgili tablodan tek tek veriler çekilerek puanlama süreci başlıyor.
11 $query = $db->query("SELECT agay.agay_id, agay.eksiz_iliskili_kelimeler, agay.durum FROM aspete_gore_ayrilmis_yorumlar agay", PDO::FETCH_ASSOC);
12 if ( $query->rowCount() ) {
13     foreach( $query as $row ){
14
15         //Eksiz ilişkili kelimeler veritabanından çekilip bir array haline getiriliyor.
16         $agay_id = $row['agay_id'];
17         $eksiziliskilikelimeler = explode (" ",trim($row['eksiz_iliskili_kelimeler']));
18
19         //Puanlama
20         $pozitif = 0;
21         $negatif = 0;
22         for ($i=0; $i < count($eksiziliskilikelimeler) ; $i++) {
23             $query = $db->query("SELECT st.synonyms, AVG(st.neg_value), AVG(st.pos_value) FROM sentiturnet st where st.synonyms = '$eksiziliskilikelimeler[$i]'");
24             if ($query) {
25                 $pozitif = $pozitif + $query['AVG(st.pos_value)'];
26                 $negatif = $negatif+ $query['AVG(st.neg_value)'];
27             }
28         }
29
30         $pozitif = $pozitif/count($eksiziliskilikelimeler);
31         $negatif = $negatif/count($eksiziliskilikelimeler);
32
33         //Puana bakarak duygu durumuna karar veriliyor ve veritabanına kaydediliyor.
34         if ($pozitif>$negatif) {
35             $durum = 'p';
36         }else if ($pozitif<$negatif) {
37             $durum = 'n';
38         }else {
39             $durum = 'o';
40         }
41
42         //Eksiz ilişkili kelimeleri Veritabanına yazıyor
43         $query = $db->prepare("UPDATE aspete_gore_ayrilmis_yorumlar SET durum2 = :durum2 WHERE agay_id = :agay_id");
44         $update = $query->execute(array("durum2" => $durum,"agay_id" => $agay_id));
45
46         if (trim($row['durum']) == $durum) {
47             $sayac++;
48         }
49     }
50 }
51

```

Şekil 62: Sözlük Tabanlı Yaklaşım Modelinin Test Edilmesi için Geliştirilen Kod Bloğu

Şekil 62’de sözlük tabanlı yaklaşım için geliştirilen kod bloğu görülmektedir. Bu kod bloğu geliştirilen modelin 1925 hedef için çalıştırılıp test edilebilmesi için PHP programlama dili kullanılarak geliştirilmiştir. Kod bloğu şu şekilde çalışmaktadır:

1. Verilerimizin yer aldığı veri tabanına bağlanır.
2. Her bir hedef için ilgili ilişkili kelimeleri veri tabanından çeker ve bunu bir diziye atar.
3. Dizinin her bir elemanı için (her bir ilişkili kelime) duygu sözlüğündeki pozitif ve negatif puanlarını tespit eder.
4. Toplam negatif ve pozitif puanların ortalaması alınır.
5. Şekil 61’de yer alan fonksiyon doğrultusunda duygu durumuna karar verir.

• BULGULAR

4.1. MAKİNE ÖĞRENMESİ YAKLAŞIMI İÇİN GELİŞTİRİLEN MODELE AİT BULGULAR

Geliştirilen model farklı katman, nöron ve eğitim sayısı (*epoch*) değerleri ile test edilmiştir.

Tablo 18: Modelin Testi için Kullanılan Farklı Katman, Nöron ve Eğitim Sayısı Değerleri

Katman	Nöron Dizilimleri	Eğitim Sayıları (Epoch)
2	2-4	5, 10, 15
	3-6	
	4-8	
	5-10	
	6-12	
3	2-4-8	
	3-6-12	
	4-8-16	
	5-10-20	
	6-12-24	
4	2-4-8-16	
	3-6-12-24	
	4-8-16-32	
	5-10-20-40	
	6-12-24-48	
5	2-4-8-16-32	
	3-6-12-24-48	
	4-8-16-32-64	
	5-10-20-40-80	
	6-12-24-48-96	

İki katmanlı yapıda 2-4, 3-6, 4-8, 5-10, 6-12 nöron dizilimleri ve 5, 10 ve 15 eğitim sayı (*epoch*) değerleri ile test edilmiştir. Üç katmanlı yapıda 2-4-8, 3-6-12, 4-8-16, 5-10-20, 6-12-24 nöron dizilimleri ve 5, 10 ve 15 eğitim sayısı (*epoch*) değerleri ile test edilmiştir. 4 katmanlı yapıda 2-4-8-16, 3-6-12-24, 4-8-16-32, 5-10-20-40, 6-12-24-48 nöron dizilimleri ve 5, 10 ve 15 eğitim sayısı (*epoch*) değerleri ile test edilmiştir. 5 katmanlı yapıda 2-4-8-16-32, 3-6-12-24-48, 4-8-16-32-64, 5-10-20-40-80, 6-12-24-48-96 nöron dizilimleri ve 5, 10 ve 15 eğitim sayısı (*epoch*) değerleri ile test edilmiştir.

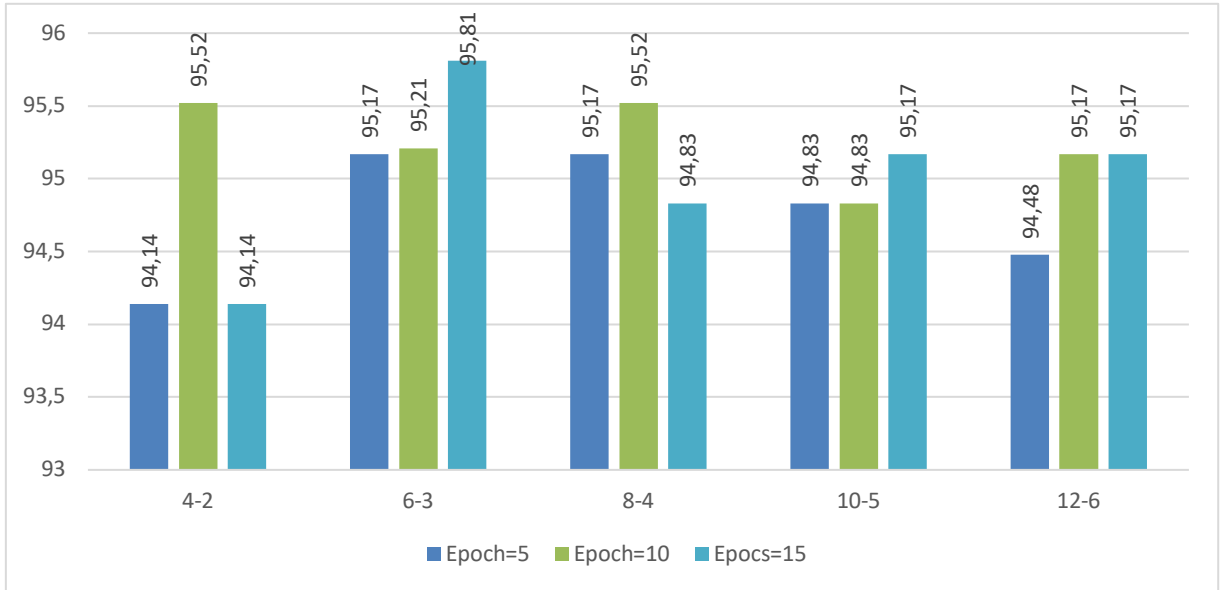
İki Katmanlı Yapıya Ait Sonuçlar

Aşağıda iki katmanlı yapıya ait modelin sonuçları yer almaktadır.

Tablo 19: İki Katmanlı Yapıda Modelin Başarım Oranlarına Ait Tablo

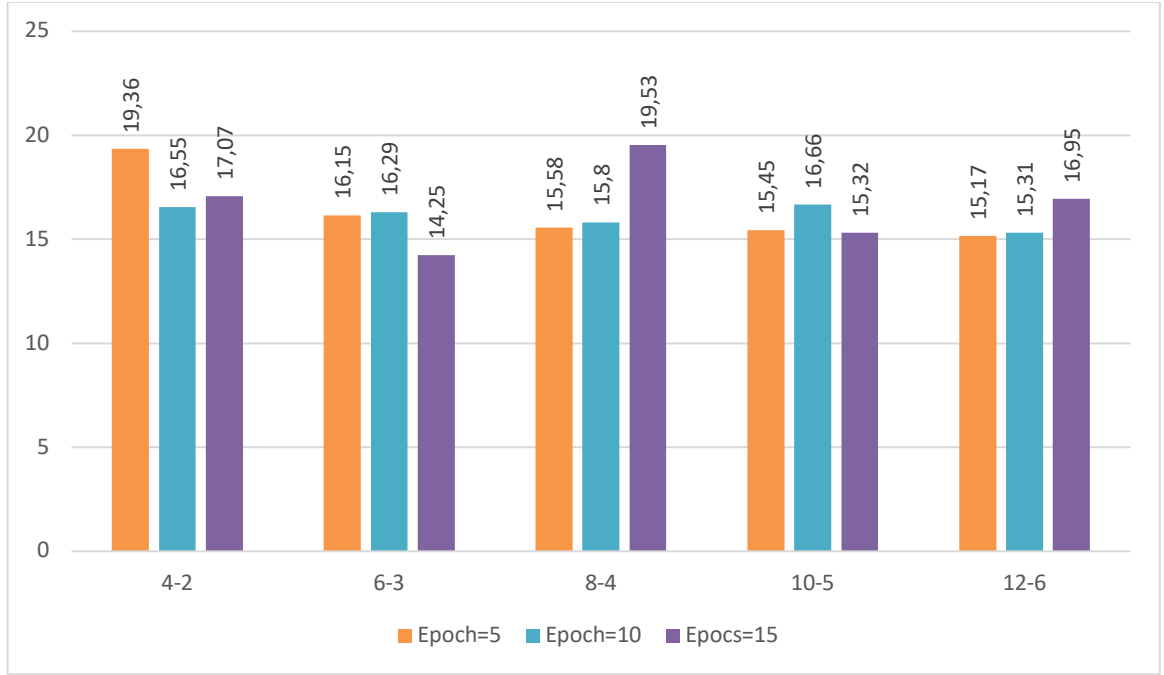
Katman	Nöron Dizilimleri	Epoch = 5	Epoch = 10	Epoch = 15
2	4-2	loss: 0.1936 - acc: 0.9414	loss: 0.1655 - acc: 0.9552	loss: 0.1707 - acc: 0.9414
	6-3	loss: 0.1615 - acc: 0.9517	loss: 0.1629 - acc: 0.9521	loss: 0.1425 - acc: 0.9581
	8-4	loss: 0.1558 - acc: 0.9517	loss: 0.1580 - acc: 0.9552	loss: 0.1953 - acc: 0.9483
	10-5	loss: 0.1545 - acc: 0.9483	loss: 0.1666 - acc: 0.9483	loss: 0.1532 - acc: 0.9517
