

T.C.  
ONDOKUZ MAYIS ÜNİVERSİTESİ  
LİSANSÜSTÜ EĞİTİM ENSTİTÜSÜ  
İLETİŞİM BİLİMLERİ ANA BİLİM DALI  
MEDYA VE İLETİŞİM BİLİMLERİ PROGRAMI



TÜRKİYE'DEKİ SURİYELİ SİĞİNMACILAR ALGISI: TÜRKÇE  
TWITTER VERİLERİ İLE DUYGU ANALİZİ

Yüksek Lisans Tezi

İlhan PARLAK

Danışman  
Doç. Dr. Ömer ÇAKIN

SAMSUN  
2022

## BİLİMSEL ETİĞE UYGUNLUK BEYANI

Hazırladığım Yüksek Lisans tezinin bütün aşamalarında bilimsel etiğe ve akademik kurallara riayet ettiğimi, çalışmada doğrudan veya dolaylı olarak kullandığım her alıntıya kaynak gösterdiğimi ve yararlandığım eserlerin Kaynaklar'da gösterilenlerden oluştuğunu, her unsurun enstitü yazım kılavuzuna uygun yazıldığını ve TÜBİTAK Araştırma ve Yayın Etiği Kurulu Yönetmeliği'nin 3. bölüm 9. maddesinde belirtilen durumlara aykırı davranılmadığını taahhüt ve beyan ederim.

Etik Kurul Gerekli mi ?

Evet  (Gerekli ise ekler kısmına ekleyiniz)

Hayır

28 / 10 / 2022  
İlhan PARLAK

## TEZ ÇALIŞMASI ÖZGÜNLÜK RAPOR BEYANI

**Tez Başlığı :** TÜRKİYE'DEKİ SURİYELİ SİĞINMACILAR ALGISI:  
TÜRKÇE TWITTER VERİLERİ İLE DUYGU ANALİZİ

Yukarıda başlığı belirtilen tez çalışması için şahsım tarafından 11/10/2022 tarihinde intihal tespit programından alınmış olan özgünlük raporu sonucunda;

Benzerlik oranı : % 9

Tek kaynak oranı : % 1 çıkmıştır.

28 / 10 / 2022  
Doç. Dr. Ömer ÇAKIN

## ÖZET

### TÜRKİYE'DEKİ SURIYELİ SİĞINMACILAR ALGISI: TÜRKÇE TWITTER VERİLERİ İLE DUYGU ANALİZİ

İlhan PARLAK

Ondokuz Mayıs Üniversitesi

Lisansüstü Eğitim Enstitüsü

İletişim Bilimleri Ana Bilim Dalı

Medya ve İletişim Bilimleri Programı

Yüksek Lisans, Ekim/2022

Danışman: Doç. Dr. Ömer ÇAKIN

Bu yüksek lisans tezinde Türkiye’de yaşayan Suriyeli sığınmacıların Türk halkı nezdindeki genel algısı, algının niteliği ve sebepleri ortaya konmaya çalışılmıştır. Bu amaç için Twitter sosyal medya platformu kapsamında makine öğrenme tekniği ile gerçekleştirilen duygu analizi yöntemi kullanılmıştır. Bu kapsamda 2021 yılı boyunca Suriyeli sığınmacılar ile ilgili olarak atılan tweet’ler toplanmış, ön işlemden geçirilmiş ve veri madenciliği algoritmalarının uygulanabileceği hale dönüştürülmüştür. Tweet’lerin bir kısmı makine öğrenmesi tekniğine uygun olarak manuel olarak sınıflandırılmış ve geri kalan tweet’lerin -manuel olarak yapılan sınıflandırmalar dikkate alınarak - belirlenen algoritmalar tarafından otomatik olarak sınıflandırılması sağlanmıştır. Sonuç olarak, Twitter kapsamında Suriyeli sığınmacılar ile ilgili olarak Türk halkı nezdindeki genel algının büyük oranda olumsuz ve “hoşnutsuzluk” niteliğinde olduğu; bu algının başlıca sebeplerinin de Suriyeli sığınmacıların kendilerine verdiklerini düşündükleri zarar, sayıca fazla olmaları, sahip oldukları başta sığınma ve ikamet hakları ile kendilerinden daha iyi durumda olduklarına dair sahip olunan algılar olduğu tespit edilmiştir.

Tezin birinci bölümünde, üzerinde çalışılan konunun önemini ortaya koyma amacı ile Suriyeli sığınmacılar meselesi masaya yatırılmıştır. Bu kapsamda, Suriyeli sığınmacıların Türkiye’ye göç etme sebepleri, Türkiye’deki nüfusu, Türkiye’ye dağılımı, eğitim durumları ve okullaşma oranları, çalışma hayatları, ekonomiye etkileri ve yapılan yardımlar, kamu hizmetlerine ve güvenliğe etkileri ile Türkiye’ye uyumları incelenmiştir. İkinci bölümde, üçüncü bölümde gerçekleştirilecek olan duygu analizinin anlaşılabilirliğine olanak verecek şekilde duygu analizine dair tüm kavram, teknik ve uygulama süreci anlatılmıştır. Bu kapsamda büyük veri, veri madenciliği, metin madenciliği, duygu analizi yaklaşımları, makine öğrenmesi, sınıflandırma algoritmaları, Twitter ve duygu analizindeki rolü ile duygu analizi yönteminin kullanıldığı daha önceki çalışmalar incelenmiştir. Üçüncü bölümde de Türkiye’de yaşayan Suriyeli sığınmacılar hakkındaki algıyı, bu algının niteliğini ve sebeplerini ortaya koymak üzere Twitter kapsamında duygu analizi gerçekleştirilmiştir.

**Anahtar Sözcükler:** Duygu analizi, Suriyeli sığınmacılar, Makine öğrenmesi, Twitter

## ABSTRACT

### THE PERCEPTION OF THE SYRIAN ASYLUM SEEKERS IN TÜRKİYE: SENTIMENT ANALYSIS WITH TURKISH TWITTER DATA

İlhan PARLAK

Ondokuz Mayıs University

Institute of Graduate Studies

Department of Communication Sciences

Media and Communication Sciences Programme

Master, October/2022

Supervisor: Assoc. Prof. Dr. Ömer ÇAKIN

With this master thesis, it is studied to be present of the general perception, the quality of the perception and the reasons of the perception of Turkish people about the Syrian asylum seekers who live in Türkiye. For this purpose, sentiment analysis method which is performed with machine learning technique, is used within the scope of Twitter social media platform. Within this scope, tweets about Syrian asylum seekers which are sent 2021 yearlong, are gathered; pre-processed and transformed to be able to perform the data mining algorithms on them. In compliance with machine learning technique, some of the tweets are classified manually and it is provided that the other tweets - considering the classifications, done manually - are classified automatically. As a result, it is determined that the general perception of Turkish people within the scope of Twitter, about Syrian asylum seekers is negative and “displeasure” as a quality to a large extent and the reasons of this perception are mainly the perceptions of the damage which is done by Syrian asylum seekers to themselves, having a large population, the rights of asylum and habitation and in being a better situation compared to themselves, which they have.

In the first section of the thesis, to present the importance of the subject which is being work on, the Syrian asylum seekers issue examined. Within this scope, the reasons of the Syrian asylum seekers immigration to Türkiye, their population in Türkiye, distribution to Türkiye, education statuses and schooling rates, working lives, effects to economy, aids, effects to public services and security and adaptation to Türkiye are examined. In the second section, all the concepts, techniques and application process of a sentiment analysis are explained, in the manner that a sentiment analysis which will be applied in the third section is letting to understand. Within this scope, big data, data mining, text mining, the approaches of sentiment analysis, machine learning, classification algorithms, Twitter and its role in sentiment analysis and the previous studies which sentiment analysis method is being used in, are examined. In the third section, a sentiment analysis is applied, to present the perception about Syrian asylum seekers in Türkiye, the quality of this perception and the reasons of this perception, within the scope of Twitter.

**Keywords:** Sentiment analysis, Syrian asylum seekers, Machine learning, Twitter

## ÖNSÖZ VE TEŞEKKÜR

Sosyal bilimler alanında yeni kullanılmaya başlanan sayısal tekniklerden bahsedilen bir yüksek lisans dersinde varlığından haberdar olduğum duygu analizi, daha ilk duyduğumda ilgimi çekmeyi başarmıştı. Zamanla kamuya açık her türlü metin üzerinde çalışabileceğimi ve böylelikle gündemdeki herhangi bir konu hakkında kamuoyunun nabzını ölçebileceğimi keşfettiğim bu keyifli teknik, kısa bir sürede bilgisayarımın kodlarla dolmasına yol açmıştı. Geldiğim bu noktada, bu tez kapsamında daha önce herhangi bir duygu analizi çalışmasında rastlamadığım şekilde konuyu üç farklı boyutta incelemeye çalıştım. Klasik olarak bir duygu analizi, konu hakkında genel görüşün ne oranda olumlu, olumsuz ya da nötr olduğunu ortaya koyar. Bu tez kapsamında ben konuyu bu klasik halinin yanı sıra, duygunun niteliği ve sebepleri ile de birlikte ortaya koymaya çalıştım. Bunun için kullanılacak olan konu olarak da her geçen gün gündemde daha da fazla yer etmeye başlayan Suriyeli sığınmacılar konusunu seçtim.

Verinin dijital ortamda, özellikle de sosyal medyada gittikçe daha fazla artan oranda bulunmaya başladığı günümüzde, sayısal tekniklerin herhangi bir alanda kullanılmaya başlanmaması artık imkansız gibi gözükmektedir. Kaldı ki toplumların eğilimlerini ölçen sosyal bilimler için bunun çok daha fazla geçerli olduğu söylenebilir.

Öncelikle beni bugünlere getiren ve eğitimim için ne gerekiyorsa yapan anne ve babama, daha sonra tez süresince benden sabrını ve hoşgörüsünü esirgemeyen sevgili eşim Nur PARLAK'a, yüksek lisansa başladığım ilk günden bu yana benden hiçbir konuda desteğini esirgemeyen ve değerli tavsiyelerine mazhar olduğum danışman hocam Doç. Dr. Ömer ÇAKIN'a ve son olarak o ilk yüksek lisans dersinde benim bu yeni sayısal tekniklerden haberdar olmamı sağlayan Dr. Ahmet Faruk ÇEÇEN hocama teşekkürü bir borç bilirim.

İlhan PARLAK

# İÇİNDEKİLER

TEZ KABUL VE ONAYI .....	i
BİLİMSEL ETİĞE UYGUNLUK BEYANI .....	ii
TEZ ÇALIŞMASI ÖZGÜNLÜK RAPORU BEYANI .....	ii
ÖZET .....	iii
ABSTRACT .....	iv
ÖNSÖZ VE TEŞEKKÜR .....	v
İÇİNDEKİLER.....	vi
SİMGELER VE KISALTMALAR .....	viii
ŞEKİLLER DİZİNİ .....	ix
TABLolar DİZİNİ .....	x
<b>1. GİRİŞ .....</b>	<b>1</b>
<b>2. SURIYELİ SİĞİNMACILAR MESELESİ.....</b>	<b>3</b>
2.1. Arap Baharı .....	3
2.2. Türkiye'ye Göç Eden Suriyeli Sığınmacılar .....	4
2.2.1. Nüfus .....	4
2.2.2. Türkiye'ye Dağılım .....	7
2.2.3. Eğitim Durumu ve Okullaşma.....	9
2.2.4. Çalışma Hayatı .....	13
2.2.5. Ekonomik Etki ve Yardımlar .....	18
2.2.6. Kamu Hizmetlerine Etkisi .....	22
2.2.7. Güvenlik .....	24
2.2.8. Uyum .....	26
<b>3. DUYGU ANALİZİ – KAVRAMLAR, TEKNİKLER VE UYGULAMA SÜRECİ . 29</b>	
3.1. Duygu Analizi .....	29
3.2. Büyük Veri .....	30
3.2.1. Büyük Veri Bileşenleri .....	31
3.2.1.1. Çeşitlilik .....	32
3.2.1.2. Hız.....	32
3.2.1.3. Hacim .....	33
3.2.1.4. Gerçeklik .....	33
3.2.1.5. Değer .....	33
3.2.2. Büyük Veri Kaynakları .....	34
3.2.3. Büyük Veri Uygulama Alanları .....	35
3.3. Veri Madenciliği .....	38
3.3.1. Veri Madenciliği Uygulama Alanları .....	38
3.3.2. Veri Madenciliği Modelleri .....	39
3.3.3. Veri Madenciliği Süreci .....	42
3.4. Metin Madenciliği .....	44
3.4.1. Metin Madenciliği Uygulama Alanları .....	46
3.4.1.1. Bilgiye Erişim .....	46
3.4.1.2. Bilgi Çıkarımı .....	46
3.4.1.3. Doğal Dil İşleme .....	47
3.4.1.4. Belge Sınıflandırma .....	48
3.4.1.5. Belge Kümeleme .....	49
3.4.1.6. Web Madenciliği .....	49
3.4.1.7. Kavram Çıkarımı .....	49
3.4.2. Metin Madenciliği Aşamaları .....	50
3.4.2.1. Veri Setinin Oluşturulması .....	50
3.4.2.2. Metin Ön İşleme .....	51
3.4.2.2.1. Dizgeciklere (Token) Ayırma .....	51
3.4.2.2.2. Filtreleme .....	51
3.4.2.2.3. Öznitelikleri Çıkarma .....	52

3.4.2.2.3.1. Gövdeleme .....	52
3.4.2.2.3.2. N-Gram Yöntemi .....	53
3.4.2.2.4. Dil Bilimsel Ön İşleme .....	54
3.4.2.3. Özniteliklerin Ağırlıklandırılması ve Yapısal Hale Getirilmesi .....	55
3.4.2.4. Boyut Azaltma .....	57
3.4.2.4.1. Öznitelik Seçimi Yöntemi ile Boyut Azaltma .....	58
3.4.2.4.2. Öznitelik Çıkarımı Yöntemi ile Boyut Azaltma .....	59
3.4.2.5. Veri Madenciliği Algoritmalarının Uygulanması .....	59
3.4.2.6. Değerlendirme .....	60
3.5. Duygu Analizi Yaklaşımları .....	60
3.5.1. Sözlüğe Dayalı Yaklaşımlar .....	60
3.5.2. Makine Öğrenmesi .....	62
3.6. Makine Öğrenmesi Yönteminde Kullanılan Sınıflandırma Algoritmaları .....	62
3.6.1. Karar Ağacı .....	62
3.6.2. K-En Yakın Komşu .....	65
3.6.3. Naive Bayes .....	66
3.6.4. Derin Öğrenme .....	67
3.6.5. Destek Vektör Makineleri .....	71
3.7. Model Performans Değerlendirme .....	73
3.7.1. Model Performans Değerlendirme Yöntemleri .....	73
3.7.1.1. Dışarıda Tutma (Holdout) .....	73
3.7.1.2. Tabakalı Öğrenme .....	73
3.7.1.3. Çapraz Geçerleme (Cross Validation) .....	74
3.7.1.4. Yerine Geri Koyarak Örneklem (Bootstrap Örneklem) .....	74
3.7.2. Model Performans Değerlendirme Ölçütleri .....	74
3.8. Makine Öğrenmesi Yöntemi ile Duygu Analizi Süreci .....	76
3.9. Twitter ve Duygu Analizindeki Rolü .....	78
3.9.1. Twitter ve Terminolojisi .....	79
3.9.2. Twitter API .....	80
3.10. Duygu Analizi Yöntemi Kullanılarak Daha Önce Yapılan Çalışmalar .....	81
<b>4. TÜRKİYE'DE YAŞAYAN SURIYELİ SİĞİNMACILAR HAKKINDAKİ ALGI ÜZERİNE DUYGU ANALİZİ .....</b>	<b>83</b>
4.1. Verileri Toplama .....	83
4.2. Verilerin Ön İşlemden Geçirilmesi .....	84
4.3. Makine Öğrenmesi Yöntemi ile Duygu Analizi .....	86
4.4. Genel Algı Analizi.....	88
4.5. Algının Niteliği Analizi .....	100
4.5.1. Olumsuz Algının Niteliği Analizi .....	101
4.5.2. Olumlu Algının Niteliği Analizi .....	103
4.6. Algının Sebep Analizi .....	108
4.6.1. Olumsuz Algının Sebep Analizi .....	108
4.6.2. Olumlu Algının Sebep Analizi .....	112
4.7. Sonuç ve Tartışma .....	117
4.8. Öneriler .....	120
<b>KAYNAKLAR .....</b>	<b>121</b>
<b>ÖZ GEÇMİŞ .....</b>	<b>131</b>

## SİMGELER VE KISALTMALAR

AFAD	: Afet ve Acil Durum Yönetimi Başkanlığı
F-Ölç.	: F-Ölçütü
İŞİD	: Irak Şam İslam Devleti
ŞEY	: Şartlı Eğitim Yardımı Programı
TF/IDF	: Terim Frekansı – Ters Doküman Frekansı Ağırlıklandırması
TGDA	: Tekstil, Giyim, Deri ve Ayakkabı



## ŞEKİLLER DİZİNİ

Şekil 2.1. Suriyeli sığınmacıların yıllara göre Türkiye’deki nüfus değişimi .....	5
Şekil 2.2. Suriyeli sığınmacıların Türkiye’ye dağılımı .....	7
Şekil 2.3. Suriyeli sığınmacıların en fazla bulunduğu iller .....	8
Şekil 2.4. Suriyeli sığınmacıların en yoğun olarak buldukları iller .....	9
Şekil 2.5. Suriyeli sığınmacıların eğitim durumu .....	10
Şekil 2.6. Suriyeli sığınmacıların eğitim durumu-2 .....	11
Şekil 2.7. Suriyeli sığınmacıların Türkiye’deki yıllara göre okullaşma durumu .....	12
Şekil 2.8. Türkiye’de okuyan 5-17 yaş arası yabancı öğrencilerin uyruklarına göre dağılımı .....	13
Şekil 2.9. Suriyeli sığınmacıların gelir dağılımı .....	17
Şekil 2.10. Suriyeli şirketlerin Türkiye’ye dağılımı .....	21
Şekil 2.11. Suriye genel durum haritası .....	25
Şekil 3.1. Veri madenciliği modelleri .....	40
Şekil 3.2. Veri madenciliği süreci .....	42
Şekil 3.3. Doküman-terim matris gösterimi .....	57
Şekil 3.4. Duygu analizi sınıflandırma teknikleri .....	60
Şekil 3.5. Karar ağacı örneği .....	63
Şekil 3.6. K-En yakın komşu sınıflandırma örneği .....	65
Şekil 3.7. Sinir hücresinin (nöron) yapısı .....	68
Şekil 3.8. Yapay nöron .....	68
Şekil 3.9. Basit bir derin sinir ağı yapısı .....	70
Şekil 3.10. Doğrusal olarak ayrılabilen veri setinde hiperdüzlem ve destek vektörleri .....	71
Şekil 3.11. Doğrusal olarak ayrılamayan veri seti .....	72
Şekil 3.12. Kernel fonksiyonu ile verinin daha yüksek bir boyuta dönüştürülmesi .....	72
Şekil 4.1. Rapidminer arayüz görünümü .....	90
Şekil 4.2. 1.Verit seti k-NN sınıflandırıcısı 4-gram (karakter) öznitelik çıkarma yöntemi genel algı analizi modeli .....	91
Şekil 4.3. Process Documents from Data operatörünün içinde yer alan operatörler .....	92
Şekil 4.4. Genel algı analizi 1.model için çapraz geçerlilik alt prosesi .....	94
Şekil 4.5. Genel algı analizi 1.model çalışma sırası .....	96
Şekil 4.6. Olumsuz algının niteliği analizi 1.modeli .....	101
Şekil 4.7. Olumlu algının niteliği analizi 1.verit seti 1.modeli .....	104
Şekil 4.8. Olumsuz algının sebep analizi 1.modeli .....	110
Şekil 4.9. Olumlu algının sebep analizi 1.modeli .....	113

## TABLolar DİZİNİ

Tablo 2.1. Suriyeli sığınmacıların yaş aralığı – cinsiyet tablosu (2022 Mart ayı itibariyle) ...	6
Tablo 2.2. Suriyeli sığınmacıların meslek grubu dağılımı .....	14
Tablo 2.3. Ekonomik faaliyete göre Suriyeli sığınmacıların dağılımı .....	15
Tablo 3.1. Karışıklık matrisi .....	74
Tablo 4.1. Tweet'lerin ilk çekildiği andaki ham haline örnekler .....	84
Tablo 4.2. Tweet'lerin ön işlemden geçtikten sonraki hallerine örnekler .....	85
Tablo 4.3. Duygu analizi manuel sınıflandırma tablosu .....	86
Tablo 4.4. Duygu analizi manuel sınıflandırma örnekleri .....	87
Tablo 4.5. Genel algı analizi eğitim setleri sınıflandırma dağılımları .....	88
Tablo 4.6. Genel algı analizi eğitim ve test setleri tweet sayıları .....	89
Tablo 4.7. Genel algı analizi 1.veri seti sınıflandırıcı – öznitelik çıkarma yöntemleri performansları .....	97
Tablo 4.8. Genel algı analizi 2.veri seti sınıflandırıcı – öznitelik çıkarma yöntemleri performansları .....	98
Tablo 4.9. Genel algı analizi 3.veri seti sınıflandırıcı – öznitelik çıkarma yöntemleri performansları .....	98
Tablo 4.10. Naive Bayes sınıflandırıcısı makro-F1 ölçütü değerleri .....	99
Tablo 4.11. Derin Öğrenme sınıflandırıcısı makro-F1 ölçütü değerleri .....	99
Tablo 4.12. Genel algı analizi en yüksek performans gösteren modellerin tahmin sonuçları .....	100
Tablo 4.13. Olumsuz algının niteliği analizi eğitim seti sınıflandırma verileri .....	101
Tablo 4.14. Olumsuz nitelik analizi sınıflandırıcı – öznitelik çıkarma yöntemleri performansları .....	102
Tablo 4.15. Olumsuz algının niteliği analizi tahmin sonuçları .....	102
Tablo 4.16. Olumlu algının niteliği analizi 1. eğitim seti sınıflandırma verileri .....	103
Tablo 4.17. Olumlu nitelik analizi 1.veri seti sınıflandırıcı – öznitelik çıkarma yöntemleri performansları .....	104
Tablo 4.18. Olumlu algının niteliği analizi 2. eğitim seti sınıflandırma verileri .....	105
Tablo 4.19. Olumlu nitelik analizi 2.veri seti sınıflandırıcı – öznitelik çıkarma yöntemleri performansları .....	105
Tablo 4.20. Olumlu algının niteliği analizi 3.eğitim seti sınıflandırma verileri .....	106
Tablo 4.21. Olumlu nitelik analizi 3.veri seti sınıflandırıcı – öznitelik çıkarma yöntemleri performansları .....	106
Tablo 4.22. Olumlu algının niteliği analizi tahmin sonuçları .....	107
Tablo 4.23. Olumsuz algının sebep analizi eğitim seti sınıflandırma verileri .....	108
Tablo 4.24. Olumsuz algının sebep analizi sınıflandırıcı – öznitelik çıkarma yöntemleri performansları .....	110
Tablo 4.25. Olumsuz algının sebep analizi tahmin sonuçları .....	111
Tablo 4.26. Olumlu algının sebep analizi 1.eğitim seti sınıflandırma verileri .....	113

Tablo 4.27. Olumlu algının sebep analizi 1.veri seti sınıflandırıcı – öznitelik çıkarma yöntemleri performansları .....	114
Tablo 4.28. Olumlu algının sebep analizi 2.eğitim seti sınıflandırma verileri .....	114
Tablo 4.29. Olumlu algının sebep analizi 2.veri seti sınıflandırıcı – öznitelik çıkarma yöntemleri performansları .....	115
Tablo 4.30. Olumlu algının sebep analizi 3.eğitim seti sınıflandırma verileri .....	115
Tablo 4.31. Olumlu algının sebep analizi 3.veri seti sınıflandırıcı – öznitelik çıkarma yöntemleri performansları .....	116
Tablo 4.32. Olumlu algının sebep analizi tahmin sonuçları .....	117



# 1. GİRİŞ

Kitle iletişim olanaklarının da artmasıyla gelişmiş medeniyetlerin yönetim şekillerini daha iyi müşahede etme fırsatı bulan Arap coğrafyası, 2010 yılından sonra birer birer hareketlenmeye başlar. İlk kıvılcımı Tunus'ta bir seyyar satıcının güvenlik güçlerinden görmüş olduğu kötü muameleden ötürü kendini ateşe vermesi ile yakılan isyan dalgası (Oğuzlu, 2011: 9) kısa sürede tüm Arap coğrafyasına yayılır ve etkisi bugün dahi devam etmekte olan ve ülkeden ülkeye farklılık gösteren değişimlere sebep olur. Tarihleri boyunca monarşi ile yönetilmiş olan Arap halklarının demokrasi, özgürlük, insan hakları ve daha müreffeh bir hayat talepleri ile başlattıkları bu ayaklanmalara "Arap Baharı" adı verilir. Arap Baharı kimi ülkelerde hükümetlerin devrilmesine, kimilerinde daha yumuşak geçişler yapılmasına, kimilerinde ise bugün dahi devam eden iç savaşlara neden olur.

Arap Baharı'nın en büyük olumsuz etkiye sebep olduğu ülkelerden biri de Suriye olur. Ayaklanmalar sonrası devlet başkanı Esed iktidarı bırakmayı reddeder ve böylelikle uzun yıllar sürecektir kanlı bir iç savaşın da fitili ateşlenmiş olur. İş savaş sonucu 600.000'in üzerinde kişi hayatını kaybeder (SOHR, 2022); 6,7 milyon kişi de ülke dışına göç etmek zorunda kalır (UNHCR, 2020a). Ülke dışına yapılan göçlerin de %80'den fazlası komşu ülkelere yapılır. Komşu ülkelere yapılan göçlerin %65'ini Türkiye'ye yapılan göçler oluşturur (UNHCR, 2020a). 2022 yılı itibari ile Türkiye'de kayıtlı olarak 3,8 milyonu aşkın Suriyeli sığınmacı bulunmakta ve bu sayı Türkiye nüfusunun %4,4'üne tekabül etmektedir (GİB, 2022a; 2022b; TÜİK, 2022a).

Bu tez kapsamında Suriyeli sığınmacılar meselesinin Türkiye için önemi ve Türk halkı nezdindeki algısı Twitter kapsamında konu ile ilgili olarak atılan tweet'ler üzerinden ortaya konmaya çalışılmıştır. Bu amaçla ilk bölümde Suriyeli sığınmacıların nüfusu, Türkiye'ye dağılımı, eğitim ve okullaşma durumları, çalışma hayatları, Türkiye'ye ekonomik etkileri, sığınmacılara yapılan yardımlar, kamu hizmetlerine etkileri, güvenliğe etkileri ve son olarak uyum durumları incelenmiştir. İkinci bölümde, üçüncü bölümde Suriyeli sığınmacılar hakkındaki algıyı ortaya çıkarma amacı ile Twitter kapsamında gerçekleştirilecek olan duygu analizi yöntemine dair tüm kavram, teknik ve uygulama süreci açıklanmıştır. Bu kapsamda büyük veri, veri madenciliği, metin madenciliği, duygu analizi yaklaşımları, makine öğrenmesi yönteminde kullanılan sınıflandırma algoritmaları, model performans

değerlendirme, makine öğrenmesi yöntemi ile duygu analizi süreci ile Twitter ve duygu analizindeki rolü incelenmiştir. Son bölümde ise Türkiye’de yaşayan Suriyeli sığınmacılar hakkındaki algıyı, bu algının niteliklerini ve sebeplerini ortaya koyma amacı ile Twitter kapsamında konu hakkında atılan Türkçe tweet’ler üzerinden duygu analizi gerçekleştirilmiş ve sonuçlar ortaya konmuştur.



## 2. SURİYELİ SIĞINMACILAR MESELESİ

### 2.1. Arap Baharı

Giriş kısmında kısaca değinilen Arap Baharı ülkeden ülkeye değişiklik göstermiş ve yaşanan her ülkede farklı şiddet ve seviyelerde cereyan etmiştir. İlk olarak Tunus'ta patlak veren ayaklanmalar devlet başkanı Zeynel Abidin Bin Ali'nin ülkeyi terk etmesine ve hükümetin devrilmesine neden olmuştur. Cezayir'de protestolar sonucu 19 yıllık olağanüstü hal kaldırılmış, Ürdün'de hükümet değişmiştir (Deniz, 2013: 70).

Mısır'da büyük çaplı protestolar sonucu 30 yıldır iktidarda olan Hüsnü Mübarek devrilmiştir. Mübarek sonrası her ne kadar demokratik seçimler gerçekleştirilip bir hükümet kurulsa da, kurulan bu hükümet de askeri bir darbe sonucu devrilmiştir. Darbe sonrası başta Mısır tarihindeki ilk seçilmiş Cumhurbaşkanı Muhammed Mursi olmak üzere birçok Müslüman Kardeşler (iktidar partisine temel teşkil eden hareket) mensubu tutuklanmış, darbe karşıtı gösteriler kanlı bir şekilde bastırılmış ve bazı darbe karşıtları idam edilmiştir. Mısır'ın kısa demokrasi deneyimi de böylelikle son bulmuştur (Çelik, 2015: 41-47; Yaşar, 2013: 15; Akyol, 2019: 347-348).

Libya'da 42 yıldır iktidarda olan Muammer Kaddafi devrilmiş ve öldürülmüştür (Yılmaz, 2012: 2). Kaddafi sonrası ülkede iç savaş patlak vermiş ve ülke Trablus ve Tobruk merkezli olmak üzere ikiye bölünmüştür. Bu ikiye bölünmüşlük halen daha devam etmektedir (Güneş, 2018: 282; Karakaya ve Çelik, 2021: 189-190).

Yemen'de başlayan protestolar sonucu Cumhurbaşkanı Ali Abdullah Salih her ne kadar bir müddet koltuğunu terk etmemek için dirense de sonunda Suudi Arabistan önderliğindeki Körfez Arap İşbirliği Konseyi'nin kendisine dayattığı yol haritasını kabul etmek zorunda kalmış ve koltuğunu yardımcısı Hadi'ye bırakmıştır. Ancak bu değişim de Yemen'deki yerel unsurları yatıştırmaya yetmemiş ve Şii Husiler başta olmak üzere birçok grup ayaklanmıştır. Yemen'in geçmişinde de iktidar ile sürekli bir çatışma halinde olan Husiler, silahlı mücadele sonucu başkent Sana da dahil olmak üzere ülkenin batı bölgelerini ele geçirmeyi başarmışlardır. Bunun üzerine Suudi Arabistan önderliğindeki 9 Arap ülkesinden oluşan koalisyon güçleri Yemen'e hava saldırıları başlatmış ve bu saldırılar ülke çapında büyük bir

yıkıma neden olmuştur. Halen daha savaşın devam ettiği Yemen büyük bir insani kriz ile karşı karşıyadır (Tinas, 2020).

Arap Baharı'nın en büyük değişime neden olduğu ülkelerden bir diğeri de Suriye'dir. Ayaklanmalar sonrası devlet başkanı Esed, iktidarı bırakmayı reddetmiş ve Rusya'nın desteği ile, silahlanan muhalifleri bastırmıştır. Diğer taraftan Suriye'nin kuzey doğu bölgelerini ele geçiren IŞİD (Irak Şam İslam Devleti)'e karşı ABD tarafından, Türkiye'nin terörist örgüt olarak kabul ettiği YPG'ye silahlı destek verilerek bu toprakların YPG tarafından ele geçirilmesi ve kontrol edilmesi sağlanmıştır. Tüm bu savaşlar ve Suriye topraklarının farklı güç merkezleri tarafından kontrol edilmeye başlanması son yüzyılda görülmemiş bir sığınmacı göçüne neden olmuştur. Suriye iç savaşı sonucu 600.000'nin üzerinde kişi hayatını kaybetmiş (SOHR, 2022); 6,7 milyon kişi de ülke dışına göç etmek zorunda kalmıştır (UNHCR, 2020a).

## **2.2. Türkiye'ye Göç Eden Suriyeli Sığınmacılar**

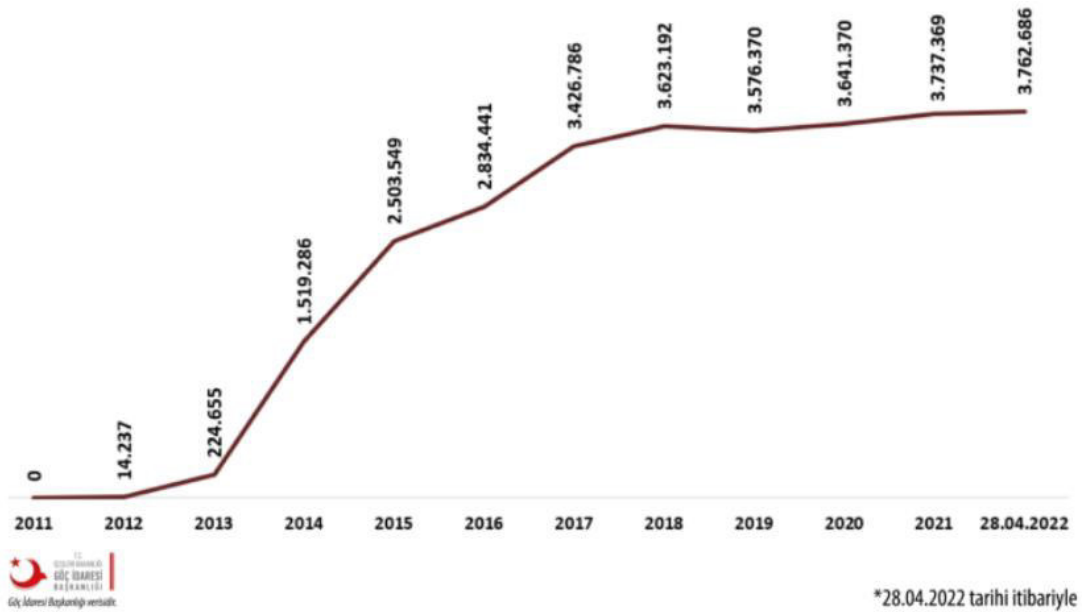
Aşağıda Türkiye'ye göç eden Suriyeli sığınmacılar ile ilgili olarak mevcut durum ortaya konulmaya çalışılmıştır. Buna göre sığınmacıların nüfusu, Türkiye'ye dağılımları, eğitimleri, Türkiye'deki hakları, ekonomiye etkileri, yapılan yardımlar, iş ve dış güvenliğe olan etkiler ve uyum konuları mercek altına alınmıştır.

### **2.2.1. Nüfus**

İç savaş sonucu Suriye nüfusunun yarıdan fazlası (13,5 milyon kişi) - ülke içi göçler ile birlikte - yaşadığı yeri terk etmek zorunda kalmıştır. Bu göçlerin yaklaşık yarısını (6,8 milyon kişi) ülke dışına yapılan göçler oluşturmaktadır. Ülke dışına yapılan göçlerin de %80'den fazlası komşu ülkelere yapılmıştır. Geri kalanı da Almanya başta olmak üzere batılı ülkelere yapılan göçlerdir. Komşu ülkelere yapılan göçlerin %65'ini Türkiye'ye yapılan göçler oluşturmaktadır (UNHCR, 2020a). Nisan 2022 itibari ile 3.762.686 Suriyelinin "geçici koruma" kapsamında Türkiye'de bulunduğu görülmektedir (GİB, 2022a). Türkiye, sadece Avrupa'dan gelen sığınmacılara mülteci statüsü vermektedir. Bu sebeple Türkiye'ye sığınan Suriyelilere mülteci statüsü verilmemiş; 2012 yılında çıkan bir genelge ile Suriyeli sığınmacılar "geçici koruma statüsü altındaki yabancılar" olarak tanımlanmıştır (Erdoğan, 2020: 8). Bu sebeple çalışma boyunca Türkiye'ye sığınan Suriyeliler için "bir ülkeye sığınmış kişi" anlamında "sığınmacı" ifadesi kullanılacaktır.

Geçici koruma kapsamındaki Suriyelilere ilaveten Türkiye’de 109.388 Suriyeli de ikamet izni ile bulunmaktadır. Her ikisi toplandığında Türkiye’de bulunan kayıtlı Suriyeli sayısının 3,8 milyonu aştığı görülmektedir (GİB, 2022b). Bu sayıya vatandaşlığa geçen 200 bin Suriyeli dahil değildir (Karadağ, 2022). Bu haliyle kayıtlı Suriyeli sığınmacıların Türkiye nüfusuna oranı %4,4 olarak hesaplanmaktadır (TÜİK, 2022a; GİB, 2022a). Buna ilaveten 2021 yılı sonu itibari ile 2011 Suriye krizi sonrası Afgan uyruklular başta olmak üzere 586.384 kişinin Türkiye’ye giriş yaparak uluslararası koruma talep ettiği görülmektedir (GİB, 2022c). Böylelikle sadece kayıtlı olarak Türkiye’de bulunan sığınmacı sayısının 4,3 milyonu aştığı, bu sayının da ülke nüfusunun %5’ini aştığı görülmektedir. Kayıtlı sığınmacı sayısı dikkate alındığında, Türkiye şu anda dünyada en fazla sığınmacıya ev sahipliği yapan ülke konumundadır. İkinci sıradaki Kolombiya 1,7 milyon sığınmacıya ev sahipliği yapmaktadır (UNHCR, 2020a). Ayrıca kayıtlı sığınmacılar dışında Türkiye’nin ciddi bir kayıt dışı göçmen sorunu da bulunmaktadır. Tüm bu veriler dikkate alındığında 2011 yılında sadece 58 bin sığınmacısı olan Türkiye için bu denli büyük bir sığınmacı göçünün oldukça ciddi etkileri olacağı aşikardır (Erdoğan, 2019: 3-4).

Yıllara göre Suriyeli sığınmacıların Türkiye’deki nüfusu aşağıdaki gibi seyretmiştir:



Şekil 2.1. Suriyeli sığınmacıların yıllara göre Türkiye’deki nüfus değişimi (GİB, 2022a)

Suriyeli sığınmacıların %53,8'i erkekler, %46,2'si kadınlardan oluşmaktadır. Yaş ortalaması ise 22,3'tür. Türkiye'nin yaş ortalaması ise 33'tür (Caro, 2020: 5). Yaş aralığı – cinsiyet tablosu aşağıdaki gibidir (Mülteciler Derneği, 2022):

Tablo 2. 1. Suriyeli sığınmacıların yaş aralığı – cinsiyet tablosu (2022 Mart ayı itibariyle)

YAŞ ARALIĞI	ERKEK	KADIN	TOPLAM
0 – 4	259.953	242.653	502.606
5-9	296.171	278.491	574.662
10-14	229.424	215.783	445.207
15-18	141.340	121.875	263.215
19-24	274.124	206.822	480.946
25-29	233.233	166.574	399.807
30-34	166.992	122.801	289.793
35-39	130.159	100.167	230.326
40-44	88.275	74.694	162.969
45-49	61.050	59.486	120.536
50-54	46.105	45.127	91.232
55-59	36.343	36.579	72.922
60-64	23.520	24.518	48.038
65-69	15.412	16.635	32.047
70-74	9.378	10.350	19.728
75+	8.910	11.647	20.557
TOPLAM	2.020.389	1.734.202	3.754.591

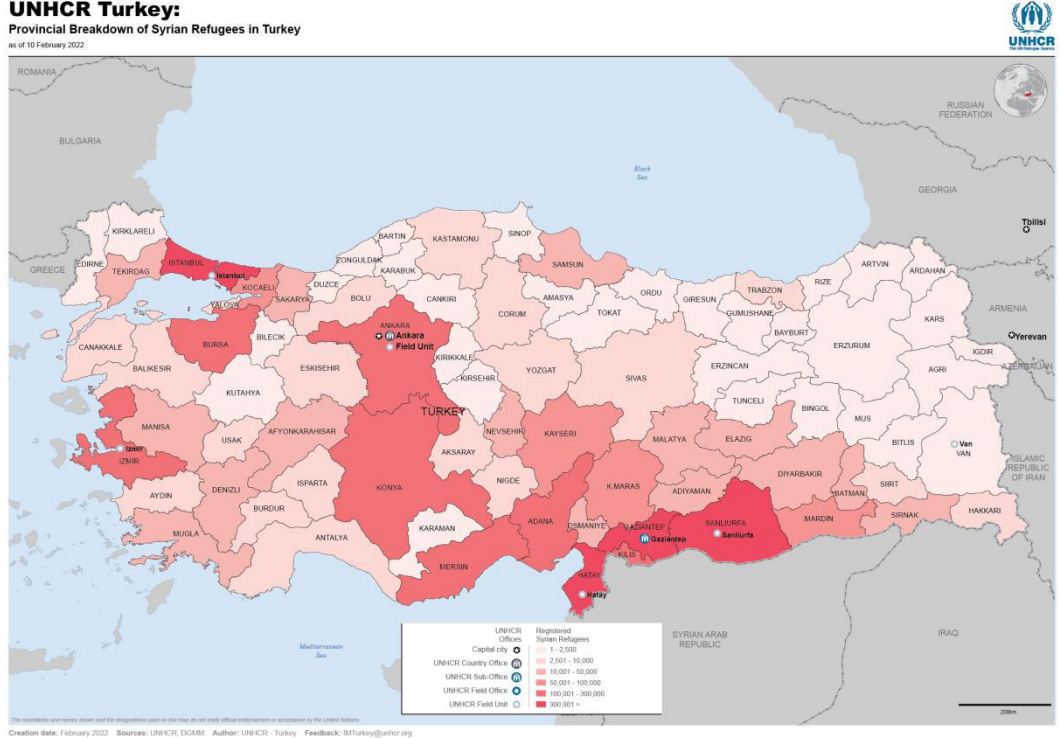
Doğurganlık hızı, bir kadının doğurganlık yaşamı süresince, yaşa özel doğurganlık hızlarının geçerli olması durumunda doğuracağı toplam çocuk sayısını ifade etmektedir. 2018 yılında Hacettepe Üniversitesi Nüfus Etütleri Enstitüsü tarafından Suriyeli sığınmacılar üzerinde örneklem alma yolu ile yapılan araştırma sonucu, Suriyeli sığınmacıların doğurganlık hızının 5,3 olduğu tespit edilmiştir. Aynı yıl Türkiye'nin ortalama doğurganlık hızı aynı kurum tarafından 2,3 olarak hesaplanmıştır. (HÜNEE, 2019b). 2021 TÜİK verilerine göre ise Türkiye'nin

ortalama doğurganlık hızı 1,7'ye gerilemiştir (TÜİK, 2022b). Bu veriler Suriyeli sığınmacıların nüfusunun Türkiye ortalamasından çok daha hızlı bir şekilde arttığını göstermektedir. Sağlık Bakanı Fahrettin Koca tarafından Mart 2022 itibari ile Türkiye’de 754.000 Suriyeli bebeğin dünyaya geldiği ifade edilmiştir (Rakipoğlu ve Kara, 2022).

Nüfus konusu ile ilgili olarak ifade edilmesi gereken bir başka husus geri dönüşlerdir. Savaş koşullarının azalması veya sonlanması sebebiyle Türkiye’den Suriye’ye geri dönüşler de yaşanmıştır. 4 Nisan 2022 tarihinde İçişleri Bakanlığı tarafından yapılan resmi açıklamaya göre toplamda 492.983 Suriyeli sığınmacı ülkesine geri dönmüştür (Mülteciler Derneği, 2022).

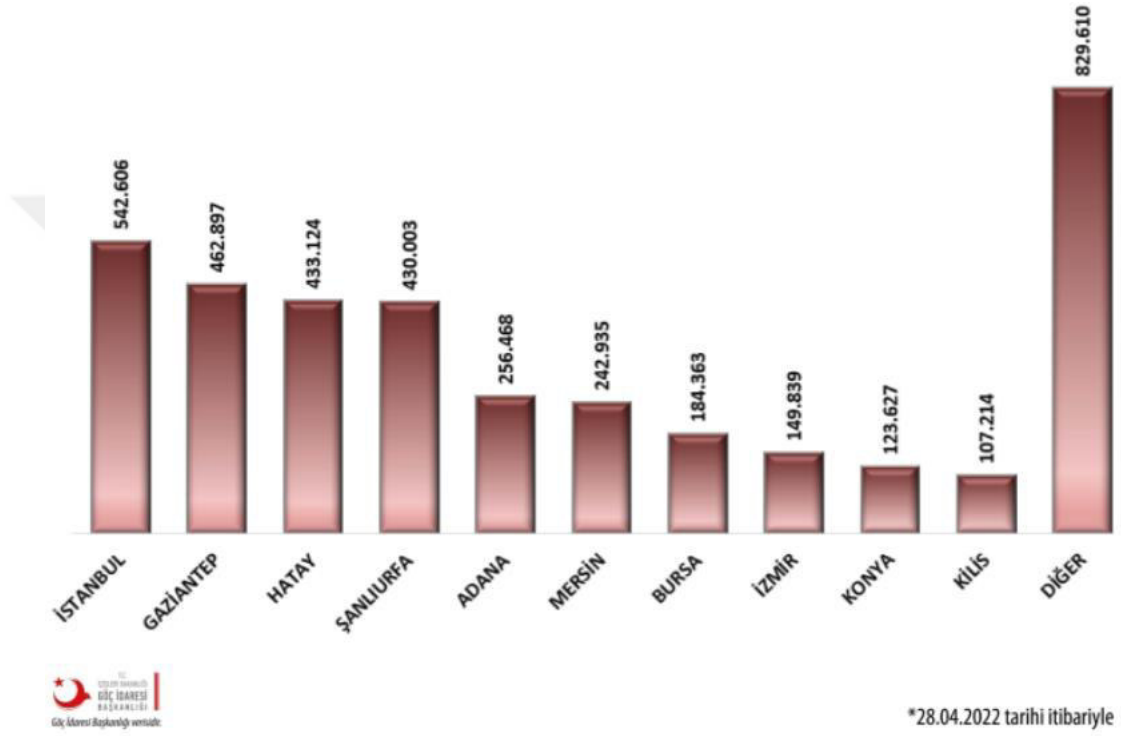
### 2.2.2. Türkiye’ye Dağılım

Suriyeli sığınmacıların Türkiye’deki dağılımı oldukça dengesiz olup, sığınmacıların özellikle Suriye sınırı boyunca ve bazı büyükşehirlerde yoğunlaştıkları görülmektedir. Geçici koruma kapsamındaki Suriyeli sığınmacıların Türkiye’deki dağılımını gösterir harita aşağıdaki gibidir:



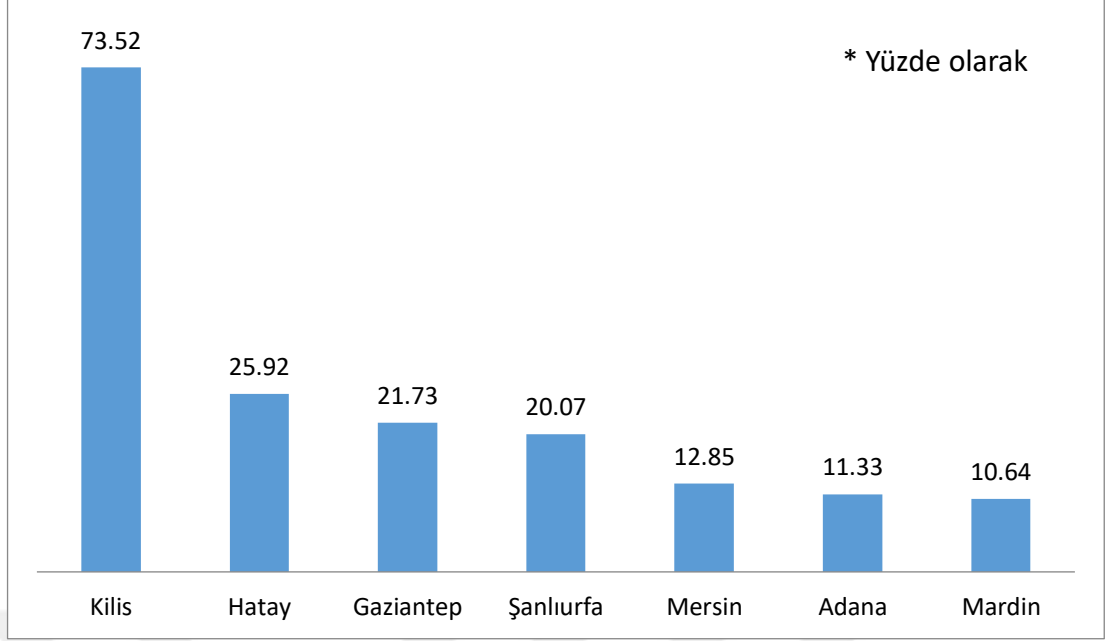
Şekil 2.2. Suriyeli sığınmacıların Türkiye’ye dağılımı (UNHCR, 2022a)

Şekil 2.2'deki haritada da görüleceği üzere Suriyeli sığınmacılar özellikle Suriye sınırı boyunca ve başta İstanbul, İzmir, Bursa ve Konya olmak üzere bazı büyükşehirlerde yoğunlaşmışlardır. Geçici koruma kapsamında en fazla Suriyeli sığınmacı bulunan il 542.606 kişi ile İstanbul; Suriyeli sığınmacı yoğunluğunun en fazla olduğu il ise %73,52'si geçici koruma kapsamındaki Suriyeli sığınmacılardan oluşan Kilis'tir (GİB, 2022a). Suriyeli sığınmacıların en fazla bulunduğu iller aşağıdaki gibidir:



Şekil 2.3. Suriyeli sığınmacıların en fazla bulunduğu iller (GİB, 2022a)

Şekil 2.3'te görüldüğü gibi Suriyeli sığınmacıların en fazla bulunduğu il İstanbul'dur. İstanbul'u sırası ile Gaziantep, Hatay ve Şanlıurfa takip etmektedir. Suriyeli sığınmacı sayısının ilin en az %10'unu oluşturduğu iller ise sırası ile şu şekildedir:



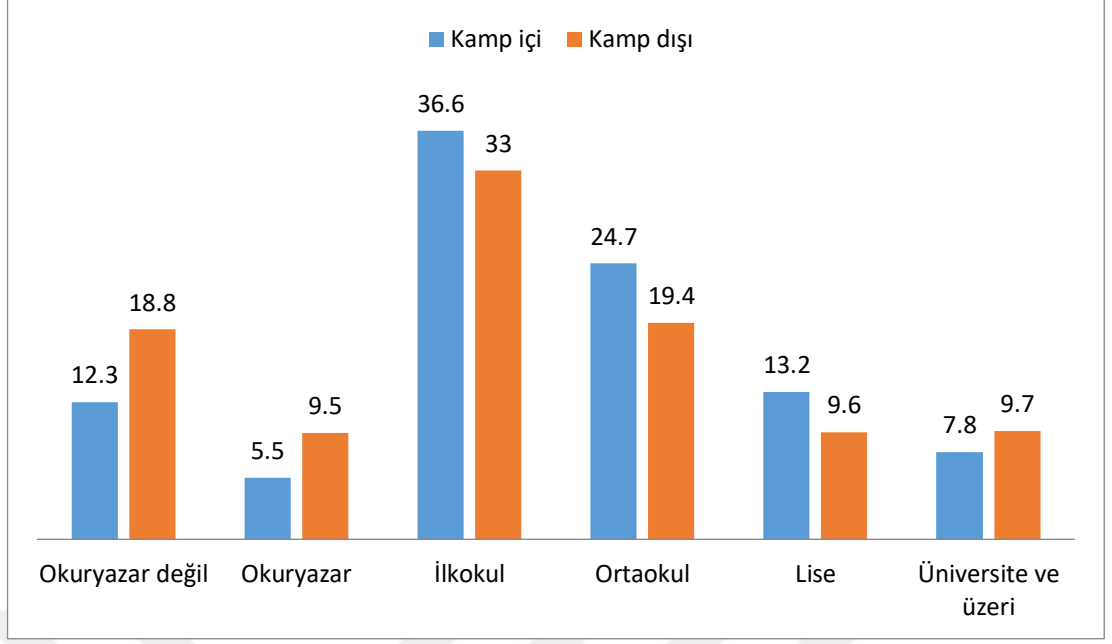
Şekil 2.4. Suriyeli sığınmacıların en yoğun olarak buldukları iller (GİB, 2022a)

Şekil 2.4’te görüldüğü gibi Suriyeli sığınmacıların en yoğun olarak bulunduğu ille %73,52 ile Kilis’tir. Kilis’i %25,92 ile Hatay, %21,73 ile Gaziantep ve %20,07 ile Şanlıurfa takip etmektedir.

### 2.2.3. Eğitim Durumu ve Okullaşma

Yapılan araştırmalar sonucu Türkiye’ye sığınan Suriyeli sığınmacıların eğitim durumlarının Türkiye ortalamasının oldukça altında olduğu görülmektedir. Suriyeli sığınmacıların eğitim seviyelerini tespit etmeye yönelik bugüne kadar çeşitli araştırmalar yapılmıştır. En önemlileri aşağıdaki gibidir:

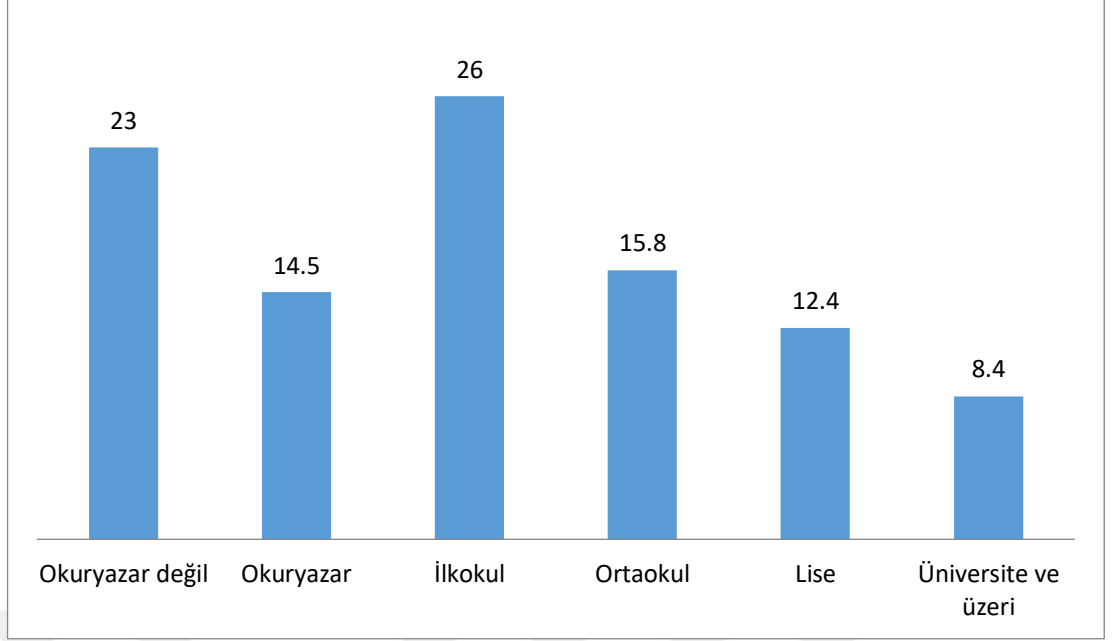
2013 yılında AFAD (Afet ve Acil Durum Yönetimi) tarafından yapılan ilk çalışmaya ve toplam sayı hakkında kesin bir bilgiye sahip olunamamakla birlikte Türkiye’ye 550.386 Suriyeli sığınmacının giriş yaptığı ifade edilmektedir. Bu sığınmacıların 200.386’sı kamplara yerleşmiş, 350.000’i ise çeşitli kentlere dağılmıştır. Buna göre 2013 yılı Ağustos ayı itibari ile Suriyeli sığınmacıların %36’sı kamplarda, %64’ü ise kamp dışında yaşamaktadır (AFAD, 2013). AFAD tarafından hem kamp hem de kentlerde yaşayan toplamda 2700 hanehalkı üzerinde yapılan anket sonuçları aşağıdaki gibidir:



Şekil 2.5. Suriyeli sığınmacıların eğitim durumu (AFAD, 2013)

AFAD'ın 2013 yılında gerçekleştirmiş olduğu çalışmaya göre özetle, kamplarda yaşayan Suriyeli sığınmacıların %54'ünün, kamp dışında yaşayanların ise %61'inin okuryazar olmayan, okuryazar olup herhangi bir okulu bitirmemiş ve ilkökul mezunlarından oluştuğu görülmektedir.

AFAD benzer bir çalışmayı 2017 yılında tekrar gerçekleştirmiş ve bu çalışmada Mayıs 2017 itibari ile Suriyeli sığınmacı sayısının 3.020.654 kişiye ulaştığı, bu sayının 247.000'inin kamplarda, 2.774.000'nin ise kamp dışında yaşadığı ifade edilmiştir. Buna göre 2017 yılı itibari ile kamplarda yaşayan Suriyeli sığınmacıların oranının yaklaşık olarak %8,2'ye indiği, geri kalan %91,2'lik kesimin ise kamp dışına yerleştiği görülmektedir. Örneklem alma yolu ile gerçekleştirilen çalışmada kamp içi ve kamp dışı toplam 10.838 kişiye ulaşılmıştır (AFAD, 2017). Bu 10.838 kişiden 6 yaş ve üzerinde olan 9235 kişinin eğitim durumlarını gösterir tablo aşağıdaki gibidir:



Şekil 2.6. Suriyeli sığınmacıların eğitim durumu-2 (AFAD, 2017)

Şekil 2.6’da görüldüğü üzere AFAD’ın 2017 yılında tespit ettiği eğitim durumu dağılımı, 2013 yılı ile benzerlik göstermektedir. Buna göre Suriyeli sığınmacıların %23’ü okur-yazar değil, %14,5’i okuryazar olup herhangi bir okulu bitirmemiş, %26’sı ise ilkokul mezunudur. Bu durumda ilkokul ve altı eğitim durumu oranı %63,5 olarak tespit edilmiştir (AFAD, 2017).

2016 yılında Kalkınma Bakanlığı tarafından hazırlanan “Türkiye’deki Geçici Koruma Statüsündeki Suriyelilere Yönelik 2016-2018 Dönemini Kapsayan Birinci Aşama İhtiyaç Analizi” adlı çalışmada, Suriyeli sığınmacıların %33,3’ünün okuryazar olmadığı, %13’ünün okuryazar olup herhangi bir okulu bitirmediği, %16,5’inin ilkokul ve dengi okul mezunu olduğu, %6,5’inin ortaokul ve dengi okul mezunu olduğu, %5,6’sının lise ve üstü mezunu olduğu, %26,6’sının ise bu konuda beyanda bulunmadığı görülmektedir (Erdoğan, 2020: 33).

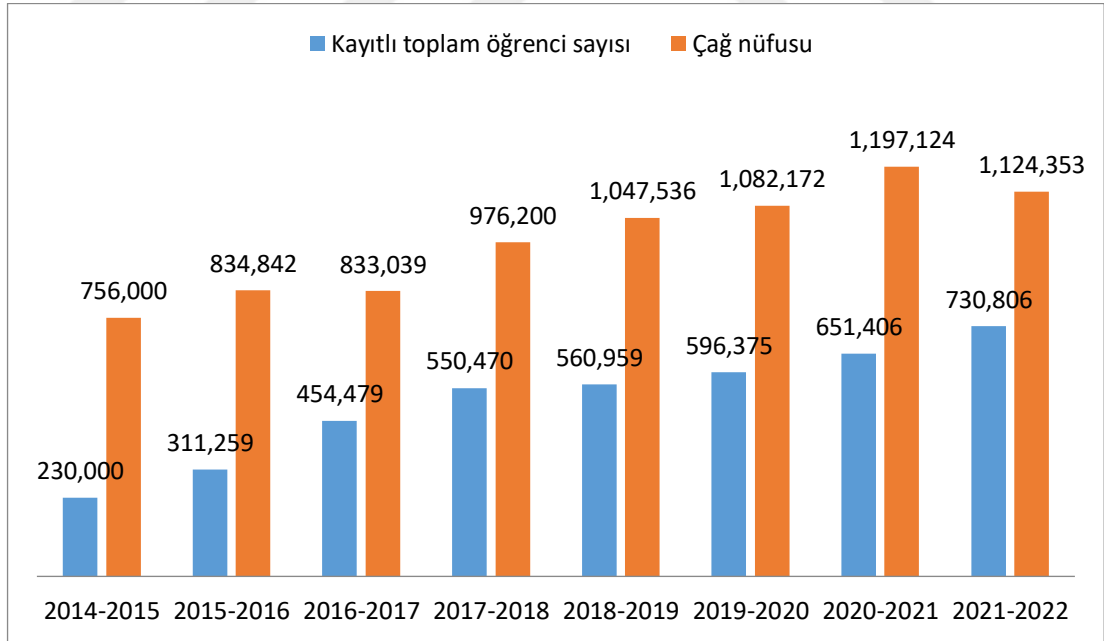
Hacettepe Üniversitesi Nüfus Etütleri Enstitüsü tarafından 2018 yılında Suriyeli sığınmacılar özelinde örneklem alma yolu ile gerçekleştirilen nüfus ve sağlık araştırmasında kadınların %40,3’ünün, erkeklerin ise %34,8’inin herhangi bir eğitimi olmadığı veya ilkokulu bitirmediği; kadınların %37,3’ünün, erkeklerin ise %38,2’sinin ilkokul mezunu olduğu sonuçlarına ulaşılmıştır. Buna göre kadınların %77,6’sı, erkeklerin ise %73’ü ilkokul ve altı eğitim durumuna sahiptirler. Aynı

çalışmada kadınlar için ortalama almış oldukları eğitim süresi 4,5 yıl, erkekler için ise 5,1 yıl olarak tespit edilmiştir (HÜNEE, 2019a: 20).

Erdoğan (2020)'ın örneklem alma yolu ile gerçekleştirdiği, “Suriyeliler Barometresi 2019” adlı çalışmasında ise, Suriyeli sığınmacıların %8,2'sinin okuryazar olmadığı, %16,7'sinin okuryazar olup herhangi bir okulu bitirmediği, %31,7'sinin ilkokul mezunu olduğu, %22'sinin ilköğretim/ortaokul mezunu olduğu, %11,4'ünün lise ve dengi okul mezunu olduğu, %2,7'sinin ön lisans, %7'sinin lisans, %0,3'ünün ise lisansüstü/doktora mezunu olduğu sonuçlarına ulaşılmıştır (Erdoğan, 2020:34). Buna göre ilkokul ve altı eğitim durumuna sahip olan Suriyeli sığınmacıların oranı %56,6 olarak tespit edilmiştir.

Suriyeli sığınmacılar ile ilgili olarak bugüne kadar yapılmış belli başlı çalışmaların hangisine bakılırsa bakılsın, hepsinde de Türkiye'ye göç eden Suriyeli sığınmacıların eğitim durumlarının Türkiye ortalamasının oldukça altında olduğu açıkça görülmektedir.

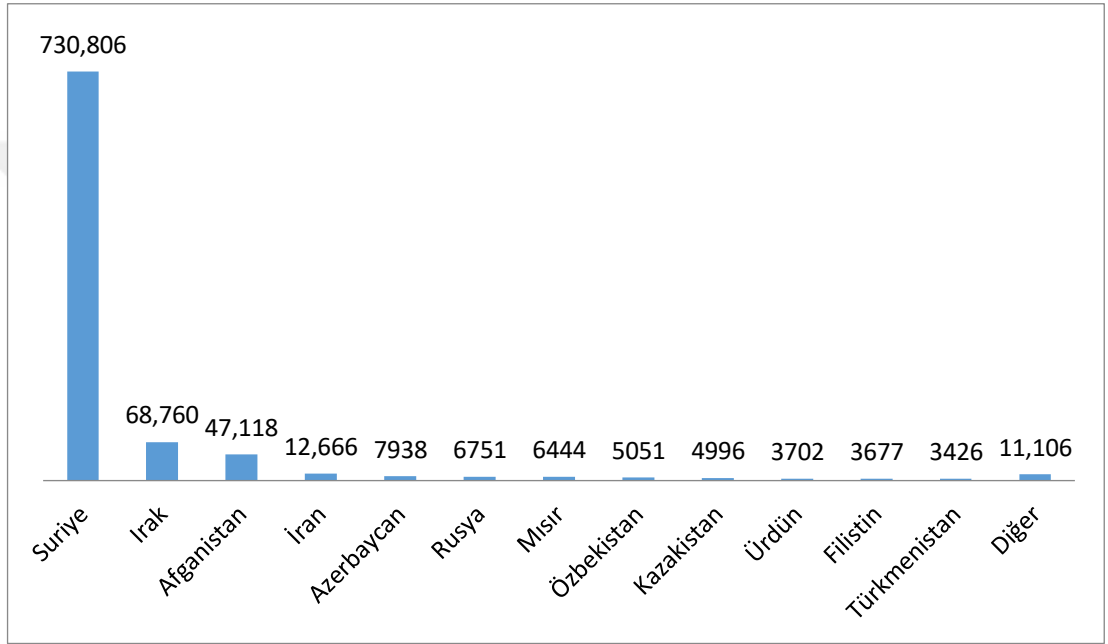
Suriyeli sığınmacılar ile ilgili olarak Türkiye'ye giriş yaptıktan sonraki okullaşma oranlarına baktığımızda ise aşağıdaki tablo ile karşılaşırız:



Şekil 2.7. Suriyeli sığınmacıların Türkiye'deki yıllara göre okullaşma durumu (MEB, 2022b)

Şekil 2.7'de görüldüğü üzere, yıllar ilerledikçe Suriyeli sığınmacı çocukların okullaşma oranı da artmış, 2021-2022 eğitim öğretim yılı itibari ile bu oran %65

olarak gerçekleşmiştir. 2021-2022 eğitim öğretim yılı itibari 5-17 yaş arası eğitim çağındaki 1.124.353 Suriyeli sığınmacıdan 730.806'sının okula gittiği görülmektedir. Bu durum aynı zamanda okula gitmesi gereken 393.547 Suriyeli sığınmacı çocuğun okula gitmediğini göstermektedir. Suriyeli sığınmacı öğrencilerin eğitim kademelerine göre okullaşma oranları ise, okul öncesi %34,34, ilkokul %75,13, ortaokul %80 ve lise %42,65 olarak gerçekleşmiştir (MEB, 2022b). Türkiye’de okuyan 5-17 yaş arası tüm yabancı öğrencilerin uyruklarına göre dağılımı ise aşağıdaki gibidir:



Şekil 2.8. Türkiye’de okuyan 5-17 yaş arası yabancı öğrencilerin uyruklarına göre dağılımı (MEB, 2022b)

Şekil 2.8’de görüldüğü gibi, Türkiye’de bulunan 5-17 yaş arası yabancı uyruklu öğrencilerin %80’ini Suriyeli sığınmacılar oluşturmaktadır.

Yükseköğretime bakıldığında ise, 2020-2021 eğitim öğretim yılı itibari ile yükseköğretimini Türkiye’de yapan yabancı uyruklu öğrenci sayısının 223.952 olduğu, bunlar içinde de 47.482 öğrenci ile Suriye uyruklu öğrencilerin birinci sırada yer aldığı görülmektedir (YBYS, 2022).

#### 2.2.4. Çalışma Hayatı

Türkiye’ye göç eden Suriyeli sığınmacıların meslekleri ile ilgili olarak AFAD’ın 2017 yılında gerçekleştirmiş olduğu çalışmaya baktığımızda aşağıdaki tablo ile karşılaşıyoruz.

Tablo 2.2. Suriyeli sığınmacıların meslek grubu dağılımı (AFAD, 2017)

Meslek grupları	Kamp İçi			Kamp Dışı		
	Erkek %	Kadın %	Toplam %	Erkek %	Kadın %	Toplam %
Mimar/Mühendis/Müteahhit	0,6	0,2	0,4	1,1	0,7	0,9
Devlet memuru	9	7,3	8,2	1	0,5	0,8
El işçiliği	8,2	0,5	4,5	41,8	34,1	38
Herhangi bir mesleği yok	68,4	90,3	79	39	61,1	49,9
Ofis çalışanı	0,5	0,3	0,4	3,3	1,1	2,2
Operatör / Şoför	2,4	0	1,2	1,2	0	0,6
Ordu mensubu	2,1	0	1,1	0,3	0	0,2
Sağlık çalışanı	0,3	0,2	0,2	0,5	0,6	0,6
Tarım ve hayvancılık	2,4	0	1,2	1,6	0,7	1,1
Zanaatkar	6,2	1,2	3,8	10,2	1,2	5,8

Tablo 2.2’de görüldüğü üzere, 2017 yılı itibari ile Suriyeli sığınmacıların kampta yaşayanlarının %79’unun, kamp dışında yaşayanlarının ise %49,9’unun herhangi bir mesleği bulunmamaktadır. Mesleği olan sığınmacılarda ise kamp dışında yaşayanlarda %38 ile “el işçiliği”; kampta yaşayanlarda ise %8,2 ile “devlet memuru” meslek grupları başı çekmektedir (AFAD, 2017).

Uluslararası Çalışma Örgütü (ILO) bünyesinde ve Caro (2020) tarafından gerçekleştirilen “Türk İşgücü Piyasasında Suriyeli Mülteciler” adlı çalışmada Türkiye’de yaşayan Suriyeli sığınmacıların çalışma hayatına dair tüm hususlar detaylı olarak ele alınmıştır. Buna göre, Türkiye’de 2017 itibari ile - 15 yaş altı Suriyeli sığınmacı çocuk işçiler ile birlikte - toplam 940.921 Suriyeli sığınmacının çalıştığı tahmin edilmektedir. Bu sığınmacıların %91,6’sı kayıt dışı çalışmaktadır. 15 yaş altında çalışan Suriyeli çocuk çalışan sayısı ise 127.140 olarak hesaplanmıştır. Bunların 109.328’i erkektir. Yukarıda “Eğitim Durumu ve Okullaşma” başlığı altında da belirtildiği üzere Suriyeli sığınmacı çocuklar ilkokulda %75,13, ortaokulda %80 ve lisede %42,65 oranında okula gitmektedirler. Bu aradaki kayıp ise burada görüldüğü üzere Suriyeli çocukların istihdamı şeklinde kendini göstermektedir.

Çalışan Suriyeli sığınmacıların kadın-erkek dağılımına bakıldığında ise erkek çalışanların büyük bir çoğunluğu oluşturduğu görülmektedir. 15-65 yaş aralığındaki Suriyeli sığınmacı erkeklerin %71'i çalışmaktadır, kadınlarda ise bu oran %11,2'dir (Caro, 2020).

15 yaş ve üstü, çalışan Suriyeli sığınmacıların sektörlere göre dağılımı aşağıdaki gibidir:

Tablo 2.3. Ekonomik faaliyete göre Suriyeli sığınmacıların dağılımı (Caro, 2020: 13)

Faaliyet	İstatistik		
	Toplam Suriyeli sayısı	Suriyeliler arasındaki pay	Sektör payı içinde
Tarım	63.110	7,8	1,2
İmalat	392.350	48,2	7,4
Tekstil, Giyim, Deri ve Ayakkabı	252.888	64,5	16,1
Diğer	139.462	35,5	3,7
İnşaat	107.389	13,2	5,2
Ulaşım ve İletişim	11.492	1,4	0,8
Ticaret ve Konaklama	144.273	17,7	2,6
Perakende	51.572	35,7	2
Yiyecek	43.363	30,1	3,5
Diğer	49.338	34,2	3
İş faaliyetleri	12.843	1,6	0,5
Eğitim	22.258	2,7	1,4
Sağlık	8990	1,1	0,8
Diğer hizmetler	51.076	6,3	5
Toplam (1, 2) / Ortalama (3)	813.781	100	2,9

Tablo 2.3'te görüldüğü üzere Suriyeli sığınmacıların yarıya yakını imalat sektöründe çalışmaktadır. İmalat sektörünü %17,7 ile ticaret ve konaklama, %13,2 ile inşaat ve %7,8 ile tarım sektörleri takip etmektedir. Dikkat edileceği üzere bu sektörler zaten Türkiye'de Suriyeli sığınmacılar gelmeden önce de kayıt dışı

istihdamın en yüksek olduğu sektörlerdendir. En fazla Suriyeli sığınmacı istihdam eden imalat sektöründe de özellikle Tekstil, Giyim, Deri ve Ayakkabı (TGDA) sektörünün yaklaşık 3'te 2 çoğunluğu oluşturduğu görülmektedir. Böylelikle Suriyeli sığınmacıların 3'te 1'inin sadece TGDA sektörü tarafından istihdam edildiği ortaya çıkmaktadır.

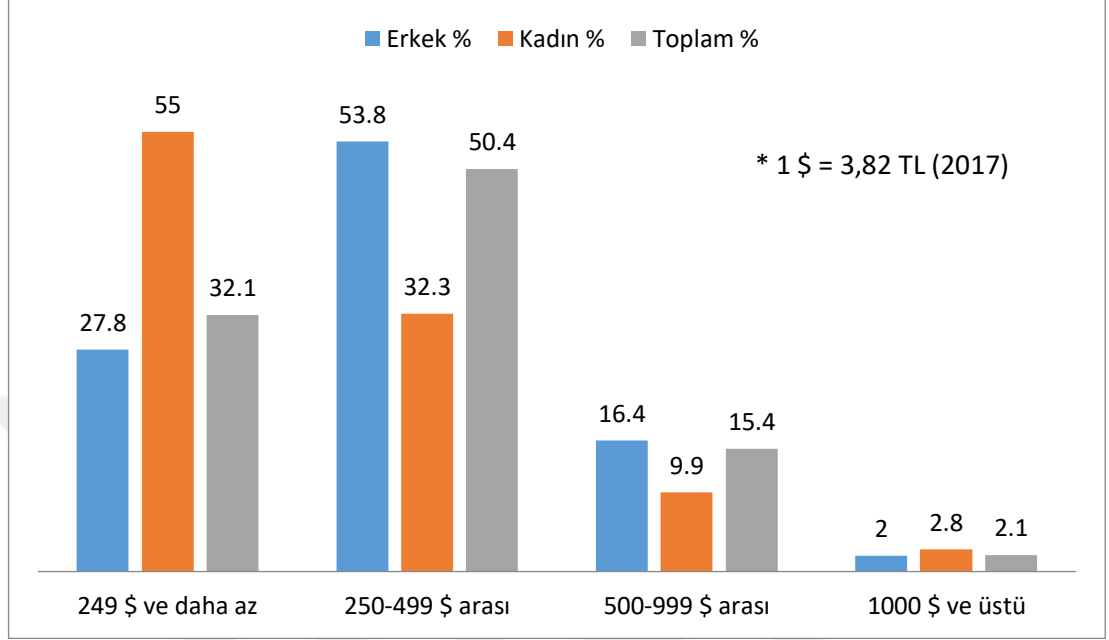
Suriyeli sığınmacıların Türkiye'de tüm çalışanlar içindeki oranı %2,9'dur. Bu oran her ne kadar küçük görünse de özellikle sığınmacıların yoğun olarak çalıştığı sektörlerde ciddi etkilere neden olmuştur. Bu etkilerin başında – büyük oranda kayıt dışı çalışıldığı için – daha düşük ücretlerle çalışılması sebebi ile işverenler açısından karlılık artışı ve aynı sektörlerde kayıt dışı çalışan T.C. vatandaşlarının işlerini kaybetme oranlarındaki artış gelmektedir (Caro, 2020: 14; Del Carpio ve Wagner, 2015).

15 Ocak 2016 itibariyle Türkiye'de yaşayan Geçici Koruma kapsamındaki Suriyeli sığınmacılara çalışma hakkı tanınmıştır. Buna göre en az 6 aydır Türkiye'de kayıt altında olan geçici koruma kapsamındaki yabancılar, bir iş yerinde çalışan T.C. vatandaşlarının 1/10'unu geçmeyecek şekilde ve kayıtlı oldukları illerde çalışma hakkına sahip olabilmektedirler. İstisna olarak, toplam çalışan sayısı 10'dan az olan işyerlerinde bir yabancıya çalışmasına izin verilmektedir ve işveren tarafından yabancıya çalıştırılacağı işi yapacak aynı nitelikte Türk vatandaşı bulunmadığının belgelendirilmesi halinde 1/10'luk istihdam kotası uygulanmayabilmektedir (Resmi Gazete, 2016). Çalışan Suriyeli sığınmacıların %58'inin kayıtlı olduğu ilin dışında çalıştığı tahmin edilmektedir (Caro, 2020: 38). Bu durumda çalışan Suriyeli sığınmacıların yarısından fazlasının çalışma izni alabilmesi mümkün görünmemektedir.

Aile, Çalışma ve Sosyal Hizmetler Bakanlığı tarafından yayınlanan "Yabancıların Çalışma İzinleri 2020" adlı rapora göre 2020 yılında 62.369 Suriye uyruklu kişiye çalışma izni verilmiştir (ÇSGB, 2020). Birleşmiş Milletler Mülteciler Yüksek Komiserliği'nin 2020 raporuna göre ise Türkiye'de çalışma izni verilen Suriyeli sığınmacı sayısı 132.497 kişidir (UNHCR, 2020b). Toplam çalışan Suriyeli sayısı düşünüldüğünde bu sayının oldukça düşük olduğu göze çarpmaktadır. 2017 itibari ile Türkiye'de 657.385 (15 yaş üzeri) Suriyeli çalışanın çalışma izninin olmadığı tahmin edilmektedir. Bu da Türkiye'de çalışan toplam kayıtlı çalışan sayısının %4,2'sine tekabül etmektedir. Çalışma izni olmadan Suriyeli sığınmacıların kayıt dışı çalışması sonucu başta SGK ve kayıt dışı çalışan Suriyeli sığınmacılar

ekonomik olarak zarara uğramaktadır. Bu durumdan en büyük fayda sağlayan kesimin ise kayıt dışı işçi çalıştıran işverenler olduğu görülmektedir (Caro, 2020).