	12-6	loss: 0.1517 - acc: 0.9448	loss: 0.1531 - acc: 0.9517	loss: 0.1695 - acc: 0.9517

Tablo 19’da görüldüğü üzere iki katmanlı yapıda en yüksek doğruluk oranı (*accuracy*) 3-6 nöron diziliminde eğitim sayısı (*epoch*) değeri 15’de elde edilmiştir. Bu varyasyonda kayıp oranı (*loss*) değeri 0,1425, doğruluk oranı (*accuracy*) değeri 0,9581 olarak elde edilmiştir.



Şekil 63: İki Katmanlı Yapıda Modelin Doğruluk Oranlarına Ait Grafik

Şekil 63'te iki katmanlı yapıda modelin doğruluk oranlarına (*accuracy*) ait grafik yer almaktadır. Grafikte de görüldüğü üzere iki katmanlı yapıda en düşük doğruluk oranı (*accuracy*) %94,14 en yüksek doğruluk oranı (*accuracy*) ise %95,81 olarak elde edilmiştir.



Şekil 64: 2 Katmanlı Yapıda Modelin Kayıp Oranlarına Ait Grafik

Şekil 64'te 2 katmanlı yapıda modelin kayıp oranlarına (*loss*) ait grafik yer almaktadır. Grafikte de görüldüğü üzere iki katmanlı yapıda en düşük kayıp oranı (*loss*) %14,25 en yüksek kayıp oranı (*loss*) ise %19,53 olarak elde edilmiştir.

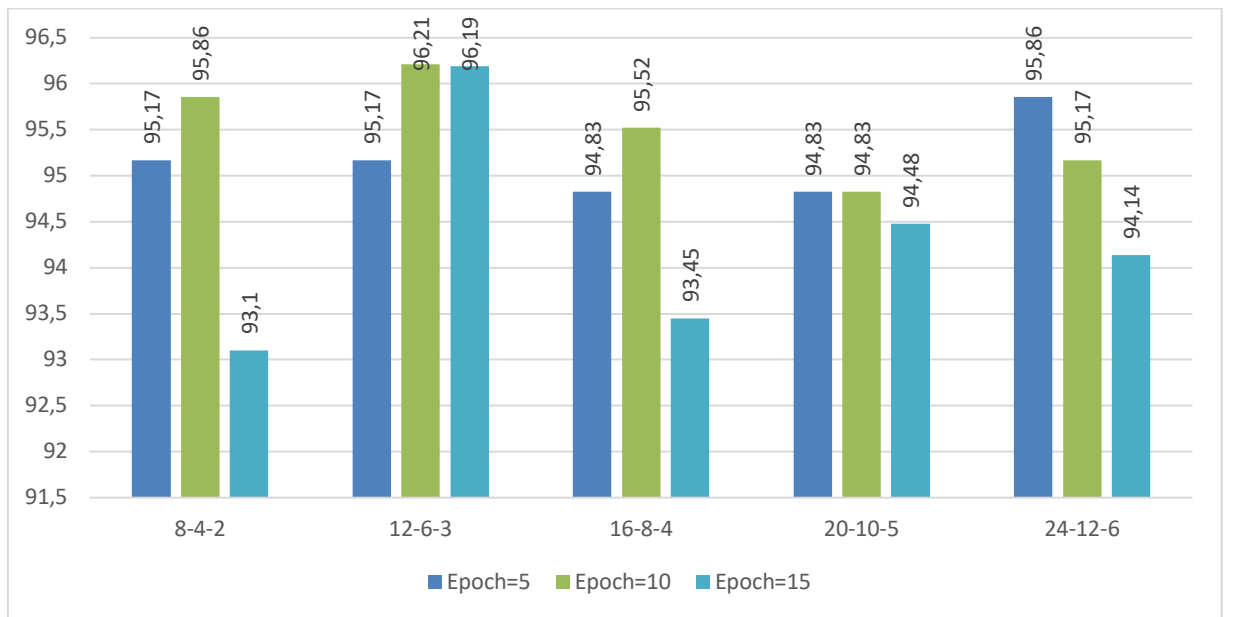
Üç Katmanlı Yapıya Ait Sonuçlar

Aşağıda üç katmanlı yapıya ait modelin sonuçları yer almaktadır.

Tablo 20: Üç Katmanlı Yapıda Modelin Doğruluk Oranlarına Ait Tablo

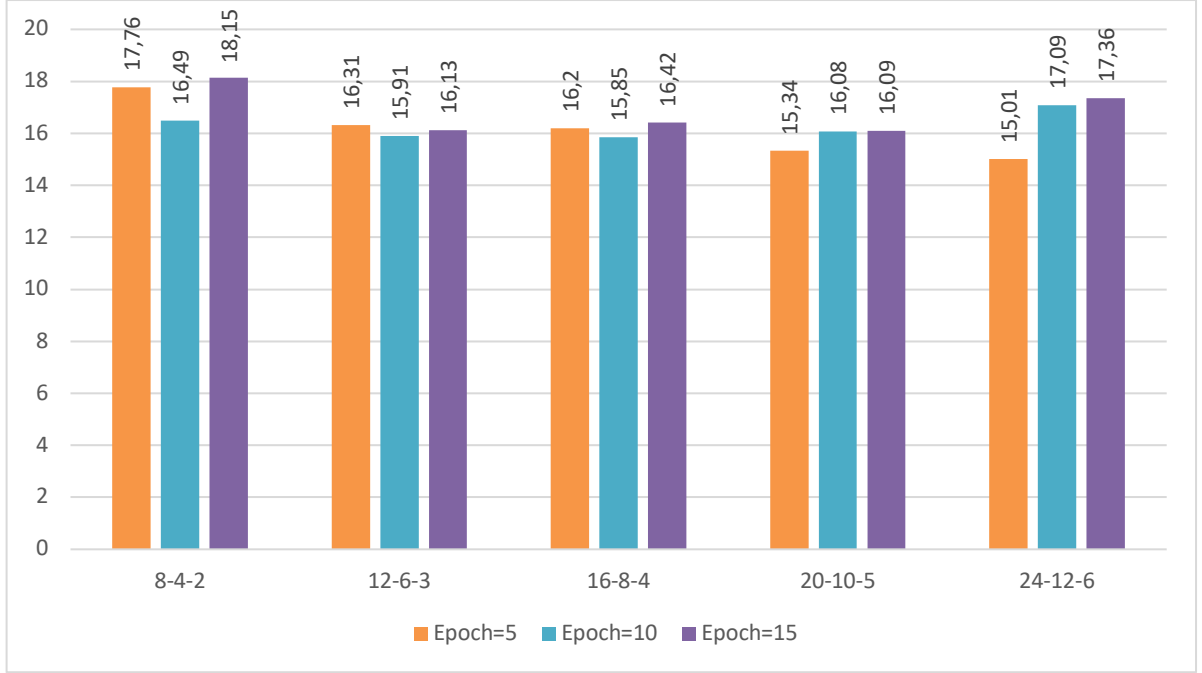
Katman	Nöron Dizilimleri	Epoch = 5	Epoch = 10	Epoch = 15
3	8-4-2	loss: 0.1776 - acc: 0.9517	loss: 0.1649 - acc: 0.9586	loss: 0.1815 - acc: 0.9310
	12-6-3	loss: 0.1631 - acc: 0.9517	loss: 0.1591 - acc: 0.9621	loss: 0.1613 - acc: 0.9619
	16-8-4	loss: 0.1620 - acc: 0.9483	loss: 0.1585 - acc: 0.9552	loss: 0.1642 - acc: 0.9345
	20-10-5	loss: 0.1534 - acc: 0.9483	loss: 0.1608 - acc: 0.9483	loss: 0.1609 - acc: 0.9448
	24-12-6	loss: 0.1501 - acc: 0.9586	loss: 0.1709 - acc: 0.9517	loss: 0.1736 - acc: 0.9414

Tablo 20’de görüldüğü üzere üç katmanlı yapıda en yüksek doğruluk oranı (*accuracy*) 3-6-12 nöron diziliminde eğitim sayısı (*epoch*) değeri 10’da gözlenmiştir. Bu varyasyonda kayıp oranı (*loss*) değeri 0,1591, doğruluk oranı (*accuracy*) değeri 0,9621 olarak elde edilmiştir.



Şekil 65: Üç Katmanlı Yapıda Modelin Doğruluk Oranlarına Ait Grafik

Şekil 65'te Üç katmanlı yapıda modelin doğruluk oranlarına (*accuracy*) ait grafik yer almaktadır. Grafikte de görüldüğü üzere üç katmanlı yapıda en düşük doğruluk oranı (*accuracy*) %94,14 en yüksek doğruluk oranı (*accuracy*) ise %95,81 olarak elde edilmiştir.



Şekil 66: Üç Katmanlı Yapıda Modelin Kayıp Oranlarına Ait Grafik

Şekil 66'da üç katmanlı yapıda modelin kayıp oranlarına (*loss*) ait grafik yer almaktadır. Grafikte de görüldüğü üzere üç katmanlı yapıda en düşük kayıp oranı (*loss*) %15,01 en yüksek kayıp oranı (*loss*) ise %18,15 olarak elde edilmiştir.

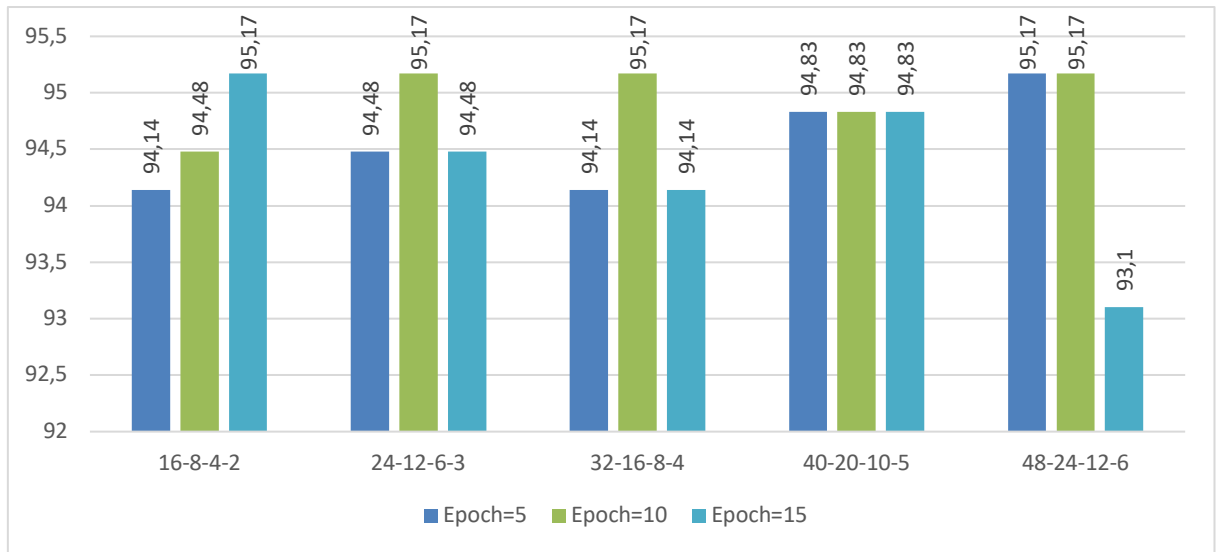
Dört Katmanlı Yapıya Ait Sonuçlar

Aşağıda dört katmanlı yapıya ait modelin sonuçları yer almaktadır.

Tablo 21: Dört Katmanlı Yapıda Modelin Doğruluk Oranlarına Ait Tablo

Katman	Nöron Dizilimleri	Epoch = 5	Epoch = 10	Epoch = 15
4	16-8-4-2	loss: 0.2507 – acc: 0.9414	loss: 0.1603 – acc: 0.9448	loss: 0.1674 – acc: 0.9517
	24-12-6-3	loss: 0.1930 – acc: 0.9448	loss: 0.1547 – acc: 0.9517	loss: 0.1792 – acc: 0.9448
	32-16-8-4	loss: 0.2403 – acc: 0.9414	loss: 0.1549 – acc: 0.9517	loss: 0.1748 – acc: 0.9414
	40-20-10-5	loss: 0.1541 – acc: 0.9483	loss: 0.1821 – acc: 0.9483	loss: 0.1674 – acc: 0.9483
	48-24-12-6	loss: 0.1516 – acc: 0.9517	loss: 0.1616 – acc: 0.9517	loss: 0.1665 – acc: 0.9310

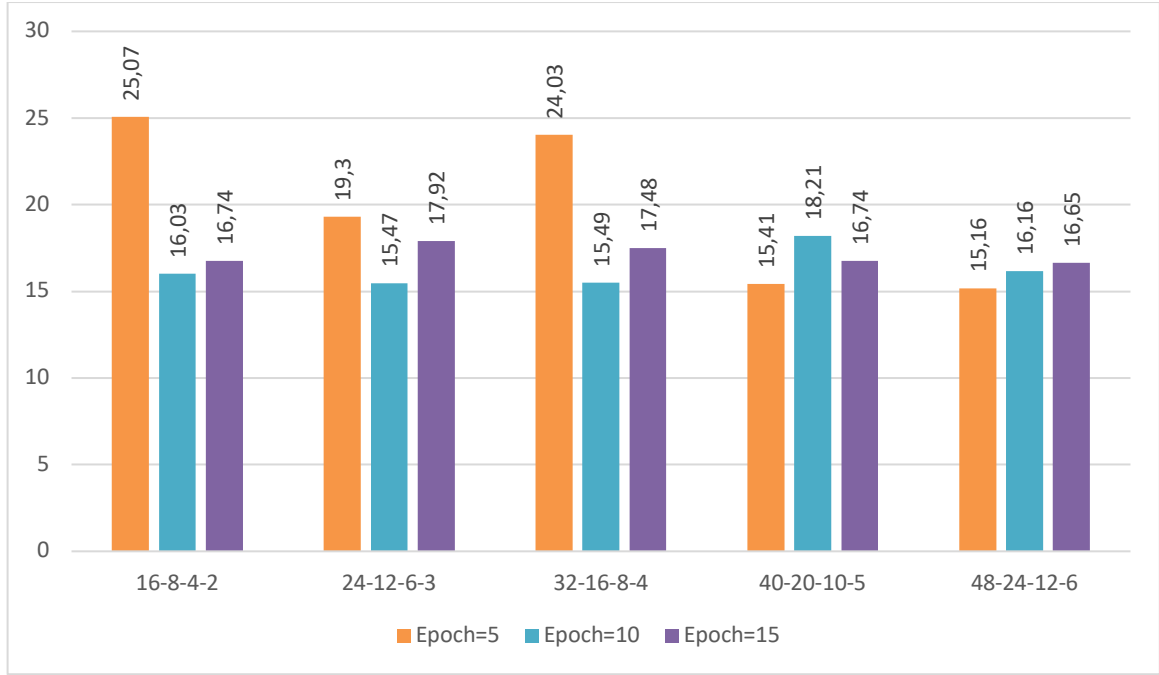
Tablo 21’de görüldüğü üzere dört katmanlı yapıda en yüksek doğruluk oranı (*accuracy*) 6-12-24-48 nöron diziliminde eğitim sayısı (*epoch*) değeri 5’de gözlenmiştir. Bu varyasyonda kayıp oranı (*loss*) değeri 0,1516, doğruluk oranı (*accuracy*) değeri 0,9517 olarak elde edilmiştir.



Şekil 67: Dört Katmanlı Yapıda Modelin Doğruluk Oranlarına Ait Grafik

Şekil 67’de dört katmanlı yapıda modelin doğruluk oranlarına ait grafik yer almaktadır.

Grafikte de görüldüğü üzere dört katmanlı yapıda en düşük doğruluk oranı (*accuracy*) %93,10 en yüksek doğruluk oranı (*accuracy*) ise %95,17 olarak elde edilmiştir.



Şekil 68: Dört Katmanlı Yapıda Modelin Kayıp Oranlarına Ait Grafik

Şekil 68’de Dört katmanlı yapıda modelin kayıp oranlarına (*loss*) ait grafik yer almaktadır.

Grafikte de görüldüğü üzere 4 katmanlı yapıda en düşük kayıp oranı (*loss*) %15,16 en yüksek kayıp oranı (*loss*) ise %25,07 olarak elde edilmiştir.

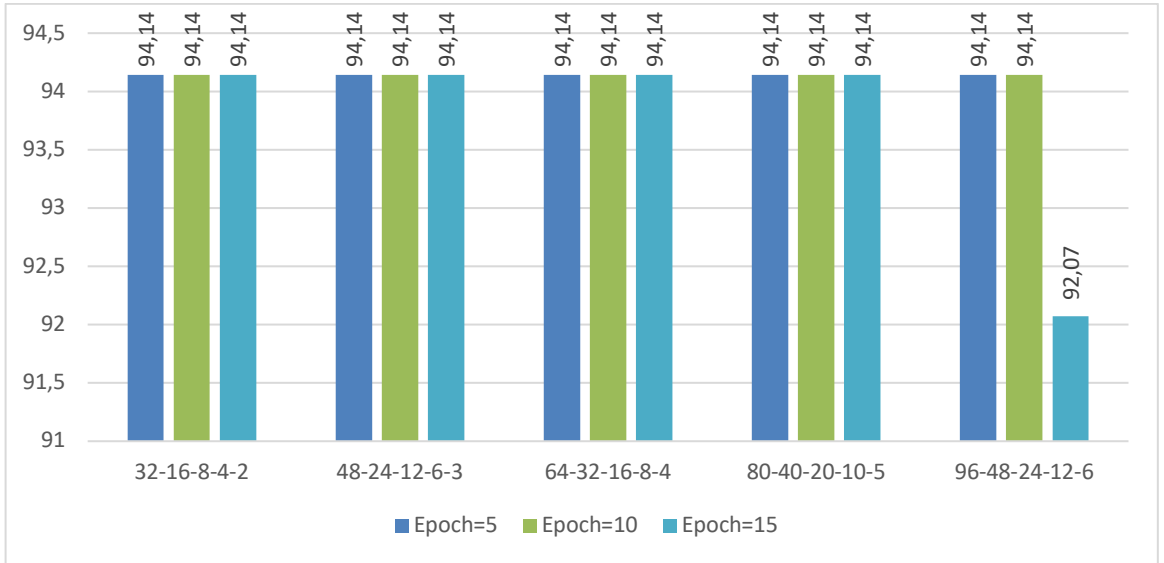
Beş Katmanlı Yapıya Ait Sonuçlar

Aşağıda beş katmanlı yapıya ait modelin sonuçları yer almaktadır.

Tablo 22: Beş Katmanlı Yapıda Modelin Doğruluk Oranlarına Ait Tablo

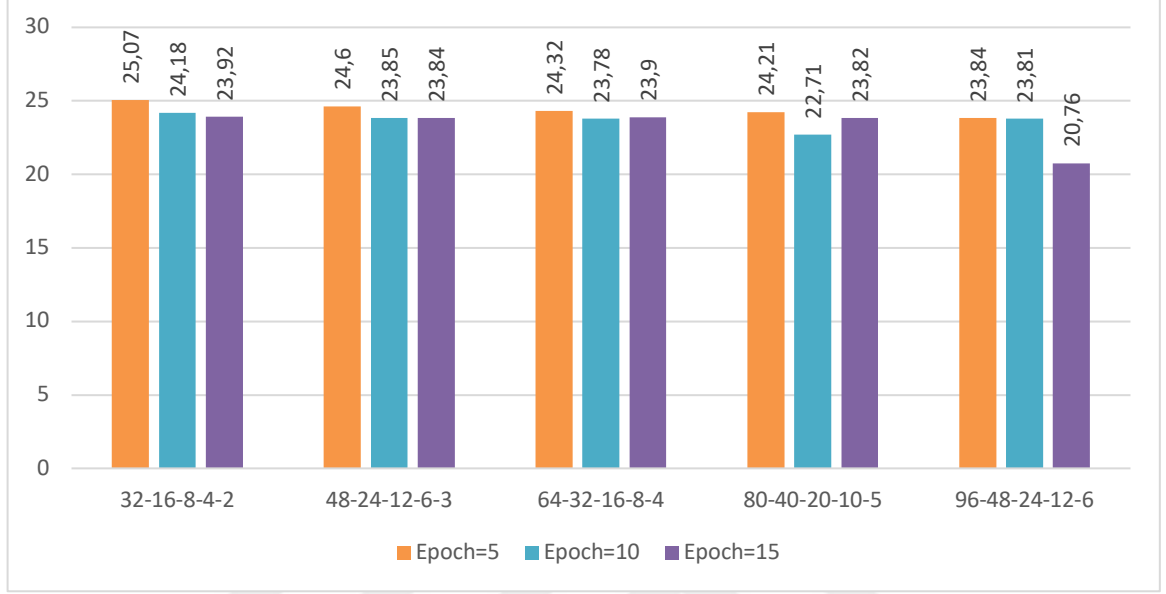
Katman	Nöron Dizilimleri	Epoch = 5	Epoch = 10	Epoch = 15
5	32-16-8-4-2	loss: 0.2507 - acc: 0.9414	loss: 0.2418 - acc: 0.9414	loss: 0.2392 - acc: 0.9414
	48-24-12-6-3	loss: 0.2460 - acc: 0.9414	loss: 0.2385 - acc: 0.9414	loss: 0.2384 - acc: 0.9414
	64-32-16-8-4	loss: 0.2432 - acc: 0.9414	loss: 0.2378 - acc: 0.9414	loss: 0.2390 - acc: 0.9414
	80-40-20-10-5	loss: 0.2421 - acc: 0.9414	loss: 0.2271 - acc: 0.9414	loss: 0.2382 - acc: 0.9414
	96-48-24-12-6	loss: 0.2384 - acc: 0.9414	loss: 0.2381 - acc: 0.9414	loss: 0.2076 - acc: 0.9207

Tablo 22’de görüldüğü üzere beş katmanlı yapıda en yüksek doğruluk oranı (*accuracy*) 5-10-20-40-80 nöron diziliminde eğitim sayısı (*epoch*) değeri 10’da gözlenmiştir. Bu varyasyonda kayıp oranı (*loss*) değeri 0,2271, doğruluk oranı (*accuracy*) değeri 0,9414 olarak elde edilmiştir.



Şekil 69: Beş Katmanlı Yapıda Modelin Doğruluk Oranlarına Ait Grafik

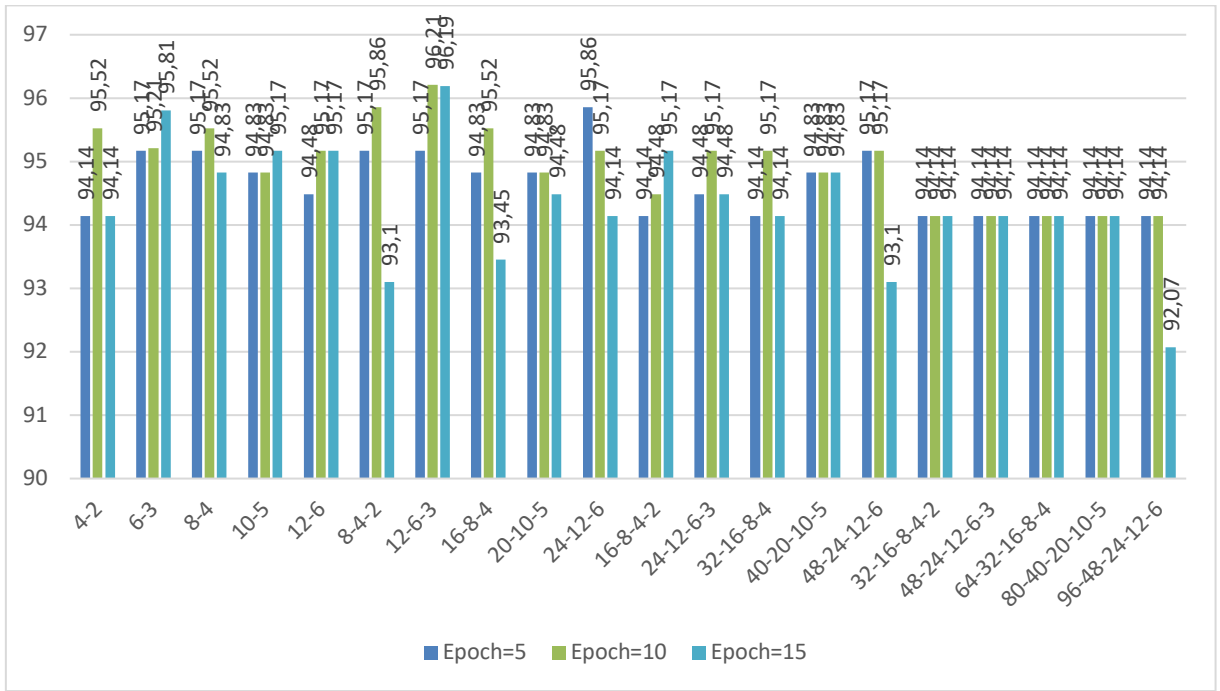
Şekil 69’da beş katmanlı yapıda modelin doğruluk oranlarına (*accuracy*) ait grafik yer almaktadır. Grafikte de görüldüğü üzere beş katmanlı yapıda en düşük doğruluk oranı (*accuracy*) %92,07, en yüksek doğruluk oranı (*accuracy*) ise %94,14 olarak elde edilmiştir.



Şekil 70: Beş Katmanlı Yapıda Modelin Kayıp Oranlarına Ait Grafik

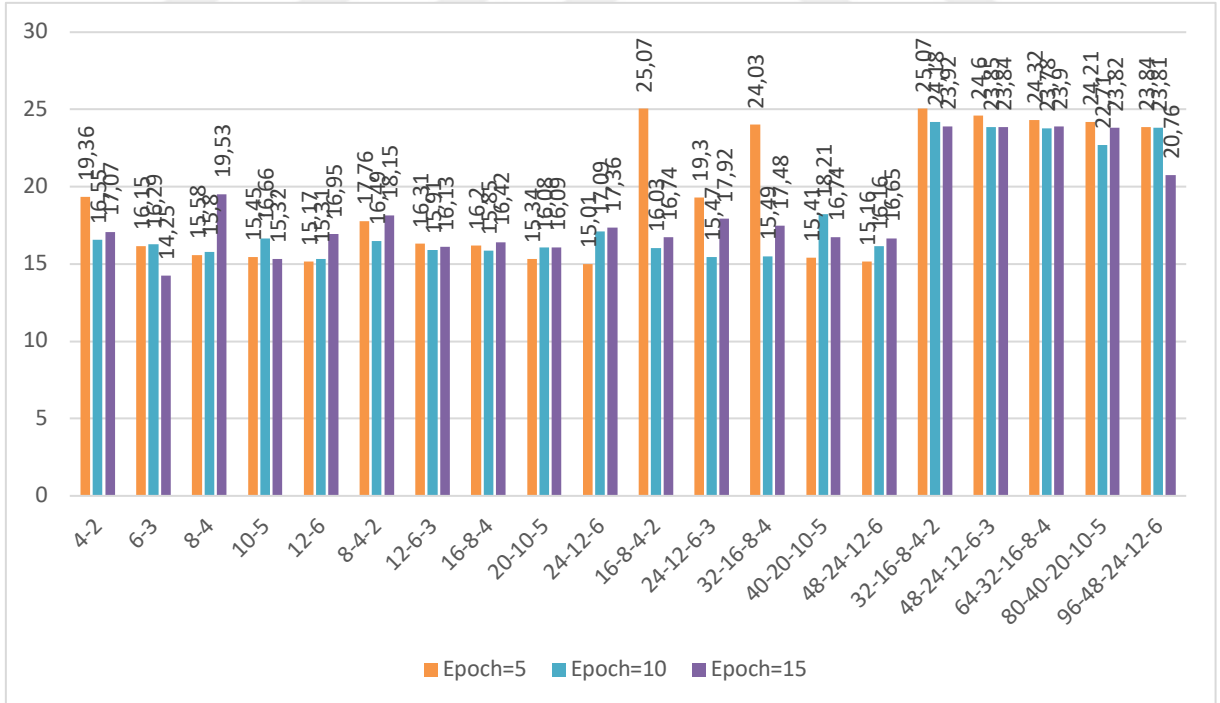
Şekil 70’te beş katmanlı yapıda modelin kayıp oranlarına ait grafik yer almaktadır. Grafikte de görüldüğü üzere beş katmanlı yapıda en düşük kayıp oranı (*loss*) %20,76, en yüksek kayıp oranı (*loss*) ise %25,07 olarak elde edilmiştir. Tüm varyasyonlar içerisinde en yüksek doğruluk oranı (*accuracy*) 3-6-12 nöron diziliminde epoch değeri 10’da gözlenmiştir. Bu varyasyonda kayıp oranı (*loss*) değeri 0,1591, doğruluk oranı (*accuracy*) değeri 0,9621 olarak elde edilmiştir. En düşük doğruluk oranı (*accuracy*) ise 6-12-24-48-98 nöron diziliminde eğitim sayısı (*epoch*) değeri 15’de gözlenmiştir. Bu varyasyonda kayıp oranı (*loss*) değeri 0,2076, doğruluk oranı (*accuracy*) değeri 0,9207 olarak elde edilmiştir.

Yukarıdaki elde edilen doğruluk oranlarında da (*accuracy*) görüldüğü üzere katman sayısı ve eğitim sayısı (*epoch*) değeri belli bir eşikten sonra artsa bile modelin başarı oranı artmamakta ve belli oranda düşmektedir. Literatürde bu durum aşırı öğrenme (*overfitting*) denilen kavram ile açıklanmaktadır. Model çok katmanlı yapıda çok fazla kez eğitildiğinde öğrenme yerine ezberleme gerçekleşmekte bu da modelin doğruluk oranını (*accuracy*) düşürmektedir (Yeom ve diğ., 2018).



Şekil 71: Modelin Doğruluk Oranlarına Ait Grafik – Tüm Varyasyonlar

Şekil 71’de modelin farklı tüm varyasyonlardaki doğruluk oranlarına ait grafik yer almaktadır.



Şekil 72: Modelin Kayıp Oranlarına Ait Grafik – Tüm Varyasyonlar

Şekil 72’de modelin farklı tüm varyasyonlardaki kayıp oranlarına ait grafik yer almaktadır.

F-Score'un Hesaplanması

Özellikle dağınık veri setlerinde modelin doğruluk oranı (*accuracy*) ile birlikte f-score'un da hesaplanması gerekmektedir (Sokolova ve diğ., 2006). F-score'un hesaplanması için şu adımlar izlenmiştir.

1. Karar Matrisi (*Confusion Matrix*) tablosu'nun oluşturulması
2. Kesinlik (*Precision*) değerinin hesaplanması
3. Duyarlılık (*Recall*) değerinin hesaplanması
4. F-score'un hesaplanması

Karar Matrisi (*Confusion Matrix*) Tablosunun Oluşturulması

Karar matrisinin oluşturulmasında aşağıda yer alan tablo kullanılır.

Gerçek Değerler	Tahmini Değerler	
	Pozitif	Negatif
	Pozitif	Negatif
Pozitif	TN	FP
Negatif	FN	TP

Şekil 73: Karar Matrisi (Ohsaki ve diğ., 2018)

Şekil 73'teki karar matrisinde:

TP (Doğru Pozitif): Doğru olarak sınıflanmış pozitif örnekleri,

FP (Yanlış Pozitif): Yanlış olarak sınıflanmış pozitif örnekleri,

TN (Doğru Negatif): Doğru olarak sınıflanmış negatif örnekleri,

FN (Yanlış Negatifler): Yanlış olarak sınıflanmış negatif örnekleri temsil etmektedir.

Toplam verilerin %30'unu oluşturan test veri seti kullanılarak geliştirilen modelin hata matrisi oluşturulmuştur.