AFAD (2017)'a göre Suriyeli sığınmacıların gelir dağılımı aşağıdaki gibidir:



Şekil 2.9. Suriyeli sığınmacıların gelir dağılımı

Şekil 2.9'a göre, Suriyeli sığınmacıların yarısının gelir dağılımı 250-499 \$ arasındadır. 2017 kuruna göre bu gelir 955-1906 TL arası bir gelire tekabül etmektedir. Cinsiyete göre bakıldığında ise gelir aralığının kadınlarda daha aşağı düştüğü ortaya çıkmaktadır. Suriyeli sığınmacı kadınların %55'inin geliri 249 \$'dan daha aşağıdadır.

Suriyeli sığınmacıların T.C. vatandaşlarından daha düşük bir gelire sahip olduğu ve düşük ücretli ve nitelik gerektirmeyen işlerde yoğunlaştıkları görülmektedir (Gençler, 2020: 132). Gençler (2020)'e göre, Suriyeli sığınmacıların yerli işgücü üzerinde ikame etkisi oluşturdukları söylenebilir. Başka bir deyişle, özellikle kayıt dışı sektörlerde çalışan T.C. vatandaşlarının yerini – daha düşük ücretlerle çalıştıkları için – Suriyeli sığınmacıların aldığı görülmektedir. Bu durum Suriyeli sığınmacıların yoğun olarak buldukları bölgelerde daha da fazla artmaktadır. Ceritoğlu vd. (2017)'ne göre bu işini kaybeden T.C. vatandaşlarının da %50'si çalışma faaliyeti dışında kalmış, %32'si işsiz statüsüne geçmiş, %18'i ise kayıtlı sektöre geçiş yapmıştır. Suriyeli sığınmacılar sebebi ile işgücü piyasasının dışında kalma oranı kadınlarda çok daha yüksektir.

Suriyeli sığınmacılar Türk Ticaret Kanunu'na göre Türkiye'de kendi işyerlerini kurma hakkına sahiptirler. Ticaret Bakanlığı tarafından yapılan açıklamaya göre 26 Şubat 2019 tarihi itibari ile en az bir ortağı Suriye uyruklu olan şirket sayısı 15.159'dur (CNN Türk, 2019). Kayıt dışı açılan firmalarla birlikte sayının daha yüksek olduğu tahmin edilmektedir.

### **2.2.5. Ekonomik Etki ve Yardımlar**

Suriyeli sığınmacıların Türkiye ekonomisine etkisi öncelikle kendini kamu harcamalarında göstermektedir. Türkiye'nin Suriyeli sığınmacılar için yapmış olduğu toplam harcamayı ifade eden son resmi açıklamalar, 2019 yılı itibari ile yapılan toplam harcamanın 40 milyar doları aştığı ve bu harcamaların eğitim, sağlık ve barınma ihtiyaçlarını karşılamak için gerçekleştirildiği şeklindedir (Burun ve Kanlı, 2019; TRT Haber, 2019). Suriyeli sığınmacı çocuklar Milli Eğitim Bakanlığı'na bağlı okullarda ücretsiz olarak eğitim görmektedirler. 2013 yılında alınan Bakanlar Kurulu kararı ile de Suriye uyruklu üniversite öğrencileri devlet üniversitelerindeki her türlü harçtan muaf tutulmaktadır (Resmi Gazete, 2013). Suriyeli sığınmacıların sağlık giderleri ise AFAD (Afet ve Acil Durum Yönetimi Başkanlığı) tarafından karşılanmakta, temel ve acil sağlık hizmetleri kapsamında alacak oldukları tedavi ve ilaçlardan katılım payı alınmamaktadır (Gençler, 2020: 126-127).

Her ne kadar 2019 yılı itibari ile yapılan toplam harcamanın 40 milyar doları aştığı belirtilmiş olsa da Almanya'nın Köln Üniversitesi'nin, bir mültecinin tüm maliyet kalemleri ile Almanya'ya olan maliyetini hesaplama amacı ile geliştirdiği formül Türkiye'ye uygulandığında yine aynı yıl itibari ile toplam maliyetin 230 milyar avroyu aştığı görülmektedir (Erdoğan, 2019: 14). Bir Suriyeli sığınmacının Almanya'ya olan bir günlük maliyeti ile Türkiye'ye olan bir günlük maliyetinin aynı olmama ihtimali ile birlikte, bugüne kadar yapılan toplam maliyet hesaplamalarında bu iki rakam arasında değişen farklı hesaplamalara da rastlamak mümkündür. Ancak tüm bu maliyet hesaplamalarına karşın, Suriyeli sığınmacılar ile ilgili olarak Türkiye'ye yapılan yardımların, hesaplanan bu maliyetlerin yanında oldukça küçük kaldığı görülmektedir. Örneğin 2019 yılında açıklanan 40 milyar dolarlık maliyete karşın, aynı açıklamada yardımların büyük bir kısmını oluşturan Avrupa Birliği kaynaklarından 2016 yılından bu yana sadece 3 milyar avroluk destek verildiği ifade edilmiştir (Burun ve Kanlı, 2019). 2022 yılı itibari ile ise Avrupa Birliği ülkeleri tarafından Türkiye'ye sığınmacılar için aktarılan toplam kaynağın 4,3 milyar avro

olduğu görülmektedir. 2016 yılı itibari ile Avrupa Birliği ülkeleri tarafından proje karşılığında gönderilmesi taahhüt edilen 6 milyar avro dikkate alındığında, Türkiye'nin proje karşılığında sığınmacılar için temin edebileceği 1,7 milyar avrosu daha bulunmaktadır. Buna ilaveten Avrupa Komisyonu Başkanı Ursula von der Leyen tarafından 2024 yılına kadar sığınmacılar için Türkiye'ye gönderilmek üzere ve yine projeye bağlı olarak ilave bir 3 milyar avro daha bütçe ayrılacağı duyurulmuştur. Sığınmacılar ile ilgili olarak Türkiye'ye yapılan yardımların büyük bir kısmını oluşturan Avrupa Birliği ülkelerinin yapmış oldukları bu yardımlar ile gerçekleştirilen projelerden bazıları aşağıdaki gibidir:

- Yabancılara yönelik sosyal uyum yardımı programı (SUY): Kızılay kart aracılığıyla belirli kriterleri sağlayan tüm sığınmacılara aylık 155 TL'lik nakit yardımınıdır. Yardımın %90'ı Suriyeli sığınmacılara yapılmaktadır ve 2021 yılı itibari ile programdan 1,3 milyon Suriyeli sığınmacı faydalanmaktadır. Program bu haliyle dünyanın en büyük yardım programıdır.
- Tamamlayıcı sosyal uyum yardımı programı (T-SUY): İş gücüne yönlendirilmesi mümkün olmayan en kırılgan bireylerin temel ihtiyaçlarını karşılayabilmeleri amacıyla faaliyete geçirilen yardım programıdır. 2021 yılı itibari ile 350 bin kişiye 250 TL'lik nakit desteği sağlamaktadır. Bu kişilerin %85'ini Suriyeli sığınmacılar oluşturmaktadır.
- PİKTES – Suriyeli çocukların Türk eğitim sistemine entegrasyonunun desteklenmesi projesi
- SIHHAT – Göçmen sağlığı merkezleri projesi
- Kayıtlı istihdama geçiş programı
- Eğitim bursları
- Geri gönderme merkezlerindeki yabancıların iletişim ihtiyaçları
- Okullar ve bu okullarda görev alan öğretmenler vs.

Avrupa Birliği ülkeleri tarafından yapılan bu yardımlar Suriyeli sığınmacıların yaklaşık olarak yarısını kapsamaktadır (Mülteciler Derneği, 2021a).

Avrupa Birliği ülkelerinin yapmış oldukları yardımlar dışında Avrupa Birliği, ABD ve Norveç tarafından okul çağındaki maddi imkanları kısıtlı ilkökul, ortaokul ve lise çağındaki sığınmacı öğrencilere 45 – 75 TL arasında değişen nakit yardımının yapıldığı Şartlı Eğitim Yardımı (ŞEY) adında bir program daha bulunmaktadır. Bu

programdan %85 oranında Suriyeli sığınmacılar yararlanmaktadır (Kızılay, 2021). 2021 yılı itibari ile programdan en az bir kez destek alan öğrenci sayısı 695.556 kişidir ve 2017- 2021 yılları arası dağıtılan toplam yardım miktarı 1 milyar 83 milyon TL'dir (Mülteciler Derneği, 2021b).

Birleşmiş Milletler Mülteciler Yüksek Komiserliği tarafından Suriye krizine bütüncül bir bakış açısı ile yaklaşmak amacı ile Suriyelilerin sığındığı beş komşu ülkenin ihtiyacı olan yardımları belirlemek ve bu yardımları organize etmek için 3RP (Regional Refugee & Resilience Plan [Bölgesel Mülteci ve Dayanıklılık Planı]) adında bir yardım planı tasarlanmıştır. Periyodik aralıklarla yenilenen bu yardım planının 2022 Mayıs raporunda, 2015'ten bu yana Türkiye'ye 4,8 milyar doları aşan bir yardımda bulunduğu belirtilmektedir (UNHCR, 2022b: 18). 3RP Planı çerçevesinde Avrupa Birliği ülkeleri tarafından yapılan yardımlar da zikredildiği için bu rapora bakarak Avrupa Birliği ülkeleri dışında, Türkiye'ye yapılan toplam dış yardımların (ŞEY programı da dahil olmak üzere) 500 milyon civarında olduğu hesaplanabilir. Bu yardımlar sadece hükümetler tarafından değil, çok sayıda uluslararası sivil toplum kuruluşu tarafından da yapılmaktadır.

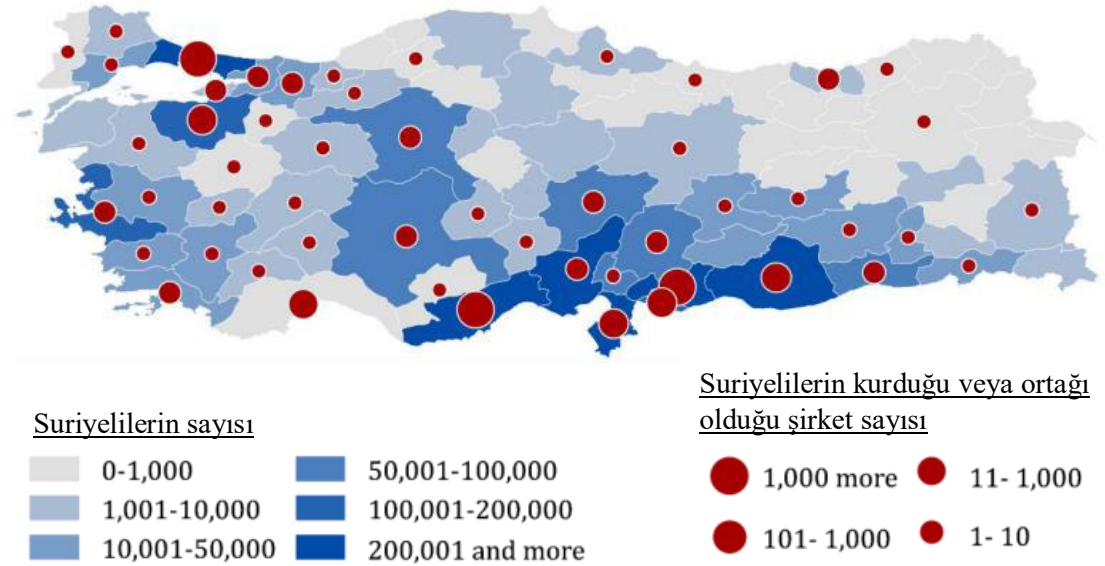
Suriyeli sığınmacılar için Türkiye'ye yapılan dış yardımların yanı sıra Türkiye'nin Suriyeli sığınmacılara yaptığı kendi iç yardımları da mevcuttur. Bu direkt yardımlar genellikle illerin valilikleri veya belediyelerin ilgili birimleri tarafından yapılmaktadır ve yardımlar şehirden şehire değişebilmektedir (BBC, 2021). Türkiye'nin kendi kurumları tarafından yapılan yardımlara bir örnek de Yurtdışı Türkler ve Akraba Toplulukları tarafından Suriye uyruklu üniversite öğrencilerine yapılan aylık 1200 TL'lik burs yardımudur. Bu yardım 2016 yılında başlamış olup 2020 yılı itibari ile toplam 12.531 Suriyeli uyruklu üniversite öğrencisine yapılmıştır (YTB, 2016; Hürriyet, 2020).

Suriyeli sığınmacıların Türkiye'ye olan ekonomik maliyeti ve yapılan yardımların dışında ekonomik olarak incelenmesi gereken bir diğer husus da Suriyeli sığınmacıların Türkiye'ye yapmış oldukları ekonomik katkıdır. Türkiye Ekonomi Politikaları Araştırma Vakfı (TEPAV)'nın 2018 yılsonu raporuna göre, 2013 – 2018 yılları arasında Suriyeli ortak sermayesi toplamda 1,2 milyar TL olan 7906 şirket kurulmuştur (TEPAV, 2019a: 1). Birleşmiş Milletler Mülteci Örgütü de 2018 sonu itibari ile en az bir ortağı Suriyeli olmak üzere kurulan şirket sayısının 7000'in üzerinde olduğunu doğrulamakta ve 2018 sonu itibari ile Suriyeli sığınmacıların

Türkiye’deki girişimcilik sermayelerinin 400 milyon doları aştığını belirtmektedir (UNHCR, 2019). Suriyelilerin Türk bankalarındaki mevduatları ise Bankacılık Düzenleme ve Denetleme Kurumu (BDDK) verilerine göre 2016 yılı itibari ile 408 milyon doları aşmış durumdadır (Kuyumcu ve Kösematoğlu, 2017: 80).

2019 sonu itibari ile “Çalışma Hayatı” başlığı altında daha önce de belirtildiği üzere Türkiye’de en az bir ortağı Suriyeli olan kayıtlı toplam şirket sayısı Ticaret Bakanlığı tarafından 15.159 olarak açıklanmıştır. Bu şirketlerde istihdam edilen kişi sayısı ise 100.000 civarında olarak tahmin edilmektedir (Özaydın ve İlğazi, 2019: 147; Yağcı, 2018).

TEPAV (2019b)’a göre, Suriyeli girişimciler Türkiye’deki 3,5 milyon Suriyeli sığınmacının %7’sine geçim kaynağı sağlamaktadır. Ortalama Suriyeli hane halkı büyüklüğünün 6 kişiden oluştuğu düşünüldüğünde, Suriyeli girişimciler tarafından kurulan bu şirketlerden yaklaşık 250.000 Suriyeli sığınmacının faydalandığı hesap edilebilir (TEPAV, 2019b). 2011 yılında yabancı sermayeli işletmelerin sadece %2’sini Suriyeliler oluştururken, bu oran 2020 yılı itibari ile %20’lere yükselmiştir (Gençler, 2020: 137).



Şekil 2.10. Suriyeli şirketlerin Türkiye’ye dağılımı (TEPAV & EBRD, 2018)

Şekil 2.10’da görüldüğü gibi, Suriye sermayeli şirketlerin Türkiye’ye dağılımı, Suriyeli sığınmacıların Türkiye’ye dağılımı ile paralellik göstermektedir. Suriyeli girişimcilerin kurdukları veya ortağı oldukları şirketler, Suriye sınır hattı boyunca ve

İstanbul başta olmak üzere yoğunlukla batı bölgelerindeki bazı büyükşehirlerde yoğunlaşmıştır.

Suriyeli sığınmacıların Türkiye ekonomisine bir diğer etkisi “Çalışma Hayatı” başlığı altında da incelendiği üzere işsizliğe olmuştur. Bu alanda yapılan çalışmaların büyük bir çoğunluğu, Suriyeli sığınmacıların özellikle kayıt dışı sektörlerde ve nitelik gerektirmeyen işlerde kayıt dışı çalışan T.C. vatandaşlarının yerlerini aldığını göstermektedir (Gençler, 2020: 132; Ceritoğlu vd., 2017: 5; Tümen, 2016: 2; Topuz, 2019: 79). Bu durum maliyetleri düşürmesi sebebi ile fiyatlara olumlu yansırken, nüfusun milyonlar seviyesinde artmasının talebi arttırması sebebi ile de olumsuz yansımaktadır. Bu alanda yapılan çalışmaların birçoğu, bu iki karşıt etkide sonuç itibari ile fiyatların artmasının daha baskın geldiğini ileri sürmektedir (Topuz, 2019: 79).

Özaydın ve İlgazi (2019)’ye göre Suriyeli sığınmacıların ekonomiye hem olumlu hem de olumsuz etkileri bulunmaktadır. Kira ve konut fiyatları başta olmak üzere genel olarak enflasyondaki artış, kayıt dışı istihdamın büyük oranda artması ve bu durumun haksız rekabete, ücretlerde ve vergi gelirlerinde düşüşe neden olması, sığınmacılar arasında çocuk işçiliğinin artması ve kamu harcamalarındaki artış olumsuz etkiler olarak sayılabilir iken; Suriyeli sığınmacıların ihracat ve vergi gelirlerinde meydana getirdiği artışlar, yeni pazarlara ulaşabilme imkanının artması, yeni iş fikirleri ve üretim modellerinin ortaya konması, döviz arzı ve doğrudan yabancı sermaye yatırımları olumlu etkiler olarak sayılabilmektedir (Özaydın ve İlgazi, 2019: 158).

#### **2.2.6. Kamu Hizmetlerine Etkisi**

Suriyeli sığınmacıların Türkiye nüfusuna oranı %4,4 olarak ifade edilir iken bu oranın 0-24 yaş grubu için %7 olması dikkat çekicidir (TÜİK, 2022a; GİB, 2022a; Mülteciler Derneği, 2022). Ayrıca Suriyeli sığınmacıların doğurganlık hızının Türkiye ortalamasından yaklaşık olarak 3 kat daha fazla olması, bu oranın her geçen gün Suriyeli sığınmacılar lehine değiştiğini göstermektedir. Bu kadar büyük orandaki bir sığınmacı nüfusunun ve bu nüfus içerisindeki yüksek genç nüfus oranının Türkiye için çok ciddi etkilere neden olacağı aşıkardır. Yukarıda ekonomi ve işgücü piyasasına etkileri incelenen Suriyeli sığınmacıları etki ettiği bir diğer başlık kamu hizmetleridir.

Suriyeli sığınmacıların kamu hizmetlerindeki etkisi öncelikli olarak ücretsiz olarak hizmet aldıkları eğitim ve sağlık hizmetlerinde kendini göstermektedir. 5-17 yaş arası 730.806 Suriyeli sığınmacı öğrenci Milli Eğitim Bakanlığı'na bağlı okullarda, 47.482 Suriye uyruklu üniversite öğrencisi de yükseköğrenim kurumlarında eğitim görmektedir. Sisteme bu denli büyük ve Türkçe bilmeyen bir öğrenci kitlesinin ani ve hızlı bir şekilde girişi özellikle Milli Eğitim Bakanlığı'na bağlı okullarda ve sığınmacıların yoğun olarak buldukları bölgelerde sınıfların ani bir şekilde kalabalıklaşmasına, okul kapasitelerinin aşılmasına ve eğitim-öğretimin kalitesinin düşmesine neden olmuştur.

Eğitim hizmetlerine benzer bir durum, sağlık hizmetlerinde de yaşanmıştır. Aynı şekilde bu denli büyük bir nüfusun sağlık sistemine ani bir şekilde girişi sığınmacıların özellikle yoğun olarak yaşadıkları bölgelerde sağlık kurumlarının bir anda kalabalıklaşmasına ve sağlık hizmetlerinde düşüşe neden olmuştur. Ortadoğu Stratejik Araştırmalar Merkezi (ORSAM)'nin Ocak 2015 raporunda Suriyeli sığınmacıların yoğun olarak buldukları bölgelerde sağlık hizmetlerinde ciddi sıkıntıların yaşandığı ve yerel halkın bu duruma olumsuz tepkiler geliştirdiği rapor edilmektedir. Sağlık kurumlarının kalabalıklaşması ve hizmet kalitesinin düşmesi dışında kamu sağlığını olumsuz etkileyen bir başka husus, Türkiye'de artık sorun olmaktan çıkmış, aşuları dahi kaldırılmış olan çocuk felci, şark çıbanı grubu, kızamık gibi bazı hastalıkların tekrar ortaya çıkmasıdır. Bu sebeple sınır illerinde 0-5 yaş arası çocukların tamamına çocuk felci aşısı yapılmıştır (Orhan ve Gündoğar, 2015: 20).

Suriyeli sığınmacılara sağlanan sağlık hizmetleri ile ilgili olarak yapılan son resmi açıklamaya göre 2022 yılı itibari ile 29 ilde 185 Göçmen Sağlığı Merkezi faaliyete geçirilmiştir. Bu merkezlerde 787 hekim, 34 diş hekimi ve 1149 hemşire olmak üzere yaklaşık 4000 Suriyeli sağlık çalışanı görev yapmaktadır. Göçmen sağlığı merkezleri dışında sağlık hizmetine ihtiyaç duyulması halinde doğrudan sağlık hizmeti verilmektedir. Bu kapsamda bugüne kadar Suriyelilere yaklaşık 97 milyon poliklinik hizmeti ve 3 milyonun üzerinde yataklı tedavi hizmeti verilmiş; 2,6 milyon ameliyat gerçekleştirilmiş ve 754.000 Suriyeli bebek sağlık tesislerinde dünyaya gelmiştir (Rakipoğlu ve Kara, 2022).

En çok etkilenen kamu hizmetlerine bir başka örnek olarak belediye hizmetleri verilebilir. Sığınmacıların bulunma oranları ile de orantılı olarak belediyelere

özellikle toplu taşıma, trafik, temizlik, çöp toplama, su temini/dağıtımını, alt yapı hizmetleri gibi konularda ekstra bir yük binmiştir (Orhan ve Gündoğar, 2015: 20).

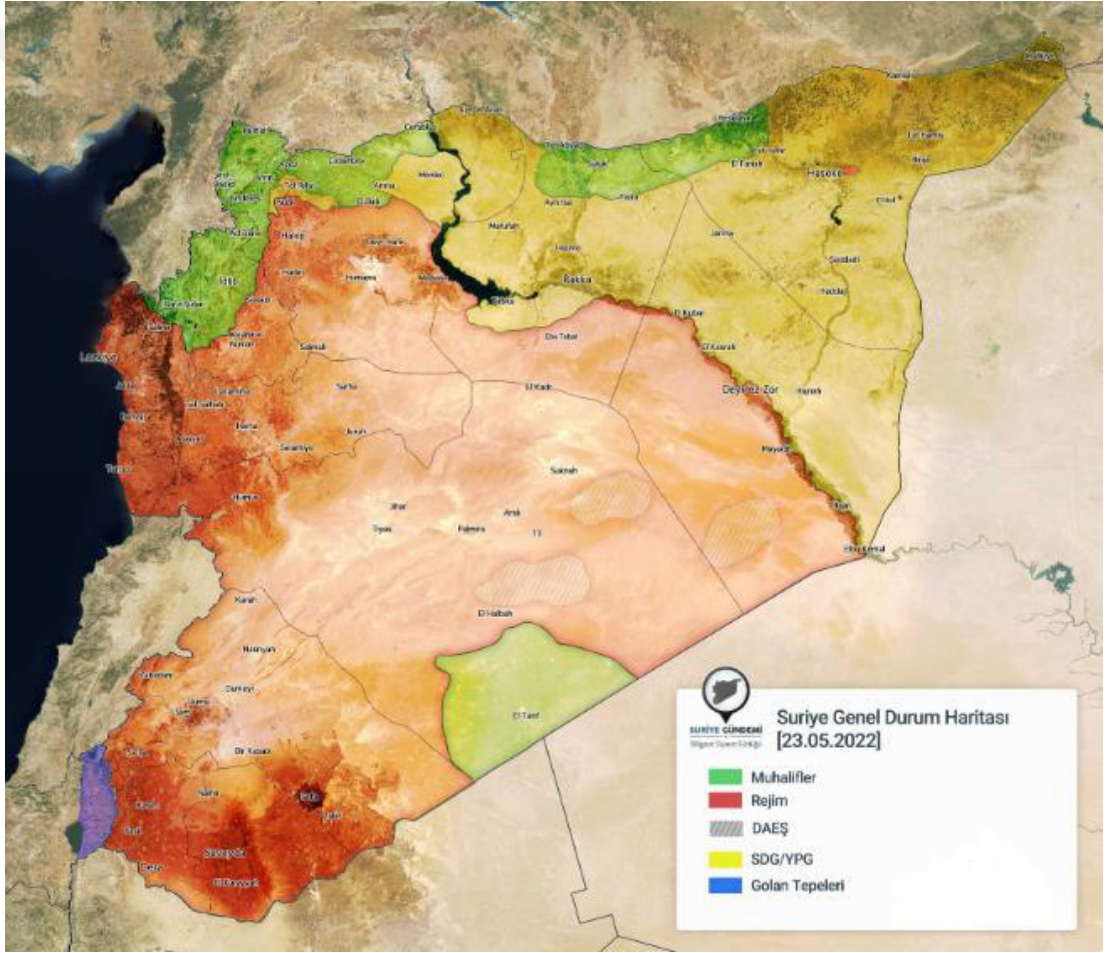
### **2.2.7. Güvenlik**

Suriyeli sığınmacıların etki ettiği bir diğer husus güvenlik meselesidir. Güvenlik hem yurt içi hem de yurt dışı güvenlik olarak düşünülebilir. Yurt içi güvenlik ile Suriyeli sığınmacıların ülke içinde bir güvenlik sorunu oluşturup oluşturmadığı; yurt dışı güvenlik ile de bu denli büyük bir Suriyeli sığınmacı kitlesinin buldukları bölgeleri terk edip Türkiye'ye göç etmesinin Türkiye açısından bir dış güvenlik sorunu oluşturup oluşturmadığına odaklanılmaktadır.

Yurt içi güvenlik meselesine odaklanıldığında ilk olarak bakılması gereken yer suç kayıtlarıdır. İçişleri Bakanlığı tarafından yapılan açıklamaya göre, 2014-2017 yılları arasında Suriyelilerin karıştığı olayların Türkiye'de gerçekleşen toplam asayiş olaylarına oranı yıllık ortalama %1,32'dir. Açıklamada ayrıca bu olayların önemli bir kısmının Suriyelilerin kendi aralarındaki anlaşmazlıklar olduğu ifade edilmiştir (İçişleri Bakanlığı, 2017). Bu oran 2018 yılında 1,46 olarak güncellenmiş ve Suriyelilerin karıştığı olayların aydınlatılma ve failerin yakalanma oranı %91,4 olarak belirtilmiştir (Bulur, 2018). 2022 yılına gelindiğinde ise Suriyeli sığınmacıların suç işleme oranı %1,3 olarak ifade edilmiştir. Aynı açıklamada T.C. vatandaşlarının suça karışma oranı ise %2,2 olarak belirtilmiştir (Aykırı, 2022). 2022 Nisan tarihi itibari ile ayrıca 2016 yılından buyana asayiş sorunları sebebi ile sınır dışı edilen Suriyeli sayısı 19.336 olarak açıklanmıştır (DHA, 2022).

Dış güvenlik meselesine gelindiğinde ise Suriye'nin – özellikle Kuzey Suriye'nin – kısa tarihine bir göz atmak gerekmektedir. Suriye'de Arap Baharı'nın etkilerinin başladığı 2011 yılına kadar örgütlenmesini tamamlamış olan ve Türkiye tarafından terör örgütü PKK'nın Suriye kolu olarak tanımlanan PYD (Demokratik Birlik Partisi) ve onun silahlı kanadı YPG, Suriye'de isyanın başlaması ve Esed rejiminin muhaliflere karşı daha güçlü durabilmek adına Suriye'nin kuzeydoğusundan çekilmesi üzerine Suriye'nin kuzeyinde yer alan Afrin, Ayn el-Arap (Kobani) ve Haseke gibi bölgelerde fiili olarak yönetimi ele geçirir. Bu dönem aynı zamanda Türkiye tarafından terör örgütü olarak tanınan bir başka örgüt olan IŞİD'in Suriye'nin önemli bir bölümünü ele geçirdiği ve YPG'nin kontrol ettiği kuzeydeki bölgelere dayandığı bir dönemdir. IŞİD'in Ayn el-Arap'ı kuşatması üzerine ABD tarafından destek verilen YPG, IŞİD'i yenmeyi başarır. Bu tarihten

itibaren IŞİD'in ele geçirdiği topraklar ABD'nin de desteği ile YPG tarafından ele geçirilmeye başlanır. Bunların içinde Arap nüfusun yoğun olarak yaşadığı Tel Abyad ve Münbiç gibi şehirler de bulunmaktadır. YPG ele geçirdiği topraklardan kendilerine muhalefet eden Arap ve Türkmenleri tehcir ederek Suriye'nin kuzeyini tek bir hat şeklinde birleştirmeyi hedefler (Acun ve Keskin, 2016: 12). Ancak Türkiye tarafından, Esed rejimine muhalif olarak kurulmuş olan Özgür Suriye Ordusu ile birlikte gerçekleştirilen Fırat Kalkanı, Zeytin Dalı ve Barış Pınarı hareketleri ile Afrin, El-Bab, Tel Abyad ve Rasulayn gibi şehirler ele geçirilerek bu plan bozulur. 2022 Haziran itibari ile henüz YPG kontrolünde bulunan Münbiç ve Tel Rıfat şehirleri için de Türk resmi makamlarınca operasyon sinyali verilmektedir.



Şekil 2.11. Suriye genel durum haritası (Suriye Gündemi, 2022)

23 Mayıs 2022 tarihi itibari ile Şekil 2.11.'deki haritada görülen yeşil kısımlar Türkiye destekli muhaliflerin, sarı kısım YPG'nin, kırmızı kısım ise Esed rejiminin kontrolünde olan bölgeleri göstermektedir. Yeşil kısımlardan en sol tarafta yer alan

İdlip, tamamen Türkiye destekli muhaliflerden oluşmamakla birlikte çok sayıda muhalif örgütün bir arada bulunduğu bir bölgedir. Sol taraftaki yeşil alanın altında kalan sarı kısım Tel Rıfat, onun sağındaki Fırat nehrine kadar olan sarı bölge Münbiç ve her iki yeşil alanın arasında kalan kısım da Ayn el-Arap şehirleridir.

Görüldüğü üzere Suriye meselesi sadece sığınmacı bakımından değil aynı zamanda milli güvenlik meselesi olarak da Türkiye'nin karşısına çıkmaktadır. Dış güvenlik meselesinin sığınmacıları ilgilendiren tarafı ise sığınmacıların bir kısmının YPG tarafından ele geçirilen bölgelerden göç etmiş olması ve hükümet politikası olarak Suriye'nin kuzeyini güvenli bir hale getirip sığınmacıların bir kısmını o bölgelere yerleştirme düşüncesidir. Böylelikle Suriye'nin kuzeyinin Türkiye açısından da daha güvenli bir hale getirilmesi planlanmaktadır.

### **2.2.8. Uyum**

Suriyeli sığınmacılar ile ilgili olarak irdelenmesi gereken bir başka husus, sığınmacıların Türkiye'ye ne oranda uyum sağladıkları ve Türk halkı tarafından ne oranda kabul gördükleridir. Bu konuyu irdellemek için daha önce yapılan belli başlı çalışmalara bakıldığında, Suriyeli sığınmacılar tarafında Türkiye'ye uyum sağlama konusunda genel olarak olumlu bir görüşün yer aldığı, ancak Türk toplumu tarafında genel olarak olumsuz bir görüşün yer aldığı görülmektedir.

Erdoğan (2020)'nin örneklem yöntemi ile çalıştığı "Suriyeliler Barometresi 2019" adlı çalışmada, Suriyeli sığınmacıların çoğunun Türk toplumuna/Türkiye'ye uyum sağladığına dair bir görüş içerisinde oldukları tespit edilmiştir. Çok az ya da hiç uyum sağlayamadığını düşünenlerin oranı %8,5'tir (Erdoğan, 2020: 165). Aynı çalışmada T.C. vatandaşlarının %64,4'ü ise Suriyeli sığınmacıların Türk toplumuna/Türkiye'ye uyum sağlayamadığını düşünmektedir (Erdoğan, 2020: 104).

AFAD'ın 2017 yılında gerçekleştirmiş olduğu benzer bir çalışmada Suriyeli sığınmacılara yöneltilen "Türkiye'deki sosyal hayata ne derece uyum sağlayabildiğinizi düşünüyorsunuz?" sorusuna sığınmacıların %50'si "uyum sağladım" veya "tamamıyla uyum sağladım", %28,6'sı "kararsızım", %21,4'ü de "uyum sağlayamadım" veya "hiç uyum sağlayamadım" cevaplarını vermişlerdir (AFAD, 2017: 115). Aynı çalışmada "Kendinizi toplumun bir parçası gibi hissediyor musunuz?" sorusuna ise, %46,4'ü "hissediyorum" veya "tamamıyla hissediyorum",

%20,6'sı “kararsızım”, %33'ü ise “hissetmiyorum” veya “hiç hissetmiyorum” cevaplarını vermişlerdir (AFAD, 2017: 118).

Suriyeli sığınmacıların genel olarak Türkiye'ye uyum sağladıklarına dair bir algıya sahip olmaları ile paralel olarak, aynı zamanda - savaş bitse dahi - Türkiye'de kalmayı tercih edeceklerine dair bir görüş içerisinde oldukları da görülmektedir. Erdoğan (2020)'a göre Suriyeli sığınmacıların %51,8'i Suriye'ye dönmeyi hiçbir şekilde düşünmemektedir. %30,3'ü savaşın bitmesi ve kendilerinin istediği şekilde bir yönetim oluşması halinde döneceğini; %5,5'i savaşın bitmesi halinde kendilerinin istediği bir yönetim oluşmasa da döneceğini; %5,9'u ise Suriye'de güvenli bölge oluşursa döneceğini ifade etmiştir (Erdoğan, 2020: 176-177). Ancak aynı çalışma T.C. vatandaşlarının Suriyeli sığınmacıların Türkiye'de kalmasına bu kadar sıcak bakmadığını göstermektedir. T.C. vatandaşlarından Suriyeli sığınmacılar ile birlikte yaşamayı arzu etmeyenlerin oranı %87,2 olarak tespit edilmiştir (Erdoğan, 2020: 97).

Sığınmacılar açısından bir başka ülkeye yapılan göç sonucu uyumun en üst seviyede gerçekleştiğini gösteren en büyük göstergelerden biri de vatandaşlıktır. Türk resmi makamlarınca 2022 yılı itibari ile Suriyeli sığınmacılardan T.C. vatandaşlığına geçenlerin sayısının 200 bini aştığı açıklanmıştır. Yine araştırmalar geri kalan Suriyeli sığınmacıların da T.C. vatandaşlığına sıcak baktığını göstermektedir (İNGEV, 2017). Ancak T.C. vatandaşlarının büyük bir bölümünün vatandaşlık konusunda da Suriyeli sığınmacılar ile aynı görüşü paylaşmadığı, Suriyeli sığınmacıların vatandaşlığına büyük oranda karşı çıktığı görülmektedir (Erdoğan, 2020: 205).

Uyumun önündeki bir diğer engel de dildir. Bilindiği üzere Suriyeli sığınmacıların ana dili Arapça'dır. AFAD (2017)'in Suriyeli sığınmacılara yönelmiş olduğu “Türkçeyi ne ölçüde konuşabiliyorsunuz?” sorusuna, %31,9'u “ne iyi ne kötü”, %24,4'ü “kötü”, %19,5'i “iyi”, %11,6'sı “çok kötü”, %8,3'ü “hiç”, %4,3'ü “çok iyi” cevaplarını vermiştir (AFAD, 2017: 119).

Aynı dinin mensubu olmak her ne kadar uyum için bir avantaj olsa da Suriyeli ve Türkler arasında belirgin kültür farklılıkları da bulunmaktadır. Küçük yaşta kız çocuklarını evlendirme örneğinin Türklerin aksine Suriyelilerde yaygın olarak görülen bir uygulamadır (Şentürk, 2020: 26). Bunun yanı sıra kültürel yaşam, sosyal hayat, gelenekler ve kılık-kıyafet bakımından da belirgin farklılıklar olduğu Suriyeli sığınmacıların kendisi tarafından ifade edilmiştir (AFAD, 2017: 117).

Suriyeli sığınmacıların büyük bir çoğunluğu Türkiye'ye göç ettiğinden beri düşük vasıflı işlerde çalışmakta ve bu da onların yoksul mahallelerde yoğunlaşmasına neden olmaktadır. Bugüne değin Türkler ve Suriyeliler arasında her ne kadar bazı çatışma vakaları söz konusu olsa da henüz kitlesel ve büyük çapta bir çatışma vaki değildir. Ancak bu denli büyük bir kitle için bu gettolaşma ve sosyo-ekonomik olarak en alt katmanda yer alma durumunun uzaması, gelecek nesiller için ciddi bir çatışma riskini de artırmaktadır (Caro, 2020: 16).



### 3. DUYGU ANALİZİ – KAVRAMLAR, TEKNİKLER VE UYGULAMA SÜRECİ

Bu bölümde tezin son bölümünde gerçekleştirilecek olan araştırma kapsamında kullanılacak olan duygu analizi yöntemine dair tüm kavram, teknik ve süreç açıklanacaktır.

#### 3.1. Duygu Analizi

Literatürde duygu analizi ile ilgili olarak yer alan çeşitli tanımlar aşağıdaki gibidir:

Duygu analizi temel olarak bir metin işleme (text processing) işlemi olup verilen metnin duygusal olarak ifade etmek istediği sınıfı belirlemeyi amaçlar (Şeker, 2016:21). Fikir madenciliği olarak da adlandırılan duygu analizi, insanların belirli bir konu, fikir, ürün, hizmet, kişi, kurum veya olay hakkındaki görüşlerini, düşüncelerini, duygularını, değerlendirmelerini veya tutumlarını analiz eden bir çalışma alanıdır (Liu, 2012: 7).

Duygu analizi, elektronik ortamdaki yazılı metin içeriklerinde, makine öğrenmesi tabanlı veya istatistiksel yöntemlerle çeşitli analizler yapılarak insani duyguların tespit edilerek sınıflandırılması işlemidir (Poria vd., 2018; akt. Ataman, 2020: 10).

Duygu analizi (görüş madenciliği), doğal dil işleme, istatistik, bilgisayar bilimleri gibi alanlardan yöntem ve tekniklerin kullanılması ile görüş sahibinin metin içerisinde belirttiği duygu, görüş, tutum gibi öznel bilgilerin belirlenmesini amaçlayan güncel bir araştırma alanıdır (Onan, 2017: 3).