```
metrics.confusion_matrix(y_test, predict)

array([[138, 11],
       [ 14, 415]], dtype=int64)
```

Şekil 74: Sklearn Kütüphanesi Kullanılarak Oluşturulan Karar Matrisi

Şekil 74'te Sklearn kütüphanesi ve eğitilen model kullanılarak test veri seti ile oluşturulan karar matrisinin çıktısı görülmektedir.

Tablo 23: Karar Matrisi – Değerlerin Eklenmiş Hali

Gerçek Değerler	Tahmini Değerler	
	Pozitif	Negatif
	Pozitif	Negatif
Pozitif	138 (TN)	11 (FP)
Negatif	14 (FN)	415 (TP)

Sklearn kütüphanesi ve eğitilen model kullanılarak, test veri setleri ile elde edilen sonuçlar Tablo 23'teki karar matrisine eklenmiştir. Tabloda da görüldüğü üzere 415 doğru pozitif, 11 yanlış pozitif, 138 doğru negatif, 14 yanlış negatif olarak sınıflanmış toplamda 578 hedefin duygu durumu yer almaktadır.

Kesinlik (*Precision*) Değerinin Hesaplanması

Kesinlik (*Precision*) değerinin hesaplanması için kullanılan formül aşağıdadır.

$$Kesinlik (Precision) = \frac{TP}{TP + FP}$$

Şekil 75: Kesinlik (*Precision*) Değerinin Hesaplanması için Kullanılan Formül (Sokolova ve diğ., 2006)

Kesinlik değeri TP (Doğru Pozitif) değerinin, TP (Doğru Pozitif) ile FP (Yanlış Pozitif) değerlerinin toplamına bölünmesi ile elde edilir.

$$Kesinlik (Precision) = \frac{411}{411 + 15} = 0,97$$

Şekil 76: Kesinlik (*Precision*) Değeri

Şekil 76’da elde edilen kesinlik (*precision*) değeri görülmektedir. Karar matrisinden alınan veriler ile elde edilen kesinlik değeri (*precision*) 0,97’dir.

Duyarlılık (*Recall*) Değerinin Hesaplanması

Duyarlılık (*Recall*) değerinin hesaplanması için kullanılan formül aşağıdadır.

$$Duyarlılık (Recall) = \frac{TP}{TP + FN}$$

Şekil 77: Duyarlılık (*Recall*) Değerinin Hesaplanması için Kullanılan Formül (Sokolova ve diğ., 2006)

Duyarlılık değeri TP (Doğru Pozitif) değerinin, TP (Doğru Pozitif) ile FN (Yanlış Negatif) değerlerinin toplamına bölünmesi ile elde edilir.

$$Duyarlılık (Recall) = \frac{411}{411 + 14} = 0,96$$

Şekil 78: Duyarlılık (*Recall*) Değeri

Şekil 78’de elde edilen duyarlılık (*recall*) değeri görülmektedir. Karar matrisinden alınan veriler ile elde edilen duyarlılık değeri 0,96’dır.

F-score’un hesaplanması

F-score değerinin hesaplanması için kullanılan formül aşağıdadır.

$$Fscore = 2 * \frac{Precision * Recall}{Precision + Recall}$$

Şekil 79: F-score’un Hesaplanması için Kullanılan Formül (Sokolova ve diğ., 2006)

Şekil 79’da görüldüğü üzere F-score değerinin hesaplanmasında kesinlik (*precision*) ve duyarlılık (*recall*) değerlerinin harmonik ortalaması alınır.

$$Fscore = 2 * \frac{0,97 * 0,96}{0,97 + 0,96} = 0,96$$

Şekil 80: F-score Değeri

Şekil 80’de elde edilen f-score değeri görülmektedir. Kesinlik (*precision*) ve duyarlılık (*recall*) değerleri ile elde edilen f-score değeri 0,96’dır.