Liu (2017)'ye göre, duygu analizi ile ilgili çalışmalar daha öncesine dayansa da “duygu analizi” tabirine ilk olarak Nasukawa ve Yi (2003)'nin “Sentiment analysis: Capturing Favorability Using Natural Language Processing (Duygu Analizi: Doğal Dil İşlemeyi Kullanarak Elverişliliği Yakalama)” adlı çalışmasında rastlandığı söylenebilir. Dil bilimi ve Doğal Dil İşleme'nin uzun bir geçmişi olsa da 2000 yılı öncesine bakıldığında insanların duygu ve düşünceleri ile ilgili olarak az sayıda çalışmaya rastlanmaktadır. 2000 yılından sonra ise bu araştırmaların büyük bir hızla arttığı görülmektedir. Bunun çeşitli sebeplerinden söz edilebilecek olsa da en önemli

sebebinin, internet ortamında insanların duygu ve düşüncelerinin gittikçe artan oranda yer almaya başlaması olduğu söylenebilir (Liu, 2012: 7).

İnsanların herhangi bir konu hakkındaki görüşleri insanlık tarihi boyunca önem arz etmiştir. Ancak demokrasinin genel olarak kabul görmesi ve kitle iletişim olanaklarının artması ile birlikte, insanların bireysel olarak görüşleri tarihin hiçbir döneminde olmadığı kadar önemli ve bilimsel olarak ölçülebilir hale gelmiştir. Ülkelerin gelişmişlik düzeylerine göre değişmekle birlikte, bugün insanların fikirlerini beyan etmek için kullandıkları platformların başında internet ve daha özelinde sosyal medyanın geldiği söylenebilir. Sosyal medya içinde de kamuya açık görüş beyan etme amacı ile yapılan kullanımlarda Twitter'ın başı çektiği görülmektedir. Bu sebeple duygu analizi araştırmalarına bakıldığında Twitter'ın diğer sosyal medya platformlarına nazaran daha fazla tercih edildiği görülmektedir.

Duygu analizi, manuel olarak bir insan tarafından – en azından makul bir süre içerisinde – işlenemeyecek büyüklükte bir verinin bilgisayar yardımı ile çeşitli programlar ve algoritmalar kullanılarak işlenmesi ve bundan anlamlı sonuçlar çıkarılmasını içermektedir. Burada sözü edilen “makul bir sürede işlenemeyecek büyüklükte veri” tabiri karşımıza “büyük veri” kavramını çıkarmaktadır. Duygu analizi çalışmalarına baktığımızda üzerinde çalışılan verinin “büyük veri” olarak tanımlanabilecek büyüklükte olduğunu görürüz. Büyük veri içerisinde amaca hizmet edecek yararlı verinin çıkarılması, işlenmesi ve anlamlı sonuçlar üretilmesi de karşımıza “veri madenciliği” kavramını çıkarmaktadır. Duygu analizi kapsamında toplanan ve işlenen verinin “metin” formatında olması hasebiyle öncelikle bu verinin yapısal bir hale getirilmesi (sayısallaştırılması) gerekmektedir ki bu da metin madenciliği yöntemlerini gerektirmektedir. Aşağıda tüm bu dile getirilen kavramlar ayrı ayrı mercek altına alınmıştır.

### **3.2. Büyük Veri**

Veri dediğimiz dijital bilgilerin hacmi çok farklı formatlarda olmak üzere her geçen gün daha fazla artmakta ve bu verilerin analiz edilerek içerisinde anlamlı bilgilerin ortaya çıkarılması her geçen gün daha da karmaşık bir hale gelmektedir. “Büyük veri” dediğimiz bu devasa bilgi yığını her gün Facebook, Youtube, Twitter gibi sosyal medya ortamlarının yanı sıra lojistik, telekomünikasyon, kamu hizmetleri, sağlık hizmetleri, dijital eğlence medyası gibi çok farklı sektörler tarafından da beslenmekte, hacmi üstel olarak artacak şekilde büyümektedir. 2022 yılı sonu

itibariyle büyük verinin hacminin 97 zettabayt (1 zettabayt  $10^{12}$  gigabayt) seviyesine ulaşacağı tahmin edilmektedir (Statista, 2022). 2010 yılında bu seviyenin 2 zettabayt olduğu dikkate alındığında büyük verinin artış hızının üstel grafiği çok daha net bir şekilde görülebilmekte ve gelecekte ulaşabileceği olası noktalar daha sağlıklı bir şekilde tahmin edilebilmektedir.

Türkiye Bilimler Akademisi tarafından hazırlanan Türkçe Bilim Terimleri sözlüğünde büyük verinin tanımı şu şekilde yapılmıştır: “Toplanıp işlenecek verilerin çok büyük hacimleri, biçim ve içerik bakımından fazlasıyla farklılık ve değişkenlik içermeleri, büyük bir hızla birikmeleri ve geçerlilik düzeylerinin de değişkenliği nedeniyle geleneksel veri işleme yazılım ve yöntemlerinin yetersiz kaldığı, ancak bir yandan içerik zenginliği nedeniyle, diğer yandan da hızla gelişen teknolojik altyapılar sayesinde 21. yüzyılda önemi gittikçe artan kitlesel ve sürekli bilgi toplama, işleme ve karar desteği sağlama yaklaşımlarına toplu halde verilen ad” (TÜBA, 2022).

Büyük veri, matematik alanında istatistik ve optimizasyon; bilgisayar bilimleri alanında algoritmalar, programlar, makine öğrenmesi, veri madenciliği; alan bilimlerinde ise işletme, görselleştirme vd. uygulamaların iç içe girdiği bir çoklu disiplinler çalışma alanıdır (Buyya vd., 2016; Chen vd., 2019; akt. Ataman, 2020: 22). Bir başka tanımıyla büyük veri, makine öğrenmesi alanları (veri, enformasyon, deneyimlenmiş bilgi, zeka) ile bulut bilişimin iç içe geçtiği bir uygulama alanıdır (Anderson, 2015; Buyya vd., 2016; Sathi, 2012; akt. Ataman, 2020: 23).

Günümüzde bilgi; emek, sermaye ve doğal kaynakların yanı sıra yeni bir üretim faktörü olarak yerini almış ve organizasyonların rekabet avantajı sağlama açısından en önemli kaynak unsuru hâline gelmiştir. Dolayısıyla günümüzde bilginin en önemli kaynağı olan büyük veri de aynı oranda önem kazanmıştır. Ancak büyük verinin faydalı bilgiye dönüşebilmesi için ileri analitik yöntemlerle işlenmesi gerekmektedir (Aktan, 2018: 1).

### **3.2.1. Büyük Veri Bileşenleri**

Büyük verinin bileşenleri ile ilgili olarak bugüne kadar çeşitli tanımlar yapılmıştır. İlk tanım Laney (2001) tarafından yapılan ve kısaca 3V olarak adlandırılan tanımdır. 3V'nin ilkinin “variety” (çeşitlilik), ikincisini “velocity” (hız) ve üçüncüsünü “volume” (hacim) oluşturur. Bunlara ek olarak çeşitli kişi ve

kurumlar tarafından günümüze kadar çeşitli ilaveler yapıldıysa da yaygın olarak kabul göreninin “veracity” (gerçeklik) ve “value” (değer) bileşenlerinin de dahil olduğu 5V şeklindeki ifade olduğu söylenebilir (Debattista, Lange, Scerri ve Auer, 2015: 92; Gandomi ve Haider, 2015: 139; Zainal, Hussin ve Nazri, 2016: 305; akt. Aktan, 2018: 4)

### **3.2.1.1.Çeşitlilik**

Büyük veri, geniş bir yelpazede herhangi bir türde ve formatta olabilmekte ve bu veri türleri arasında herhangi standart bir dizi veya kural bulunmamaktadır (Aktan, 2018: 4). Bu çeşitlilik en genel anlamıyla yapısal, yarı yapısal ve yapısal olmayan olarak üçe ayrılmaktadır.

1. Yapısal veri: İşlenmesi ve saklanması en kolay veri türüdür. Veri tabanı yönetim sistemlerinden üretilen veriler yapısal veriye örnek olarak verilebilir (Ataman, 2020: 24).
2. Yarı yapısal veri: Bir yönüyle yapısal bir yönüyle de yapısal olmayan özellikler gösteren veridir. XML (Extensible Markup Language) ve JSON (JavaScript Object Notation) programlama dilleri yarı yapısal veriye örnek olarak verilebilir.
3. Yapısal olmayan veri: Büyük bir çeşitlilik gösterirler ve analizi diğer veri türlerine göre daha zordur. Hem insanlar hem de makineler tarafından üretilebilirler. Twitter, Facebook, Youtube vb. sosyal medya platformları vasıtasıyla üretilen tüm metin, fotoğraf, video, ses formatındaki veriler, mobil cihazlar veya bilgisayarlar vasıtasıyla elle yazılan tüm dokümantasyonlar, web sitesi içerikleri insanlar tarafından üretilen yapısal olmayan verilere; uydu görüntü verileri, sismik cihaz veya atmosfer verileri gibi bilimsel veriler, güvenlik veya trafik kontrolünde kullanılan kameralardan üretilen video veya resim verileri, radar sistemlerinden üretilen veriler makineler tarafından üretilen yapısal olmayan verilere örnek olarak verilebilir (Ataman, 2020: 24). Büyük verinin %95’ini yapısal olmayan veri oluşturmaktadır (Gandomi ve Haider, 2015: 143).

### **3.2.1.2.Hız**

Büyük verinin bir diğer özelliği de büyük bir hızla üretilmesidir ve gün geçtikçe de bu hız üstel olarak artmaktadır. Örneğin 2022 yılı itibari ile Facebook’ta günde 5,7 milyar kez beğenme butonuna tıklanmakta ve 1 milyar hikaye paylaşılmakta, Instagram’da günde 4,2 milyar kez beğenme butonuna tıklanmakta ve 500 milyon hikaye paylaşılmakta, Twitter’da günde 500 milyon tweet atılmaktadır

(Bagadiya, 2022). Whatsapp'ta her gün 100 milyardan fazla mesaj atılmakta (Dean, 2022) ve 2022 yılı itibari ile günde ortalama 333,2 milyar e-posta atılacağı tahmin edilmektedir (Johnson, 2021).

### **3.2.1.3. Hacim**

Verinin bellekte kapladığı alanı ifade eder. Daha önce belirtildiği üzere büyük veri artık geleneksel veri işleme teknikleri ile işlenemeyecek bir hacme ulaşmıştır ve bu hacim üstel olarak artmaya devam etmektedir. Bu hacmi besleyen kaynaklar günümüzde artık sadece bilgisayar ve cep telefonlarından kaynaklanmamakta, bir robot süpürge bile büyük verinin daha da büyümesine katkı sağlayabilmektedir. İnternete veri aktarımı yapabilen cihazların günümüzde büyük bir çeşitlilik kazanması, bilgisayarların internetinden nesnelerin internetine (IoT - Internet of Things) doğru bir geçişin yaşanmasına neden olmuştur.

### **3.2.1.4. Gerçeklik**

Gerçeklik ile verinin doğruluğu ve güvenilirliği ifade edilmektedir. Veri ne kadar doğru ve güvenilir ise veri üzerinde çalışmak ve içinden anlamlı ve faydalı bilgilere ulaşmak da o denli kolaydır ve ulaşılan bilgi de o denli güvenilirdir. Ancak büyük veri, içerisinde doğru olmayan, eksik veya yanlış çok sayıda veriyi barındırmaktadır. Örneğin bir alışveriş sitesinde satın alınan bir ürün hakkında yapılan değerlendirmeler kişisel görüşler içermektedir ve doğruluğundan kesin olarak emin olmak mümkün değildir. Gürültü seviyesi yüksek bu tür veriler üzerinde veri madenciliği araçları ve analiz yöntemlerinin kullanılarak çalışılması da büyük verinin başka bir yönünü yansıtmaktadır (Gandomi ve Haider, 2015: 139).

### **3.2.1.5. Değer**

Büyük veri, işlenip kişi veya kurumlar açısından faydalı bir bilgiye dönüşebildiği oranda değerlidir. Bu noktada dikkat edilmesi gereken bir başka husus, büyük verinin işleme maliyetinin, işleme sonrası elde edilecek bilginin getireceği faydadan büyük olmaması gerektiğidir. Başka bir deyişle büyük veri, kendisinden sağlanacak faydanın, kendisini işlemek için gereken maliyetten fazla olması oranında da değerlidir (Narasimhan ve Bhuvaneshwari, 2014: 351).

Büyük veri genellikle “düşük yoğunluklu değeri” ifade etmektedir. Hacmi büyük ancak değeri hacmine oranla düşük bir değere sahiptir. Büyük veriden yüksek

bir deęer elde edilebilmesi için iřlenmesi gerekmektedir (Gandomi ve Haider, 2015: 139).

### 3.2.2. Büyük Veri Kaynakları

Yukarıda bazı kaynakları zikredilen büyük veri, çok farklı kaynaklardan gittikçe de artan oranda beslenmektedir. Yine yukarıda bahsedildięi üzere internet ortamına veri aktarabilen cihaz sayısının ve çeřitlilięinin artması bunda önemli rol oynamıřtır. Günümüzde bilgisayarlar ve telefonlar dıřında da çok sayıda cihaz internete bağlanabilmekte ve veri aktarımı yapabilmektedir. Bilgisayarların internetinden nesnelere internetine geçiř olarak adlandırılan bu süreçte, daha önce manuel olarak çalıřan cihazların günümüzde sensörleri vasıtasıyla topladıkları verileri ortak aęa iletebildikleri görülmektedir.

Akıllı evler, akıllı řehirler, akıllı yollar vs. gibi ibarelerin gün geçtikçe daha sık telaffuz edildięini görmekteyiz. Artık bir telefon üzerinden evler, iř yerleri, araçlar vs. kontrol edilebilmektedir. Sadece profesyonel iř alanlarında deęil, insanların günlük hayatlarının bir parçası olan neredeyse tüm unsurların yakın bir gelecekte internet ortamına bağlanabilip uzaktan kontrol edilebilir hale geleceęi öngörülmektedir. Bu da doęal olarak gelecekte büyük verinin üstel olarak artmaya devam edeceęine iřaret etmektedir.

Büyük verinin toplandıęı internet ortamına ve onun en büyük kaynaklarından biri olan sosyal medya 2022 verilerine baktığımızda, dünya çapında internet kullanıcılarının sayısının 4,95 milyar ile %62,5 oranına, aktif sosyal medya kullanıcı sayısının ise 4,62 milyar ile %58,4 oranına ulařtıęı görülmektedir. İnternet kullanıcılarının internette bir günde geçirdikleri süre dünya ortalaması 6 saat 58 dk, sosyal medyada bir günde geçirilen süre dünya ortalaması ise 2 saat 27 dk'dır. Görüldüęü üzere hem internet hem de sosyal medya kullanımı dünya genelinde oldukça yüksek bir seviyededir ve geçmiş yıllara bakıldığında bu seviyenin her geçen yıl daha da yukarıya çıktıęı görülmektedir. Aynı verilere ülkemiz için bakıldığında ise internet kullanım oranının %82, sosyal medya kullanım oranının %80,8; internet kullanıcılarının internette bir günde geçirdikleri süre 8 saat ve sosyal medyada bir günde geçirilen süre 2 saat 59 dk olduęu görülmektedir (We Are Social & Hootsuite, 2022). Görüldüęü üzere ülkemiz hem internet hem de sosyal medya kullanım oranlarında dünya ortalamasının üzerindedir.

Sosyal medyanın büyük veriye katkısına birkaç somut örnek vermek gerekirse, 2022 yılı itibari ile Facebook'ta günde 1 milyar hikaye paylaşılmakta, 4 milyardan fazla video görüntülenmektedir ve 350 milyardan fazla fotoğraf yüklenmiş durumdadır. Instagram'da her gün 500 milyon hikaye paylaşılmakta ve 100 milyondan fazla fotoğraf yüklenmektedir. Twitter'da günde 500 milyon tweet atılmakta ve günde 2 milyar video izlenmektedir (Bagadiya, 2022). Youtube'a dakikada 500 milyon saat video yüklenmekte ve 1 milyar saatlik video izlenmektedir (Sarika, 2022).

Yukarıda sözü edilen unsurlar dışında RFID, CCTV (Closed Circuit Television – Kapalı Devre Televizyon) kameraları, GPS (Global Positioning System – Global Konumlandırma Sistemi) gibi sistemler de büyük veriye kaynak oluşturmaktadır (Goes, 2014: 4; Miah, Vu, Gammack ve McGrath, 2017: 772).

### **3.2.3. Büyük Veri Uygulama Alanları**

Bazı alanları yukarıda da bahsedildiği üzere, büyük verinin kullanım alanları büyük bir çeşitlilik göstermektedir. Günümüzde ekonomik ve ticari faaliyetlerden kamu yönetimine, ulusal güvenlikten bilimsel araştırmalara kadar birçok alanda, büyük veri ve analitiğinden yararlanılmaktadır (Aktan, 2018: 1). Büyük veri ve analitiğini kullanmanın amaçlarından bazıları tüketici deneyimlerinin iyileştirilmesi, maliyetlerin düşürülmesi, pazarlama stratejilerinin belirlenmesi, iş süreçlerinin etkinliğinin artırılması, bireylerin veya toplumun herhangi bir konu hakkındaki eğilimini ölçme, veri güvenliğini sağlama şeklinde sıralanabilir. Büyük verinin başlıca uygulama alanları arasında bankacılık, iletişim, medya ve eğlence sektörü, sağlık hizmetleri, eğitim, üretim, devlet hizmetleri, sigortacılık, perakendecilik ve ticaret, ulaşım, enerji sektörü ve kendi kendine ölçüm verisinin analiz edilmesi yer almaktadır (Aktan, 2018: 7).

Bankacılık alanında büyük veri analitiğine daha çok geçmiş verileri kullanarak para hareketlerini ve müşteri satın alma davranışlarını ortaya koymak, soygun ve finansal suçlara karşı tedbir geliştirmek, risk yönetimi, düzenlemelere uyum yönetimi gibi amaçlar için başvurulmaktadır (Naik ve Joshi, 2017: 119; Srivastava ve Gopalkrishnan, 2015: 643).

İletişim ve medya alanında son yıllarda dijitalleşme adı altında büyük değişimler göze çarpmaktadır. Radyo-TV, gazetecilik, sinema gibi alanlar büyük bir

hızla dijital dönüşüme uğramakta, yeni iş modelleri ortaya çıkmaktadır. Herhangi bir verinin dijitalleşmesi o verinin veri madenciliği algoritmalarının kullanılarak analiz edilmesini kolaylaştırmakta ve iş sürecini çok daha verimli bir hale getirmektedir. Medya alanındaki dijitalleşme de doğal olarak büyük veri analitiğinin kullanımına daha uygun bir ortamın oluşmasını sağlamıştır. Bu alandaki akla gelen ilk uygulamalar, müşterilerin içerik tercihlerinin analiz edilerek ona uygun içeriklerin sunulmasıdır. “Kişiselleştirme” adı altındaki bu uygulamalar büyük veri kullanımı ile hemen her alanda kendini göstermeye başlamıştır.

Sosyal medya daha önce de bahsedildiği üzere büyük veri analitiğinin en fazla kullanıldığı alanların başında gelmektedir. 2022 yılı itibari ile tüm dünyada 4,62 milyar aktif sosyal medya kullanıcısı bulunmaktadır. Telefon ve sosyal medya artık günümüzde bağımlılık kapsamında incelenen alanlardan biri haline gelmiştir. Doğal olarak böyle bir alanda da veri hacminin insanların tek başlarına başa çıkamayacakları ölçüde devasa boyutlara ulaştığı söylenebilir. Sosyal medya özelinde büyük veri uygulamalarına bakıldığında daha çok ürün, firma, kişiler, kurumlar ya da gündemdeki olaylar ile ilgili olarak insanların algılarının ölçülmesi amacıyla yapılan analizler göze çarpmaktadır. Bu, “müşterilerin ürün hakkındaki değerlendirmelerini öğrenme ve ona göre bir strateji geliştirme” gibi bir pazarlama faaliyeti olabileceği gibi, bir siyasi partinin veya gündemdeki herhangi bir konunun kamuoyu nezdinde algısının ölçümü de olabilmektedir.

Büyük verinin bir başka kullanım alanı sağlık hizmetleridir. Hasta ve hastalıklar ile ilgili verilerin büyük veri analitiği ile değerlendirilmesi örüntülerin ortaya koyulması anlamında büyük faydalar sağlamaktadır. Hasta merkezli hizmet sunumu, hastanelerin kalitesinin izlenmesi, tedavi yöntemlerinin iyileştirilmesi, bulaşıcı hastalıkların erkenden tespiti gibi birçok alanda büyük veriden yararlanılmaktadır (Archenaa ve Anita, 2015: 408-409). Örneğin Twitter’da atılan tweet’lerin analizi ile bulaşıcı hastalıkların ortaya çıkışı ve kaynağının tespit edildiği ve erken teşhis için avantaj sağlandığı çok sayıda çalışma mevcuttur. Elkin, Topal ve Bebek (2017)’in Twitter’den grip hastalığıyla ilgili tweetleri ve bu tweetleri atan kullanıcıların yerlerini belirleyip, mevcut salgınları saptadıkları ve geliştirdikleri modellerle virüsün ne tarafa doğru yayılacağını tespit ettikleri “Network based model of social media big data predicts contagious disease diffusion” adlı çalışmalarına buna iyi bir örnektir (Albayrak, Topal ve Altıntaş, 2017).

Büyük verinin kullanıldığı alanlardan bir diğeri de eğitimidir. Öğrencilerin gerçekleştirdikleri eğitim faaliyetleri ve eğitime dair verileri analiz edilmekte ve daha verimli bir eğitim modeli ortaya konabilmektedir. Ülkemizde de yürürlüğe giren EBA Akademik Destek sistemi öğrencilerin eğitsel anlamdaki tercihleri ve mevcut seviyelerini göz önünde bulundurarak kendilerine hangi konuyu ne kadar çalışmaları gerektiğini söyleyen akıllı bir yönlendirme sistemi olarak faaliyet göstermektedir (MEB, 2022a). Büyük veri ile eğitimi buluşturan bir başka konu uzaktan eğitimidir. Özellikle Covid-19 pandemisinden sonra büyük bir artış gösteren uzaktan eğitim faaliyetleri ile öğrencilerin eğitimlerine dair verilerinin dijital ortama taşınması açısından büyük bir artış olmuş ve büyük veri analitiği ile bu verilerin arasındaki örüntülerin ortaya çıkarılması ve daha iyi bir eğitim modeli ortaya konulması açısından yeni fırsatlar doğmuştur. Bu anlamda büyük veri, eğitim yapısının, kapsamının, teknolojilerinin ve metotlarının yeniden yapılandırılmasında merkezî bir rol oynamaktadır (Yu, Yang ve Feng, 2017: 291-292).

Üretim alanında daha verimli bir iş modeli kurma amacıyla büyük verinin grafiksel, zamansal, coğrafi ve metinsel unsurlarının kullanılarak bilgi çıkarıldığı uygulamalar gerçekleştirilmektedir (Naik ve Joshi, 2017: 120). Ayrıca üretimde kullanılan cihazların daha verimli kullanılabilmelerini sağlayan “aktif önleyici bakım” adı altında bir büyük veri analitiği uygulaması bulunmaktadır. Bu uygulama ile iş sürecinde cihazlar ile ilgili veriler toplanarak olası arızalar gerçekleşmeden önce tahmin edici modeller vasıtasıyla önceden tahmin edilebilmektedir (Wan vd., 2017: 2040).

Kamu kurum ve kuruluşları, ihtiyaç sahiplerinin tespiti ve sosyal yardımların ulaştırılması, trafiğin kontrol edilmesi (akıllı trafik sistemleri) ve güvenlik gibi birçok alanda büyük veri analitiğinden yararlanmaktadır.

Taşımacılıkta araçlar, sürücü davranışları, yol bilgileri vs. gibi veriler analiz edilerek çok daha verimli bir iş planı oluşturulabilmektedir. Enerjide müşterilerin anlık olarak enerji tüketimini ölçebilen akıllı sayaçların kullanımı tüketim hakkında çok daha detaylı bilgiler verebilmekte ve dolayısıyla müşteriye daha uygun tekliflerin sunulabilmesine zemin hazırlamaktadır (Naik ve Joshi, 2017: 121).

Büyük veriyi oluşturan bir diğeri veri çeşidi de kendi kendine ölçüm verisidir. Günümüzde insanlar akıllı saat, bileklik, telefon vs. gibi cihazlar vasıtasıyla günlük hareketlerini, sağlık bilgilerini, egzersizlerini vs. kayıt altına alabilmektedirler ve bu

veriler kendi kendine ölçüm verisini oluşturmaktadır. Bu veriler de büyük veri analitiğine tabi tutularak insanlar için daha sağlıklı bir hayat planlaması yapılabilmesine katkı sağlamaktadırlar (Aktan, 2018: 10).

Büyük veri, yukarıda da çok sayıda örneği verildiği üzere, doğru veri madenciliği yöntemleri ile işlendiğinde hemen her alanda değerli sonuçların üretilebilmesine olanak tanımaktadır.

### **3.3. Veri Madenciliği**

Veri madenciliği, insanlar tarafından işlenemeyecek büyüklükteki veri içerisinden çeşitli bilgisayar programları ve algoritmalar kullanılarak faydalı ve anlamlı bilgilerin çıkarılması işlemi ifade eder. “Büyük Veri” başlığı altında ifade edildiği gibi günümüzde dijital ortamlarda yer alan, hacmi her an büyüyen ve sürekli olarak güncellenen devasa büyüklükte bir veri yığını bulunmaktadır. Bu yığın bu haliyle kişi ve kurumlar için anlamlı bir bilgi sunmayabilir ancak işlendiği zaman değerli bilgilere ulaşılabilmesi mümkündür. Bunu gerçekleştirmek için günümüzde başvurulan en pratik ve verimli yolların başında da veri madenciliği yöntemleri gelmektedir.

Bir başka tanım olarak veri madenciliği, çok büyük miktarda bilginin depolandığı veri tabanlarından, amaç doğrultusunda gelecek ile ilgili tahminler yapılabilmesini sağlayacak, anlamlı olan veriye ulaşma ve veriyi kullanma işidir (Savaş, Topaloğlu ve Yılmaz, 2012: 4). Veri madenciliği, istatistik, makine öğrenimi, yapay zeka, veri tabanı teknolojisi, örüntü tanımlama ve veri görselleştirme gibi çok farklı alanları bünyesinde ihtiva eden çok disiplinli bir alandır (Özekes, 2003: 66).

#### **3.3.1. Veri Madenciliği Uygulama Alanları**

Veri madenciliği yöntemlerine büyük verinin olduğu her alanda başvurulabilir. Yaygın olarak kullanıldığı alanlar, bankacılık/finans, perakende, müşteri ilişkileri yönetimi/tüketici analizi, sigortacılık, üretim, sağlık ve eğitim olarak ifade edilebilir.

Bankacılık ve finans sektöründeki veriler güvenilir olarak kabul edilir ve bu da veri madenciliği sonucu elde edilecek bilginin kalitesini artırmaktadır. Veri madenciliği bu alanda kredi risk analizi, farklı pazarlama yöntemlerinin uygulanması amacıyla müşterilerin kategorilendirilmesi ve finansal suçların tespit edilmesi gibi amaçlar için kullanılmaktadır (Amanet, 2017: 8).

Perakende sektöründe satışlar ve müşterilerle ilgili olarak büyük miktarda veri toplanmaktadır ve bu da veri madenciliği için gerekli ortamı hazırlamaktadır. Veri madenciliği bu alanda kampanya etkinlik analizi, pazar sepet analizi, ürün tavsiyesi oluşturma, müşterilerin satın alma örüntülerinin belirlenmesi, müşterileri elde tutma, yeni müşteri kazanma, müşteri değer analizi ve satış tahmini gibi amaçlar için kullanılmaktadır (Köktürk ve Dirsehan, 2012: 7).

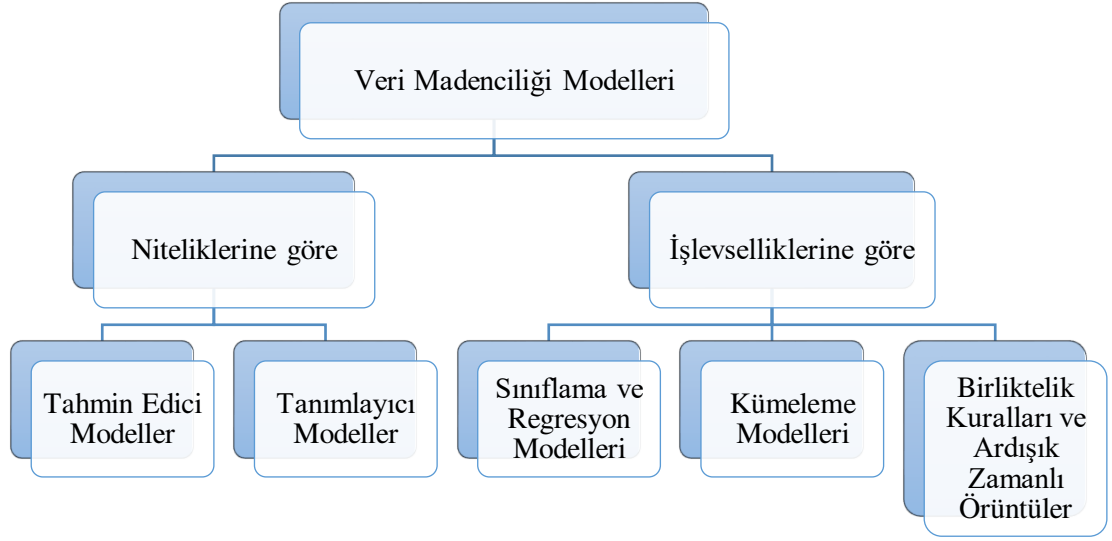
Sağlık sektöründe hastalık örüntülerinin ortaya çıkarılması, çeşitli hastalıkların önceden teşhisi, tedavi yöntemlerinin etkililiğini ölçme, daha iyi bir sağlık hizmeti politikası planlama gibi amaçlar için veri madenciliği yöntemlerine başvurulmaktadır.

Eğitim alanında eğitim politikalarının değerlendirilmesi, öğretim yöntem ve tekniklerinin değerlendirilmesi, mevcut öğretim metotlarının analizi ve öğrencilerin özellikleri ile aralarındaki örüntülerin ortaya çıkarılması, yeni eğitim politikalarının muhtemel etkilerinin tahmini gibi amaçlar için veri madenciliği yöntemlerine başvurulabilir.

Sigortacılıkta, risk değerlendirmesi, sigorta dolandırıcılıklarının tespiti, müşteriye özel poliçe teklifinin hazırlanması gibi amaçlar için veri madenciliği yöntemlerine başvurulabilir (Değer, 2017: 37).

### **3.3.2. Veri Madenciliği Modelleri**

Veri madenciliği modelleri, niteliklerine göre ve işlevselliklerine göre olmak üzere iki farklı sınıflandırmaya tabi tutulabilirler. Niteliklerine göre veri madenciliği modelleri, tahmin edici ve tanımlayıcı modeller olarak ikiye; işlevselliklerine göre ise sınıflama ve regresyon, kümeleme ve birliktelik kuralları ve ardışık zamanlı örüntüler olmak üzere üçe ayrılırlar (Köktürk ve Dirsehan, 2012: 6).



Şekil 3.1. Veri madenciliği modelleri

Tahmin edici modellerde, sonuçları bilinen verilerden hareket edilerek bir model geliştirilmesi ve kurulan bu modelle sonuçları bilinmeyen verilerin sonuçlarının tahmin edilmesi amaçlanmaktadır. Örneğin bir bankanın daha önce vermiş olduğu kredilere dair verilere bakarak verecek olduğu kredilerin ödenme ihtimalini hesaplayan bir model tahmin edici bir modeldir (Özekes, 2003: 67).

Tanımlayıcı modeller ise karar vermede faydası olabilecek, mevcut veriler arasındaki örüntüleri ortaya koymayı amaçlamaktadır. Örneğin  $X - Y$  gelir aralığındaki çocuklu bir aile ile  $X - Y$  gelir aralığından daha düşük bir gelire sahip çocuksuz bir ailenin satın alma örüntülerinin birbirine benzediğini ortaya koyan bir model tanımlayıcı bir modeldir (Özekes, 2003: 67).

Kantardzic (2011), tahmin edici ve tanımlayıcı modelleri “öngörme” ve “betimleme” olarak ifade etmiştir. Ona göre veri madenciliği pratikte öngörme ve betimleme olmak üzere iki öncelikli amacı gerçekleştirmeye eğilimlidir. Öngörme, veri setindeki bazı bilinen değişkenlerin, bilinmeyen değerlerin veya gelecekteki ilgili değişkenlerin değerlerinin öngörülmesi için kullanılmasını; betimleme ise verideki örüntülere yoğunlaşıp veriyi insanlar tarafından yorumlanabilecek şekilde betimlemeyi ifade etmektedir (Kuş, 2019: 198).

İşlevselliklerine göre yapılan sınıflandırmada sınıflama ve regresyon, mevcut verilerin kullanılarak geleceğin tahmin edilmesinde kullanılır. Sınıflamada kategorik verilerin tahmini yapılırken, regresyonda süreklilik gösteren verilerin tahmini yapılır.

Örneğin bir banka için uygulanacak olan kredi analizinde kredinin güvenli veya riskli olduğunu ortaya koyan bir model, sınıflama; müşteriye dair verilerin analiz edilerek belirli bir alan ile ilgili olarak bu müşterinin ne kadarlık bir harcama yapacağını tahmin eden bir model regresyona örnek olarak verilebilir (Özekes, 2003: 68).

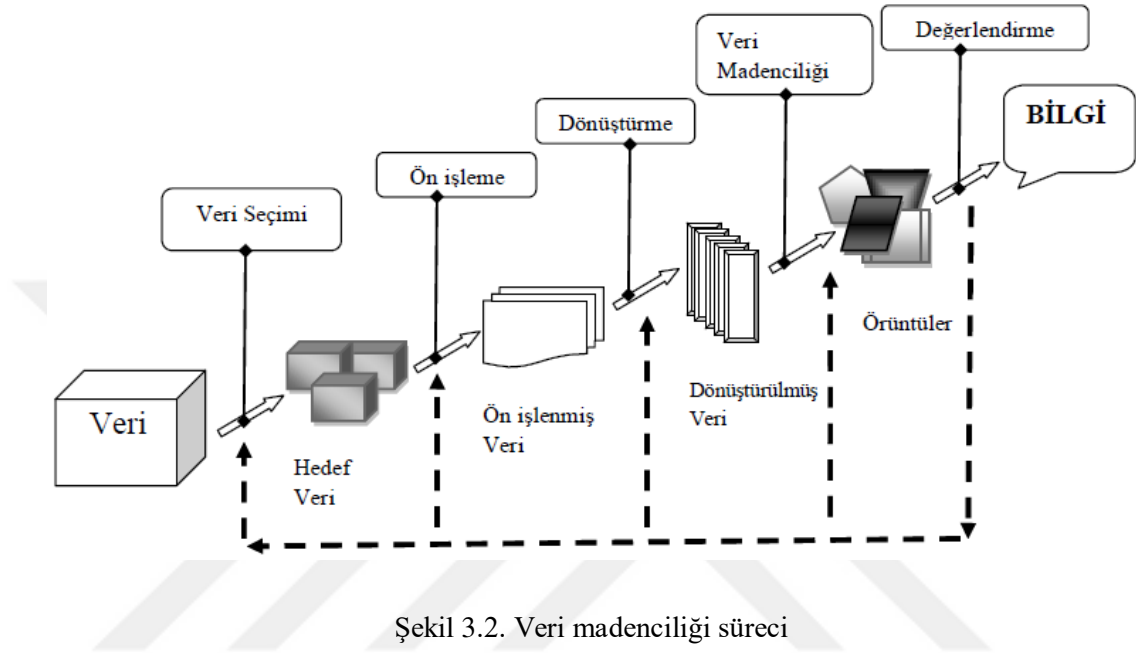
Sınıflama yapılırken sınıfları bilinen bir veri seti, eğitim ve test seti olmak üzere ikiye ayrılır. Eğitim seti kullanılarak modelin öğrenmesi sağlanır. Öğrenilen model test setine uygulanır ve ne oranda doğru tahminde bulunduğu bakılır. Özetle sınıflama, sınıfları bilinen bir veri setini analiz ederek sınıfları belli olmayan bir veri setinin sınıflarını tahmin etme işi anlamına gelmektedir. Regresyon ise bağımlı bir değişkenin değerinin bağımsız bir veya daha fazla değişkenin değerlerine bağlı olarak nasıl değişeceğini tahmin eden bir fonksiyonu ifade eder (Değer, 2017: 42).

Kümeleme verinin benzer gruplarda toplanmasıdır. Sınıflamada olduğu gibi verinin sınıfları önceden belli değildir. Aynı kümede yer alan veriler birbirleriyle benzerlik gösterirken farklı kümelerde yer alan veriler ile farklılaşırlar. Sınıflamada sınıfların önceden belirlenmesinin güç olduğu durumlarda bir ön işlem olarak kümeleme uygulanabilmektedir. Müşterilerin kategorilere ayrılması, atm için şehirde uygun bir yer seçme, biyolojide bitki ve hayvanların sınıflandırılması gibi amaçlar için kümeleme modeline başvurulabilir (Özekes, 2003: 71; Değer, 2017: 44).

Birliktelik kuralları ve ardışık zamanlı örüntüler işlevselliğine göre bir diğer veri madenciliği modelidir. Birliktelik kurallarında verilerin birlikte gerçekleşme durumlarına bakılır. Başka bir deyişle birliktelik kuralları beraber gözlenen olayların korelasyonudur. Örüntü ise, veriler arasındaki belirli bir düzeni ifade eder. Sayılar, şekiller, çizimler arasında olabileceği gibi yapılan işler arasında da bir örüntü olabilir. İnsanların farklı zamanlarda düzenli olarak yaptığı satın alma davranışları da örüntüye bir örnektir ve aldığı ürünler de zaman içinde ardışık olan bir örüntüyü oluşturur (Silahtaroglu, 2013: 53-55). “Pazar sepet analizi” olarak bilinen ve insanların hangi ürünleri birlikte satın aldığını ortaya koyan veri madenciliği uygulaması, klasik bir birliktelik kuralları ve ardışık zamanlı örüntüler modeli örneğidir (Akpınar, 2000: 70). Pazar sepeti analizi ile insanların satın aldıkları bir üründen sonra sırayla hangi ürün veya ürünleri satın alabilecekleri ortaya konulur ve bu bilgi pazarlama faaliyetlerine yön verme amacı ile kullanılır (Özkan, 2013: 49).

### 3.3.3. Veri Madenciliği Süreci

Veri madenciliği, verinin içinden amaca uygun verilerin seçimi, verilerin ön işlemden geçirilmesi, dönüştürülmesi, veri madenciliği modelinin uygulanması, modelin değerlendirilmesi ve bilgiye ulaşma aşamalarını içermektedir (Savaş vd., 2012: 8).