```
from sklearn.metrics import classification_report
target_names = ['class 0', 'class 1']
print(classification_report(y_test, predict, target_names=target_names))
```

	precision	recall	f1-score
class 0	0.91	0.93	0.92
class 1	0.97	0.96	0.96
accuracy			0.96

Şekil 81: F-score Sonuçlarına Ait Sklearn Kütüphanesi ile Elde Edilen Sonuçlar

Şekil 81’de Sklearn kütüphanesi ile elde edilen f-score değerleri görülmektedir. *Class 0* negatif kutbu, *Class 1* ise pozitif kutbu temsil etmektedir. Şekilde de görüldüğü üzere negatif kutup için kesinlik (*precision*) değeri 0,91, duyarlılık (*recall*) değeri 0,93 bu değerlere bağlı olarak da f-score değeri 0,92 olarak elde edilmiştir.

Pozitif kutup için kesinlik (*precision*) değeri 0,97, duyarlılık (*recall*) değeri 0.96 bu değerlere bağlı olarak da f-score değeri 0,96 olarak elde edilmiştir.

4.2. SÖZLÜK TABANLI YAKLAŞIM İÇİN GELİŞTİRİLEN MODELE AİT BULGULAR

Sözlük tabanlı yaklaşım kapsamında geliştirilen modelin test edilmesi için, SentiTurknet ve SentiWordNet duygu sözlükleri kullanılmıştır.

Modelin doğruluk oranını hesaplamada aşağıdaki doğruluk (*accuracy*) formülü kullanılmıştır.

$$Doğruluk = \frac{TP + TN}{TN + TP + FN + FP}$$

Şekil 82: Sözlük Tabanlı Yaklaşımın Doğruluk Oranını (*Accuracy*) Hesaplamada Kullanılan Formül

Şekil 82’de sözlük tabanlı yaklaşımın doğruluk oranını (*accuracy*) hesaplamada kullanılan formül yer almaktadır. Modelin doğruluk oranını (*accuracy*) hesaplamak için TP (Doğru Pozitif) ve TN (Doğru Negatif) toplamının, toplam hedef sayısına bölünmesi ile elde edilmektedir.

Tablo 24: SentiTurkNet Sözlüğü ile Elden Edilen Doğruluk Oranı

Sözlük Adı	Barındırdığı Kelime Sayısı	Veri Seti İçinde Yer Alan Aspect Sayısı	Duygu Durumu Doğru Tespit Edilen Aspect Sayısı	Başarı Oranı
SentiTurkNet	14.795	1925	1689	%87,7

Tablo 24’te görüldüğü üzere 14.795 kelime barındıran SentiTurkNet duygu sözlüğünde, sözlüğe puanlama için gönderilen 1925 hedefin 1689’unu doğru sınıflamıştır. %87,7’lik doğruluk oranı (*accuracy*) elde edilmiştir.

Tablo 25: SentiWordNet-TR Sözlüğü ile Elden Edilen Doğruluk Oranı

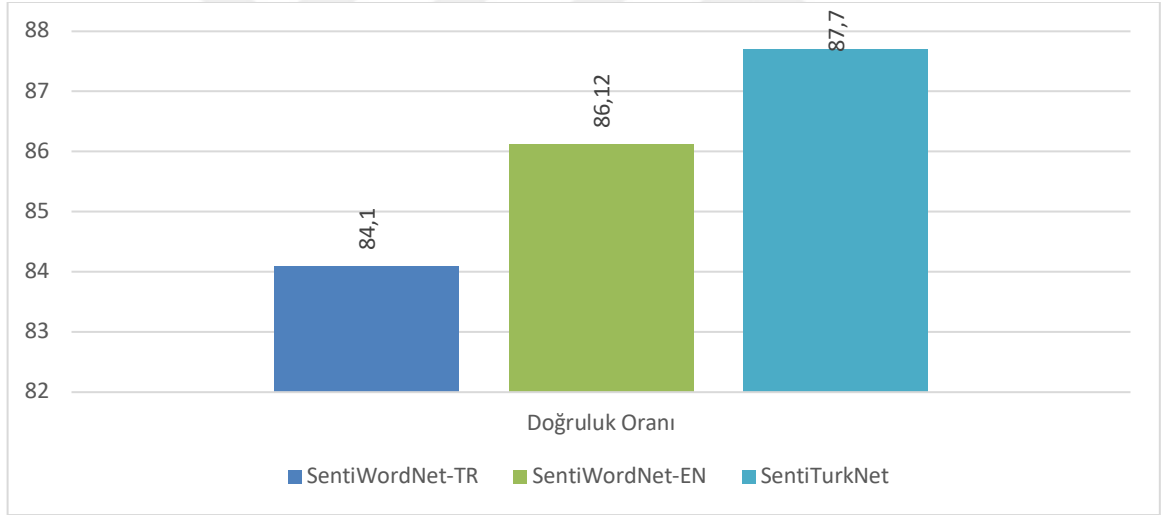
Sözlük Adı	Barındırdığı Kelime Sayısı	Veri Seti İçinde Yer Alan Aspect Sayısı	Duygu Durumu Doğru Tespit Edilen Aspect Sayısı	Başarı Oranı
SentiWordNet TR	73.386	1925	1620	%84,1

Tablo 25’te görüldüğü üzere 73.386 kelime barındıran SentiWordNet’in Türkçeye çevrilmiş duygu sözlüğüne gönderilen 1925 hedefin 1620’sini doğru sınıflamıştır. %84,1’lik bir doğruluk oranı (*accuracy*) elde edilmiştir.

Tablo 26: SentiWordNet-EN Sözlüğü ile Elden Edilen Doğruluk Oranı

Sözlük Adı	Barındırdığı Kelime Sayısı	Veri Seti İçinde Yer Alan Aspect Sayısı	Duygu Durumu Doğru Tespit Edilen Aspect Sayısı	Başarı Oranı
SentiWordNet EN	117.659	1925	1658	%86,12

Tablo 26’da görüldüğü üzere 117.659 kelime barındıran SentiWordNet sözlüğüne İngilizceye çevrilerek gönderilen ilişkili kelimelerden oluşan 1925 hedefin 1658’ini doğru sınıflamıştır. %86,12’lik doğruluk oranı (*accuracy*) elde edilmiştir.



Şekil 83: Sözlük Tabanlı Yaklaşım ile Elden Edilen Doğruluk Oranlarına Ait Grafik

Şekil 83’de sözlük tabanlı yaklaşım ile elden edilen doğruluk oranlarına (*accuracy*) ait grafik yer almaktadır. Grafikte de görüldüğü üzere en yüksek doğruluk oranı SentiTurkNet sözlüğü ile elde edilmiştir. Doğruluk oranı (*accuracy*) %87,7 olarak tespit edilmiştir. En düşük oran SentiWordNet sözlüğünün Türkçeye çevrilmesi ile elde edilen sözlük ile elde edilmiştir. Doğruluk oranı (*accuracy*) %84,1 olarak tespit edilmiştir. SentiWordNet sözlüğünü İngilizce hali ile kullanıp, hedefi niteleyen ilişkili kelimelerin İngilizceye çevrilerek puanlandığı senaryoda ise %86,12’lik bir doğruluk oranı (*accuracy*) tespit edilmiştir.

• TARTIŞMA VE SONUÇ

5.1. ELDE EDİLEN SONUÇLARIN DEĞERLENDİRİLMESİ

Bilgi ve iletişim teknolojilerindeki gelişmeler yapısal olmayan bir veri türü olan metinsel veri üretimimizi artırmıştır. Bu verilerin işlenmesi ve bir katma değer sağlaması kurum ve kuruluşlar için günümüzde çok daha önemli hale gelmiştir. Finanstan eğitime, sağlıktan ulaşımaya hemen hemen her alanda veriye dayalı öngörü ve karar destek sistemleri yönetimin her kademesinde büyük bir ihtiyaçtır. (Aytekin ve Sütçü, 2018; Çetin ve Eryiğit, 2018; Hark ve diğ., 2017; Xia ve diğ., 2011).

Yapay zeka çalışmalarının en önemli alt kollarından biri de Doğal Dil İşlemedir (DDİ). İnsan zekası doğal dili kullanarak kendini ifade edebilmekte ya da başkalarının ne söylediğini rahat bir şekilde anlayabilmektedir; başka bir ifade ile iletişim kurabilmektedir (Chomsky, 1986). İnsana özgü iletişim kurma becerilerini bilgisayar sistemlerine kazandırabilmek için ise DDİ yöntem ve metotlarından faydalanılmaktadır. Web siteleri, sosyal medya platformları, forum vb. web teknolojileri ile birlikte çok büyük miktarda metinsel veriler üretmekteyiz. Yapısal olmayan bu verilerin işlenip anlamlı hale getirilebilmesi için ise insanüstü güce ihtiyaç bulunmaktadır. Bu noktada DDİ yöntem ve metodları ile geliştirilen yapay zeka algoritmaları çok büyük verileri isteğimiz doğrultusunda sınıflayabiliyor, özetini çıkarabiliyor ya da farklı bir dile saniyeler içerisinde çevirebiliyorlar (Adalı, 2012; Fessini ve diğ., 2014). Çağrı merkezlerinde insan gücüne dayalı yapılan çağrı hizmetleri DDİ yöntem ve metotları sayesinde geliştirilen akıllı chatbotlar ile insan gücü olmadan rahatlıkla yapılabilir hale gelmiştir. Telefonlarımızın içerisine giren akıllı asistanlar sayesinde istediğimiz bilgiye telefonumuza dokunmadan ulaşabiliyor, sesli komutlar ile yemek siparişi bile verebiliyoruz. Bu ve benzeri gelişmeler DDİ disiplinine olan ilgiyi her geçen gün arttırmakta ve bu disiplini daha popüler bir hale getirmektedir (Hark ve diğ., 2017).

DDİ çalışmalarının en önemli alt kollarından biri de duygu analizi çalışmalarıdır. Duygu analizi çalışmaları ile metinsel bir dokümanın içerdiği duygu durumlarını tespit edebiliriz. Bu duygu durumları “olumlu (pozitif)”, “olumsuz (negatif)”, “tarafsız (nötr)” olabileceği gibi “mutlu”, “neşeli”, “hüzünlü”, “heyecanlı” şeklinde de olabilir (Burcu ve Şimşek, 2018).

Duygu analizi çalışmaları doküman, cümle ve hedef olmak üzere üç farklı seviyede yapılmaktadır. Doküman ve cümle seviyesinde yapılan duygu analizi çalışmalarında dokümanın veya cümlelerin ortalama duygu durumu tespit edilmektedir. Bundan dolayı duygu durumu kayıpları yaşanabilmektedir.

Hedef seviyesinde duygu analizi çalışmaları ise son yıllarda popüler olmuş bir yaklaşımdır. Bu yaklaşıma göre bir metinsel ifadenin duygu durumuna karar verebilmek için önce metin içerisindeki hedeflerin belirlenmesi gerekir. Her hedefin duygu durumu ayrı ayrı tespit edilmelidir. Bu sayede duygu durumu kaybı minimize edilir.

Bu okulun eğitim sistemi çok iyi ancak fiziksel imkanları yeterli değil.

Şekil 84: Birden Fazla Hedef Barındıran Bir Kullanıcı Yorumu

Şekil 84’te yer alan kullanıcı yorumunda da görüldüğü üzere cümle içerisinde “Eğitim Sistemi” ve “Fiziksel İmkanlar” olmak üzere iki farklı hedef ve duygu durumu vardır. Hedeflerden ilki olan “Eğitim Sistemi” “çok iyi” olarak (*pozitif*) belirtilmiş buna karşılık “Fiziksel İmkanları” ise “yeterli değil” (*negatif*) olarak belirtilmiş. Hedef tabanlı duygu analizi her bir hedefi ve duygu durumunu tespit etmeye çalıştığı için metnin duygu durumu hakkında doküman ve cümle seviyesinde yapılan çalışmalara göre daha iyi sonuçlar vermektedir. Bununla birlikte hedef seviyesinde yapılan çalışmalarda etiketli veri setleri oluşturmak, doküman ve cümle seviyesindeki çalışmalara göre daha karmaşık ve zaman alıcıdır.

Bu çalışma kapsamında sözlük tabanlı ve makine öğrenmesi yaklaşımları kullanılarak okul.com.tr sitesinden elde edilen kullanıcı yorumları üzerinde hedef seviyesinde duygu analizi çalışması yapılmıştır.

Veri seti üzerinde ilk olarak hedefler ve hedeflerin duygu durumları tespit edilmiştir. Hedefi belirleyen ilişkili kelimeler belirlenmiş, bu kelimeler üzerinde sözlük tabanlı yaklaşım ve makine öğrenmesi yaklaşımı ile geliştirilen modeller test edilmiştir.

Makine Öğrenmesi Yaklaşımında Elde Edilen Sonuçların Değerlendirilmesi

Makine öğrenmesi yaklaşımı kapsamında çok katmanlı yinelenen yapay sinir ağları (Recurrent Neural Network-RNN) kullanılmıştır. Bu doğrultuda 2,3,4,5 katmanlı ve farklı nöron dizilimlerinden oluşan yapılar oluşturulmuştur. Bu yapıların oluşturulmasında Geçitli Tekrarlayan Birim (Gated Recurrent Unit-GRU) kullanılmıştır.

En yüksek doğruluk oranı üç katmanlı 3-6-12 nöron dizilimli yapıda eğitim sayısı (*epoch*) değeri 10 ile elde edilmiştir. Elde edilen en yüksek doğruluk oranı (*accuracy*) %96,12'dir. En düşük doğruluk oranı ise 5 katmanlı 96-48-24-12-6 nöron dizilimli yapıda eğitim sayısı (*epoch*) değeri 15 ile elde edilmiştir. Elde edilen en düşük doğruluk oranı (*accuracy*) ise %92,07'dir. Geliştirilen modelde katman ve eğitim sayısı arttıkça başarı oranı artmış ancak belli bir noktadan sonra aşırı öğrenmeden dolayı (*overfitting*) sabit kalarak düşmeye başlamıştır.

Geliştirilen modelde her bir duygu durumu için f-score değerleride hesaplanmıştır. Pozitif kutup için kesinlik (*precision*) değeri 0,97, duyarlılık (*recall*) değeri 0,96 ve bu doğrultuda f-score değeri 0,96 olarak elde edilmiştir. Negatif kutup için ise kesinlik (*precision*) değeri 0,91, duyarlılık (*recall*) değeri 0,93 ve bu doğrultuda f-score değeri 0,92 olarak elde edilmiştir. Bu değerler geliştirilen modelin her bir duygu kutbunu (pozitif-negatif) yüksek oranda doğru sınıflayabildiğini göstermektedir.

F-score hesaplaması özellikle eşit dağılım göstermeyen dağınık veri setlerinde önemlidir. Bu tür veri setlerinde sadece doğruluk (*accuracy*) yeterli olmayabilir. Modelin, her bir kutbu ne kadar oranda doğru sınıfladığına bakılmalıdır.

Sözlük Tabanlı Yaklaşımında Elde Edilen Sonuçların Değerlendirilmesi

Sözlük tabanlı yaklaşım kapsamında üç farklı duygu sözlüğü kullanılmıştır. Bunlar SentiWordNet-TR, SentiWordNet – EN ve SentiTurkNet'dir.

SentiWordNet-TR, SentiWordNet sözlüğünün Türkçeye çevrilmesi ile elde edilmiştir. Sözlükte 73.386 Türkçe kelime bulunmaktadır. Bu sözlük kullanılarak geliştirilen modelin doğruluk oranı (*accuracy*) test edilmiştir. %84,1'lik bir doğruluk oranına (*accuracy*) ulaşılmıştır.

SentiWordNet–EN, WordNet sözlüğü kullanılarak geliştirilmiş bir duygu sözlüğüdür. Sözlükte 117.659 İngilizce kelime bulunmaktadır. Bu sözlük kullanılarak geliştirilen modelin doğruluk oranı (*accuracy*) test edilmiştir. %86,12’lik bir doğruluk oranına (*accuracy*) ulaşılmıştır.

SentiTurkNet sözlüğü Türkçe dilinde geliştirilmiş bir duygu sözlüğüdür. Sözlükte 14.795 kelime bulunmaktadır. Bu sözlük kullanılarak geliştirilen modelin doğruluk oranı (*accuracy*) test edilmiştir. %87,7’lik bir doğruluk oranına (*accuracy*) ulaşılmıştır.

Sonuçlardan da görüldüğü üzere en yüksek doğruluk oranına (*accuracy*) SentiTürkNet sözlüğü ile ulaşılmıştır. Sözlükteki kelime sayısı diğer sözlüklere oranla daha az olmasına rağmen daha yüksek bir sonuç alınmıştır. Bunun en önemli nedeni sözlüğün Türkçe dilinde geliştirilmiş olmasıdır.

SentiWordNet-TR sözlüğünde kelime sayısı daha fazla olmasına rağmen, sözlük Türkçe diline çevrilirken duygu kayıpları yaşandığından dolayı %84,1’lik bir doğruluk oranına (*accuracy*) erişebilmiştir.

SentiWordNet-EN sözlüğü Türkçeye çevrilmeden, hedefi niteleyen ilişkili kelimelerin İngilizceye çevrilerek sözlüğe gönderilmesi ve puanlanması senaryosunu içermektedir. Bu senaryoda ise %2’lik bir artış olmuş ve %86,12’lik bir doğruluk oranı (*accuracy*) elde edilmiştir.

5.2. SÖZLÜK TABANLI YAKLAŞIM VE MAKİNE ÖĞRENMESİ TEKNİKLERİNİN KARŞILAŞTIRILMASI

Sözlük tabanlı yaklaşım ve makine öğrenmesi teknikleri ile yapılan hedef tabanlı duygu analizi çalışmasının güçlü ve zayıf yanları bir swot analizi ile ifade edilmiştir.

Şekil 27’de makine öğrenmesi yaklaşıma ait swot analizi tablosu yer almaktadır. Tabloda modelin, güçlü yanları, zayıf yanları, fırsatlar ve tehditler olmak üzere dört farklı başlık yer almaktadır.

Tablo 27’de görüldüğü üzere, doğruluk oranlarının (*accuracy*) yüksek olması, hızlı, performanslı olması ve klasik programlama tekniklerinin kısıtlı yapısı içerdiği handikapları

içermemesi makine öğrenmesi yaklaşımlarının güçlü yanları arasındadır. Bununla birlikte teknik süreçlerin daha zor ve karmaşık olması, donanım yatırımı gerektirmesi, etiketli veri ihtiyacının olması, model için bir eğitim süreci içermesi ve geliştirilen modelin tek bir alan (domain) için çalışması makine öğrenmesi yaklaşımının zayıf yanları arasındadır.

Tablo 27: Makine Öğrenmesi Yaklaşımına Ait Swot Analizi Tablosu

Güçlü Yanları	Zayıf Yanları
1. Doğruluk oranlarının (<i>accuracy</i>) sözlük tabanlı yaklaşıma göre daha yüksek olması. 2. Daha hızlı ve yüksek performanslı olması. 3. Geleneksel programlama tekniklerinin içermiş olduğu problemleri içermemesi.	1. Teknik süreçlerin daha zor olması. 2. Etiketli veri ihtiyacının olması. 3. Model için bir eğitim sürecine ihtiyacı olması. 4. Geliştirilen modelin tek bir alan (domain) için çalışması.
Fırsatlar	Tehditler
1. Yeni bir alan olmasından dolayı model ve yaklaşım oluşturma konusunda fırsatlar (imkanlar) içermesi.	1. Model geliştirilirken çoğunlukla hazır kütüphanelerin kullanılmasından dolayı dış kaynak ve geliştiricilere bağımlı olması.

Makine öğrenmesi yaklaşımı, sözlük tabanlı yaklaşıma göre yeni bir yaklaşım olması sebebiyle yeni model ve yaklaşımlar oluşturma konusunda fırsatlar barındırmaktadır. Bununla birlikte model oluşturma sürecinde genellikle hazır kütüphanelerin kullanılması dış kaynak ve geliştiricilere olan bağımlılığı artırmaktadır.

Tablo 28’de sözlük tabanlı yaklaşıma ait swot analizi tablosu yer almaktadır. Tabloda modelin, güçlü yanları, zayıf yanları, fırsatlar ve tehditler olmak üzere dört farklı başlık yer almaktadır.

Tablo 28’de de görüldüğü üzere teknik süreçlerin nispeten kolay olması, geliştirilen model için bir eğitim sürecine gerek duyulmaması ve alandan (*domain*) bağımsız olması sözlük tabanlı

yaklaşımın güçlü yanları arasındadır. Bununla birlikte sözlüğün kapasitesine bağlı kalınması, yapılan çalışmalar neticesinde görüldüğü üzere doğruluk oranlarının makine öğrenmesi yaklaşımlarına göre daha düşük olması ve Türkçe duygu sözlüklerinin sayı ve kapasite bakımından yeterli olmaması sözlük tabanlı yaklaşımın zayıf yanları arasındadır.