Şekil 3.2. Veri madenciliği süreci

Veri madenciliği sürecine başlamadan önce veri madenciliğinin hangi amaç için yapıldığının, içinde çalışılan alanın ve verinin, sağlıklı bir veri madenciliği uygulaması için iyi analiz edilmesi gerekir (Savaş vd., 2012: 7). Veri madenciliği süreci büyük veri içerisinde ihtiyaç duyulan verilerin seçimi ile başlar. Üzerinde çalışılacak olan verinin seçimi veri madenciliği süreci sonunda ulaşılmak istenen bilginin doğruluğuna direkt olarak etki etmesi açısından büyük önem arz etmektedir. Bu sebeple seçilecek olan verilerin amaca uygun olarak, doğru kaynaklardan ve doğru miktarda seçilmesi gerekmektedir.

Bir sonraki aşamada seçilen veriler bazı ön işlemlere tabi tutulurlar. Veri ön işleme veri madenciliği sürecinin genellikle en çok zaman alan aşamasıdır. Bu aşamada öncelikle veri temizleme işlemi yapılır. Veri temizleme işlemi ile veri üzerindeki gürültüler, düzensizlikler ve tutarsızlıklar giderilerek veri kalitesi artırılır (İnan, 2015: 16). Eksik veriler var ise veri setinden çıkarma, yerine sabit bir değer atama, ortalama bir değer atama gibi seçeneklerden biri uygulanabilir (Değer, 2017:

46). Gürültülü verileri tespit etmek için histogram, kümeleme analizi ve regresyon gibi metotlar kullanılabilir (Oğuzlar, 2003: 71).

Veri ön işleme aşamasında uygulanabilecek bir başka işlem eğer veriler farklı kaynaklardan geliyor ise veri bütünleştirmedir. Veri bütünleştirme, farklı kaynaklardan gelen verilerin birbirine uyumlu hale getirilerek tek bir veri ambarında toplanmasını ifade eder. Bu noktada uygulanabilecek bir başka işlem veri indirgemedir. Veri indirgeme, verinin boyutunu sonucu etkilemeyecek şekilde azaltmayı ifade eder. Veri indirgeme için birleştirme veya veri küpü, genelleme, örnekleme, sıkıştırma ve boyut indirgeme yöntemleri kullanılabilir (Özkan, 2013: 41). Birleştirme, farklı veri kümelerini aynı veri setinde toplama olarak ifade edilebilir. Çeyrek dönemlik satış verilerinin yer aldığı ayrı tabloların tek bir tabloda birleştirilmesi buna örnek olarak verilebilir. Veri küpleri ise çok değişkenli verilerin veri küpü denilen tek bir alanda birleştirilmesini ifade eder. Örneğin bir firmanın yaptığı satışa ait tutar, satışın yapıldığı yer, zaman gibi bilgiler tek bir veri küpünde gösterilebilir (Oğuzlar, 2003: 74). Genelleme, ortak noktaları bulunan verilerin aynı başlık altında toplanmasını ifade eder. Örnekleme, veri setinin içinden kendisini temsil edebilecek daha küçük hacimli bir veri setinin çıkarılmasıdır. Temsil kabiliyeti örneklemede büyük önem arz etmektedir. Sıkıştırma, verilerin aynı anlama gelecek, ancak daha az yer kaplayacak şekilde kodlarla ifade edilmesi anlamına gelir (Çölkesen, 2014: 439). Son yöntem olarak veri içerisindeki sonuca etki etmeyecek gereksiz veriler çıkarılarak boyut indirgenebilir.

Veri dönüştürme verilerin veri madenciliği yöntemlerinin uygulanabilmesi için uygun formlara dönüştürülmesini ifade eder (Oğuzlar, 2003: 73). Veri dönüştürme için uygulanabilecek yöntemler normalleştirme, standartlaştırma, düzleştirme, birleştirme, genelleme ve özellik inşası olarak sıralanabilir (Şeker, 2013: 109). Bunların içinden daha özelinde minimum-maksimum normalleştirmesi, ondalık ölçekleme normalleştirmesi ve z skor standartlaştırmasının daha yaygın olarak kullanıldığı söylenebilir. Minimum-maksimum normalleştirmesinde veri setinde yer alan değerler, 0 – 1 arasında ifade edilecek şekilde dönüştürülürler. Z skor standartlaştırma yönteminde verilerin ortalaması ve standart sapması dikkate alınarak dönüştürme işlemi gerçekleştirilir. Veri setindeki değerlerin ortalamaları ve varyansları birbirinden çok farklı olduğu durumlarda sistematik bir hatanın önüne geçmek için bu yöntem başvurulabilir (Değer, 2017: 49). Ondalık ölçekleme

yönteminde ise veriler -1 ile 1 arasında yer alacak şekilde dönüştürülürler (Dondurmacı, 2011: 11).

Veri madenciliği algoritmalarının uygulanmasına modelleme denilir (Değer, 2017: 49). Bu süreç, amaca uygun çok sayıda modelin kurulmasını ve test edilmesini içermektedir. Sonuç itibariyle en yüksek performans gösteren model seçilir. Bu sebeple amaca uygun ne kadar çok model kurulur ve test edilirse, performansın yükselme ihtimali de o oranda artacaktır. Bu süreç aynı zamanda veri ön işleme ve hatta veri seçme aşamasına kadar geri dönme ihtimalini de barındırmaktadır. En yüksek performansı gösteren modele ulaşana kadar süreç bu şekilde geriye de dönülebilecek şekilde devam eder. En yüksek performansı gösteren model tespit edildikten sonra veriler bu model kullanılarak işlenir ve sonuç verileri elde edilir. Bu verilerin değerlendirilmesi sonucunda da ihtiyaç duyulan bilgiye ulaşılması beklenir.

### **3.4. Metin Madenciliği**

Her ne kadar birçok çalışma kapsamında metin madenciliği veri madenciliğinin bir alt disiplini olarak gösterilse de veri madenciliğinde olmayan bazı önemli farklılıklara da haizdir. Bu farklılıkların en önemlisi metin madenciliğinin üzerinde çalıştığı verinin yapısal olmayan ve bulanık olarak tabir edilebilecek bir veri olmasıdır (Budak, 2021: 17). Metin madenciliğinin üzerinde çalıştığı veri çoğunlukla sosyal medyadan, müşteri yorumlarından, web sayfalarından vs. çıkarılmış metinlerden oluşmaktadır. Doğal olarak bu metinler de konuşma dili ile yazılabilmekte, birçok yazım yanlışı, kısaltma ve ironi barındırabilmekte, dolayısıyla içerdikleri anlamların ortaya çıkarılması zor olabilmektedir. Metin madenciliği bu her türde yazılan metinlerin içindeki anlamların ortaya çıkarılması gibi güç bir problem ile uğraşmaktadır.

Bugün dünya üzerinde yer alan verinin büyük bir kısmı metin formatında bulunmaktadır (Hamde, 2018: 30; Değer, 2017: 55; Kızılkaya, 2018: 4) ve dolayısıyla insanlar tarafından manuel olarak işlenemeyecek büyüklükte bir metni işleyip içinden işe yarar bilgilere ulaşılmasını sağlayan metin madenciliği teknikleri büyük bir önem arz etmektedir. Büyük veri olarak nitelendirilebilecek bir metne veri madenciliği tekniklerinin uygulanabilmesi için metnin bir dizi ön işlemde geçerek yapılandırılması (sayısallaştırılması) gerekmektedir. Bu noktada devreye doğal dil işleme, istatistik, makine öğrenimi gibi alanlar da girmektedir (Delen ve Crossland, 2008: 1711; Değer, 2017: 55). Metin madenciliği, bu farklı alanların bir kesişimi

olarak nitelendirilebilir. Örneğin doğal dil kullanılan bir metin üzerinde çalışılırken metnin içinde yer alan sözcüksel ve dilsel kullanım kalıpları saptanmaya çalışır ve buradan anlamlı ve kullanışlı bilgi elde etme hedeflenir (Hamde, 2018: 27). Bunu yaparken de doğal dil işleme yöntemlerinden yararlanır. Metin madenciliğinde sınıflandırma amaçlı olarak çokça kullanılan makine öğrenimi ise metnin bir kısmının insan tarafından sınıflandırılarak modele öğretilmesi, geri kalanının ise öğrenen model tarafından otomatik olarak sınıflandırılmasına dayanmaktadır. İstatistik de metin madenciliği sürecinde metni oluşturan unsurların yan yana gelme olasılıklarının hesaplanması noktasında sıkça başvurulan alanlardan bir diğeridir.

Metin madenciliğinin ortaya çıkışı ve tarihi seyrine bakıldığında, öncelikle kütüphanecilik alanında sınıflama ve özetleme çalışmalarına rastlanmaktadır. Daha sonra bilgisayarların ortaya çıkışı ile metinlerin otomatik olarak aranması ve metin içeriklerine ulaşılması çalışmaları görülmektedir. Son olarak doğal dil işleme ve makine öğrenimi yöntemlerinin kullanılması ile otomatik olarak metnin içindeki anlamın ortaya çıkarıldığı bir sürece evrilmiştir (Ağdeniz, 2017: 72). Yazılı belgelerin sınıflandırılması, müşteri yorumlarının değerlendirilmesi ve dil çevirileri gibi birçok işlem günümüzde metin madenciliği yöntemleri kullanılarak hızlı bir şekilde gerçekleştirilebilmektedir (Pekin, 2020: 22). Özetle erişim seviyesi ile başlayan süreç günümüzde analiz ve keşif seviyesine ulaşmıştır (Jhanji ve Garg, 2014: 1642; akt. Hamde, 2018: 27).

Metin madenciliği için bugüne kadar yapılan birçok tanımdan bazıları aşağıdaki gibidir:

Metin madenciliği metin içinde bulunan değerli bilgiden yararlanmak için otomatikleştirilmiş yöntemlerin kullanımınıdır (Hamde, 2018: 32).

Metin madenciliği “metinleri sayılara dönüştürme” teknolojileridir (Miner vd., 2012: 30).

Metin madenciliği metin verilerinden yüksek kalitede bilginin elde edilmesi amacıyla yapılan anlam analizidir (Gaikwad, Chaugule ve Pramod, 2014: 42).

Bazı kaynakları yukarıda belirtilen metin madenciliği dijital ortamdaki çokça kaynaktan beslenmektedir. Günümüzde metin madenciliğini besleyen en büyük kaynağın sosyal medya olduğu söylenebilir. Onun dışında e-postalar, haber siteleri, işletme raporları, dijital formattaki kitaplar, akademik yayınlar, sağlık kayıtları vs.

gibi yazılı metinlerin tümü metin madenciliği için kaynak olarak kullanılabilirler (Kızılkaya, 2018: 5).

### **3.4.1. Metin Madenciliği Uygulama Alanları**

Miner ve arkadaşlarına göre (2012: 38), metin madenciliğinin yedi farklı uygulama alanı bulunmaktadır. Bunlar, bilgiye erişim, bilgi çıkarımı, doğal dil işleme, belge sınıflandırma, belge kümeleme, web madenciliği ve kavram çıkarımıdır.

#### **3.4.1.1. Bilgiye Erişim**

Basitçe, bilginin getirilmesini sağlayan sistemlerdir. Google gibi arama motorları, bilgiye erişim sistemlerine iyi bir örnektir. Anahtar sözcükler kullanılarak ihtiyaç duyulan bilgiyi içeren belgelerin daraltılması sağlanır. Bilgiye erişim sistemleri, büyük metin veri tabanından ihtiyaç duyulan bilgiye sahip olma ihtimali olan tüm belgelere erişip, diğerlerini filtreleyen sistemler olarak tanımlanabilir (Oğuzlar, 2011: 15). Bilgiye erişim, kullanıcının ihtiyaç duyduğu bilgiye ulaşma yolunda başvurduğu metin madenciliği yöntemlerinin ilkidir.

#### **3.4.1.2. Bilgi Çıkarımı**

Bilgi çıkarımı, büyük metin belgeleri içinden ihtiyaç duyulan bilgileri çıkarır. Bilgi çıkarımı bazen - her ikisinin de bilgiye ulaşma amacı ile hareket etmesi sebebi ile - bilgiye erişim ile karıştırılabilmektedir. Ancak her iki sistem birbirinden büyük farklılıklar içermektedir. Bilgiye erişimde büyük bir metin koleksiyonu içerisinde ihtiyaç duyulan bilgiye sahip belgelerin getirilmesi hedeflenir fakat bilgiye direkt olarak ulaşma söz konusu değildir. Bilgi çıkarımında ise metin belgeleri içinden daha spesifik olarak ihtiyaç duyulan bilgilere ulaşma hedeflenir. Bu, bilgilere ulaşım bunları bir şablona yazmaya benzetilebilir. Örneğin otomobillere ait bazı özelliklerin yer aldığı metin belgeleri içinden otomobillerin markası, modeli, yılı, fiyatı gibi özelliklerin otomatik olarak çıkarılması bir bilgi çıkarımı işlemidir. Bilgi çıkarımı işlemi sonucunda yapısal olmayan veriler yapılandırılmış olur. Başka bir deyişle, bilgi çıkarımı öncesi makine tarafından sadece okunabilir veriler, bilgi çıkarımı sonrası makine tarafından işlenebilir hale getirilmiş olur (Oğuzlar, 2011: 18). Bilgi çıkarımı sistemlerinde doğal dil işleme ve makine öğrenmesi yöntemlerinden yararlanılır (Değer, 2017: 65).

### 3.4.1.3. Doğal Dil İşleme

Doğal dil işleme, makinelerin insanlar ile konuşma dili üzerinden iletişim kurabilmesi üzerine çalışan bir mühendislik alanıdır. Bu iletişim yazılı ya da sesli olabilmektedir. Doğal dil işleme teknikleri kullanılarak otomatik olarak bir dilden başka bir dile çeviri yapılabilen, yazılı ya da sözlü soru-cevap sistemleri oluşturulabilmekte, metin özetleme, metin imla düzeltme, metin içerisinde istenilen bilgiyi çıkarma, metni seslendirme, metni anlayıp doğal dille karşılık verme gibi birçok işlem gerçekleştirilebilmektedir (Ağdeniz, 2017: 78; Değer, 2017: 60; Çabuk, Yüksel, Mocan, Diri ve Amasyalı, 2003: 679).

Doğal dil işleme, doğal dili anlama ve üretme olarak iki ayrı başlık altında incelenebilir. “Anlama”, sisteme girdi olarak girilen dilin analizini ve anlaşılabilir şekilde ifade etmektedir. “Üretme” ise anlaşılabilir dille karşılık aynı dille karşılık verilmesini ifade eder. “Anlama” sürecinde makinenin fonksiyonu bir okuyucu ya da dinleyiciye; “üretme” sürecinde ise bir yazar veya konuşmacıya benzetilebilir (Liddy, 2001; akt. Pekin, 2020: 24).

Doğal dil işleme süreci her dilin kendine özgü morfolojik yapısı olması sebebiyle üzerinde çalışılan dile bağımlı bir şekilde ilerler (Aksu, 2021: 25) ve doğal olarak süreç içerisinde karşılaşılan zorluklar da dilden dile değişiklik gösterir. Ancak bir genelleme yapılması gerekirse, doğal dil işleme sürecinde yaşanan en büyük zorluğun anlamdaki belirsizlik olduğu söylenebilir. Örneğin, sondan eklemeli dillerde sona eklenen ekler ile anlam tamamen değişebilmektedir. Kelimelerin cümle içerisinde yer değiştirmesi yine anlamı değiştirebilmektedir (Değer, 2017: 60). Yazım yanlışları, bir kurala bağlı olmayan kısaltmalar, jargonlar vs. yine anlaşılmayı güçleştiren diğer unsurlardır.

Doğal dil işlemenin çalışma seviyelerine göre dört ana başlıkta toplandığı görülmektedir. Bunlar, kelime bilimi (morphological-lexical), söz dizimsel (syntactic), anlamsal (semantic) ve söylevidir (pragmatic-discourse) (Şeker, 2015: 15).

Bir metin madenciliği uygulamasında yapılması gereken en önemli işlerden biri, metnin en küçük birimlerine kadar ayrıştırılması işlemidir. Doğal dil işleme yöntemlerine başvuruluyor ise de bu en küçük birimler anlam ifade eden kelimeler olur. Böyle bir uygulama için metin önce kelimelere ayrılır, kelimeler de eklerinden

arındırılarak köklerine kadar ayrıştırılır. Bu işlemlerin hepsi kelime bilimi seviyesinde gerçekleştirilen işlemlerdir. Ayrıca bir kelimenin birden fazla niteliği bulunabilir. Örneğin bir kelime hem isim hem de çoğul olma özelliği taşıyabilir. Kelimelerin bu birden fazla niteliğini ortaya çıkarma da yine kelime bilimi seviyesinde gerçekleştirilir. Kelime bilimi seviyesine bir başka örnek, kelimelerin yanlış yazılışlarının tespiti ve düzeltilmesi olarak verilebilir (Değer, 2017: 61).

Söz dizimsel seviye kelimelerin cümle içindeki dizilişi ile ilgilendir (Şeker, 2015: 15). Kelimelerin birbirleri ile olan ilişkilerine bakılır ve kurallara uygun bir cümle yapısının oluşup oluşmadığı kontrol edilir (Değer, 2017: 61).

Anlamsal seviye cümlelerin anlamları ile ilgilendir. Makinenin doğru bir işlem yapabilmesi için cümlenin anlamını doğru olarak algılaması gerekmektedir (Şeker, 2015: 15).

Söylev seviyesi, bir konuşmada kullanılan kelimeler ve anlamları ile anlaşılabilir konuşmaya doğru kelimeler kullanılarak karşılık verilmesi üzerine çalışır (Şeker, 2015: 15). Eş anlamlı kelimelerin hangi anlamda kullanıldığının tespiti söylev seviyesi kapsamındadır (Değer, 2017: 63).

Doğal dil işleme tüm bu işlemleri gerçekleştirirken, bilgisayar ve bilişim bilimleri, matematik, dil bilimi, elektrik-elektronik mühendisliği, psikoloji, yapay zeka ve robotik gibi çok sayıda disiplinden yararlanmaktadır (Chowdhury, 2003: 52).

#### **3.4.1.4. Belge Sınıflandırma**

Belge sınıflandırma, metin belgelerinin daha önceden belirlenmiş olan sınıflara otomatik olarak yerleştirilmesidir. Bu işlem için genellikle makine öğrenmesi yöntemlerine başvurulur. Bunun için öncelikle belirli bir kritere göre sınıflar belirlenir ve belgelerin bir kısmı her sınıf için yeteri kadar örnek oluşacak şekilde manuel olarak sınıflandırılır. Yapılan sınıflandırma makine öğrenmesi için kullanılan algoritma tarafından öğrenilir ve metnin geri kalanının sınıflandırılması algoritma tarafından otomatik olarak gerçekleştirilir. Bu yöntemde insan tarafından eğitim amaçlı olarak manuel bir sınıflandırma gerçekleştirildiği için, bu yönteme ayrıca denetimli öğrenme adı da verilir.

Belge sınıflandırmaya, istenmeyen e-postaların tespiti, duygu analizi ve haber metinlerinin sınıflandırılması gibi uygulamalar örnek olarak verilebilir (Aksu, 2021: 24).

### **3.4.1.5. Belge Kümeleme**

Belge kümeleme, metin belgelerinin birbirleri ile olan benzerlik ya da farklılıklarına göre otomatik olarak kümelendiğini ifade eder. Benzer belgeler aynı kümelere toplanırlar. Sınıflandırmada olduğu gibi sınıflar ya da kümeler önceden belirli değildir. Bu yöntem ayrıca denetimsiz öğrenme adı da verilir (Tunalı, 2011: 11). Firmaların insanlar tarafından tasnif edilmesi, nispeten uzun zaman alacak metin belgelerinin bu şekilde tasniflenmesi belge kümelemeye örnek olarak verilebilir.

### **3.4.1.6. Web Madenciliği**

Web üzerinde çok farklı tipte veri çeşidi bulunmaktadır. Bunlar, web sayfaları, erişim günlüğü (Access log) dosyaları, kullanıcı kayıt bilgileri, oturum ve hareket bilgileri ile site yapısı ve içeriği şeklinde sıralanabilir. Web madenciliği, bu farklı tipteki web verilerini incelemek ve aralarındaki örüntüyü ortaya çıkarmak için veri madenciliği tekniklerinin kullanılması olarak tanımlanabilir. Bir internet alışveriş sitesinin satışlarını artırmak için siteyi kullanan kişilerin hareket ve işlemlerini incelemesi ve ona göre yeni satış stratejileri belirlemesi web madenciliğine örnek olarak verilebilir (Baykal ve Çoşkun, 2009: 797).

Web madenciliği üç başlık altında incelenmektedir. Bunlar, web içerik madenciliği, web yapı madenciliği ve web kullanım madenciliğidir. İçerik madenciliği, metin, resim, video, ses, meta veri gibi web içeriklerinden yararlı bilgiler elde etmeyi hedeflemektedir. Kullanıcıların web sitelerinde istedikleri bilgiye ulaşabilip ulaşamadıkları, içerik unsurlarının yerleşimlerinin doğru olup olmadığı, az ziyaret edilen bölümlerin sayfa yapısı ile ilişkisi web içerik madenciliği kapsamındaki konulardır (Baykal ve Çoşkun, 2009: 797). Web yapı madenciliği web siteleri ve web sayfaları arasındaki bağlantılara odaklanmaktadır. Web kullanım madenciliği ise web sitelerini ziyaret eden kullanıcıların hareketlerine odaklanır. Bu veriler kullanılarak kullanıcılar tarafından gerçekleştirilen işlemler analiz edilebilir ve bu işlemler arasındaki örüntüler tespit edilebilir (Aksu, 2021: 22).

### **3.4.1.7. Kavram Çıkarımı**

Türk Dil Kurumu (TDK, 2022)'na göre kavram, bir nesnenin veya düşüncenin zihindeki soyut ve genel tasarımı anlamına gelmektedir. Metin madenciliği kapsamında kavram çıkarımı, metnin ana temasını ortaya çıkarmayı amaçlar (Ağdeniz, 2017: 79), metni oluşturan unsurların hangi ortak ad etrafında toplandığını

bulmaya çalışır. Bunu yaparken de makine öğrenimi ve istatistik gibi alanlardan yararlanır (Aydn, Erkan, Güngör ve Takçı, 2013: 1)

### **3.4.2. Metin Madenciliği Aşamaları**

Metin madenciliği aşamaları, veri madenciliği aşamaları ile benzerlik göstermektedir. Metin madenciliğinde farklı olarak verinin çok çeşitli kaynaklardan yapısal olmayan türde gelmesi ve veri madenciliği algoritmalarının uygulanabilmesi için öncelikle yapısal hale dönüştürülmesi gerekliliğidir.

Metin madenciliği aşamaları veri setinin oluşturulması, metin ön işleme, özniteliklerin çıkarılması, özniteliklerin ağırlıklandırılması, metnin yapısal hale dönüştürülmesi, boyut azaltma, veri madenciliği algoritmalarının uygulanması ve değerlendirme aşamaları olarak sıralanabilir.

#### **3.4.2.1. Veri Setinin Oluşturulması**

Metin madenciliği aşamalarının ilki metin madenciliği yöntemlerinin uygulanacağı bir veri setinin oluşturulmasıdır. Metin madenciliği için kullanılacak olan veriler çok çeşitli kaynaklardan ve çok farklı formatta olabilmektedirler. Günümüzde metin madenciliği olarak kullanılacak olan verilerin çoğunlukla internet ortamından elde edildiği söylenebilir. Özellikle Google başta olmak üzere arama motorları ile Twitter, Facebook, Instagram, Youtube gibi sosyal medya platformlarının bu konuda başı çektiği görülmektedir (Oğuzlar ve Kızılkaya, 2019: 24). Sosyal medya platformları dışında web sayfaları, e-postalar, bloglar, dijital veya fiziksel kütüphaneler, makaleler, haber metinleri, kurumlara ait dokümanlar, ses dosyalarının yazıya dökülmüş halleri gibi çok farklı kaynaklar da metin madenciliğinde kullanılabilir (Han, Kamber ve Pei, 2012: 596; Aksu, 2021: 19). İnternet ortamından metinlere - erişime açıksa - direkt olarak ulaşılabileceği gibi çeşitli programlar vasıtasıyla da ulaşmak mümkündür.

Metin madenciliğinde üzerinde çalışılacak metin formatları da büyük bir çeşitlilik gösterebilir. Word, excel, csv, pdf, text dosyası gibi farklı formatlarda olabilen bu metinlerin öncelikle aynı formata dönüştürülmesi gerekmektedir. Toplanan bu metinlerin tümüne derlem adı da verilmektedir (Ağdeniz, 2017: 80).

### 3.4.2.2. Metin Ön İşleme

Metin ön işleme süreci metni dizgeciklere (token) ayırma, filtreleme ve gövdeleme aşamaları ile ihtiyaç halinde uygulanabilecek çeşitli doğal dil işleme uygulamalarından oluşmaktadır. Metin ön işleme her metin madenciliği uygulaması için standart olmamakla birlikte ortak amacın veri madenciliği algoritmalarının uygulanması aşamasının en hızlı ve verimli olmasına imkan verecek şekilde metin içerisindeki gereksiz karakter ve kelimelerin temizlenmesi ve metin içerisindeki öznitelik adı verilen üzerinde işlem yapılacak en küçük birimlerinin çıkarılması olduğu söylenebilir. Metin ön işleme metin madenciliği sürecinin en fazla zaman alan aşamasıdır. Aksoy, Çelik ve Gülseçen (2020)'e göre tüm sürecin %50-%80'lik kısmını kapsamaktadır.

Dizgeciklere ayırma işlemine başlamadan önce metindeki büyük harfler küçük harflere dönüştürülür, böylelikle büyük-küçük harf farkından dolayı aynı sözcüklerin farklı bir sözcükmüş gibi algılanmalarının önüne geçilmiş olur. Ayrıca metin içerisinde analize faydası olmayacak web sitesi linkleri bulunuyorsa bunlar da metinden temizlenir. Bu işlemler dışında çalışmaya has olarak birbirinin kopyası içerikleri temizleme, eksik değerleri temizleme ya da dönüştürme gibi işlemler de uygulanabilir.

#### 3.4.2.2.1. Dizgeciklere (Token) Ayırma

Dizgeciklere ayırma işlemi ile metin okunabilir olmayan karakterlerden, noktalama işaretlerinden, tab ve satır sonu karakterlerinden temizlenir (Tunalı, 2011: 13). Eğer web tabanlı bir veri üzerinde çalışılıyorsa ilaveten XML ve HTML gibi uzantılar da temizlenir (Çelikyay, 2010: 60). XML ve HTML uzantıları web dili için bir anlam ifade eden ancak metnin içeriğine dair herhangi bir bilgi barındırmayan etiketlerdir. Dizgeciklere ayırma işlemi metin madenciliği için bu gereksiz verilerden metni temizleyerek metin içerisindeki anlamlı sözcüklere ulaşmayı hedefler. Dizgecik ya da “token” adı verilen yapılar metin içerisindeki anlamlı sözcükleri ifade etmektedir (Değer, 2017: 67)

#### 3.4.2.2.2. Filtreleme

Metin içerisinde çok sık tekrarlanan ancak kendi başına bir anlam ifade etmeyen sözcüklerin analize de bir katkıları olmamakta ve hatta boyutu artırdıkları için verimi düşürmektedirler. Literatürde “stop words” (durak sözcükleri) adı verilen

bu sözcüklerin filtrelenmesi tercih edilir ve metin içerisinden temizlenirler. Genellikle bu sözcükler edat, bağlaç ya da zamir olurlar. “ile”, “bile”, “ama”, “böyle”, “bu”, “da”, “için”, “şey”, “ve”, “gibi” sözcükleri durak sözcüklerine örnek olarak verilebilir. Her dil için hazırlanmış çeşitli durak sözcükleri listeleri bulunmaktadır. Çalışmadan çalışmaya bu listeler değişiklik gösterebilmektedir. Belirtilen türde sözcükler dışında çalışmaya özel olarak metin içerisinde çok sık veya çok seyrek geçen sözcükler de filtrelenebilir (Hotho, Andreas ve Gerhard, 2005: 25). Burada amaç ayırım gücü en yüksek sözcüklerin analiz için seçilmesidir (Değer, 2017: 68).

### **3.4.2.2.3. Öznitelikleri Çıkarma**

Öznitelikler metnin analiz için kullanılacak olan en küçük birimlerini ifade eden yapılarıdır. Öznitelik çıkarmak için temel olarak iki yol bulunmaktadır. Bunlardan biri doğal dil işleme yöntemlerini kullanarak sözcüklerin köklerine inme (gövdeleme), diğeri de istatistik yöntemlerinin kullanıldığı n-gram yöntemidir.

#### **3.4.2.2.3.1. Gövdeleme**

Metin madenciliği bir bilgisayar ve çeşitli programlar kullanılarak gerçekleştirilen bir süreçtir. Dolayısıyla sürecin bazı kısıtları bulunmaktadır. Bilgisayarın donanımsal olarak işlem hızı ve hafıza kapasiteleri bunların en başta gelenleridir. Örneğin metin içerisinde yer alan “sözcük”, “sözcükler”, “sözcüklerden” gibi aynı anlama gelen bir sözcüğün çeşitli ek almış halleri bulunabilir. Bu sözcüklerin hepsinin analize dahil edilmeye kalkışılması durumunda işlem hacminin büyümesi, işlem hızının düşmesi, hafızada daha fazla yer kaplaması gibi teknik sorunların yanı sıra oluşturulan modelin bu sözcüklerin hepsini farklı bir sözcük olarak algılaması ve sonucu olumsuz etkilemesi gibi bir durum da meydana gelecektir. Bu gibi durumların önüne geçmek için aynı sözcüğün anlamı değiştirmeyen ek almış farklı versiyonlarının tümü tek bir kök sözcüğe dönüştürülür. Bu işleme gövdeleme denmektedir. Gövdeleme ile fiiller mastar hallerine, çoğul sözcükler de tekil hallerine dönüştürülürler (Hotho vd., 2005: 25).

Gövdelemede genellikle iki temel sorun ile karşılaşılır. Bunlardan biri sözcükten yeteri kadar ekin atılamayarak köke ulaşılamaması, diğeri de gereğinden fazla ek atılarak anlam olarak farklı bir sözcüğün köküne ulaşılmasıdır (Kannan ve Gurusamy, 2014: 10). Yukarıdaki örnekten hareket edilecek olursa, “sözcüklerden”

sözcüğü sadece “sözcükler” sözcüğüne kadar gövdelenirse eksik; “söz” sözcüğüne kadar gövdelenirse fazla gövdelenmiş olur ve yanlış sonuçlara yol açar.

Gövdeleme literatürde “stemming”(kök bulma) ve “lemmetization” (anlamsal köke inme) olarak ikiye ayrılmaktadır. “Stemming” işleminde sözcüklerin son eklerinin atılarak köke inmeye çalışılması gibi daha yüzeysel bir yol tercih edilirken, “lemmetization”da bir sözlük kullanılarak ve sözcüklerin anlamları da analiz edilerek köke inmeye çalışılır. Bu haliyle “lemmetization” “stemming”e göre daha gelişmiş ve detaylı bir işlemdir (Manning, Raghavan ve Schütze, 2009: 32). Örneğin İngilizce “meeting” sözcüğü “stemming” işleminde sözcüğün anlamına bakılmadan sadece “meet” olarak ele alınırken, “lemmatization” işleminde sözcüğün fiil hali “meet”in yanı sıra isim hali “meeting” sözcüğü de analize dahil edilir (Miner vd, 2012: 47).

Dünya üzerinde çok çeşitli diller için kullanılan çeşitli gövdeleme algoritmaları mevcuttur. Bunların içinden Porter tarafından başta İngilizce için oluşturulan “Snowball” algoritmasının diğerlerine göre daha fazla öne çıktığı söylenebilir (Feinerer, Hornik ve Meyer, 2008: 24). “Snowball” gövdeleme algoritması günümüzde Türkçe de dahil olmak üzere çok sayıda dilin gövdelenmesi işlemini gerçekleştirmektedir. Bunun yanı sıra Türkçeye has olarak oluşturulan “Zemberek” doğal dil işleme kütüphanesi de Türkçe çalışmalarda sıklıkla kullanılmaktadır. Zemberek, kök bulmak için kök ve ek sözlüğü kullanan ve sözcüklerin türlerini de belirleme özelliği bulunan açık kaynak kodlu bir gövdeleme algoritmasıdır. Snowball’da ise sözcük türlerini bulma özelliği bulunmamaktadır. Son olarak İstanbul Teknik Üniversitesi Doğal Dil İşleme çalışma grubu tarafından geliştirilen “İTÜ doğal dil işleme yazılım zinciri” adındaki platform da doğal dil işleme çalışmaları için başvurulabilecek seçeneklerden bir diğeridir (Aksaya, 2021: 25).

Türkçe için gövdeleme algoritmalarının henüz tam olarak istenilen seviyeye ulaştığı söylenemez. Bunda Türkçenin sondan eklemeli bir dil olması, sert sessizlerin benzeşmesi/yumuşaması, harf düşmesi vb. gibi kendine has durumlarının söz konusu olması, yapım eki-çekim eki tespitinin zorluğu gibi etkenler etkili olmaktadır (Değer, 2017: 71).

#### **3.4.2.2.3.2. N-Gram Yöntemi**

N-gram yöntemi karakter ve terim tabanlı olmak üzere ikiye ayrılmaktadır. Karakter tabanlı n-gram yönteminde sözcükler karakterlerine ayrıştırılarak analiz

edilirler. En çok kullanılanları karakter tabanlı 2,3 ve 4-gram yöntemleridir. Karakter tabanlı 2-gram yönteminde sözcükler ikişerli karakterler olarak parçalara ayrılmaktadır. Örneğin “sözcük” sözcüğü, “sö”, “öz”, “zc”, “cü” ve “ük” şeklinde parçalara ayrılır. Aynı sözcük 3-gram yönteminde ise “söz”, “özc”, “zcü” ve “cük” şeklinde parçalarına ayrılır. Sistem bu şekilde devam etmektedir. Görüldüğü üzere n-gram yönteminde sözcüklerin ve ayrıştırıldıkları parçaların anlamlarına bakılmaz, sistem tamamen bu parçaların yan yana gelme durumları üzerine çalışan istatistik hesaplamalarına dayalıdır. Karakter tabanlı n-gram yönteminin en büyük avantajı metnin hangi dile ait olduğunun herhangi bir öneminin olmaması ve metin içerisinde yer alan yazım hataları, noktalama işaretleri gibi gürültülerden etkilenmemesidir (Değer, 2017: 73).

Terim tabanlı n-gram yönteminde ise metin n sayısı kadar terimlerine ayrılarak analiz edilir. En çok kullanılanları terim tabanlı 2 ve 3-gram yöntemleridir. Örneğin terim tabanlı 2-gram yöntemine göre “tez yazmak oldukça zahmetli bir işmiş” cümlesi, “tez yazmak”, “yazmak oldukça”, “oldukça zahmetli”, “zahmetli bir” ve “bir işmiş” ibarelerine ayrılır. Terim tabanlı 3-gram yöntemi göre ise aynı cümle, “tez yazmak oldukça”, “yazmak oldukça zahmetli”, “oldukça zahmetli bir” ve “zahmetli bir işmiş” şeklinde ibarelere ayrılırlar.

N-gram yönteminde n sayısı arttıkça işlem hacmi büyüyeceğinden analiz teknik olarak zorlaşacaktır. Bu sebeple yüksek n sayıda analiz gerçekleştirilmesi pek tercih edilmez. N-gram yöntemi metin madenciliğinde özellikle duygu analizi çalışmalarında doğal dil işleme yöntemlerinin kullanılmadığı durumlarda sıklıkla kullanılmaktadır. (Değer, 2017: 72-73).

#### **3.4.2.2.4. Dilbilimsel Ön İşleme**

Metin madenciliğinde bazı durumlarda sadece metin ön işleme yeterli olmamakta, bunun yanı sıra metnin anlamsal olarak da incelendiği dilbilimsel ön işleme metotları uygulanmaktadır (Hotho vd., 2005: 27). Bunların ilki sözcük türü etiketleme (Part of Speech [POS] Tagging) yöntemidir.

Sözcük türü etiketleme yönteminde sözcüklerin cümle içindeki türleri ortaya çıkarılır; isim, fiil, sıfat, zamir vs. şeklinde etiketleme yapılır. Sözcüklerin türlerinin belirlenmesi analiz için daha sağlıklı sonuçlara ulaşılmasına olanak sağlamaktadır. Örneğin “kazan” sözcüğü isim olabileceği gibi, “kazanmak” fiilinin emir kipi de

olabilir. Sözcük türünün tam olarak ne olduğunun bilinmesi gövdelemenin daha doğru olarak yapılmasına imkan vermektedir.

Bu yöntemde genellikle makine öğrenmesi metodu kullanılır. Önce sözcüklerin manuel olarak tipleri etiketlenerek bir eğitim seti oluşturulur. Daha sonra algoritma bu eğitim setini dikkate alarak geri kalan sözcüklerin türlerini belirler. Sözcük türlerini belirlemek için istatistiksel yöntemler kullanılır. Sözcüklerin eğitim veri setindeki sıralanışları (hangi sözcüğün hangi sözcükten sonra geldiği) dikkate alınarak yeni sözcüklerin tipleri tahmin edilmeye çalışılır. Sözcük tipleri belirlendikten sonra bu bilgi yukarıda da belirtildiği üzere doğru gövdeleme yapmak için kullanılır. Bunun dışında bu sözcüklerin kullanım sıklıkları Bag of Words (BOW)” adı verilen tek bir havuzda toplanır ve daha sonra yapılacak olan ağırlıklandırma işleminde sözcüklerin türlerine uygun ağırlıklar alması sağlanır (Hotho vd., 2005: 29).

Makine öğrenmesi metodunun kullanılmadığı durumlarda kural tabanlı yöntemler kullanılır. Ancak kural tabanlı yöntemlerde olasılık tabanlı modeller çalıştırılmaz (Değer, 2017: 76).

Sözcük türü etiketlemeden sonra öbek tanıma (phrase recognition) işlemi gerçekleştirilebilir. Öbek tanıma ile metin içerisinde yer alan kişi isimleri, yer isimleri, kısaltmalar, semboller, fiil ve edat öbekleri belirlenir (Değer, 2017: 76).

Bir diğer dilbilimsel ön işleme yöntemi ise sözdizimsel analizdir (parsing). Çoğunlukla öbek tanıma işlemi ile birlikte uygulanır ve öbek tanımaya yardımcı olur. Cümle içerisindeki söz diziminin anlamlı olup olmadığını ortaya çıkarmayı hedefler. Bunu gerçekleştirmek için cümle içindeki her bir sözcük, türüne göre ayrıştırılır ve cümlenin öğeleri (özne, nesne vs.) ortaya çıkarılır (Değer, 2017: 77).