Tablo 28: Sözlük Tabanlı Yaklaşımına Ait Swot Analizi Tablosu

Güçlü Yanları	Zayıf Yanları
1. Teknik süreçlerin nispeten kolay olması. 2. Model için bir eğitim sürecine gerek olmaması. 3. Alandan bağımsız olması.	1. Sözlüğün kabiliyeti ve kapasitesine bağlı kalınması. 2. Doğruluk oranlarının (<i>accuracy</i>) makine öğrenmesi yaklaşımına göre daha düşük olması. 3. Türkçe duygu sözlüklerinin sayı ve kapasite olarak diğer dillere göre yeterli olmaması.
Fırsatlar	Tehditler
1. Türkçe sözlüklerin az olması. Bu alanda çalışma ihtiyacı olduğunu göstermektedir.	1. Dilin yaşayan bir organizma olmasından dolayı sözlüklerin belirli bir zaman sonra güncelliklerini yitirebilmesi. 2. Sözlükleri geliştiren ekiplerin dağılması ve yeni versiyonların çıkarılmaması.

Türkçe sözlüklerin sayıca az ve kapasite olarak yeterli olmaması bu alanda çalışma ihtiyacı olduğunu göstermektedir. Bununla birlikte dilin yaşayan bir organizma olması sözlüklerin belli bir zaman sonra güncelliklerini yitirebilmesine neden olabilmektedir. Ayrıca sözlükleri geliştiren ekiplerin dağılması ve yeni versiyonların çıkarılmaması gibi riskler barındırmaktadır.

5.3. ÖNERİLER

Çalışma sonucunda hem yapılan çalışma hem de alan için birtakım önerilerde bulunulmuştur.

1. Çalışma kapsamında yorum metnindeki hedeflerin tespit edilmesi ve etiketlenmesi süreci insan gücü ile yapıldı. Bu sürecin otomatikleştirilmesine dair çalışmalar yapılmalıdır. Hedef tabanlı duygu analizi çalışmalarında hedeflerin otomatik tespit edilmesine dair çalışmalar yapılmalı, yaklaşım ve yöntemler geliştirilmelidir.
2. Yeni bir yaklaşım olan hedef tabanlı duygu analizi çalışması için yapay sinir ağları dışında kalan diğer makine öğrenmesi teknikleri de kullanılabilir, doğruluk oranları kıyaslanabilir.
3. Yorum veri seti kapasite olarak artırılmalı, nispeten daha kirli veriler üzerinde de model test edilmelidir.
4. Türkçe dilinde duygu ve kelime sözlükleri geliştirilmelidir. Yabancı dillerden çevrilen sözlüklerde duygu ve anlam kaybı yaşanabilmektedir.
5. Akademik ve ticari çalışmalarda kullanılmak üzere Türkçe etiketli veri setleri oluşturulmalıdır. Google'ın "Captcha" sistemi gibi (İnternet sitelerinde form girişi yapılırken, insan-robot tespiti sürecinde kullanılan görsel seçme-etiketleme sistemi) kazan-kazan sistemi ile çalışan sistemler geliştirerek etiketli veri setleri oluşturma süreci otomatize edilmelidir.
6. Türkçe DDİ süreçlerinde yer alan metin ön işleme süreçleri için kullanılan Zemberek ve İTÜ DDİYZ gibi araçlar geliştirilmeli ve desteklenmelidir.
7. Model oluşturma sürecinde kullanılan yapay zekâ kütüphanelerinin yerelde geliştirilmesi üzerine çalışmalar yapılmalıdır.

KAYNAKLAR

- Adalı, E. (2012). Doğal dil işleme. *Türkiye Bilişim Vakfı Bilgisayar Bilimleri ve Mühendisliği Dergisi*, 5.
- Agarwal, B., and Mittal, N. (2016). *Prominent feature extraction for sentiment analysis*. Berlin: Springer International Publishing, 1-115.
- Agrawal, R., and Srikant, R. (1995, March). Mining sequential patterns. *In Proceedings of the eleventh international conference on data engineering*. IEEE, 3-14.
- Aiken, M., and Balan, S. (2011). An analysis of Google Translate accuracy. *Translation journal*, 16(2), 1-3.
- Akın, A. A., and Akın, M. D. (2007). Zemberek, an open source nlp framework for Turkic languages. *Structure*, 10(2007), 1-5.
- Al-Bnd, A., and Salman, M. (2015). *Sentiment analysis and opinion mining via microblogging in social media like: Twitter*, Doctoral dissertation, Graduate School of Natural and Applied Sciences.
- Antonic, S., and Krstev, C. (2008). Serbian Wordnet for biomedical sciences. In *INFORUM*, 28-30.
- Atkinson, A., and Riani, M. (2000). *Robust diagnostic regression analysis*. Springer Science and Business Media, 17-23.
- Ayata, D., Saraçlar, M., and Özgür, A. (2017, May). Turkish tweet sentiment analysis with word embedding and machine learning. In *2017 25th Signal Processing and Communications Applications Conference (SIU)* (pp. 1-4). IEEE.
- Ay Karakuş, B., Talo, M., Hallaç, İ. R., and Aydın, G. (2018). Evaluating deep learning models for sentiment classification. *Concurrency and Computation: Practice and Experience*, 30(21), e4783.
- Aytekin, Ç., ve Sütçü, C. S. (2012). Yeni medya ve ağ nesli: Türkiye örneği. *Marmara Sosyal Araştırmalar Dergisi*, 1-13.
- Ayetkin, Ç., Sütçü, C. S., ve Özfıdan, U. (2018). Kara Ağacı Algoritması ile Metin Sınıflandırma: Müşteri Yorumları Örneği. *Journal of International Social Research*, 55.
- Ayvaz, S., and Shiha, M. O. (2018, August). A scalable streaming big data architecture for real-time sentiment analysis. In *Proceedings of the 2018 2nd International Conference on Cloud and Big Data Computing*, 47-51.
- B. Liu, "Sentiment analysis and subjectivity", In *Handbook of Natural Language Processing*, Cilt 2, Chapman and Hall/CRC, 2010. 627-666

- Bakay, Ö., Ergelen, Ö., Sarmış, E., Yıldırım, S., Arıcan, B. N., Kocabalcıoğlu, A., ... and Yıldız, O. T. (2021, January). Turkish wordnet kenet. In *Proceedings of the 11th global wordnet conference*, 166-174.
- Barbosa, L., and Feng, J. (2010, August). Robust sentiment detection on twitter from biased and noisy data. In *Coling 2010: Posters*, 36-44.
- Bergamaschi, S., and Leiba, B. (2010). Information overload. *IEEE Internet Computing*, 1-13.
- Bilgin, O., Çetinoğlu, Ö., and Oflazer, K. (2004). Building a wordnet for Turkish. *Romanian Journal of Information Science and Technology*, 7(1-2), 163-172.
- Bird, S. (2006, July). NLTK: the natural language toolkit. In *Proceedings of the COLING/ACL 2006 Interactive Presentation Sessions*, 69-72.
- Brill, E. (1992). A simple rule-based part of speech tagger. In: *Proc. 3rd Conference on Applied Natural Language Processing (ANLC)*, Trento, Italy, 152-155.
- Bose, R., Aithal, P. S., and Roy, S. (2020). Sentiment analysis on the basis of Tweeter comments of application of drugs by customary language toolkit and textblob opinions of distinct countries. *International Journal*, 8.
- Burcu, A., ve Şimşek, U. (2018). Sosyal medya analitiği ile değer yaratma: Duygu analizi ile geleceğe yönelim. *Mehmet Akif Ersoy Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi*, 5(3), 797-811.
- Cambria, E., Li, Y., Xing, F. Z., Poria, S., and Kwok, K. (2020, October). SenticNet 6: Ensemble application of symbolic and subsymbolic AI for sentiment analysis. In *Proceedings of the 29th ACM international conference on information ve knowledge management* 105-114.
- Cambria, E., and White, B. (2014). Jumping NLP curves: A review of natural language processing research. *IEEE Computational Intelligence Magazine*, 9(2), 48-57.
- Can, U., and Alatas, B. (2017). Duygu analizi ve fikir madenciliği algoritmalarının incelenmesi. *International Journal of Pure and Applied Sciences*, 3(1), 75-111.
- Catal, C., and Nangir, M. (2017). A sentiment classification model based on multiple classifiers. *Applied Soft Computing*, 50, 135-141.
- Cetin, F. S., ve Eryiğit, G. (2018). Türkçe hedef tabanlı duygu analizi için alt görevlerin incelenmesi-hedef terim, hedef kategori ve duygu sınıfı belirleme. *Bilişim Teknolojileri Dergisi*, 11(1), 43-56.
- Chomsky, N. (1986). *Knowledge of language: Its nature, origin, and use*. Greenwood Publishing Group.
- Çankırı, S., Kartal, E., Yıldırım, K., ve Gülseçen, S. (2011). *Organizasyonlarda bilgi yönetimi sürecinde veri madenciliği yaklaşımı*. ÜNAK 2009, 148.

- Çetinoğlu, Ö., Bilgin, O., and Oflazer, K. (2018). Turkish Wordnet. In *Turkish Natural Language Processing*, Springer, Cham, 317-336.
- Davenport, T. H., and Prusak, L. (1998). *Working knowledge: How organizations manage what they know*. Harvard Business Press.
- Dehkharghani, R., Saygin, Y., Yanikoglu, B., and Oflazer, K. (2016). SentiTurkNet: a Turkish polarity lexicon for sentiment analysis. *Language Resources and Evaluation*, 50(3), pp. 667-685.
- Delibas, A. (2008). *Doğal dil işleme ile Türkçe yazım hatalarının denetlenmesi*, Doktora Tezi, İ.T.Ü. Fen Bilimleri Enstitüsü.
- Denecke, K. (2008, April). Using sentiwordnet for multilingual sentiment analysis. In *2008 IEEE 24th international conference on data engineering workshop*, IEEE, 507-512.
- Doğan, K., ve Arslantekin, S. (2016). Büyük veri: önemi, yapısı ve günümüzdeki durum. *Ankara Üniversitesi Dil ve Tarih-Coğrafya Fakültesi Dergisi*, 56.
- Egrioglu, E., Aladag, C. H., Yolcu, U., Uslu, V. R., and Basaran, M. A. (2009). A new approach based on artificial neural networks for high order multivariate fuzzy time series. *Expert Systems with Applications*, 36(7), 10589-10594.
- Elmas, Ç. (2003). *Yapay sinir ağları (kuram, mimari, eğitim, uygulama)*. Ankara: Seçkin Yayıncılık.
- Ergezer, H., Dikmen, M., ve Özdemir, E. (2003). *Yapay sinir ağları ve tanıma sistemleri*. *PiVOLKA*, 2(6), 14-17.
- Eryiğit, G. (2014, April). ITU Turkish NLP web service. In *Proceedings of the Demonstrations at the 14th Conference of the European Chapter of the Association for Computational Linguistics* 1-4.
- Eryiğit, G., and Torunoğlu-Sfelamet, D. (2017). Social media text normalization for Turkish. *Natural Language Engineering*, 23(6), pp. 835-875.
- Eryiğit, G., Nivre, J., and Oflazer, K. (2008). Dependency parsing of Turkish. *Computational Linguistics*, 34(3), 357-389.
- Fan, W., Wallace, L., Rich, S., and Zhang, Z. (2006). Tapping the power of text mining. *Communications of the ACM*, 49(9), 76-82.
- Fersini, E., Messina, E., and Pozzi, F. A. (2014). Sentiment analysis: Bayesian ensemble learning. *Decision support systems*, pp. 26-38.
- Fersini, E., Messina, E., and Pozzi, F. A. (2016). Expressive signals in social media languages to improve polarity detection. *Information Processing and Management*, 52(1), 20-35.

- Finn, A., and Kushmerick, N. (2006). Learning to classify documents according to genre. *Journal of the American Society for Information Science and Technology*, 57(11), 1506-1518.
- Fyfe, C. (2000). *Artificial neural networks and information theory*. University of Paisley.
- Gahi, Y., Guennoun, M. and Mouftah, H. T. (2016, 27-30 June). Big Data analytics: Security and privacy challenges. *2016 IEEE Symposium on Computers and Communication (ISCC)*, Messina, Italy, 952-957.
- Gardner, M., Grus, J., Neumann, M., Tafjord, O., Dasigi, P., Liu, N., ... and Zettlemoyer, L. (2018). Allennlp: A deep semantic natural language processing platform. *arXiv preprint arXiv:1803.07640*.
- Ganesh, S. (2002). Data Mining: Should it be included in the 'Statistics' Curriculum. *In The Sixth International Conference on Teaching Statistics*, Cape Town, South Africa.
- Ghag, K., and Shah, K. (2013, January). Comparative analysis of the techniques for sentiment analysis. In *2013 International Conference on Advances in Technology and Engineering (ICATE)*, IEEE, 1-7.
- Go, A., Bhayani, R., and Huang, L. (2009). Twitter sentiment classification using distant supervision. *CS224N project report, Stanford*, 1(12), 2009, 12-15.
- Göker, H., ve Tekedere, H. (2017). FATİH projesine yönelik görüşlerin metin madenciliği yöntemleri ile otomatik değerlendirilmesi. *Bilişim Teknolojileri Dergisi*, 10(3), 291-299.
- Guo, L., and Wan, X. (2012). Exploiting syntactic and semantic relationships between terms for opinion retrieval. *Journal of the american society for information science and technology*, 63(11), 2269-2282.
- Gülseçen, S., Özdemir, Ş., Çelik, S., Uğraş, T., ve Özcan, M. (2014). Dijital dünyadan yansımalar: Bilgide ve vatandaşlıkta değişim. *XIX. Türkiye'de İnternet Konferansı*, 27-29.
- Gürbüz, S., ve Şahin, F. (2014). *Sosyal bilimlerde araştırma yöntemleri*. Ankara: Seçkin Yayıncılık.
- Gürsaka, N. (2014). Büyük Veri (Genişletilmiş 2. Baskı). *Bursa: Dora Yayınevi*.
- Han, J., and Fu, Y. (1999). Mining multiple-level association rules in large databases. *IEEE Transactions on knowledge and data engineering*, 11(5), 798-805.
- Hand, D. J. (1998). Data mining: statistics and more?. *The American Statistician*, 52(2), 112-118.
- Hark, C., Myo, A., Seyyarer, A., Myo, G., Uçkan, T., Myo, B., ve Karci, A. (2017, September). Doğal dil işleme yaklaşımları ile yapısal olmayan dökümanların benzerliği. In *2017 International Artificial Intelligence and Data Processing Symposium (IDAP)*, IEEE, 1-6.

- Haykin, S. O. (2008). *Neural Networks and Learning Machines (3rd edition.)*. New York: Pearson.
- Horák, A., and Smrž, P. (2004). VisDic–wordnet browsing and editing tool. In *Proceedings of the Second International WordNet Conference–GWC* 136-141.
- Horák, A., Pala, K., Rambousek, A., and Povolný, M. (2006). DEBVisDic-first version of new client-server Wordnet browsing and editing tool. In *Proceedings of the Third International WordNet Conference-GWC*, 325-328.
- İlhan, S., Duru, N., Karagöz, Ş., ve Sağır, M. (2008). Metin madenciliği ile soru cevaplama sistemi. *Elektronik ve Bilgisayar Mühendisliği Sempozyumu (ELECO)*, Bursa, 26-30.
- İTÜ DDİYZ, Erişim Adresi: <http://tools.nlp.itu.edu.tr/>, [Ziyaret tarihi: 6 Eylül 2021].
- Jacobs, P. (1999). Data mining: What general managers need to know. *Harvard Management Update*, 4(10), 8.
- Jensen, P. E. (2005). A contextual theory of learning and the learning organization. *Knowledge and Process Management*, 12(1), 53-64.
- Kartal, E., ve Özen, Z. (2017). *Dengesiz veri setlerinde sınıflandırma. Mühendislikte yapay zeka ve uygulamaları*, Sakarya Üniversitesi Kütüphanesi Yayınevi, 109–131.
- Karaöz, B., ve Gürsoy, U. T. (2018). Adaptif öğrenme sözlüğü temelli duygu analiz Algoritması önerisi. *Bilişim Teknolojileri Dergisi*, 11(3), 245-253.
- Katz, G., Ofek, N., and Shapira, B. (2015). ConSent: Context-based sentiment analysis. *Knowledge-Based Systems*, 84, 162-178.
- Kavzoğlu, T., ve Çölkesen, İ. (2010). Karar ağaçları ile uydu görüntülerinin sınıflandırılması: Kocaeli örneği. *Harita Teknolojileri Elektronik Dergisi*, 2(1), 36-45.
- Kearney, C., and Liu, S. (2014). Textual sentiment in finance: A survey of methods and models. *International Review of Financial Analysis*, 33, 171-185.
- Khurana, Diksha, et al. (2017). "Natural language processing: State of the art, current trends and challenges." *arXiv preprint arXiv:1708.05148*.
- Kim, M., Zimmermann, T., DeLine, R., and Begel, A. (2017). Data scientists in software teams: State of the art and challenges. *IEEE Transactions on Software Engineering*, 44(11), 1024-1038.