### **3.4.2.3. Özniteliklerin Ağırlıklandırılması ve Yapısal Hale Getirilmesi**

Gövdeleme veya n-gram yöntemleri sonucu elde edilen özniteliklerin hepsi analiz için aynı derecede öneme sahip olmayabilir. Bu önem derecesi özniteliklerin metin içerisinde geçme sıklığının diğer özniteliklerin toplam geçme sıklıklarına oranına göre belirlenebilir. Terim ağırlıklandırma adı da verilen bu işlem sonucunda öznitelikler ayrıca veri madenciliği algoritmalarının uygulanabilmesine imkan verecek şekilde yapısal hale de dönüştürülürler. En sık kullanılan terim ağırlıklandırma yöntemleri aşağıdaki gibidir:

İkili Ağırlıklandırma: Terim dokümanda geçiyorsa 1, geçmiyorsa 0 olarak ağırlıklandırılır. Bu yöntem ile sadece terimin dokümanda geçip geçmediği bilgisine ulaşılabilir, terimlerin birbirleri ile olan ilişkilerine dair bilgilere ulaşılamaz (Demir, 2013: 13).

Terim Frekans Ağırlıklandırması (TF): Metin içerisinde daha sık geçen terimlerin daha fazla önem teşkil ettiği yaklaşımına dayanır. Terimler metin içerisinde geçme sıklıklarına göre ağırlıklandırılırlar.

Ters Doküman Frekans Ağırlıklandırması (IDF [Inverse Document Frequency]): Terim frekansı ağırlıklandırmasının tersi olarak, metin içerisinde daha az sıklıkta geçen terimlerin ayırt edici güçlerinin daha yüksek olduğu yaklaşımına dayanır. Bu yöntemde daha az sıklıkta geçen terimler daha fazla ağırlıklandırılırlar.

Terim Frekansı – Ters Doküman Frekans Ağırlıklandırması (TF/IDF): Terim frekansı ile ters doküman frekansı ağırlıklandırma yöntemlerinin birleşimidir. Metin madenciliğinde metin genellikle dokümanlara bölünmüş bir haldedir. Başka bir deyişle metin madenciliğinin en küçük çalışma birimi dokümanlardır. TF/IDF ağırlıklandırma yöntemi bir terimin bir dokümanda geçme sıklığının fazla oluşunun o terimin önemini arttırdığı, ancak aynı terimin diğer tüm dokümanlarda da geçme sıklığının yüksek oluşunun terimin önem derecesini / ayırt etme gücünü düşürdüğü yaklaşımına dayanır. Dolayısıyla terimler dokümanda ne kadar fazla geçiyorsa ağırlıkları o oranda arttırılır, diğer dokümanlarda geçme sıklıkları yükseldikçe de ağırlıkları düşürülür (Değer, 2017: 80). TF/IDF ağırlıklandırma yöntemi özellikle metin sınıflandırma işlemlerinde diğer ağırlıklandırma yöntemlerine göre daha fazla tercih edilmektedir.

Özniteliklerin ağırlıklandırılması işlemi ile birlikte aynı zamanda öznitelikler üzerlerinde veri madenciliği algoritmalarının uygulanabilmesine imkan verecek şekilde yapısal hale de dönüştürülürler. Başka bir deyişle metin formatındaki veri sayısallaştırılır. Bunu gerçekleştirmenin çeşitli yolları bulunsa da en çok tercih edileni “vektör uzay modeli”dir. Metin madenciliğinde çalışma birimi daha önce de belirtildiği üzere “doküman (belge)”dır. Başka bir deyişle metin dokümanlardan oluşmaktadır. Dokümanlar da cümleler, sözcükler veya karakterlerden müteşekkil özniteliklerden oluşmaktadır. Vektör uzay modelinde her doküman içinde barındırdığı öznitelik sayısı kadar boyutlu vektör ile temsil edilir. Her bir dokümanı

temsil eden vektörler de birleşerek tüm dokümanları temsil eden bir matris oluştururlar. Bu matrise doküman-terim (belge-terim) matrisi adı verilir.

$$\begin{array}{c|cccc}
 & T_1 & T_2 & \dots & T_t \\
 D_1 & d_{11} & d_{12} & \dots & d_{1t} \\
 D_2 & d_{21} & d_{22} & \dots & d_{2t} \\
 \dots & \dots & \dots & \dots & \dots \\
 D_n & d_{n1} & d_{n2} & \dots & D_{nt}
 \end{array}$$

Şekil 3.3. Doküman-terim matris gösterimi

Şekil 3.3'teki doküman-terim matris gösteriminde her bir satır doküman vektörünü gösterirken, her bir sütun da terim vektörünü göstermektedir (Pekin 2020: 28). Bir öznitelik bu matris içerisinde  $d_n$  dokümanının  $t_t$  terimi olarak gösterilir. Örneğin  $d_{12}$  vektörü,  $D_1$  dokümanının  $T_2$  terimini ifade eder. Bu vektörlerdeki özniteliklerin sayısal değerleri kullanılan ağırlıklandırma yöntemine göre belirlenir (Aksoy vd., 2020: 141).

Görüldüğü üzere dokümanlardaki terim sayıları ne kadar fazla olursa doküman-terim matrisi de o kadar büyür ve bu durumda işlem hacmi de aynı oranda büyüyecektir. Her ne kadar metin formatındaki veriler vektör uzay modeli kullanılarak sayısallaştırıldığında aynı zamanda boyut olarak da bir azalma gerçekleşse de yine de matrisin büyümesi aynı zamanda teknik sınırların daha fazla zorlanması anlamına gelecektir. Bu sebeple analizden maksimum verim alacak şekilde öznitelik sayısının mümkün olduğu kadar düşük tutulması büyük önem arz etmektedir. Bu, metin sayısallaştırılmadan önce metin ön işleme aşamasında dile getirilen yöntemlerin uygulanması yolu ile yapılabileceği gibi metin sayısallaştırıldıktan sonra uygulanabilecek boyut azaltma yöntemleri ile de gerçekleştirilebilir.

Sonuç olarak vektör uzay modeli ile hem dokümanlar hem de öznitelikler birbirleri ile karşılaştırılabilme imkanına sahip olmakta ve birbirleri ile benzerlik ve farklılık ilişkileri saptanabilmektedir (Değer, 2017: 96).

#### 3.4.2.4. Boyut Azaltma

Metin formatındaki veriler özniteliklerine ayrıştırılıp yapısal hale getirildikten sonra yukarıda da belirtildiği üzere çeşitli boyut azaltma yöntemleri uygulanarak

boyutları azaltılabilir. Bu, bazı çalışmalar için özellikle teknik kapasite sebeplerinden dolayı elzem olabilmektedir.

Metin madenciliğinde boyut azaltma iki farklı şekilde yapılabilmektedir: öznitelik seçimi (feature selection) ve öznitelik çıkarımı (feature extraction). Öznitelik seçiminde analize daha fazla katkı sağlayacak özniteliklerin seçilmesi ve geri kalanlarının analiz dışında bırakılarak boyut azaltılması söz konusu iken; öznitelik çıkarmada özniteliklerin daha az boyut kaplayacak şekilde dönüştürülmeleri söz konusudur.

#### **3.4.2.4.1. Öznitelik Seçimi Yöntemi ile Boyut Azaltma**

Daha önce belirtildiği üzere teknik sınırlamalar sebebi ile boyut azaltma yöntemleri uygulanabilmektedir. Ancak boyut azaltmanın tek sebebi bu değildir. Boyut azaltma ile aynı zamanda analize katkısı olmayan veya nispeten az olan öznitelikler de analiz kapsamı dışına çıkarılabilir ve bu yolla verim artırılabilir ki bu yöntem öznitelik seçim yöntemleri kapsamına girmektedir. Boyut azaltmak için uygulanabilecek çeşitli öznitelik seçim yöntemleri mevcut olmakla birlikte, bunların içinden bilgi kazanımı ve ki-kare yöntemlerinin yaygın olarak tercih edilen iki yöntem olduğu söylenebilir (Topaçan, 2016: 90).

Bilgi kazanımı yöntemi her bir özneliğin dokümanı ayırt etme gücünün hesaplanarak en yüksek ayırt etme gücüne sahip özniteliklerin seçilmesi esasına dayanmaktadır. Bir özneliğin bilgi kazanımı ne kadar yüksekse ayırt etme gücü de o kadar yüksektir. Bilgi kazanımını hesaplamak için entropinin hesaplanması gerekmektedir. Entropi, bir sistemin düzensizliğini ifade etmek için kullanılan bir terimdir. Bir özneliğin bilgi kazanımı veri setinin entropisinden özneliğin entropisinin çıkarımıdır. Bir özneliğin bilgi kazanımı o özneliğin entropide meydana getireceği azalma hesaplanarak bulunur (Topaçan, 2016: 92).

Ki-kare testi değişkenler arasında bir ilişki olup olmadığını, varsa bu ilişkinin derecesini ortaya koymak için uygulanan bir metottur. Boyut azaltmada kullanılırken, hesaplanan ki-kare değeri ne kadar yüksekse özneliğin dokümanı ayırt etme gücünün de o kadar yüksek olduğu varsayılır (Topaçan, 2016: 92).

Pratikte hem bilgi kazanımı hem de ki-kare yöntemi uygulanırken, tüm özniteliklerin bilgi kazanımı veya ki-kare değerleri hesaplandıktan sonra uygulayıcı tarafından çalışmaya göre istenilen sayıda veya belli bir eşik değerin üzerinde, en

yüksek ayırt etme gücüne sahip öznelikler seçilip geri kalanlar analize dahil edilmeyerek boyut azaltılmış olur.

#### **3.4.2.4.2. Öznelik Çıkarımı Yöntemi ile Boyut Azaltma**

Pratikte uygulanan çeşitli öznelik çıkarımı yöntemleri bulunmakla birlikte temel bileşenler analizi (principal component analysis) ve tekil değer ayrışımı (singular value decomposition) yöntemlerinin yaygınlıkla tercih edilen iki yöntem olduğu söylenebilir.

Temel bileşenler analizi, veri setinin temel özelliklerini seçerek farklı bir formatta ve daha az veri ile orijinal veri setini ifade etme esasına dayanan denetimsiz, istatistiksel bir yöntemdir (Durmaz, 2011: 13). Veri setindeki değişkenler sayıca azaltılarak ve temel bileşenlerine dönüştürülerek boyut azaltılmaktadır (Alkan, 2008: 12; Albayrak, 2006: 88). Temel bileşenler analizi ile veri seti için önemli bilgiler çıkarılmakta ve sadece bu önemli bilgileri içerecek şekilde veri seti sıkıştırılmaktadır (Durmaz, 2011: 13). Veri setinde aralarında korelasyon bulunan değişkenler, hem sayıca daha az hem de aralarında korelasyon bulunmayacak şekilde orijinal değişkenlerin doğrusal bileşenlerine dönüştürülmüş olurlar (Johnson ve Wichern, 2002: 425). Temel bileşenler analizinde orijinal veri setine göre bilgi kaybı söz konusu olsa da oluşturulan temel bileşenlerin orijinal değişkenleri temsil etme yeteneğine sahip olduğu varsayılır.

Tekil değer ayrışımı yöntemi ise matrislerin ayrıştırılarak daha küçük bir veri uzayı ile temsil edilmeleri esasına dayanmaktadır. Vektör uzay modeli başlığı altında ifade edildiği üzere metin formatındaki veriler doküman-terim matrisi adı verilen bir matris ile ifade edilmektedirler. Metin madenciliğinde tekil değer ayrışımı yöntemi kullanılarak bu matris ayrıştırılmakta ve bu şekilde daha düşük bir boyutla temsil edilmektedir (Durmaz, 2011: 12). Tekil değer ayrışımı yöntemi kullanılarak veri setine dair en etkili özellikler muhafaza edilerek diğerleri çıkarılmaktadır (Abidin, Yusuf ve Umran, 2010; akt. Durmaz, 2011: 12).

#### **3.4.2.5. Veri Madenciliği Algoritmalarının Uygulanması**

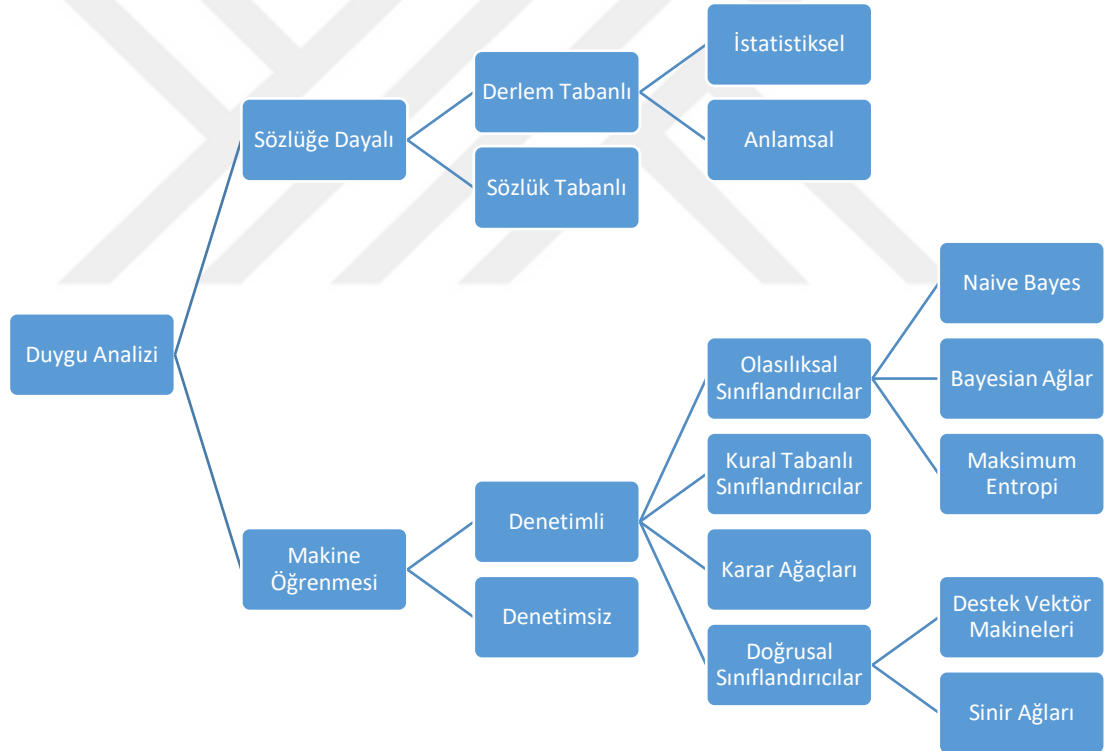
Bu aşamada yapısal hale dönüştürülen veriye sınıflandırma, kümeleme, bilgiye erişim gibi veri madenciliği algoritmaları uygulanır (Sumathy ve Chidambaram, 2013: 30).

### 3.4.2.6. Değerlendirme

Bu aşamada analiz sonucu elde edilen bilgiler veri bilimciler ve kurumların karar alıcıları tarafından değerlendirilir. Ayrıca sonuçların görselleştirme işlemi de bu aşamada yapılır (Polat, 2021: 140).

### 3.5. Duygu Analizi Yaklaşımları

Duygu analizi yaklaşımları temelde makine öğrenmesi ve sözlüğe dayalı yaklaşımlar olmak üzere ikiye ayrılmaktadır. Makine öğrenmesi yöntemleri verinin algoritmaya tanıtılarak öğrenmesinin sağlanması ve öğrenen algoritmanın yeni veriler için otomatik olarak sınıflandırma yapması esasına dayanırken; sözlüğe dayalı yaklaşımlar her bir sözcük için belirli bir duygu skorunun yer aldığı sözlükler kullanılarak bir duygu sınıflandırması yapılması esasına dayanmaktadır.



Şekil 3.4. Duygu analizi sınıflandırma teknikleri (Medhat, Hassan ve Korashy, 2014: 1095)

#### 3.5.1. Sözlüğe Dayalı Yaklaşımlar

Sözlüğe dayalı yaklaşımlar sözlük tabanlı ve derlem tabanlı olmak üzere ikiye ayrılmaktadır. Sözlük tabanlı yaklaşım, her sözcüğün bir duygu skoruna sahip olduğu bir sözlük oluşturulması esasına dayanmaktadır. Uygulayıcı bu sözlüğü manuel olarak kendisi oluşturabileceği gibi daha önce hazırlanmış çok sayıda sözlükten

birini de kullanabilir. Tahmin edileceği üzere herhangi bir duygu analizi çalışmasında kullanılacak bir sözlük oluşturma oldukça zahmetli bir iştir. Özellikle sondan eklemeli bir dil olan Türkçe için bu çok daha zordur. Ayrıca belirli bir alana özgü bir metin için genel bir sözlüğün kullanımı sağlıklı sonuçlar doğurabilmektedir. Bu sebeple çalışmaya özgün sözlük oluşturma şeklinde bir yöntem tercih edilebilmektedir ki bu da derlem tabanlı yaklaşım anlamına gelmektedir. Derlem tabanlı yaklaşımda içeriğe özgün sözcükler bir araya getirilerek büyük bir derlem oluşturulur. Metinde yer alan sözcükler bu oluşturulan derlemde yer alan sözcükler ile karşılaştırılarak duygu tespiti yapılmaya çalışılır. Derlem tabanlı yaklaşım, istatistiksel ve anlamsal olmak üzere ikiye ayrılmaktadır. İstatistiksel yöntemde sözcük derlem içinde yer almıyorsa sözcük ile birlikte kullanılan diğer sözcüklere bakılır. Bu yöntemde bir sözcüğün sıklıkla birlikte kullanıldığı diğer sözcükler ile aynı duygu kutbunda yer aldığı varsayılır. Anlamsal yöntem ise anlamsal olarak birbirine yakın sözcüklerin aynı duygu kutbunda yer aldığı esasına dayanmaktadır (Kızılkaya, 2018: 45).

Sözlüğe dayalı yöntemlerde sadece sözcükler değil, emoji de duygu skorları ile ifade edilirler. Sözcükler pozitif duygu belirtiyor ise “+”, negatif duygu belirtiyor ise “-“ değer alır. Bir cümlenin pozitif mi negatif mi duygu ifade ettiğini tespit etmek için her bir sözcüğün duygu skoru toplanır, çıkan sonuç pozitif ise cümlenin pozitif, negatif ise cümlenin negatif duygu ifade ettiği sonucuna ulaşılır. Sonuç “0” ise cümlenin herhangi bir duygu belirtmediği, “nötr” olduğunu sonucuna ulaşılır (Kızılkaya, 2018: 44-45).

Duygu analizi çalışmaları için kullanılan çok sayıda sözlük mevcuttur. 2004 yılında Bing Liu tarafından geliştirilmeye başlanan sözlük son haliyle 2006 adet pozitif, 4783 adet negatif sözcük içermektedir (Hu ve Liu, 2004). Harvard General Inquirer adlı sözlükte, aynı sözcüğün farklı anlamlarının da ayrı sayıldığı toplam 11787 adet sözcük yer almaktadır (Harvard, 2015; akt. Topaçan, 2016: 54). SentiWordNet, WordNet adlı sözcük veritabanında yer alan sözcük kümeleri (synset) için duygu skoru üreten bir başka uygulamadır (Esuli ve Sebastiani, 2006: 417).

Sözlüğe dayalı yöntemlere makine öğrenmesi için gerekli olan eğitim setinin oluşturulamadığı durumlarda pratik bir çözüm olarak da başvurulabilir (Deng, Sinha ve Zhao, 2017: 66).

### 3.5.2. Makine Öğrenmesi

Makine öğrenmesi yöntemleri denetimli ve denetimsiz olmak üzere ikiye ayrılmaktadır. Denetimli öğrenmede uygulayıcı tarafından öncelikle manuel olarak metnin bir kısmının sınıflandırması yapılır ve bu şekilde bir eğitim seti oluşturulur. Metnin bir kısmı da test seti için ayrılır ve belirlenen algoritma eğitim setini analiz edip buna göre test setini sınıflandırır. Bu şekilde çeşitli sınıflandırıcı algoritmalar kullanılarak oluşturulan modeller çalıştırılır ve performansları ölçülür. Son olarak en yüksek performansı gösteren model kullanılarak tüm metnin sınıflandırması yapılır. Denetimsiz öğrenmede ise uygulayıcı tarafından yapılan manuel bir sınıflandırma yoktur. Metin herhangi bir sınıflandırma bilgisi olmaksızın algoritmaya girer, algoritma metni oluşturulan unsurların birbirlerine olan benzerliklerine göre bir kümeleme işlemi gerçekleştirir. Bazı çalışmalarda denetimsiz öğrenme denetimli öğrenme öncesi bir ön işlem olarak uygulanır.

Denetimli öğrenme, Şekil 2.4'te de görüldüğü gibi kendi içinde çok sayıda sınıfa ayrılmaktadır. Medhat vd. (2014)'ne göre, denetimli öğrenme kendi içinde olasılıksal, kural tabanlı, karar ağaçları ve doğrusal sınıflandırıcılar şeklinde dörde ayrılmaktadır. Yine bu sınıflar da kendi içinde alt sınıflara ayrılabilir.

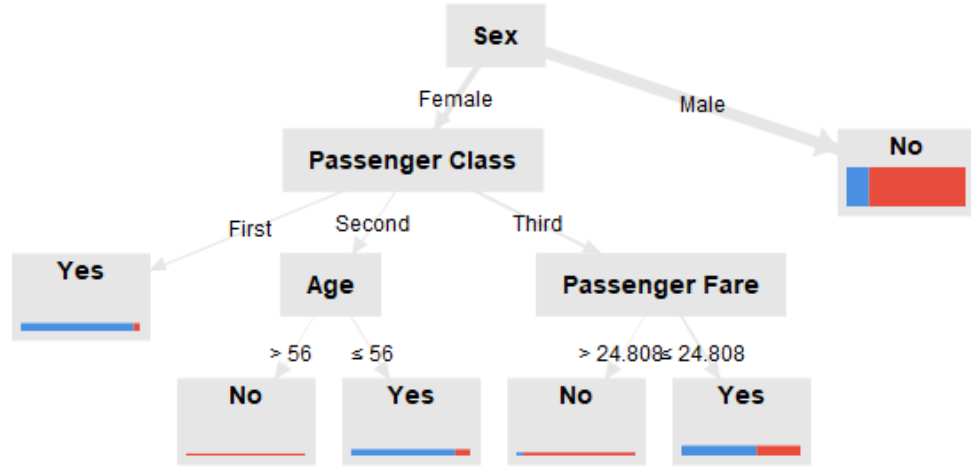
### 3.6. Makine Öğrenmesi Yönteminde Kullanılan Sınıflandırma Algoritmaları

Algoritmalar duygu analizi için oluşturulan modellerin en önemli parçalarıdır (Değer, 2017: 109). Bu başlık altında tez kapsamında tercih edilen sınıflandırma algoritmalarına yer verilmiştir.

#### 3.6.1. Karar Ağacı

Tarihi 1980'lere kadar dayanan karar ağacı (decision tree) algoritması, kullanımının ve yorumlanmasının kolay olması, güvenilirliğinin ve performansının yeterli seviyede olması, veritabanları ile entegrasyonunun kolay olması gibi sebeplerden ötürü bugün dahi geçerliliğini koruyan en eski veri madenciliği algoritmalarından biridir (Williams, 2011: 205; Ersöz, 2019: 70; Çalış, Kayapınar ve Çetinyokuş, 2014: 5). Karar ağacı, bir kök düğüm ile başlayıp kök düğümün dallara ayrılması, dalların da ya yeni bir düğüm ve akabinde yeni dallar ya da bir yaprak ile sonlanması şeklinde gelişen bir yapıdadır. Karar ağacı tüm dallar bir yaprak ile sonlanıncaya dek dallanmaya devam eder (Emel ve Taşkın, 2005: 225). Burada

düğümler veri setinin özniteliklerini, dallar özniteliklerin sınıflarını, yapraklar da tahmini yapılan özniteliğin sınıflandırma sonucunu göstermektedir.



Şekil 3.5. Karar ağacı örneği

Şekil 3.5'te RapidMiner adlı veri madenciliği aracında yer alan bir eğitim uygulamasına ait bir karar ağacı görünmektedir. Modelde, makine öğrenmesi yönteminde olması gerektiği gibi iki farklı veri seti kullanılmaktadır. Eğitim seti olan ilkinde Titanik gemisi yolcularının %80'ine dair cinsiyet, yolcu sınıfı, bilet ücreti, yaş, birlikte bindiği kişi sayısı ve kazadan kurtulup kurtulmadığına dair bilgiler (öznitelikler) yer almaktadır. Diğer veri seti olan test setinde ise yolcuların kalan %20'sine dair eğitim setinde yer alan bilgilerden sadece kazadan kurtulup kurtulmadığına dair bilgi dışındaki geri kalan tüm bilgiler yer almaktadır. Model, eğitim setindeki bilgileri analiz ederek test setindeki yolcuların kazadan kurtulup kurtulmayacağına karar ağacı algoritması kullanılarak tahmin edilmesini sağlayacak şekilde kurulmuştur. Sonuç ise Şekil 3.5'te görüldüğü gibi bir karar ağacıdır. Buna göre karar ağacı algoritmasının yapmış olduğu analiz sonucu en büyük kırılmayı oluşturacak olan öznitelik "cinsiyet"tir. Modelin yapmış olduğu tahmin sonucu yolcu eğer "erkek" ise kurtulma şansı %18,8'dir. Yolcu eğer kadın ise burada "yolcu sınıfı" özniteliği karşımıza çıkmaktadır. 1.sınıf (first class) yolcu ise kurtulma şansı %94,8 olarak tespit edilmiştir. 2.sınıf ise "yaş" özniteliği karşımıza çıkmaktadır. 3.sınıf ise "bilet ücreti" özniteliği önem kazanmaktadır. Karar ağacı algoritması görüldüğü gibi kendisine verilen, tüm bilgilerin yer aldığı bir eğitim setini analiz ederek, test setindeki eksik bilgiyi tahmin etmiştir.

Şekil 3.5'teki karar ağacında kök düğüm “cinsiyet (sex)”tir. “Cinsiyet” düğümü “erkek (male)” ve “kadın (female)” olmak üzere dallara ayrılmış; bunlardan “kadın” dalı “yolcu sınıfı (passenger class)” adında başka bir düğüm ile sonuçlanırken, “erkek” dalı bir “sınıflandırma sonucu” olan “yaprak” ile sonuçlanmıştır. Karar ağaçlarında her bir dal bir sınıflandırma sonucunu ifade eden bir “yaprak” ile sonuçlanana dek devam eder. Karar ağacının derinliği (inebileceği maksimum dal sayısı), düğüm olacak özniteliklerin hangi kritere göre belirleneceği, bir düğümün oluşturacağı minimum yaprak sayısı gibi özellikler modelin oluşturulma aşamasında belirlenebilir. Burada karar ağacının yapısını etkileyecek en önemli unsur kök düğüm de dahil olmak üzere düğüm olarak seçilecek özniteliklerin hangi kritere göre belirleneceğidir (Budak, 2021: 8). Karar ağaçları algoritmasına dayalı olarak geliştirilen çok sayıda kriter belirleme algoritması mevcuttur ve bu algoritmalar birbirlerinden kök, düğüm ve dallanma kriteri seçimlerinde izledikleri yollar bakımından ayrılmaktadırlar (Silahtaroglu, 2004: 41).

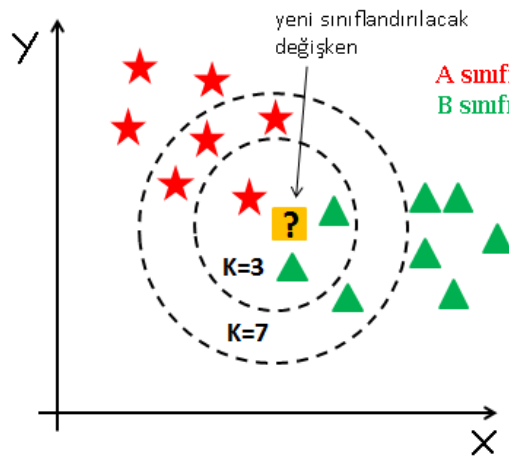
Bilgi kazanımı, bilgi kazanımı oranı, gini indeksi, ki-kare gibi algoritmalar yaygın olarak kullanılan algoritmalar olmakla birlikte bunların içinden bilgi kazanımı oranı algoritmasının diğerlerine göre daha fazla tercih edildiği söylenebilir. İlk olarak 1993 yılında Quinlan tarafından ortaya ve C4.5 algoritması olarak da bilinen bilgi kazanımı oranı algoritması, bilgi kazanımı algoritmasından türetilmiştir ve bilgi kazanımı algoritmasına göre bazı üstünlükleri bulunmaktadır (Silahtaroglu, 2004: 43). Hem bilgi kazanımı hem de bilgi kazanımı oranı algoritmalarında öznitelik seçimi bilgi kazanımı hesaplama yöntemine göre yapılmaktadır. Bilgi kazanımının nasıl hesaplandığı, “Boyut Azaltma” başlığı altında daha önce ifade edildiği için burada tekrar edilmesine gerek görülmemiştir. Sonuç olarak kök düğümden başlamak üzere düğümler veri setinin tüm öznitelikleri arasından en yüksek bilgi kazanımına sahip öznitelikler seçilerek belirlenir. Bu şekilde belirlenen öznitelikler aynı zamanda sınıflama için ayırıştırma gücü en yüksek özniteliklerdir (Mitchell, 1997; akt. Silahtaroglu, 2004: 41).

Karar ağacı algoritmasının dezavantajları ise, eğitim setindeki küçük değişikliklerin algoritmanın karar mantığında büyük değişikliklere neden olabilmesi, karar ağacındaki dalların sayısı arttıkça yorumlamanın zorlaşması (Lantz, 2013: 125) ve veri setlerinin yeteri kadar büyük olmaması durumunda hatalı sonuç verme ihtimalinin yükselmesi (Değer, 2017: 111) şeklinde ifade edilebilir.

### 3.6.2. K-En Yakın Komşu

K-En yakın komşu (K-Nearest Neighbor [kNN]) algoritması, özniteliklerin birbirlerine olan uzaklıklarının hesaplanarak, sınıfı belirsiz bir özneliğin sınıfını belirlemek için ona uzaklık olarak en yakın komşu özniteliklerin sınıfına bakılarak karar verilmesi esasına dayanmaktadır. “k” burada sınıfı belirsiz özneliğin hangi sınıfa ait olduğunu belirlemek için bakılacak olan en yakın komşu öznelik sayısıdır. K sayısı yaygın olarak 1-9 arasında seçilmektedir ve genellikle tek sayı seçilir. Tek sayı seçilmesindeki sebep, sınıfı belirlenecek olan özneliğin kendisine eşit mesafede ve farklı sınıfta en yakın komşu özniteliklerinin bulunması ihtimalidir. Örneğin k değerinin 2 seçildiğini, sınıfı belirlenecek olan özneliğin en yakınındaki iki komşu özneliğe eşit mesafede olduğunu ve bu iki komşu özneliğin de farklı sınıflarda olduğunu düşünelim. Bu durumda algoritma temel mantığına göre sınıflandırma yapamayacaktır. Pratikte yaygın olarak 1-9 arasındaki farklı tek sayı k değerleri modelde denenerek en yüksek performans göstereni seçilir. (Ağdeniz, 2017: 88). K değerinin yüksek olması, birbirine benzemeyen özniteliklerin aynı sınıfta toplanmasına; düşük olması ise birbirine benzediği halde özniteliklerin aynı sınıfta toplanamamasına neden olmaktadır.

Özniteliklerin birbirlerine olan uzaklıklarını hesaplamak için genellikle öklit hesaplama yöntemi kullanılır. Algoritma, sınıfı belirlenecek olan özneliğin k sayısı kadar en yakın komşu özneliğinin sınıfına bakılıp, çoğunluk komşu özneliğinin sınıfı hangi sınıf ise o sınıfa atanması esasına göre çalışır. K-En yakın komşu algoritması çalışma mantığı en basit algoritmalarından biridir (Ağdeniz, 2017: 88).



Şekil 3.6. K-En yakın komşu sınıflandırma örneği (Gokte, 2020)

Şekil 3.6’da görüldüğü gibi, veri uzayına yeni eklenen bir değişken, k değeri kadar kendisine en yakın komşu değişkenlerin sınıfına bakılarak sınıflandırılmaktadır. Kendisine en yakın komşu değişkenlerin çoğunun sınıfı ne ise yeni değişken de o sınıfa atanmaktadır.

### 3.6.3. Naive Bayes

Naive Bayes (Basit Bayes) sınıflandırıcısı, Bayes teoremine dayanan olasılıksal bir sınıflandırma yöntemidir (Değer, 2017: 115). Yöntem, daha önce sınıflandırılmış verileri kullanarak, yeni verilerin mevcut sınıflardan herhangi birine girme olasılığını hesaplayarak çalışır (Silahtaroglu, 2004: 48). Algoritmanın sınıflandırma yapmak için kullandığı Bayes teoremi aşağıdaki gibi formüle edilebilir:

$$P(C|x) = \frac{P(x|C)P(C)}{P(x)} \quad (3.1)$$

Metin sınıflandırma çalışmaları için formüldeki;

- “C”, dokümanın sınıfını; “x” ise dokümanın öznitelik vektörünü temsil etmektedir. n sayıda dokümanın bulunduğu bir veri setinde dokümanlar,  $x_1, x_2, \dots, x_n$  şeklinde; m sayıda sınıf ise  $C_1, C_2, \dots, C_m$  şeklinde tanımlanır.
- $P(C|x)$ , x dokümanının C sınıfında olması olasılığıdır.
- $P(x|C)$ , C sınıfında x dokümanına ait özniteliklerin bulunması olasılığıdır.
- $P(C)$ , veri setinde C sınıfının bulunma olasılığıdır.
- $P(x)$ , veri setinde x dokümanının bulunma olasılığıdır.

Bir dokümanın hangi sınıfta olacağına karar verilmesi için her bir sınıf için  $P(C|x)$  olasılıkları hesaplanır, hangi olasılık daha büyük ise doküman o sınıfa atanır. Örneğin iki sınıfın olduğu sınıflandırma için,  $P(C_1|x) > P(C_2|x)$  ise, x dokümanı  $C_1$  sınıfına atanır. Buna göre bu karşılaştırmayı formülde yerine yazarsak, aşağıdaki eşitsizliğe ulaşmış oluruz.

$$\frac{P(x|C_1)P(C_1)}{P(x)} > \frac{P(x|C_2)P(C_2)}{P(x)} \quad (3.2)$$

Yukarıdaki eşitsizlikte görülebileceği gibi, eşitsizliğin her iki tarafının da paydaları aynıdır. Buna göre her iki tarafta da  $P(x)$  değerleri silinebilir. Güncellenen formül şu şekilde yazılabilir (Topaçan, 2016: 98):

$$P(C|x) = P(x|C)P(C) \quad (3.3)$$

Buna göre bir dokümanın hangi sınıfta olduğunu hesaplamak için tüm sınıflar için  $P(x|C)$  ve  $P(C)$  olasılıklarının hesaplanması yeterlidir.  $P(x|C)$  ve  $P(C)$  olasılıklarının hesaplanması, eğitim setinde yapılan sınıflandırmalara bakılarak yapılır.  $P(x|C)$ , C sınıfında x dokümanındaki her bir özneliğin bulunma olasılıklarının çarpımına eşittir. C sınıfında x dokümanındaki bir özneliğin bulunma olasılığı da o özneliğin C sınıfında bulunma sayısının C sınıfındaki tüm özniteliklerin sayısına bölümüdür. Bu hesaplamada sonucun 0 (sıfır) çıkmasını önlemek için  $\alpha$  düzeltme terimi kullanılır.  $P(C)$  ise, C sınıfına ait doküman sayısının toplam doküman sayısına bölümüdür.

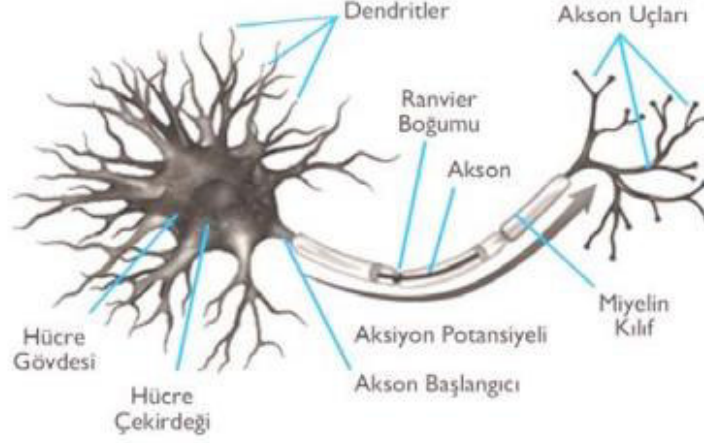
Naive Bayes sınıflandırma süreci şu şekilde işler:

Sınıflandırması yapılacak dokümandaki tüm özniteliklerin tüm sınıflar için  $P(x|C)$  olasılık değerleri ve yine tüm sınıflar için  $P(C)$  olasılık değerleri hesaplanır. Tüm sınıflar için bu iki değer çarpılarak sınıfların  $P(C|x)$  değerleri hesaplanır. Hangi sınıfın  $P(C|x)$  değeri daha büyük ise doküman o sınıfa atanır.

Naive Bayes sınıflandırıcısı her özneliğin eşit öneme sahip olduğunu ve bir özneliğin bir sınıf üzerindeki etkisinin diğer özniteliklerden bağımsız olduğunu varsayar (Han vd., 2012: 350; Balaban ve Kartal, 2015: 69). Bu sebeple veri setindeki özniteliklerin birbirileri ile olan bağımlılık seviyelerinin düşük olduğu durumlarda daha başarılı sonuçlar üretir. Ancak gerçek hayatta veri setindeki özniteliklerin birbirileri ile bağımlılık seviyeleri yüksektir, bu da Naive Bayes sınıflandırıcısının dezavantajlı tarafıdır (Polat, 2021: 123). Yeni eklenen dokümanlar ile eğitim setinin sürekli olarak yenilenmesi ile gürültülü ve eksik verilerin sınıflandırıcının başarısını olumsuz etkilememesi avantajlı taraflarıdır (Lantz, 2013: 95).

### **3.6.4. Derin Öğrenme**

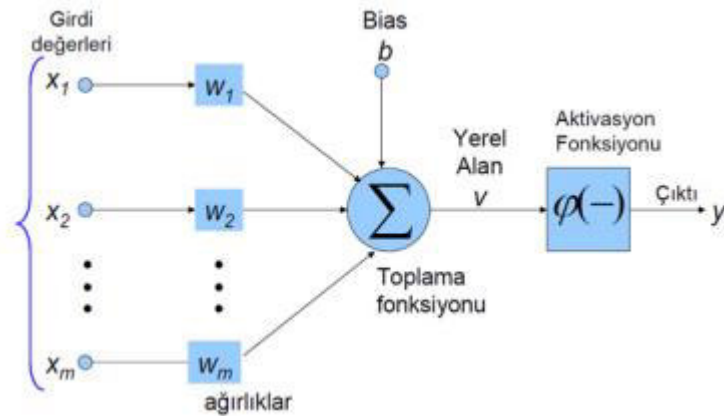
Derin öğrenme (deep learning), yapay sinir ağları başlığı altında yer alan bir makine öğrenmesi algoritmasıdır. Algoritma, insan sinir hücresi olan nöronların çalışma yapısı esas alınarak oluşturulmuştur.