- Koyuncugil, A. S. (2007). Veri madenciliği ve sermaye piyasalarına uygulanması. *Sermaye piyasası kurulu araştırma raporu*, 1-17.
- Linguistic knowledge in natural language processing. Erişim adresi: <https://towardsdatascience.com/linguistic-knowledge-in-natural-language-processing-332630f43cel> [Ziyaret tarihi: 26 Ağustos 2018].

- Liu, B. (2010). Sentiment analysis and subjectivity. *Handbook of natural language processing*, 2(2010), 627-666.
- Liu, B. (2012). Sentiment analysis and opinion mining. *Synthesis lectures on human language technologies*, 5(1), 1-167.
- Manning, C. D., Surdeanu, M., Bauer, J., Finkel, J. R., Bethard, S., and McClosky, D. (2014, June). The Stanford CoreNLP natural language processing toolkit. In *Proceedings of 52nd annual meeting of the association for computational linguistics: system demonstrations* 55-60.
- Matsumoto, S., Takamura, H., and Okumura, M. (2005, May). Sentiment classification using word sub-sequences and dependency sub-trees. In *Pacific-Asia conference on knowledge discovery and data mining*. Springer, Berlin, Heidelberg, 301-311.
- Medhat, W., Hassan, A., and Korashy, H. (2014). Sentiment analysis algorithms and applications: A survey. *Ain Shams engineering journal*, 5(4), 1093-1113.
- Meena, A., and Prabhakar, T. V. (2007). Sentence level sentiment analysis in the presence of conjuncts using linguistic analysis. In *European conference on information retrieval*, Springer, Berlin, Heidelberg, 573-580.
- Miller, G. A. (1995). WordNet: a lexical database for English. *Communications of the ACM*, 38(11), 39-41.
- Miller, G. A., Beckwith, R., Fellbaum, C., Gross, D., and Miller, K. J. (1990). Introduction to WordNet: An on-line lexical database. *International journal of lexicography*, 3(4), 235-244.
- Nassar, A., & Sezer, E. (2019, March). Multilevel Sentiment Analysis In Arabic. In 2019 IEEE 7th Palestinian International Conference on Electrical and Computer Engineering (PICECE), IEEE, 1-10.
- Nikfarjam, A., Sarker, A., O'connor, K., Ginn, R., and Gonzalez, G. (2015). Pharmacovigilance from social media: mining adverse drug reaction mentions using sequence labeling with word embedding cluster features. *Journal of the American Medical Informatics Association*, 22(3), pp. 671-681.
- Nizam, H., ve Akın, S. S. (2014). Sosyal medyada makine öğrenmesi ile duygu analizinde dengeli ve dengesiz veri setlerinin performanslarının karşılaştırılması. *XIX. Türkiye'de İnternet Konferansı*, 1-6.
- Nystrom, N., Clarkson, M. R., and Myers, A. C. (2003, April). Polyglot: An extensible compiler framework for Java. In *International Conference on Compiler Construction*, Springer, Berlin, Heidelberg, 138-152.
- Oflazer, K. (2016). Türkçe ve doğal dil işleme. *Türkiye Bilişim Vakfı Bilgisayar Bilimleri ve Mühendisliği Dergisi*, 5(2).

- Ohsaki, M., Wang, P., Matsuda, K., Katagiri, S., Watanabe, H., and Ralescu, A. (2017). Confusion-matrix-based kernel logistic regression for imbalanced data classification. *IEEE Transactions on Knowledge and Data Engineering*, 29(9), 1806-1819.
- Onan, A., and Korukoğlu, S. (2017). A feature selection model based on genetic rank aggregation for text sentiment classification. *Journal of Information Science*, 43(1), 25-38.
- Ott, M., Cardie, C., and Hancock, J. (2012, April). Estimating the prevalence of deception in online review communities. In *Proceedings of the 21st international conference on World Wide Web* 201-210.
- Ozyurt, B., and Akcayol, M. A. (2021). A new topic modeling based approach for aspect extraction in aspect based sentiment analysis: SS-LDA. *Expert Systems with Applications*, 168, 114231.
- Özçelik, M., Arıcan, B. N., Bakay, Ö., Sarıış, E., Ergelen, Ö., Bayezit, N. G., and Yıldız, O. T. (2021, January). HisNet: A Polarity Lexicon based on WordNet for Emotion Analysis. In *Proceedings of the 11th Global Wordnet Conference* 157-165.
- Pang, B., and Lee, L. (2005). "Seeing stars: Exploiting class relationships for sentiment categorization with respect to rating scales. *arXiv preprint cs/0506075*, 115-124.
- Pang, B., Lee, L., and Vaithyanathan, S. (2002). Thumbs up? Sentiment classification using machine learning techniques. *arXiv preprint cs/0205070*.
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., ... and Duchesnay, E. (2011). Scikit-learn: Machine learning in Python. *the Journal of machine Learning research*, 12, 2825-2830.
- Pennacchiotti, M., and Popescu, A. M. (2011). A machine learning approach to twitter user classification. In *Fifth international AAAI conference on weblogs and social media*, 281-287.
- Pervan, N., ve Keleş, Y. (2019). *Derin öğrenme yaklaşımları kullanarak Türkçe metinlerden anlamsal çıkarım yapma*, Yüksek Lisans Tezi, Ankara Üniversitesi Fen Bilimleri Enstitüsü.
- Pontiki, M., Galanis, D., Papageorgiou, H., Androutsopoulos, I., Manandhar, S., Al-Smadi, M., ... and Eryiğit, G. (2016). Semeval-2016 task 5: Aspect based sentiment analysis. In *International workshop on semantic evaluation*, 19-30.
- Pozzi, F. A., Fersini, E., Messina, E., and Liu, B. (2016). *Sentiment analysis in social networks*. Morgan Kaufmann.
- Prytherch, Ray. (2005). Harrod's Librarians' glossary and reference book : A dictionary of over 10,200 terms. 10. Hampshire: Ashgate Publishing Limited.

- Puri, R., and Catanzaro, B. (2019). Zero-shot text classification with generative language models. *arXiv preprint arXiv:1912.10165*.
- Reyes, A., and Rosso, P. (2012). Making objective decisions from subjective data: Detecting irony in customer reviews. *Decision support systems*, 53(4), 754-760.
- Sahin, M., Sulubacak, U., and Eryigit, G. (2013). Redefinition of Turkish morphology using flag diacritics. In *Proceedings of The Tenth Symposium on Natural Language Processing (SNLP-2013)*, Phuket, Thailand.
- Sagayam, R., Srinivasan, S. ve Roshni, S. (2012). A survey of text mining: Retrieval, extraction and indexing techniques. *International Journal of Computational Engineering Research*, 2(5), 1443.
- Saleh, M. R., Martín-Valdivia, M. T., Montejo-Ráez, A., and Ureña-López, L. A. (2011). Experiments with SVM to classify opinions in different domains. *Expert Systems with Applications*, 38(12), 14799-14804.
- Saquete, E., Tomás, D., Moreda, P., Martínez-Barco, P., and Palomar, M. (2020). Fighting post-truth using natural language processing: A review and open challenges. *Expert systems with applications*, 141, 112943.
- Schmitt, X., Kubler, S., Robert, J., Papadakis, M., and LeTraon, Y. (2019). A replicable comparison study of NER software: StanfordNLP, NLTK, OpenNLP, SpaCy, Gate. In *2019 Sixth International Conference on Social Networks Analysis, Management and Security (SNAMS)*, IEEE, 338-343.
- Schönberger, V. M., ve Cukier, K. (2013). *Büyük Veri - yaşama, çalışma ve düşünme şeklimizi dönüştürecek bir devrim*. Çeviren: B. Erol, Paloma Yayınları, İstanbul.
- Schütze, H., and Manning, C. D. (1999). *Foundation of statistical natural language processing*.
- Sebastiani, F., and Esuli, A. (2006). Sentiwordnet: A publicly available lexical resource for opinion mining. In *Proceedings of the 5th International Conference on Language Resources and Evaluation*, 417-422.
- Seker, G. A., and Eryigit, G. (2017). Extending a CRF-based named entity recognition model for Turkish well formed text and user generated content. *Semantic Web*, 8(5), 625-642.
- Şeker, Ş. E. , (2013). *İş zekası ve veri madenciliği* . İstanbul: Cinius Yayınları.
- Serrano-Guerrero, J., Olivas, J. A., Romero, F. P., and Herrera-Viedma, E. (2015). Sentiment analysis: A review and comparative analysis of web services. *Information Sciences*, 311, 18-38.
- Sevindi, B. İ. (2013). *Türkçe metinlerde denetimli ve sözlük tabanlı duygu analizi yaklaşımlarının karşılaştırılması*, Yüksek Lisans Tezi, Gazi Üniversitesi Fen Bilimleri Enstitüsü.
- Silahtaroglu, G. (2008). *Veri madenciliği*. Papatya Yayınları, İstanbul.

- Sokolova, M., Japkowicz, N., and Szpakowicz, S. (2006, December). Beyond accuracy, F-score and ROC: a family of discriminant measures for performance evaluation. In *Australasian joint conference on artificial intelligence* (pp. 1015-1021). Springer, Berlin, Heidelberg.
- Song, H. A., and Lee, S. Y. (2013). Hierarchical representation using NMF. In *International conference on neural information processing*, Springer, Berlin, Heidelberg, 466-473.
- Stamou, S., Nenadic, G., and Christodoulakis, D. (2004). Exploring Balkanet Shared Ontology for Multilingual Conceptual Indexing. In *LREC*.
- Strawn, G. (2016). Data scientist. *IT Professional*, 18(3), 55-57.
- Strother, J. B., and Ulijn, J. M. (2012). The challenge of information overload. Professional Communication Conference (IPCC), IEEE International, IEEE, 1-3.
- Suat, A. T. A. N., ve ÇINAR, Y. (2019). Borsa İstanbul'da finansal haberler ile piyasa değeri ilişkisinin metin madenciliği ve duygu (sentiment) analizi ile incelenmesi. *Ankara Üniversitesi SBF Dergisi*, 74(1), pp. 1-34.
- Taboada, M., Brooke, J., Tofiloski, M., Voll, K., and Stede, M. (2011). Lexicon-based methods for sentiment analysis. *Computational linguistics*, 37(2), 267-307.
- Tang, D., Qin, B., Feng, X., and Liu, T. (2015). Effective LSTMs for target-dependent sentiment classification. *arXiv preprint arXiv:1512.01100*.
- Tetsuya, N. (2003). Jeonghee Yi. Sentiment analysis: capturing favorability using natural language processing. In *Proceedings of the 2nd International Conference on Knowledge Capture, America*.
- Torunoğlu-Selamet, D., and Eryiğit, G. (2014). A cascaded approach for social media text normalization of Turkish. In *Proceedings of the 5th Workshop on Language Analysis for Social Media (LASM)*, 62-70.
- Tunalı, V. (2011). Metin madenciliği için iyileştirilmiş bir kümeleme yapısının tasarımı ve uygulaması. Doktora Tezi, Marmara Üniversitesi Fen Bilimleri Enstitüsü.
- Türkmenoglu, C., & Tantug, A. C. (2014). Sentiment analysis in Turkish media. In *International Conference on Machine Learning (ICML)*.
- Uçan, A. (2014). Otomatik duygu sözlüğü çevirimi ve duygu analizinde kullanımı. Yüksek Lisans Tezi, Hacettepe Üniversitesi Fen Bilimleri Enstitüsü.
- Wang, D., Zhu, S., and Li, T. (2013). SumView: A Web-based engine for summarizing product reviews and customer opinions. *Expert Systems with Applications*, 40(1), 27-33.
- Weiss, S. M., Indurkha, N., Zhang, T., and Damerau, F. (2010). *Text mining: predictive methods for analyzing unstructured information*. Springer Science and Business Media.

- Wilson, T., Wiebe, J., and Hoffmann, P. (2005). Recognizing contextual polarity in phrase-level sentiment analysis. In *Proceedings of human language technology conference and conference on empirical methods in natural language processing*, 347-354.
- Xia, R., Zong, C., and Li, S. (2011). Ensemble of feature sets and classification algorithms for sentiment classification. *Information sciences*, 181(6), 1138-1152.
- Yang, L., Li, Y., Wang, J., and Sherratt, R. S. (2020). Sentiment analysis for E-commerce product reviews in Chinese based on sentiment lexicon and deep learning. *IEEE Access*, 8, 23522-23530.
- Yengi, Y., and Omurca, S. İ. (2016). Distributed recommender systems with sentiment analysis. *Avrupa Bilim ve Teknoloji Dergisi*, 4(7), 51-57.
- Yeom, S., Giacomelli, I., Fredrikson, M., and Jha, S. (2018, July). Privacy risk in machine learning: Analyzing the connection to overfitting. In *2018 IEEE 31st Computer Security Foundations Symposium (CSF)*, IEEE, 268-282.
- Yılmaz, İ. G., Aygün, D., and Tanrikulu, Z. (2017). Social media's perspective on industry 4.0: A Twitter analysis. *Social Networking*, 6(04), 251.
- Yılmaz, M. (2017). Enformasyon ve bilgi kavramları bağlamında enformasyon yönetimi ve bilgi yönetimi. *Ankara Üniversitesi Dil ve Tarih-Coğrafya Fakültesi Dergisi*, 49(1).
- Young, T., Hazarika, D., Poria, S., and Cambria, E. (2018). Recent trends in deep learning based natural language processing. *IEEE Computational intelligence magazine*, 13(3), 55-75.
- Yu, L. C., Wu, J. L., Chang, P. C., and Chu, H. S. (2013). Using a contextual entropy model to expand emotion words and their intensity for the sentiment classification of stock market news. *Knowledge-Based Systems*, 41, 89-97.
- Zhao, Y., and Zhang, Y. (2008). Comparison of decision tree methods for finding active objects. *Advances in Space Research*, 41(12), 1955-1959.

EKLER

Ek 1: İTÜ DDİYZ Servisi İçin İmzalanan Kullanım Sözleşmesi

05/09/2018

İTÜ Turkish Language Processing API Non-transferable License Agreement

By requesting access to use the İTÜ Turkish Language Processing API services¹ ("*Services*"), you agree to be bound by this license agreement ("*Agreement*"), and acknowledge that you have read and understood the terms of the *Agreement*. If you do not agree to the terms of the *Agreement*, your request will not be considered and you may not use any of the *Services*.

- Your license to use the *Services* lasts only as long as you conform to this agreement. We reserve the right to revoke your license without prior notice at any time we deem that the agreement is violated.
- You may only use the *Services* for non-commercial, academic, and/or research purposes. You may not use the *Services* or distribute any output thereof for commercial purposes.
- You may not copy, reproduce, republish, make available to the public, create derivative works from, or otherwise use our *Website* content in any way except for your own personal, non-commercial use.
- On the acceptance of your application, you will be given a username, a password and an API access token. These are exclusively for your own personal use of the *Services*, and you may not share them with, or transfer them to, third parties.
- Our *Services* are currently in beta testing stage and we frequently apply updates on our server. We are not responsible for any inconvenience caused by unscheduled downtimes, unannounced changes in the backend, or any software bugs there may be on our *Services*, nor indemnifying you for any damage caused thereby.

¹ Including the tokenization, deaccification, vocalization, spelling correction, normalization, language validation, morphological analysis, morphological disambiguation and syntactic analysis services as demonstrated on <http://tools.nlp.itu.edu.tr> ("*Website*").

Ek 2: Makine Öğrenmesi Modeli İçin Geliştirilen Yapıya Ait Kod Blogu