Şekil 3.7. Sinir hücresinin (nöron) yapısı (Mijwel, 2017: 1)

İnsan beyni bilginin nöronlar vasıtasıyla taşınması suretiyle öğrenmeyi gerçekleştirir. Bilgi, nöronlar vasıtasıyla elektriksel bir sinyal olarak taşınmaktadır. Bir nöron, dendritler, hücre gövdesi ve aksondan oluşmaktadır. Dendritler bilgiyi toplayan çok sayıda girişten oluşmaktadır. Dendritlerden gelen bilgiler hücre gövdesinde birleştirilir. Toplanan bu bilgiye aksiyon potansiyeli adı verilir. Birleştirme işleminden sonra hücre gövdesinde aksiyon potansiyeline bakılarak bir çıkış sinyali üretilip üretilmemesine karar verilir. Bir çıkış sinyali üretilecek ise aksiyon potansiyeli aksonlar üzerinden taşınarak diğer nöronlara aktarılır (Yavuzkılıç, 2022: 21).

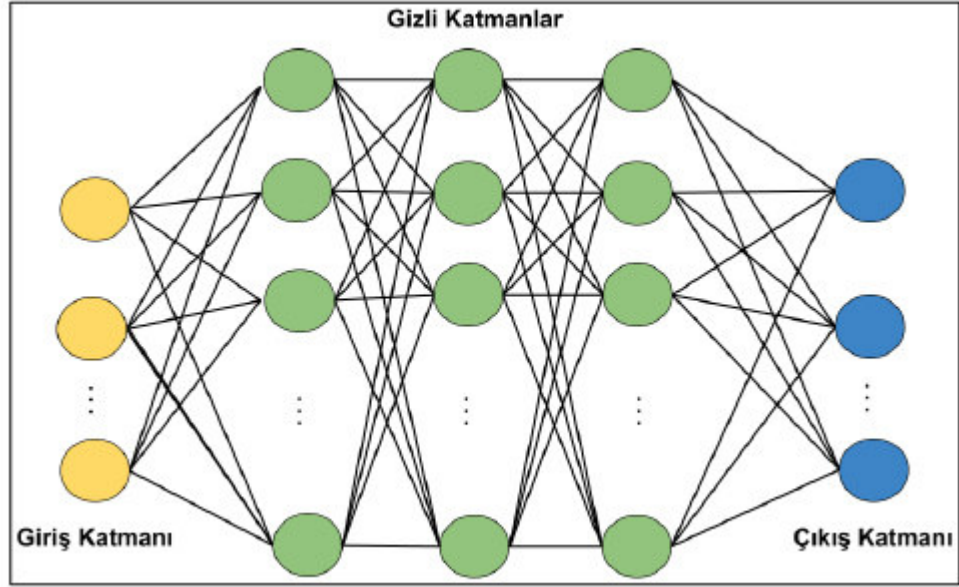
Yapay bir sinir hücresi de insan sinir hücresinin çalışma yapısına benzer bir şekilde tasarlanmıştır. Buna göre dendritlerin yerini girdiler, hücre gövdesinin yerini girdilerin birleştirilmiş hali olan fonksiyonlar ve aksonların yerini de çıktılar almaktadır.



Şekil 3.8. Yapay nöron (Mijwel, 2017: 1)

Şekil 3.8’de yapay bir sinir hücresi modeli görünmektedir. Burada girdiler özniteliklerdir. Öznitelikler farklı ağırlıklandırılarak toplanırlar.  $x$  değerlerinin sıfır olması sonucu ağırlık değerlerinin anlamsız olması ihtimaline karşın toplama bir “bias (yanlılık)” değeri eklenmektedir (Doğan, 2021: 38). Son olarak toplama fonksiyonuna aktivasyon fonksiyonu uygulanarak  $y$  çıktısı oluşturulur. Aktivasyon fonksiyonu toplama fonksiyonunu bir sonraki katmanın giriş sinyali olarak kullanılabilir şekilde çıkış sinyaline dönüştüren özel fonksiyonlardır. Aktivasyon fonksiyonları doğrusal ve doğrusal olmayan fonksiyonlar olarak ikiye ayrılmaktadır (Altuncu, 2021: 43). Gerçek hayatta veriler çoğunlukla doğrusal olmayan veri olduğu için doğrusal olmayan aktivasyon fonksiyonları daha yaygın olarak kullanılmaktadır. Aktivasyon fonksiyonları ayrıca bir sonraki nöronun aktive edilip edilmeyeceğini belirler. Bir sonraki nöronun aktive edilebilmesi için, başka bir deyişle çıkış sinyalinin gönderilebilmesi için aktivasyon fonksiyonunun belli bir eşik değerini aşması gerekmektedir (Silahtaroglu, 2004: 55).

Yapay bir sinir ağı yukarıda yapısı açıklanan yapay bir sinir hücresinden çok sayıda bulunan ağa denmektedir ve en ilkel hali ile en az iki katmandan oluşmaktadır. Bunlar giriş ve çıkış katmanlarıdır. Ancak günümüzde gerçekleştirilecek olan işlemlerin karmaşıklığı, öğrenme kapasitesi ihtiyacının yüksekliği, veri boyutunun büyüklüğü gibi sebeplerden ötürü iki katmanlı bir yapay sinir ağının pratikte kullanımı kalmamıştır. Onun yerine giriş ve çıkış katmanları arasında gizli katmanların da yer aldığı ağlar kullanılmaktadır. Gizli katmanların zamanla yüksek sayılara ulaşması “derin sinir ağı” ve “derin öğrenme” kavramlarını ortaya çıkarmıştır. Başka bir deyişle modelin derinliği sahip olduğu gizli katman sayısı ile orantılıdır.



Şekil 3.9. Basit bir derin sinir ağı yapısı (Altuncu, 2021: 43)

Şekil 3.9'daki her bir daire yapay bir sinir hücresini (nöron) temsil etmektedir. İnsan beynine benzer şekilde nöronlar birbirine bağlıdır. Bir nöron birden fazla giriş sinyali alabilir, çıktı olarak tek bir sayısal değer çıkarır ve bu değeri birden fazla nörona aktarabilir. İlk katman görüldüğü gibi giriş katmanıdır ve özniteliklerden oluşmaktadır. Doğal olarak giriş katmanında öznitelik sayısınca nöron bulunmaktadır.

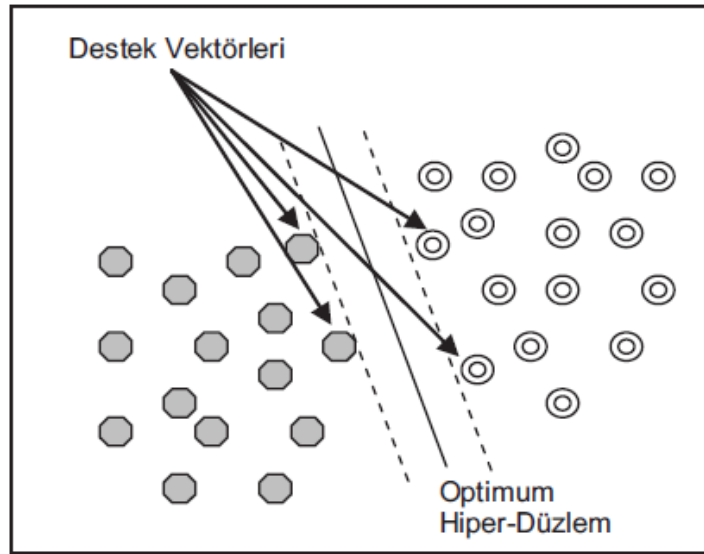
Derin öğrenme algoritmasında öğrenme, Şekil 3.8'de "w" olarak ifade edilen ağırlıkların belirlenmesi ve gerektiğinde güncellenmesi sürecinde gerçekleştirilir. Bir sınıflandırma işlemi için ilk olarak eğitim setinde bu ağırlıklar hesaplanır ve test seti üzerinde test edilir. Test sonucu ağırlık değerleri en düşük hata seviyesine ulaşmaya dek güncellenmeye devam eder. En düşük hata seviyesine ulaştığı anda öğrenme tamamlanmış olur ve yeni gelen bir veri için belirlenen ağırlıklar kullanılarak sınıflandırma işlemi yapılır (Silahtaroglu, 2004: 57). Çıkış katmanı sınıflandırma sonuç bilgisinin yer aldığı katmandır.

Derin öğrenme, diğer sınıflandırma yöntemlerine göre çalışma yapısı itibari ile daha karmaşık, öğrenme süreci daha uzun zaman alan ve daha yüksek bir teknik kapasiteye sahip donanım gerektiren, istikrarlı ve yüksek performansa sahip bir sınıflandırma yöntemidir. Algoritma kendisine tanıtılan verileri analiz ettikten sonra daha önce karşılaşmadığı veriler için de sınıflandırma yapabilir. Bunun için karşılaştığı veriye en yakın daha önce tanıtılmış veriyi bulur ve ona göre işlem yapar.

Bu şekilde algoritma eksik ve gürültülü veriler için çözüm üretebilir (Değer, 2017: 119).

### 3.6.5. Destek Vektör Makineleri

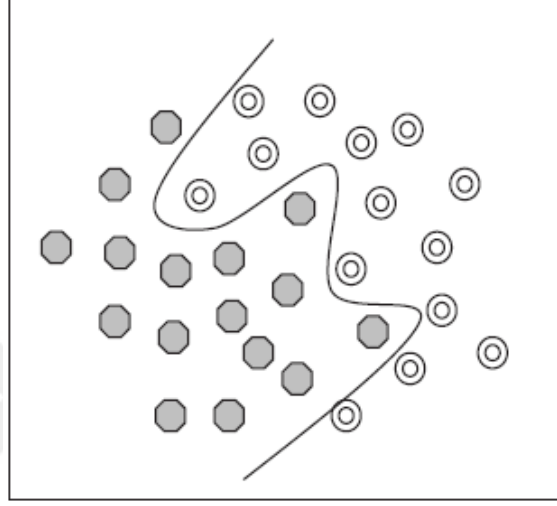
Destek vektör makineleri (support vector machines), verileri iki boyutlu uzayda bir doğru, üç boyutlu uzayda bir düzlem ve çok boyutlu uzayda bir hiperdüzlem kullanarak en az iki ayrı gruba ayırmaya yarayan istatistiksel bir makine öğrenimi algoritmasıdır (Güran, Uysal ve Doğrusöz, 2014: 87; Budak, 2021: 35). Başlangıçta sadece iki sınıflı doğrusal verileri sınıflandırabilirken zamanla çok sınıflı ve doğrusal olmayan verileri de sınıflandırabilecek şekilde geliştirilmiştir (Cemaloğlu ve Duykuluoğlu, 2020: 475). Algoritmayı geliştiren Vapnik (1995)'e göre, destek vektör makinelerinin çalışma prensibi iki sınıfı birbirinden en uygun şekilde ayırabilen hiperdüzlemin tanımlanması esasına dayanmaktadır (Kavzoğlu ve Çölkesen, 2010: 75).



Şekil 3.10. Doğrusal olarak ayrılabilen veri setinde hiperdüzlem ve destek vektörleri  
(Kavzoğlu ve Çölkesen, 2010: 76)

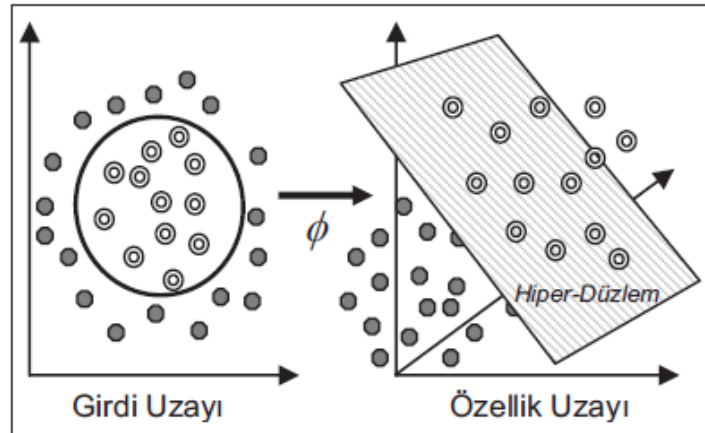
İki sınıflı doğrusal olarak ayrılabilen bir veri seti destek vektör makineleri algoritması kullanılarak Şekil 2.7'deki gibi ayrılabilir. Burada eğitim verisi ile elde edilen bir karar fonksiyonu yardımı ile ayırım yapılmaktadır. Amaç, iki sınıfa ait verileri birbirinden maksimum uzaklıkta ayıracak hiperdüzlemin bulunmasıdır ki bu hiperdüzleme optimum hiperdüzlem adı verilmektedir. Hiperdüzlemin verileri birbirinden maksimum uzaklıkta olacak şekilde ayırması gerekliliği, veri setine

eklenecek yeni verilerin optimum hiperdüzlemi değiştirmemesi gerekliliğinden kaynaklanmaktadır (Değer, 2017: 126). Her iki sınıfa da ait sınırdaki noktalara destek vektörleri, destek vektörleri arasındaki mesafeye de marj (genişlik) denilmektedir. Optimum hiperdüzlem, kendisine paralel ve sınırları oluşturan hiperdüzlemlerin bulunması vasıtası ile elde edilir (Kavzoğlu ve Çölkesen, 2010: 75).



Şekil 3.11. Doğrusal olarak ayrılamayan veri seti (Kavzoğlu ve Çölkesen, 2010: 77)

Veri setinin Şekil 3.11’de olduğu gibi girdi uzayında doğrusal olarak ayrılabilmesi mümkün değil ise, kernel fonksiyonu yardımı ile veri seti “özellik uzayı” olarak adlandırılan yüksek boyutlu bir uzayda görüntülenecek şekilde dönüştürülür ve böylelikle doğrusal olarak ayırılabilir hale gelir. Şekil 3.12’de veri setinin kernel fonksiyonu ile daha yüksek bir boyuta dönüştürülmüş hali görülmektedir (Kavzoğlu ve Çölkesen, 2010: 77).



Şekil 3.12. Kernel fonksiyonu ile verinin daha yüksek bir boyuta dönüştürülmesi

Destek vektör makineleri güvenilirliği ve performansı yüksek, karmaşık veri setlerindeki ilişkilerin ortaya çıkarılmasında etkili bir algoritmadır (Akpınar, 2014: 268; Cemaloğlu ve Duykuluoğlu,2020: 477).

### **3.7. Model Performans Değerlendirme**

Duygu analizinin son aşaması, oluşturulan modelin değerlendirilme aşamasıdır. Bu aşamada kullanılacak olan çeşitli yöntemler ve bilinmesi gereken bazı ölçütler bulunmaktadır.

#### **3.7.1. Model Performans Değerlendirme Yöntemleri**

Duygu analizi için oluşturulan modelin değerlendirilmesinde kullanılan çeşitli yöntemler mevcuttur. Aşağıda yaygın olarak kullanılan yöntemlerin bazıları açıklanmıştır.

##### **3.7.1.1. Dışarıda Tutma (Holdout)**

Dışarıda tutma yönteminde veri seti eğitim ve test seti olmak üzere ikiye ayrılır. Bu ayrımındaki oranlar kesin olmamakla birlikte yaygın uygulama veri setinin 3'te 2'sinin eğitim, 3'te 1'inin de test seti için ayrılmasıdır. Makine öğrenmesi yöntemine uygun olarak eğitim setindeki veriler modelin öğrenmesi, test seti de modelin performansının ölçülmesi için kullanılır. Modelde veri seti yalnızca bir kez eğitim ve test seti olarak ikiye ayrıldığı için bir veri ya eğitim ya da test setinde yer almakta ve bu haliyle performans ölçümü yapılmaktadır. Ancak verinin diğer sette yer alması halinde performansın nasıl değişeceğine dair herhangi bir bilgi sahibi olunulamamaktadır. Dışarıda tutma yönteminin en büyük dezavantajı da budur. Bir başka dezavantajı ise düşük boyutlu veri setlerinde eğitim veya test seti için yeteri kadar veri ayrılabilmesi durumudur (Değer, 2017: 84; Kızılkaya, 2018: 53).

##### **3.7.1.2. Tabakalı Öğrenme**

Tabakalı öğrenmenin kullanılabilmesi için hedef verinin nominal olması gerekmektedir, hedef veri nümerik ise kullanılamaz. Nominal verilerde bazen bazı sınıfların veri sayısı diğerlerine göre oldukça az olabilir. Bu gibi durumlarda dengeli sınıf sayıları elde edebilme amacı ile tabakalı öğrenme yöntemi kullanılır (Kızılkaya, 2018: 53). Tabakalı öğrenme her sınıfa dengeli bir dağılım olacak şekilde örneklem alınması esasına dayanmaktadır. Bu sınıfların her birine tabaka denmektedir.

Tabakaların kendi içerisindeki değişim az, tabakalar arası değişim fazla olmalıdır (Değer, 2017: 85).

### 3.7.1.3. Çapraz Geçerleme (Cross Validation)

Günümüzde en yaygın olarak kullanılan performans değerlendirme yöntemidir. Bu yöntemde de dışarıda tutma yönteminde olduğu gibi veri seti eğitim ve test seti olmak üzere ikiye ayrılır ancak dışarıda tutma yönteminden farklı olarak veri seti belirlenen k değeri kadar eşit parçaya bölünür ve her bir parça bir kez test seti olacak şekilde k kadar performans ölçümü gerçekleştirilir. Sonuç olarak bu performansların ortalaması alınarak genel performans hesaplanır. k değeri yaygınlıkla 10 olarak seçilir.

### 3.7.1.4. Yerine Geri Koyarak Örnekleme (Bootstrap Örnekleme)

Bir veri setinden örneklem alma yolu ile çok sayıda veri seti oluşturma esasına dayanmaktadır. Ancak örneklem alınırken bir eğitim seti için kullanılan veri başka eğitim setleri için de kullanılabilir. Matematikteki permütasyon mantığı gibi işler. Eğitim seti oluşturulduktan sonra, test seti için eğitim setinde kullanılmayan veriler kullanılır. Küçük boyutlu veri setleri ile çalışmanın zorunlu olduğu durumlarda kullanımı uygun bir yöntemdir. Oluşturulan çok sayıda örneklemin performans ölçümlerinin ortalaması alınarak genel performans hesaplanır (Değer, 2017: 86).

## 3.7.2. Model Performans Değerlendirme Ölçütleri

Oluşturulan modelin performans ölçümü için kullanılan çeşitli ölçütler mevcuttur. Bunların içinde en yaygın olarak kullanılan doğruluk (accuracy)'tur. Duyarlılık (recall), kesinlik (precision) ve bu iki değerden türetilen f-ölçütü de doğruluk dışında en fazla dikkat edilen performans değerlendirme ölçütlerinin başında gelmektedir. Bu ölçütlerin tümü karışıklık matrisi (confusion matrix) adı verilen bir tablodan elde edilmektedir. Karışıklık matrisi aşağıdaki gibidir:

Tablo 3.1. Karışıklık matrisi

		Gerçekte olan	
		Pozitif	Negatif
Tahmin edilen	Pozitif	Doğru Pozitif (DP)	Yanlış Pozitif (YP)
	Negatif	Yanlış Negatif (YN)	Doğru Negatif (DN)

Tablo 3.1’de iki sınıflı bir sınıflandırma uygulamasının karışıklık matrisi görülmektedir. Sınıflandırma uygulaması sonucu pozitif tahmin edilip de gerçekte de pozitif olan dokümanlar doğru pozitif (DP), pozitif tahmin edilip de gerçekte negatif olan dokümanlar yanlış pozitif (YP), negatif olarak tahmin edilip de gerçekte pozitif olan dokümanlar yanlış negatif (YN), negatif olarak tahmin edilip de gerçekte de negatif olan dokümanlar doğru negatif olarak adlandırılmaktadır. Buna göre,

- Doğruluk, doğru sınıflandırılmış doküman sayısının toplam doküman sayısına bölümüdür.

$$\text{Doğruluk} = \frac{DP + DN}{DP + DN + YP + YN} \quad (3.4)$$

Doğruluk, daha önce de ifade edildiği üzere modelin performans değerlendirmesinde öncelikle dikkat edilen ölçüttür. Basitçe, modelin hangi doğruluk oranında sınıflandırma yaptığını vermektedir.

- Duyarlılık, pozitif sınıflı dokümanların doğru tahmin edilme oranıdır. Doğru tahmin edilen pozitif sınıftaki dokümanların gerçekte pozitif sınıflı tüm dokümanlara bölümü ile bulunur.

$$\text{Duyarlılık} = \frac{DP}{DP + YN} \quad (3.5)$$

- Kesinlik, doğru tahmin edilen pozitif sınıftaki dokümanların pozitif tahmin edilen tüm dokümanlara bölümüdür. Kesinlik, yapılan pozitif sınıflı tahminlerin doğruluk oranını vermektedir.

$$\text{Kesinlik} = \frac{DP}{DP + YP} \quad (3.6)$$

- F-ölçütü, duyarlılık ve kesinlik değerlerinin harmonik ortalamasıdır. Doğruluk (accuracy) oranından sonra en fazla dikkat edilen ölçüttür. Sınıflandırmada her bir sınıfın doğru sınıflandırılma oranına aynı derecede dikkat edilmesi amacıyla kullanılmaktadır. Örneğin iki sınıflı bir sınıflandırmada bir sınıfın doğru sınıflandırılma oranı çok yüksekken, diğeri çok düşük olabilir. Eğer doğru sınıflandırılma oranı yüksek olan sınıfa ait dokümanlar sayıca da diğeri sınıfa göre çok daha fazla ise bu sefer genel doğruluk oranı yüksek çıkacaktır. Ancak sayıca düşük olan sınıfın doğru

sınıflandırılma oranı gerçekte düşüktür. Bu gibi durumları daha sağlıklı değerlendirebilmek için f-ölçütü kullanılmaktadır.

$$F - \text{ölçütü} = 2 \frac{\text{Kesinlik} \times \text{Duyarlılık}}{\text{Kesinlik} + \text{Duyarlılık}} \quad (3.7)$$

### 3.8. Makine Öğrenmesi Yöntemi ile Duygu Analizi Süreci

Makine öğrenmesi yöntemi ile gerçekleştirilecek olan duygu analizi çalışmalarında çalışılacak olan konu hakkında bilgi sahibi olunması büyük önem arz etmektedir. Bunun öncelikli sebebi doküman sınıflandırmasının öncelikle uygulayıcı tarafından manuel olarak yapılması gerekliliğidir.

Duygu analizi öncelikle üzerinde çalışılacak olan dokümanların toplanması ile başlar. Günümüzde bu dokümanların toplanması için kullanılan kaynakların başında sosyal medya gelmektedir. Dolayısıyla duygu analizi çalışmalarının da büyük bir kısmı sosyal medya verileri kullanılarak gerçekleştirilmektedir. Bu verileri toplamak için çeşitli dijital araçlar mevcuttur. Bu konuda en sık tercih edilen yöntemler Python, R gibi programlama dilleri ile Rapidminer, Weka, Knime gibi komplike veri madenciliği araçlarıdır.

Dokümanlar toplandıktan sonra çoğu zaman çalışmanın amacına uygun olarak bir seçim ve düzenleme işlemi gerçekleştirilir. Seçim ile çalışma için gerekli olmayan dokümanların elenmesi, düzenleme ile de seçilen dokümanların içinde yer alan verilerden gereksiz olanların silinmesi veya verilerin duygu analizi yapılabilmesine imkan sağlayacak şekilde yeniden düzenlenmesi kast edilmektedir. Örneğin dokümanlar birbiri içine geçmiş ve karmaşık bir halde toplanmış olabilir. Bu gibi durumlarda doküman seviyesinde (çalışma birimi doküman olacak şekilde) duygu analizi yapılabilmesi mümkün değildir. Duygu analizi için kullanılacak olan aracın her bir dokümanı ayrı ayrı algılayabilmesine imkan verecek şekilde dokümanların yeniden düzenlenmesi gerekir.

Seçim ve düzenleme işlemi bittikten sonra sıra eğitim setinin oluşturulmasına gelir. Uygulayıcı tarafından dokümanların bir kısmı kullanılarak manuel olarak bir sınıflandırma yapılması suretiyle bir eğitim seti oluşturulur. Oluşturulacak olan eğitim setinin büyüklüğü çalışmadan çalışmaya farklılık gösterecektir. Ancak ne kadar fazla sınıflandırma yapılırsa modelin farklı örnekleri tanıma ve öğrenme ihtimali de o kadar yükselecektir. Eğitim seti oluşturulduktan sonra, modelin

sınıflandırma başarısının test edilmesi için bir de test seti oluşturulur. Eğitim ve test seti büyüklük oranlarının kesin olarak sınırları çizilmemiş olmakla birlikte, literatürde yaygın olarak eğitim setinin %70-80 oranında, test setinin ise %20-30 oranında olacak şekilde tercih edildiği görülmektedir (Paulino, Guimaraes ve Shiguemori, 2019: 170).

Eğitim ve test setleri oluşturulduktan sonra eğitim setindeki sınıf sütunu ya da sınıf sütunlarından biri hedef özellik olarak belirlenir. Bu şekilde duygu analizi sonucu tahmini yapılacak olan veri sütunu belirlenmiş olur. Daha sonra her iki setteki dokümanların içerdiği metinler bir dizi ön işleme tabi tutulurlar. Metin ön işleme safhasında gerçekleştirilecek olan işlemler çalışmadan çalışmaya farklılık gösterebilir. Ancak her duygu analizi çalışması için gerekli olan metin ön işlemleri dizgeciklere (token) ayırma ve filtrelemedir. Bu işlemler öncesi büyük harfler küçük harflere çevrilir. Çalışma gerektiriyorsa birbirinin aynısı dokümanların kopyaları silinip analiz için sadece bir tanesi bırakılabilir. Eksik verilerin analiz sonucunu olumsuz etkilememesi için eksik verilere sahip dokümanlar ya silinir ya da eksik olan değerlerine başka bir değer atanır. Bu işlemler sonrası uygulanacak olan dizgeciklere ayırma işlemi ile metin sözcüklerine ayrılır. Bu işlem ile aynı zamanda metin içerisinde yer alan noktalama işaretleri, sözcük vasfı taşımayan anlamsız karakterler ve satır sonu karakterleri de silinir. Filtreleme ile metin içerisinde sıkça geçen ancak analize katkısı olmayan durak sözcükleri temizlenir. Bu noktada sözcükler aynı zamanda belirli bir karakter uzunluğu ile de sınırlandırılabilir. Örneğin 4 – 25 karakter uzunluğu dışındaki sözcükler analiz dışı bırakılabilir. Bu şekilde yapılan bir filtrelemenin performansı yükseltebildiği gözlemlenmiştir.

Metin ön işleme sonrası sıra öznitelikleri çıkarmaya gelmiştir. Öznitelikler sözcüklerin içinden çıkarılan ve analiz için kullanılacak olan en küçük birimlerdir. Duygu analizi yapılacak olan dil açısından bir anlam ifade edebilecekleri gibi, sadece karakterlerin bir araya gelmesi ile oluşan anlamsız yapılar da olabilirler. Bu noktada başvurulacak olan öznitelik çıkarma yöntemleri temelde ikiye ayrılmaktadır. Birincisi doğal dil işleme yöntemleri kullanılarak sözcüklerin anlam olarak değişmeyecek şekilde köklerine inilmesi, eklerinden arındırılması anlamına gelen gövdeleme, diğeri ise istatistik yöntemlerini kullanan ve öznitelikte bir anlam aramayan n-gram yöntemidir. Özniteliklerin çıkarılması için uygulanacak olan yöntem belirlendikten sonra özniteliklerin ağırlıklandırma yöntemi belirlenir. Her bir

öznitelik belirlenen ağırlıklandırma yöntemine göre farklı ağırlıklar alarak aynı zamanda üzerlerinde veri madenciliği algoritmalarının uygulanabileceği şekilde sayısal bir hale dönüştürülürler. Bu dönüşüm işlemi için yaygın olarak vektör uzay modeli tercih edilir.

Şu ana kadar anlatılan işlemlerin - eğitim setinde hedef özellik belirleme işlemi dışında - hepsi hem eğitim hem de test setine uygulanan işlemlerdir. Bu noktadan sonra eğitim setindeki özniteliklere farklı olarak boyut azaltma işlemi uygulanabilir. Bu işlem zorunlu olmamakla birlikte duygu analizi için kullanılan teknik donanımın oluşturulan modeli işlemek için kapasite olarak yeterli olmadığı durumlarda başvurulabilir. Belirlenen boyut azaltma işlemi de uygulandıktan sonra sıra bir sınıflandırıcı algoritma vasıtası ile eğitim setindeki sınıflandırmalar dikkate alınarak test setindeki dokümanların sınıflandırılmasına gelmiştir. Bunun için öncelikle bir sınıflandırıcı algoritma seçilir ve algoritmaya eğitim setindeki sınıflandırmalar tanıtılarak öğrenmesi sağlanır. Bu sırada aynı zamanda modelin sınıflandırma performansı da ölçülür. Öğrenen model daha sonra test setine uygulanır ve test setindeki dokümanların otomatik olarak sınıflandırılması sağlanır.

Bu süreç boyunca yukarıda görüleceği üzere çok sayıda seçim yapılmaktadır. Bu seçimlerin - modele ana şeklini vermeleri sebebiyle - en önemlileri, öznitelik çıkarma yöntemi, öznitelikleri ağırlıklandırma ve sınıflandırma algoritmasının seçimidir. Yukarıda her defasında tek bir seçim yapılarak tamamlanan duygu analizi süreci aslında başta bu belirtilen ana seçimler ve bunların parametrelerinin seçimi olmak üzere çok sayıda değişkenin değiştirilerek performansın ölçülmesi ve en sonunda en yüksek performansa ulaşan modelin tespit edilmesi işlemlerinin tamamını kapsamaktadır. En yüksek performansın elde edildiği modelin yeni verilerin sınıflandırılması için kullanılmasıyla süreç tamamlanır.

### **3.9. Twitter ve Duygu Analizindeki Rolü**

Günümüzde kitlelerin herhangi bir konu hakkındaki duygu ve düşüncelerini ölçme amaçlı gerçekleştirilen araştırmalarda artık başvuru birincil kaynağın internet ve özellikle sosyal medya olduğu söylenebilir. Yüz yüze görüşme, anket, telefonla görüşme gibi yöntemler hala varlıklarını sürdürseler de, çok daha kısa sürede çok daha fazla kişiye ulaşabilme, daha ekonomik, verimli ve hızlı olması gibi özelliklerinden ötürü internet ve sosyal medyanın sözü edilen klasik yöntemlerden büyük oranda ayrıştığı görülmektedir.

Gündem hakkındaki duygu ve düşüncelerin paylaşıldığı platformların başında internet ve sosyal medyanın gelmesi durumu her ne kadar ülkelerin teknolojik ve demokratik gelişmişlik seviyeleri ile de alakalı da olsa, internet ve sosyal medya kullanım verileri bize dünya geneli için bu tespitin yine de geçerli olduğunu göstermektedir. İnsanlar tarihin hiçbir döneminde olmadığı kadar kolay bir şekilde bir araya gelebilmekte ve herhangi bir konu hakkındaki düşüncelerini - yukarıda ifade edilen kısıtlar çerçevesinde - birbirlerini bağlayan ortak ağda paylaşabilmektedirler.

Toplumları ilgilendiren ortak konular hakkındaki görüşlerin kamuya açık bir şekilde ifade edilmesi söz konusu olduğunda sosyal medya platformları içinde de Twitter'ın özel bir yeri olduğu, hatta ilk sırada yer aldığı görülmektedir. Twitter, günde yapılan 500 milyonu aşan paylaşım (tweet) ile şu anda dünyanın en popüler mikroblog uygulamasıdır (Atılğan ve Yoğurtcu, 2021: 32). 2022 yılı itibariyle dünya genelinde Twitter kullanıcısı sayısı 436,4 milyona ulaşmıştır. Türkiye ise 16,1 milyonluk kullanıcısıyla dünya sıralamasında 7.sıradadır (We Are Social & Hootsuite, 2022). Türkiye'nin toplam nüfusu ve Twitter kullanabilecek yaştaki nüfusu (Twitter tarafından +13 olarak belirlenmiştir) göz önünde bulundurulduğunda Türkiye'deki Twitter kullanıcısı olma oranının oldukça yüksek bir seviyede olduğu görülmektedir. Tüm bu sebeplerden ötürü duygu analizi çalışmaları için hem dünya hem de Türkiye'de veri kaynağı olarak Twitter ilk sırada tercih edilmektedir.

### **3.9.1. Twitter ve Terminolojisi**

Twitter, 2006 yılında Jack Dorsey tarafından, ortak bir platformda kısa metin şeklindeki mesajların (tweet) atılabilmesine imkan verecek şekilde geliştirilen mikroblog türünde bir sosyal medya aracıdır. Günümüzde, atılacak olan tweet'lerin karakter sınırı 280 olmakla birlikte metnin yanı sıra sabit veya hareketli görsellerin de paylaşılabilmesi mümkündür. Twitter, tweet dışında "hashtag" (etiket), "mention" (bahseden), "retweet" (yeniden tweet atmak), "reply" (yanıt), "alıntı tweet" gibi çok sayıda yeni terimi de beraberinde getirmiştir.

"Hashtag", tweet'lerin belirli başlıklar altında toplanabilmesi amacıyla kullanılan ve "hashtag" olarak belirlenecek metnin başına "#" işareti konarak oluşturulan etiketlerdir. Bu sayede kullanıcılar istedikleri başlık altındaki tweet'lere rahatlıkla ulaşabilmekte veya kendileri yeni bir başlık açabilmektedirler. Hashtag kullanılarak atılan tweet'ler o hashtag'i aratan veya o hashtag'e tıklayan tüm

kullanıcılar tarafından görülebilir. Başka bir deyişle kullanıcılar Twitter’da bir hashtag arattıklarında veya bir hashtag’e tıkladıklarında o hashtag kullanılarak atılan tweet’lere ulaşabilirler. Böylelikle aynı konu hakkında konuşmak isteyen insanlar daha kolay bir şekilde bir araya gelebilmektedirler. Hashtag’i oluşturan karakterler arasında boşluk olmamalıdır.

“Mention” İngilizce “bahsetmek” anlamına gelmektedir. Twitter terminolojisinde atılan tweet içerisinde başka bir kullanıcı adının etiketlenerek atılan tweet’lerdir. Bunun için kullanıcı adının önüne “@” işaretinin konulması gerekmektedir. Böylelikle atılan tweet’lerin özellikle bir kullanıcı tarafından o kullanıcının bildirimler sekmesinde görülebilmesi sağlanabilir.

“Retweet”, atılan bir tweet’in yeniden gönderilmesi; “Reply”, tweet’e verilen yanıt; alıntı tweet ise tweet’in yorum eklenerek retweet edildiği tweet’e denmektedir.

### **3.9.2. Twitter API**

Duygu analizinde kullanmak üzere Twitter verilerini çekmek için birkaç yol bulunmaktadır. Bunlardan biri tweet’leri manuel olarak Twitter sayfalarından çekmektir ki bu, yüksek sayıda tweet çekilen duygu analizi çalışmaları için oldukça zahmetli bir yöntem olacaktır. Bir başka yöntem Twitter verilerini çekmek için yazılmış hazır programları kullanmaktır ki bu yöntemde de Twitter tarafından resmi olarak yetki verilmediği için çeşitli problemler yaşanabilmektedir. En sağlıklı yöntem ise Twitter verilerini çekmek için Twitter’ın izin verdiği uygulama programlama arayüzlerini (application programming interface), yani API’leri kullanmaktır. API, Twitter verilerine üçüncü parti (Twitter dışındaki) uygulamaların erişmesine imkan sağlayan programlama ara yüzleridir (Sevli, 2016: 54). Python gibi programlama dillerinde ya da Rapidminer gibi veri madenciliği araçlarında Twitter tarafından temin edilen API kullanılarak sağlıklı bir şekilde tweet çekmek mümkündür. Bu noktada API kullanarak tweet çekme yöntemlerinin temelde ikiye ayrıldığı söylenebilir. Bunlardan biri akan veriyi, yani Twitter’da o anda akmakta olan veriyi çekme anlamına gelen “stream”, diğeri ise Twitter’da geçmişte atılmış olan verileri çekme anlamına gelen “rest”tir. Duygu analizi çalışanları genellikle “stream” olarak, Twitter’da akmakta olan veriyi çekmeyi tercih etmektedirler. Bunun sebebi stream yöntemi ile gerçek zamanlı olarak – Twitter’ın da API için belirlemiş olduğu kısıtlar dahilinde – o anda akan tüm tweet’leri çekebilmesidir. Ancak bunun da bir limiti söz konusudur. Twitter’ın kullanıcılarına sağlamış olduğu API’lerin

çeşitleri vardır ve araştırmacılara ücretsiz olarak sağladığı API'lerin belirli bir süre içerisinde belirli sayıda tweet çekebilme gibi limitleri bulunmaktadır. Bu sebeple birim zamanda, girilen anahtar sözcüğü içeren çok sayıda tweet atıldığında, tüm tweet'lerin çekilememesi gibi bir ihtimal doğmaktadır.

API'leri kullanmak için Twitter tarafından kullanıcıya çeşitli şifreler verilmektedir. Kullanıcı da yukarıda ifade edilen üçüncü parti programlara bu şifreleri girerek Twitter'a erişim sağlamaktadır. Tweet çekme, bu dile getirilen üçüncü parti programlarında anahtar sözcüklerin sorgulanması sureti ile gerçekleştirilmektedir. Sorgu sonucu içinde anahtar sözcüğü içeren tweet'ler programlar vasıtası ile çekilmektedir. Anahtar sözcük dışında, çekilecek olan tweet'leri sayı, dil, tarih, lokasyon gibi özelliklerine göre de sınırlamak mümkündür.

### **3.10. Duygu Analizi Yöntemi Kullanılarak Daha Önce Yapılan Çalışmalar**

Duygu analizi günümüzde hem dünyada hem de ülkemizde oldukça popüler olan bir yöntemdir ve kullanım oranı her geçen gün artmaktadır. Duygu analizi yöntemi kullanılarak gerçekleştirilen çalışmalardan bazıları aşağıdaki gibidir:

Kuş (2019)'un kendi deyimi ile "Suriyeli mülteciler" hakkındaki algıyı Twitter kapsamında ele aldığı çalışmasında makine öğrenmesi yöntemi ile duygu analizi gerçekleştirilmiştir. Analiz sonucu %57'lik bir doğruluk oranı ile, ele alınan 13595 satır tweet'in %29'unun Suriyeli mültecilere yönelik olarak pozitif veya nötr algı taşıdığı; %13'ünün eşitsizlik hissi taşıdığı; %30'unun Suriyeli mültecilerin toplumsal sorunlar yarattığına dair bir algıya sahip olduğu; %28'inin ise herhangi bir sebep belirtmeden Suriyeli mülteciler ile ilgili olarak negatif söylem/nefret söylemi içerdiği sonuçlarına ulaşılmıştır. Çalışmada, en büyük orana sahip olan toplumsal sorunları dile getiren tweet'lerin ise %47'sini toplumsal güvenlik, %23'ünü ekonomi, %18'ini sosyal/kültürel gerekçeler, %12'sini ise politik gerekçelerin oluşturduğu görülmüştür (Kuş, 2019: 217-218).