```

1  #Kütüphaneler Ekleniyor.
2  import numpy as np
3  import pandas as pd
4  import json
5  from tensorflow.python.keras.models import Sequential
6  from tensorflow.python.keras.layers import Dense, GRU, Embedding, CuDNNGRU
7  from tensorflow.keras.optimizers import Adam
8  from sklearn.utils import shuffle
9  from sklearn import metrics
10 from sklearn.metrics import classification_report
11
12 #Veri Seti Ekleniyor
13 dataframe = pd.read_csv('veri.csv', sep = ";")
14 dataframe = shuffle(dataframe)
15 target = dataframe['durum'].values.tolist()
16 data = dataframe['iliskili_kelimeler'].values.tolist()
17
18 #Veri Seti Eğitim ve Test Amaçlı Bölünüyor.
19 cutoff = int(len(data) * 0.70)
20 x_train, x_test = data[:cutoff], data[cutoff:]
21 y_train, y_test = target[:cutoff], target[cutoff:]
22
23 #Oluşturulan Tokenizer Dosyası Yükleniyor
24 with open('tokenizer.json') as json_dosyasi:
25     json_tokenizer = json.load(json_dosyasi)
26
27 #Veriler Tokenleştiriliyor.
28 def tokenlestir(yorumListesi):
29     y_yorumlar = []
30     for yorum in yorumListesi:
31         y_yorum = []
32         for kelime in yorum.split():
33             if (len(y_yorum) < 20 and kelime in json_tokenizer):
34                 y_yorum.append(json_tokenizer[kelime])
35
36         if (len(y_yorum) < 20):
37             sifirlar = list(np.zeros(20 - len(y_yorum), dtype = int))
38             y_yorum = sifirlar + y_yorum
39
40         y_yorumlar.append(y_yorum)
41     return np.array(y_yorumlar, dtype = np.dtype(np.int32))
42 egitim_kume = tokenlestir(x_train)
43 test_kume = tokenlestir(x_test)

```

```

#Model Oluşturuluyor.
model = Sequential()
embedding_size = 20

model.add(Embedding(input_dim=10000,
                    output_dim=embedding_size,
                    input_length=20,
                    name='embedding_layer'))

model.add(GRU(units=12, return_sequences=True))
model.add(GRU(units=6, return_sequences=True))
model.add(GRU(units=3))
model.add(Dense(1, activation='sigmoid'))

optimizer = Adam(lr=1e-3)

model.compile(loss='binary_crossentropy',
              optimizer=optimizer,
              metrics=['accuracy'])

model.summary()

#Model Eğitiliyor
model.fit(egitim_kume, y_train, epochs = 5, batch_size = 10)

#Model Test Ediliyor.
model.evaluate(test_kume, y_test)
sonuc = model.predict(test_kume)

#Karar Matrisi Oluşturuluyor
metrics.confusion_matrix(y_test, predict)

#Sınıflama Raporu Oluşturuluyor.
target_names = ['class 0', 'class 1']
print(classification_report(y_test, predict, target_names=target_names))

```

Ek 3: Makine Öğrenmesi Modeli İçin Elde Edilen Doğruluk Oranlarına Ait Tüm Tablo

Katman Sayısı	Nöron Dizilimleri	Epocs = 5	Epocs = 10	Epocs = 15
2	4-2	loss: 0.1936 - acc: 0.9414	loss: 0.1655 - acc: 0.9552	loss: 0.1707 - acc: 0.9414
	6-3	loss: 0.1615 - acc: 0.9517	loss: 0.1629 - acc: 0.9521	loss: 0.1425 - acc: 0.9581
	8-4	loss: 0.1558 - acc: 0.9517	loss: 0.1580 - acc: 0.9552	loss: 0.1953 - acc: 0.9483
	10-5	loss: 0.1545 - acc: 0.9483	loss: 0.1666 - acc: 0.9483	loss: 0.1532 - acc: 0.9517
	12-6	loss: 0.1517 - acc: 0.9448	loss: 0.1531 - acc: 0.9517	loss: 0.1695 - acc: 0.9517
3	8-4-2	loss: 0.1776 - acc: 0.9517	loss: 0.1649 - acc: 0.9586	loss: 0.1815 - acc: 0.9310
	12-6-3	loss: 0.1631 - acc: 0.9517	loss: 0.1591 - acc: 0.9621	loss: 0.1613 - acc: 0.9619
	16-8-4	loss: 0.1620 - acc: 0.9483	loss: 0.1585 - acc: 0.9552	loss: 0.1642 - acc: 0.9345
	20-10-5	loss: 0.1534 - acc: 0.9483	loss: 0.1608 - acc: 0.9483	loss: 0.1609 - acc: 0.9448
	24-12-6	loss: 0.1501 - acc: 0.9586	loss: 0.1709 - acc: 0.9517	loss: 0.1736 - acc: 0.9414
4	16-8-4-2	loss: 0.2507 - acc: 0.9414	loss: 0.1603 - acc: 0.9448	loss: 0.1674 - acc: 0.9517
	24-12-6-3	loss: 0.1930 - acc: 0.9448	loss: 0.1547 - acc: 0.9517	loss: 0.1792 - acc: 0.9448
	32-16-8-4	loss: 0.2403 - acc: 0.9414	loss: 0.1549 - acc: 0.9517	loss: 0.1748 - acc: 0.9414
	40-20-10-5	loss: 0.1541 - acc: 0.9483	loss: 0.1821 - acc: 0.9483	loss: 0.1674 - acc: 0.9483
	48-24-12-6	loss: 0.1516 - acc: 0.9517	loss: 0.1616 - acc: 0.9517	loss: 0.1665 - acc: 0.9310
5	32-16-8-4-2	loss: 0.2507 - acc: 0.9414	loss: 0.2418 - acc: 0.9414	loss: 0.2392 - acc: 0.9414
	48-24-12-6-3	loss: 0.2460 - acc: 0.9414	loss: 0.2385 - acc: 0.9414	loss: 0.2384 - acc: 0.9414
	64-32-16-8-4	loss: 0.2432 - acc: 0.9414	loss: 0.2378 - acc: 0.9414	loss: 0.2390 - acc: 0.9414
	80-40-20-10-5	loss: 0.2421 - acc: 0.9414	loss: 0.2271 - acc: 0.9414	loss: 0.2382 - acc: 0.9414
	96-48-24-12-6	loss: 0.2384 - acc: 0.9414	loss: 0.2381 - acc: 0.9414	loss: 0.2076 - acc: 0.9207

Ek 4: Sözlük Tabanlı Yaklaşım İçin Geliştirilen Yapıya Ait Kod Blogu

```

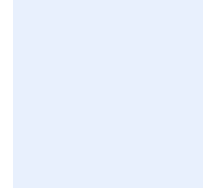
1  <?php
2      ini_set('display_errors', '1');
3      ini_set('display_startup_errors', '1');
4      error_reporting(E_ALL);
5      require 'vendor/autoload.php';
6
7      //Veritabanı bağlantısı yapılıyor.
8      ini_set('MAX_EXECUTION_TIME', '-1');
9      $db = new PDO("mysql:host=localhost;dbname=phd5;charset=utf8", "root", "");
10
11      //Kullanılacak değişkenler tanımlanıyor.
12      $sayac =0;
13
14      //İlgili tablodan tek tek veriler çekilerek puanlama süreci başlıyor.
15      $query = $db->query("SELECT agay.agay_id, agay.eksiz_iliskili_kelimeler_en,
16      *
17      agay.durum FROM aspete_gore_ayrilmis_yorumlar agay", PDO::FETCH_ASSOC);
18      if ( $query->rowCount() ){
19          foreach( $query as $row ){
20
21              //Eksiz ilişkili kelimeler veritabanından çekilip bir array haline
22              *
23              getiriliyor.
24              $agay_id = $row['agay_id'];
25              $eksizIliskiliKelimeler = explode
26              *
27              ("",trim($row['eksiz_iliskili_kelimeler_en']));
28
29              //Puanlama
30              $pozitif = 0;
31              $negatif = 0;
32              for ($i=0; $i < count($eksizIliskiliKelimeler) ; $i++) {
33                  $kelime = $eksizIliskiliKelimeler[$i].'#';
34                  $query = $db->query("SELECT AVG(es.pstv_score) as pztzf,
35                  *
36                  AVG(es.ngtv_score) as ngtf FROM en_source es where es.word like
37                  *
38                  '$kelime%'")->fetch(PDO::FETCH_ASSOC);
39                  if ($query) {
40                      $pozitif = $pozitif + $query['pztzf'];
41                      $negatif = $negatif+ $query['ngtf'];
42                  }
43              }
44          }
45      }

```

```
$pozitif = $pozitif/count($eksizIliskiliKelimeler);  
$negatif = $negatif/count($eksizIliskiliKelimeler);  
  
//Puana bakarak duygu durumuna karar veriliyor ve veritabanına  
kaydediliyor.  
if ($pozitif>$negatif) {  
    $durum = 'p';  
}else if ($pozitif<$negatif) {  
    $durum = 'n';  
}else {  
    $durum = 'o';  
}  
  
if (trim($row['durum']) == $durum) {  
    $sayac++;  
}  
  
}  
?  
?>
```

ÖZGEÇMİŞ

Kişisel Bilgiler	
Adı Soyadı	Harun AKSAYA
Doğum Yeri	
Doğum Tarihi	Tarih girmek için tıklayın veya dokununuz.
Uyruğu	☒ T.C. ☐ Diğer:
Telefon	
E-Posta Adresi	
Web Adresi	



Eğitim Bilgileri	
Lisans	
Üniversite	Marmara Üniversitesi
Fakülte	Atatürk Eğitim Fakültesi
Bölümü	Bilgisayar ve Öğretim Teknolojileri Eğitimi
Mezuniyet Yılı	2010

Yüksek Lisans	
Üniversite	Marmara Üniversitesi
Enstitü Adı	Sosyal Bilimler Enstitüsü
Anabilim Dalı	Bilişim Anabilim Dalı
Programı	Bilişim Programı

Doktora	
Üniversite	İstanbul Üniversitesi
Enstitü Adı	Fen Bilimleri Enstitüsü
Anabilim Dalı	Enformatik Anabilim Dalı
Programı	Enformatik Programı

Makale ve Bildiriler
Aksaya, H., and Gülseçen, S. (2021). <i>Comparison of Dictionary-Based Approach and Machine Learning Techniques in Aspect-Based Sentiment Analysis Studies: The Sample of Educational Institutions</i>. Uluslararası Yapay Zeka Uygulamaları Konferansı (ICIDAAI), İstanbul, Turkey.

- Aksaya, H., Gülseçen, S. (2021). Klinik Karar Destek Sistemleri. Bozbuğa, N., and Gülseçen, S. (Ed.). (2021). *Tıp Bilişimi*. İstanbul University Press, 219–316.
<https://doi.org/10.26650/B/ET07.2021.003>
- Aksaya, H., and Gülseçen, S. (2020). *Developing a Text Mining Based System for Educational Institutions: Implementation of Machine Learning Techniques in Aspect-Based Sentiment Analysis Studies*. 8th International Conference on Future Learning and Informatics: Data Revolution, İstanbul, Turkey.
- Aksaya, H., Gülseçen, S. (2020). Doğal Dil İşleme. Gülseçen, S., İnal, M., Torkuli O., Selvi, H., Erol, Ç., Çağıl, G., Reis, Z. Uçar, K.,(Ed.) *Mühendislikte Yapay Zeka ve Uygulamaları* 3. Sakarya Üniversitesi Yayınevi. 165–174.
- Aksaya, H., Arşık, A., Elibol, M., Olgun K.B., Özcan, M. and Ayvaz Reis, Z. (2015). *Hobim İçin Öğreniyorum: Akvaristler Ne Kadar Memnun*. In Proceedings of International Educational Technology Conference., İstanbul, Turkey.
- Benzer, A., ve Aksaya, H. (Ed.). (2012). *Bilgisayar Okuryazarlığı 1 - 2*. Pegem A Yayıncılık - Akademik Kitaplar.