Aksu (2021)'nin, sosyal medyadaki ziyaretçi yorumları ve fotoğrafların ele alınarak otomatik olarak ziyaretçilerin duygu durumlarının, bu duygu durumlarının zamana göre değişiminin ve bu değişimin altında yatan sebeplerin tespitini yaptığı çalışmasında hem metin tabanlı hem de görsel duygu analizi yöntemleri kullanılmıştır. Artvin ili özelinde gerçekleştirilen çalışmada turistik mekanlara yönelik olarak duygu skorlarının 8.aya kadar artış eğiliminde olduğu, bu aydan sonra

duygu skorlarının düşüşe geçtiği; bunun sebebinin ise kamp yasakları ve yol çalışmaları ile alakalı olduğu tespit edilmiştir (Aksu, 2021: V).

Değer (2017)'in havacılık sektöründe faaliyet gösteren bir firma ile ilgili olarak müşterilerin genel algısını ölçme amacı ile Twitter kapsamında makine öğrenmesi yöntemlerini kullanarak gerçekleştirdiği duygu analizinde en yüksek performansa derin öğrenme algoritması ve harf tabanlı 4-gram yöntemi kullanılarak %79,76'lık bir doğruluk oranı ile ulaşılmıştır. Analiz sonucu, 2015 yılı ile 2016 yılları kıyaslandığında, olumlu tweet'lerin %19,5'ten %14,5'e düştüğü, olumsuz tweet'lerin ise %28,6'dan %24,6'ya düştüğü tespit edilmiştir (Değer, 2017: 152-153).

Topaçan (2016)'ın banka, beyaz eşya, elektronik, gsm-telekom ve otomotiv sektörlerine ait firma veya ürünlerine yönelik olarak Twitter kapsamında gerçekleştirmiş olduğu duygu analizinde emoji de analize dahil edilmiş ve isim, sıfat, fiil gibi cümle öğelerinin tespiti yapılarak morfolojik analiz de gerçekleştirilmiştir. Makine öğrenmesi yaklaşımı ile gerçekleştirilen duygu analizinde en başarılı sınıflandırmayı yapan modele, 1-3 gram öznitelik çıkarma yöntemi, ikili kodlama ile öznitelik ağırlıklandırma, bilgi kazanımı yöntemi ile boyut azaltma ve Naive Bayes sınıflandırma algoritması ile ulaşılmıştır. Çalışma sonucu sınıflandırmaya en fazla katkı sağlayan sözcük türleri isim, sıfat, fiil ve emoji olarak tespit edilmiş, beş sektör için de atılan tweet'lerin pozitif, negatif ve nötr olarak sınıflandırılması yapılmış ve her sektör için sınıflandırma gücü en yüksek terimler tespit edilmiştir (Topaçan, 2016).

Ayan ve arkadaşları (2019), Twitter'da İngilizce olarak atılan islamofobik tweet'lerin tespitini yapmayı hedefledikleri çalışmada makine öğrenmesi yaklaşımı ile duygu analizi yöntemini kullanmışlardır. En yüksek performans gösteren modele lineer ridge regresyonu ile %96,3 doğruluk oranında ulaşılmıştır (Ayan, Kuyumcu ve Ceylan, 2019: 496).

Atılgan ve Yoğurtcu (2021), kargo firmaları ile ilgili olarak Twitter kapsamında müşteri memnuniyetini ölçmeyi hedefledikleri çalışmalarında sözlük tabanlı yaklaşım ile duygu analizi yöntemine başvurmuşlardır. Buna göre çalışmanın belirlemiş olduğu tarihler arasında dikkate alınan ve kargo firmaları ile ilgili olarak atılan tweet'lerin %47,6'sı negatif, %31,4'ü nötr ve %21'i pozitif olarak tespit edilmiştir (Atılgan ve Yoğurtcu, 2021: 35).

## 4. TÜRKİYE’DE YAŞAYAN SURIYELİ SĞINMACILAR HAKKINDAKİ ALGI ÜZERİNE DUYGU ANALİZİ

Tezin bu bölümünde Türkiye’de yaşayan Suriyeli sğınmacılar hakkındaki algı Twitter sosyal medya platformunda bu konuda atılan Türkçe tweet’ler üzerinden duygu analizi yöntemi kullanılarak incelenecektir. Araştırma üç farklı boyutta gerçekleşecektir:

1. Genel Algı Analizi: Suriyeli sğınmacılar hakkındaki genel algıyı olumlu, olumsuz ve nötr şeklinde ortaya koyacak duygu analizi.
2. Algının Niteliği Analizi: Suriyeli sğınmacılar hakkındaki olumlu ve olumsuz algının ayrı ayrı niteliklerini ortaya koyacak duygu analizleri
3. Algının Sebepleri Analizi: Suriyeli sğınmacılar hakkındaki olumlu ve olumsuz algının ayrı ayrı sebeplerini ortaya koyacak duygu analizleri.

### 4.1. Verileri Toplama

Araştırmada veri olarak Twitter sosyal medya platformunda Suriyeli sğınmacılar ile ilgili olarak atılan tweet’ler kullanılmıştır. Bunun için öncelikle Python programlama dilinde yazılan bir kodla “suriyeli” anahtar sözcüğünü içeren tweet’ler 2021 yılı boyunca rastlantısal olarak ve “stream” (akan veri) yöntemiyle çekilmiştir. 2021 yılı sonu itibari ile çekilen tweet’lerin 220.418 adete ulaştığı görülmüştür.

“Suriyeli” sözcüğünü içeren tüm tweet’lerin Suriyeli sğınmacıları karşılamayabileceği göz önünde bulundurularak, bu durumun neden olabileceği olumsuz sonuçların önüne geçmek için duygu analizi kısmında “suriyeli” sözcüğünü içeren fakat Suriyeli sğınmacıları karşılamayan tweet’leri ayrıştırabilecek - ileride detayları ile anlatılacak olan - bir modelleme yapılmıştır.

Anahtar sözcüğün “suriyeli” olarak belirlenmesindeki amaç, “suriyeli” sözcüğünün, Suriyeli sğınmacıları ifade eden, “Suriyeliler”, “Suriyeli mülteciler”, “Suriyeli sğınmacılar”, “Suriyeli göçmenler” gibi en yaygın olarak kullanılan hitapların alt kümesi olmasıdır. Bu durum araştırmanın başında yapılan deneme tweet çekme işlemleriyle de desteklenmiş, en sağlıklı sonuca “suriyeli” anahtar sözcüğü ile ulaşıldığı görülmüştür. Ayrıca konu ile ilgili olarak daha önce yapılan benzer çalışmalarda da “suriyeli” anahtar sözcüğüne başvurulduğu gözlemlenmiştir.

Kuş (2019)'un Suriyeli Mültecilerin Sosyal Entegrasyonunda Halkla İlişkilerin Rolü adlı çalışması buna örnektir.

Tablo 4. 1. Tweet'lerin ilk çekildiği andaki ham haline örnekler

date	user	is_ retweet	is_ quote	text / quoted_text
2021-04-04 09:02:56	biroolsen17	False	False	Fotoğrafa 3 saniye bakınca Suriyeliler Ceylanpınar sınır kapısından oluk oluk memleketlerine gidiyorlar... <a href="https://t.co/8j7YaHdwVS">https://t.co/8j7YaHdwVS</a>
2021-07-25 08:58:26	TurgayHoca 2013	True	False	RT @theburaakk: Daha bu sabah Suriyeli genç erkeklerin Türkiye'nin en güzel sahillerinde nargile içerek tatil yaptığını haykırırken, bu gec...
2021-09-02 19:57:54	_logrenci_	False	True	Bu bayrağı tutan bir Suriyeli kardeşimiz vardı Tutuklanmıştı sanırım İsmi Muhammed'ti Ne olduğunu bilen var mı?, Kör dünyanın göbeğine hak yol İslâm yazacağız. <a href="https://t.co/OSTIDzta5M">https://t.co/OSTIDzta5M</a>
2021-11-11 18:06:49	MicoGostiva ri	True	True	RT @ORHANOSMANOGLU: Hizmete devam.,Suriyeli Afaf kızımız ve Ömer oğlumuz ile alış verişimizi birlikte yaptık... <a href="https://t.co/7FTNJecKNe">https://t.co/7FTNJecKNe</a>
2021-12-11 16:20:47	ismailcengiz	False	False	5 milyon Suriyeli yerine 5 milyon inek besleseydik, etin kilosu 20 ₺, sütün kilosu 2 ₺ olurdu

Tablo 4.1'de görüldüğü üzere, tweet'lerin ilk çekildiği hallerinde her bir tweet'e dair tarih/saat, tweet'i atan hesabın kullanıcı adı, tweet'in bir retweet olup olmadığı, tweet'in bir alıntıya sahip olup olmadığı ve tweet veya retweet metninin kendisine dair bilgiler yer almaktadır.

#### 4.2. Verilerin Ön İşlemden Geçirilmesi

Çekilen 220.418 adet tweet'in içindeki "retweet", alıntı, yanıt ve "mention"lar Python'da bir program yazılarak temizlenmiş ve 16.186 adet tweet'e ulaşılmıştır. Bu işlemi yapmadaki öncelikli amaç Twitter kullanıcılarının Suriyeli sığınmacılar ile

ilgili olarak başkalarından bağımsız olarak ve konu bağlamından uzaklaşmadan sadece kendi görüşlerini dikkate almaktır. Herhangi bir tweet'e verilen yanıtların "suriyeli" sözcüğünü içerse dahi tek başına değerlendirildiğinde çoğu zaman herhangi bir anlam ifade etmediği görülmüştür. Retweet'lerin retweet'lenen tweet'i destekler nitelikte olup olmadığı belirsizdir ve bir kampanya faaliyetinin parçası olma ihtimali bulunmaktadır. Alıntılarının yine alıntı yapılan metin, fotoğraf, video veya linkten bağımsız değerlendirilmesi sağlıklı sonuçlar doğuracaktır. "Mention"lar bir başkasına hitaben veya o kişi veya kişiler etiketlenerek yazıldığı için konunun genelliği ile uyuşmamaktadır. Tüm bu sebepler göz önünde bulundurularak araştırma kapsamında, sadece Suriyeli sığınmacılar ile ilgili olarak atılan ilk tweet'ler dikkate alınmıştır.

Bilindiği üzere Twitter başta olmak üzere çeşitli sosyal medya platformları çeşitli sosyal, siyasal veya ticari kampanyaların hedefi olabilmektedir. Araştırmayı bu tür kampanya ve manipülasyonlardan koruma amacı ile ayrıca birebir aynı olan tweet'lerden sadece birinin dikkate alınması sağlanmış ve tek bir hesabın konu ile ilgili olarak tek bir tweet'i (son tweet'i) dikkate alınmıştır. Bu işlemler de yine Python'da iki farklı program yazılarak gerçekleştirilmiştir. Bu işlemlerden sonra ulaşılan tweet sayısı 12.854 olmuş ve duygu analizine bu sayıda tweet ile gidilmiştir.

Tweet'ler üzerinde duygu analizine geçilmeden önce yapılan son işlemler tweet'lerde yer alan büyük harflerin küçük harflere çevrilmesi ve linklerin temizlenmesi olmuştur. Büyük harfler duygu analizinde kelimenin veya karakterin farklı olarak tanımlanmasına neden olacağından küçük harflere çevrilmiş, linkler ise duygu analizine herhangi bir katkı sağlamayacağı ve analiz için ekstra bir yük oluşturacağı için silinmiştir. Bu işlemler de Python'da bir program yazılarak gerçekleştirilmiştir.

Tablo 4. 2. Tweet'lerin ön işlem den geçtikten sonraki hallerine örnekler

Tweet
arkadaşlar ukrayna rusya arası savaş olacak lütfen ukrayna lı mülteci alabilir miyiz lütfen suriyeliler pek olmadı
hatay ilinde yaşayan #suriyeliler, afgan sığınmacıları hatay'da istemiyorlarmış.. trajikomik fıkra ancak bu kadar olur.
şu suriyeliler bir yerinde durmuyorlar. bu ne gürültü ses ! ülkesini bırakıp gelmişler

---

burda millete cektiriyorlar. biz olsak sesimiz cikmaz be.

suriyeliler de yıllar sonra memleketlerine gidip “ya habibi buralar çok ayvaa, türkiye çok kötü. düzenimiz olmasa çoktan dönerdik” diyecekler mi acaba?

ya bu suriyeliler pubgde savaştıkları kadar kendi ülkelerinde savaşsalardı ülkeyi kaybetmezlerdi

---

### 4.3. Makine Öğrenmesi Yöntemi ile Duygu Analizi

Yukarıda bahsedildiği üzere makine öğrenmesi yöntemi kullanılarak gerçekleştirilen duygu analizi modelinde öncelikle metin uygulayıcı tarafından manuel olarak sınıflandırılır ve sınıflandırılan metin programa öğretilerek geri kalan metnin program tarafından otomatik olarak sınıflandırılması beklenir.

Daha önce bahsedildiği üzere duygu analizi 3 farklı boyutta gerçekleştirilecektir. Bunlardan ilki genel algı analizidir. Bunu sırayla nitelik ve sebep analizleri takip edecektir. Bu sebeple öncelikle tweet’ler sırası ile genel algı analizi için olumlu, olumsuz veya nötr olarak; nitelik analizi için hoşnutsuzluk, kızgınlık, nefret vs. gibi olumsuz nitelik veya sempati besleme, acıma/üzülme, sevme vs. gibi olumlu nitelik olarak; son olarak sebep analizi için de “Suriyeliler bize zarar veriyor”, “Suriyelilerin kötü özellikleri var” gibi olumsuz sebep olarak veya “Yardıma ihtiyaçları var”, “Ortak yanlarımız” gibi olumlu sebep olarak sınıflandırılmışlardır. Tüm analiz birimleri için sınıflandırma tablosu aşağıdaki gibidir:

Tablo 4. 3. Duygu analizi manuel sınıflandırma tablosu

---

Genel Algı Analizi	Nitelik Analizi	Sebebe Analizi
Olumlu	Sempati besleme	Yardıma ihtiyaçları olma / mağdur / mazlum olma durumları
	Kabul etme / kabul görme	Suriyelilerin iyi özellikleri
	Sevme	Ortak yanlarımız
	Hak verme	Suriyelilerin faydalı olma durumu
	Acıma / üzülme	Suriyelilerin zararlarının olmaması
	Beğenme	Suriyeliler ile arkadaş olma durumu
	Saygı duyma	Diğer

---

Değer verme		
Olumsuz	Hoşnutsuzluk	Suriyeliler bize zarar veriyor
	Nefret	Suriyelilerin Türkiye'deki hakları
	Kızgınlık	Suriyeliler bizden daha iyi durumdadır / mevcut durumları
	Alay etme	Suriyelilerin nüfusu
	Aşağılama	Suriyelilerin kötü özellikleri var
	Üzgün olma	Kişisel düşünceler / Diğer
Nötr	Nötr	Nötr

Tablo 4.3'te yer alan nitelik ve sebep sınıflandırmaları çalışmanın başında belirlenmemiş olup, sınıflandırma yaptıkça çeşitlenmiş, yer yer gruplandırılmışlardır. Özellikle olumsuz sebep sınıflandırmaları Algının Sebepleri başlığı altında daha detaylı olarak anlatılacağı üzere birçok detay sebebin başlıkları olarak şekillenmişlerdir. Aşağıda sınıflandırma sürecine ait örnekler yer almaktadır.

Tablo 4. 4. Duygu analizi manuel sınıflandırma örnekleri

Tweet	Genel Algı	Nitelik	Sebebi
suriyeliler bizden mutlu, adamlar el üstünde tutuluyor.	Olumsuz	Hoşnutsuzluk	Suriyelilere daha çok değer veriliyor ("Suriyeliler bizden daha iyi durumdadır / mevcut durumları" başlığı altında)
şu suriyeliler bir yerinde durmuyorlar. bu ne gürültü ses ! ülkesini bırakıp gelmişler burda millete cektiriyorlar. biz olsak sesimiz cikmaz be.	Olumsuz	Kızgınlık	Rahatsız ediyorlar ("Suriyeliler bize zarar veriyor" başlığı altında)
sehirler bombalar yağardı her gece biz durmadan sevisirdik sozu suriyeliler için yazılmıştır.	Olumsuz	Hoşnutsuzluk	Çok fazla çocuk yapıyorlar ("Suriyelilerin nüfusu" başlığı altında)
suriyeliler ekonomiye pozitif etki ediyor. ucuz işgücü sağlıyorlar, kimseye kiralanmayacak berbat	Olumlu	Sempati besleme	Suriyelilerin faydalı olma durumu

koşullarda yaşıyor birçoğu. daha ne istiyorsunuz. bu tavrınız yüzünden 30 sene sonra ülke içi ciddi problemlerle karşılaşacağız.			
bu görüntüleri görüp hâlâ suriyeliler evine dönsün diyen varsa yarın mahşerde bunun hesabını veremez	Olumlu	Acıma/üzülme	Yardıma ihtiyaçları olma / mağdur / mazlum olma durumları
suriyeliler gidene kadar kalsaydın bari.	Nötr		
rüyamda gustavo' yu suriyeliler kaçıyordu, yardım eden de üniversiteden arkadaşım. günaydın...	Nötr		

#### 4.4. Genel Algı Analizi

Tablo 4.4'te örnekleri görüldüğü üzere ilk olarak 3024 adet tweet bu şekilde olumlu, olumsuz veya nötr olarak sınıflandırılmıştır. Sınıflandırılan tweet'lerin seçiminde 2021 yılının tamamını yansıtacak şekilde tarih açısından dengeli bir seçim yapılmasına dikkat edilmiştir. Bu şekilde yapılan sınıflandırma sonucu 3024 adet tweet'in 2115'inin olumsuz, 547'sinin nötr, 362'sinin ise olumlu olarak sınıflandığı görülmüştür. Görüldüğü üzere tweet'lerin çekiliş sırasına göre rastgele yapılan bir seçim sonucu sınıflandırmada olumsuz algı içeren tweet'lerin sayısı, nötr ve olumlu algı içeren tweet'lerin toplamından bile yüksek çıkmaktadır ki bu da eğitim setinin oldukça dengesiz bir hal almasına neden olmaktadır. Bu noktada duygu analizini eğitim setinin dağılımının da ne oranda etkileyeceğini görme amacı ile büyük oranda sadece olumlu ve nötr tweet'lerin sınıflandırılmasına devam edilmiş ve bu şekilde iki adet daha, daha dengeli eğitim seti elde edilmiştir. Aşağıda elde edilen üç adet eğitim setine dair sayılar yer almaktadır:

Tablo 4.5. Genel algı analizi eğitim setleri sınıflandırma dağılımları

Eğitim seti	Olumlu	Olumsuz	Nötr	Toplam
1	362	2115	547	3024
2	496	2115	835	3446
3	1003	2119	1644	4766

Makine öğrenmesi yöntemi ile yapılan bir duygu analizinde eğitim seti belirlendikten sonra yapılması gereken ilk iş test setini belirlemektir. Eğitim setinde yapılan sınıflandırmalara göre öğrenen model test setinde yer alan tweet'leri otomatik olarak sınıflandıracaktır. Literatürde tüm duygu analizi çalışmaları için belirlenmiş ideal bir eğitim – test seti oranı olmamakla birlikte %70-80 eğitim seti, %20-30 test seti oran aralıklarının daha yaygın olarak kabul gördüğü görülmektedir (Paulino vd., 2019: 170). Genel algı analizi kapsamında bu oran benzer ve güncel bir çalışma olan Mujahid vd. (2021: 8)'nin “Sentiment Analysis and Topic Modeling on Tweets about Online Education during COVID-19” adlı çalışmasında tercih edilen %75 eğitim seti - %25 test seti olarak tercih edilmiştir. Buna göre her üç veri seti için de tercih edilen eğitim ve test seti tweet sayıları aşağıdaki gibi belirlenmiştir:

Tablo 4.6. Genel algı analizi eğitim ve test setleri tweet sayıları

Veri seti	Eğitim setleri	Test setleri
1	3024	975
2	3446	1177
3	4766	1622

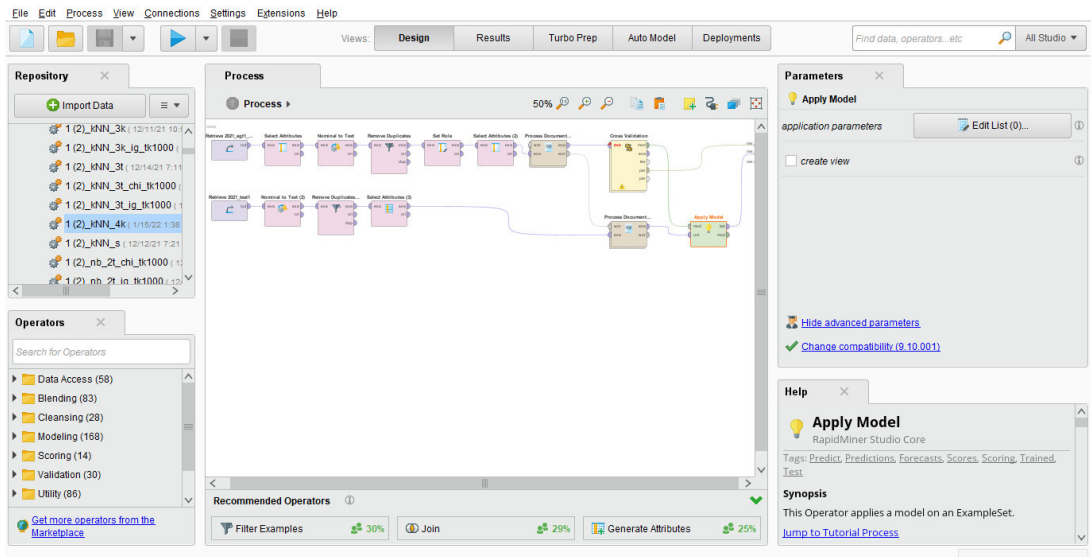
Her üç veri seti için de eğitim ve test setleri belirlendikten sonra duygu analizi uygulamasına geçilmiştir. Duygu analizi için RapidMiner Studio programı (9.10 versiyonu) tercih edilmiştir. Bu programın öncelikle denenmesi ve kullanıma karar verilmesinde benzer çalışmalardaki yaygın kullanımı etkili olmuştur. Değer (2017)'in “Sosyal Medya Mesajlarında Veri Madenciliği ile Duygu Analizi” ve Kuş (2019)'un “Suriyeli Mültecilerin Sosyal Entegrasyonunda Halkla İlişkilerin Rolü” adlı çalışmaları buna örnek olarak verilebilir.

RapidMiner A.B.D'de bulunan Yale Üniversitesi tarafından Java programlama dili kullanılarak geliştirilmiş açık kaynak kodlu bir veri madenciliği programıdır. Veriye ulaşma, analiz etme ve ondan anlamlı sonuçlar çıkarma amacıyla kullanılan veri madenciliği programları günümüzde pazarlama, bankacılık, sanayi, finans, eğitim, sağlık, güvenlik ve akademi gibi çok sayıda farklı alanda kullanılabilir. Örneğin Rapidminer programı kullanılarak pazarlama alanında müşteri kaybı olasılık tahminlemesi, doğrudan pazarlama modellemesi, pazar sepeti analizi; sanayide ekipmanların sağlıklı bir şekilde çalışabilmelerine imkan veren ve

herhangi bir hasar oluşması durumunu önceden tahminleyen önleyici bakım modellemesi, bankacılık ve finans alanında kredi risk modellemesi, fiyat risk kümeleme analizi; metin madenciliği kapsamında duygu analizi ve web analizi gibi çok sayıda uygulama gerçekleştirilebilir.

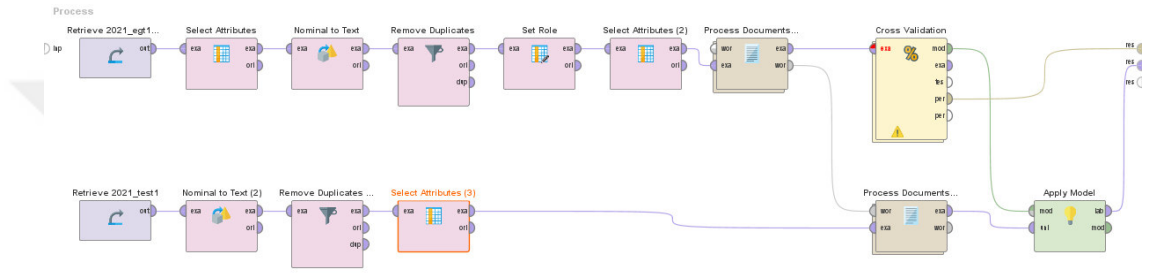
Rapidminer programının öğrenme, adaptasyon ve kullanımının nispeten kolay olması, içerisinde 1500’den fazla operatörün bulunması, arayüzünün kullanışlı ve görsel olarak göze hoş gelecek bir tasarıma sahip olması, grafik ve görselleştirmelerinin zenginliği, görece yüksek sayıda dosya formatını desteklemesi, programın gövdesinde yer almayan bazı özelliklerin eklentiler vasıtasıyla programa ilave edilebilir olması, eklentiler de dahil olmak üzere programın, “RapidMiner Community” adlı topluluk platformuna üye olunması halinde ücretsiz olması ve son olarak içerisinde bir gövdeleme algoritması olan Snowball’un Türkçe versiyonuna sahip olması tercih edilme sebeplerinin başında gelmektedir. Bu tez kapsamında da programın ana gövdesinde yer almayan ancak duygu analizi için kullanılması gereken metin madenciliği operatörleri, “Text Processing (9.3.001 versiyonu)” eklentisi yüklenerek programa ilave edilmiştir.

Rapidminer programında duygu analizi amaçlı kurulan ilk modelin arayüz görünümü aşağıdaki gibidir:



Şekil 4.1. Rapidminer arayüz görünümü

Program arayüzünde görünen sol üst taraftaki “Repository (Depo)” adlı kısımda veri setleri (eğitim ve test setleri), prosesler (modeller), örnek setleri (sonuçlar) ve performans sonuçları gibi veriler; sol alt taraftaki “Operators (Operatörler)” adlı kısımda modeller kurulurken kullanılan ve her birinin ayrı bir fonksiyonu olan operatörler; sağ üst taraftaki “Parameters (Değişkenler)” adlı kısımda seçili operatöre ait ayarlar yer almaktadır. “Process (İşlem)” adlı orta kısım ise modellerin kurulduğu alandır. “Process” alanında görünen kutucuklar ya veri ya da operatörlerdir. Aşağıda “Process” alanında oluşturulan modelin görüntüsü yer almaktadır.



Şekil 4.2. 1. Veri seti k-NN sınıflandırıcısı 4-gram (karakter) öznitelik çıkarma yöntemi genel algı analizi modeli

Modelde yer alan her bir operatörün detaylı olarak açıklaması aşağıdaki gibidir:

Retrieve 2021\_egt1



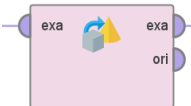
Retrieve 2021\_egt1: En üst soldan 1.sırada yer alan bu operatör programa aktarılan “2021\_egt1” adındaki 1.eğitim setini çağırma yapar. 1.eğitim seti yukarıda belirtildiği üzere 2021 yılına ait 3024 adet tweet’in ve her bir tweet’e ait genel algı, nitelik ve sebep sınıflarının yer aldığı veri setidir.

Select Attributes



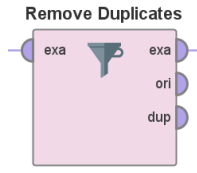
Select Attributes (Özellik Seçme): Bu operatör eğitim setinde yer alan kolonlardan (sütunlardan) üzerinde çalışılmak istenilenleri seçmeye yarar. Oluşturulan bu model kapsamında sadece genel algı analizi yapılacağı için bu operatör seçilip “Parameters” alanından sadece “Tweet” ve “Genel algı” kolonları seçilir.

Nominal to Text

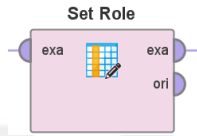


Nominal to Text: “Select Attributes” operatörü ile üzerinde çalışılacak olan kolonları belirlenen ve “nominal” tipinde gözükten verilerin hepsi bu operatör ile üzerinde metin madenciliği

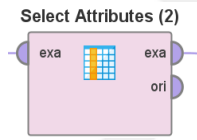
işlemlerinin yapılabilmesi için “text (metin)” tipine çevrilir.



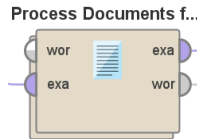
**Remove duplicates (Kopyaları temizle):** Bu operatörün “Parameters” alanından sadece “tweet” kolonu seçilerek bu operatöre aktarılan verilerden sadece tweet’lerin varsa birebir aynılarının temizlenmesi işlemi yapılır. Metin ön işleme aşamasında hatırlanacağı üzere zaten tweet’lerin birebir aynılarının temizlenmesi işlemi gerçekleştirilmişti ancak verilerin parça parça aktarılması sebebi ile her ihtimale karşı bu noktada da bir temizleme işlemi gerçekleştirilmektedir.



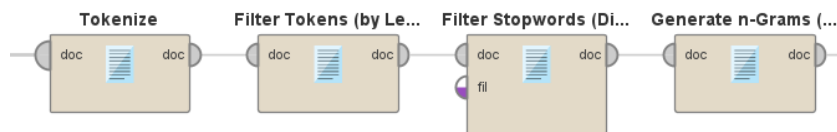
**Set Role (Rol Belirleme):** Bu operatör duygu analizi için hangi kolonun tahmininin yapılacağını belirlediği operatördür. “Parameters” alanından duygu analizi için “Genel algı” kolonu seçilir ve rolü “label (etiket)” olarak belirlenir. Artık model hangi kolonu tahmin edeceğini bilmektedir.



**Select Attributes (2):** 2.kez seçilen bu operatör ile bu kez herhangi bir kayıp ya da boş değer silinme işlemi yapılır. Bu şekilde tweet olmayan veya sınıflandırılmamış satırların silinmesi işlemi gerçekleştirilir. Her ne kadar yapılan kontrollerde tweet barındırmayan veya sınıflandırması yapılmamış satırın olmadığı teyit edilmiş olsa da yine de operatör herhangi bir ihtimale karşı kullanılmıştır.



**Process Documents from Data (Dokümanları Datadan İşle):** Bu operatöre aktarılan veriler operatörün içinde bulunan diğer operatörler tarafından işlenir. Sol taraftaki “exa (example set)” / veri seti girişinden kendisine aktarılan verileri alır, sağ tarafındaki “exa” çıkışından ise içinde uygulanan işlemler sonucu oluşan veriyi aktarır. “wor (word list)” / sözcük listesi çıkışından ise test seti için kullanılacak olan “Process Documents from Data” operatörüne eğitim setine ait özniteliklerini aktarır. Genel algı analizi 1.model için operatörün içine yerleştirilen operatörler şu şekildedir:



Şekil 4.3. Process Documents from Data operatörünün içinde yer alan operatörler



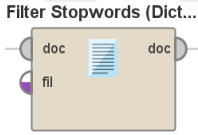
temizlenmiş olur.

Tokenize (Dizgeciklere Ayırma): Tweet metnini ayrı ayrı sözcüklere ayırmaya yarar. Bu süreçte noktalama işaretleri, sözcük vasfı taşımayan anlamsız karakterler ve satır sonu karakterleri de



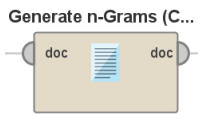
Filtreleme yapar. Tez kapsamında bu filtre 4 – 25 karakter aralığı olarak belirlenmiştir. Başka bir deyişle dizgeciklerden sadece 4 – 25 karakter uzunluğuna sahip olanlar işleme alınır. Bu, 4 karakterden daha kısa sözcüklerin analize tabi tutulmaması anlamına gelir. Yapılan testlerde 4 karakterin altındaki sözcüklerin analizden çıkarılmasının performansı artırdığı gözlemlenmiştir.

Filter Tokens (by Length) (Dizgecikleri Uzunluklarına göre Filtre): Bu operatör dizgecikleri karakter uzunluğu olarak



sağlamayacağı, hatta ekstra bir yük oluşturacağı için temizlenmesi beklenir. Bu operatör daha önce oluşturulmuş bir durak sözcükleri listesini (sözlük) kullanarak bu sözcükleri dizgecikler arasından temizleme işlemi yapar. Tez kapsamında kullanılan sözlük, Aksoy (2016) tarafından oluşturulmuş bir durak sözcükleri listesi referans alınarak ve analiz kapsamında yararı olabileceği düşünülen sözcüklerin bu listeden çıkarılması suretiyle güncellenerek oluşturulmuştur.

Filter Stopwords (Dictionary) (Durak Sözcüklerini Sözlük Kullanarak Filtre): 2.Bölümde anlatıldığı üzere durak sözcükleri bir dilde sıkça kullanılan fakat tek başına pek bir anlam ifade

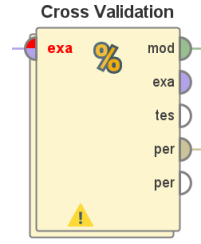


Özneliklerin çıkarılması ile artık metin formatındaki veriler, üzerinde veri madenciliği algoritmalarının kullanılabileceği sayısal vektörlere çevrilmiş olur. Bu noktada bahsedilmesi gereken bir başka husus, çıkarılan özneliklerin hangi

Generate n-Grams (Characters) (Karakter Tabanlı n-Gram Oluştur): Bu noktada dizgeciklerden çapraz geçerliliğe aktarılmak üzere

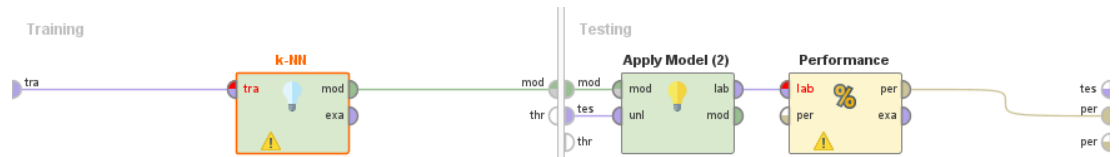
özneliklerin çıkarılması işlemi gerçekleştirilmektedir. Bu operatör ise öznelikleri karakter tabanlı 4-Gram yöntemine göre çıkarır. Genel algı analizinin ilk modeline bu yöntem ile başlanmış ancak sırayla karakter tabanlı 3-Gram, terim tabanlı 2-Gram ve 3-Gram ile Stem (Snowball) yöntemleri de uygulanarak performansları kayıt altına alınmıştır. Bu öznelik çıkarma yöntemlerinin tercih edilmesinin sebebi ise test analizlerinde en yüksek performansı göstermeleridir.

ağırlıklandırma yöntemine göre ağırlıklandırılacaktır. Bu seçim de bu operatörün içinde bulunduğu “Process Documents from Data” operatörünün “Parameters” alanından yapılır. Bu tez kapsamında bu yöntem TF-IDF (Terim Frekansı – Ters Doküman Frekansı Ağırlıklandırması) olarak seçilmiştir. Test analizlerinde diğer ağırlıklandırma yöntemleri de denenmiş, en yüksek performansa TF-IDF ağırlıklandırma yöntemi ile ulaşıldığı görülmüştür.



Cross Validation (Çapraz Geçerlilik): Bu operatör seçilen sınıflandırıcı vasıtasıyla öğrenmenin sağlandığı bir modelin kurulduğu ve modelin test edilerek performansının ölçüldüğü operatördür. Çapraz geçerlilik yönteminde, 2.Bölümde de detaylı olarak anlatıldığı üzere, öznitelikler belirlenen k sayısı kadar eşit parçaya ayrılırlar. Bu parçalardan her seferinde bir parçası test seti için, geri kalanı ise eğitim seti için kullanılır (Bu haliyle operatörün çalışma şekli duygu analizinin genel yapısına benzemektedir). Operatör k sayısı kadar, her seferinde test seti diğer bir parça olacak şekilde çalıştırılır ve her bir modelin performansı ölçülür. k kadar performans ölçüldükten sonra bu performansların ortalaması alınarak modelin genel performansı hesaplanır. k sayısı operatörün “Parameters” alanında belirlenir ve tez kapsamında literatürde de en sık kullanıldığı haliyle 10 olarak belirlenmiştir. Operatörün “mod” çıkışından içerisinde kurulan model, “per” çıkışından ise performans sonuçları aktarılır.

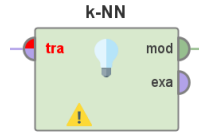
Genel algı analizi 1.model için operatörün içinde kurulan model aşağıdaki gibidir:



Şekil 4.4. Genel algı analizi 1.model için çapraz geçerlilik alt prosesi

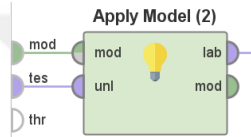
Şekil 4.4’te görüldüğü gibi çapraz geçerlilik 2 bölümden oluşmaktadır. “Training” (eğitim) adlı 1.bölümde, çapraz geçerlilik için oluşturulan eğitim seti sınıflandırıcıya tabi tutulur ve modelin öğrenmesi sağlanır; “Testing” kısmında ise öğrenen model kullanılarak test setinin sınıflandırılması yapılır. Bu işlem yukarıda anlatıldığı şekilde 10 kez tekrarlanır ve sonuç olarak öğrenmiş bir model ve modelin

performansı elde edilir. Fark edileceği üzere burada hem eğitim hem de test için ayrılan veri, genel algı analizinin daha başında belirlenen 1.eğitim setine ait veridir. Dolayısıyla bu verinin sınıflandırma bilgileri zaten mevcuttur. “Cross Validation” operatörünün burada yaptığı iş, sınıflandırması zaten bilinen veriyi ikiye ayırarak bir kısmını seçilen sınıflandırıcı vasıtasıyla modelin öğrenmesi için kullanmak, diğer kısmını kurulan modelin tahmin etmesini sağlamak ve mevcut sınıflandırma bilgisi ile sınıflandırma tahminlerini karşılaştırarak performansı ölçmektir.



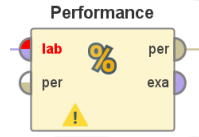
k-NN: k-Nearest Neighborhood (k-En Yakın Komşu) sınıflandırıcısı. “tra (training)” girişinden eğitim seti girişi yapılır.

“mod (model)” çıkışından ise model aktarılır.



Apply Model (Modeli Uygula): Kendisine aktarılan modeli uygular. Burada “mod” girişinden modeli, “unl (unlabeled)” girişinden de test verisini alır. “lab (labelled)” çıkışından da

veriyi sınıflandırılmış olarak çıkarır.

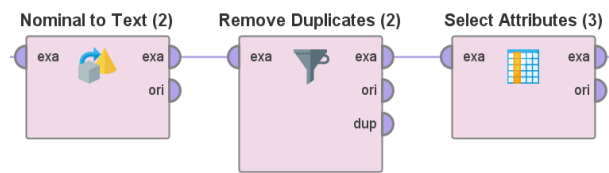


Performance (Performans): Modelin performansını ölçen operatördür. Ölçme yöntemi “Cross Validation” operatörü başlığı altında detaylı olarak anlatılmıştır.



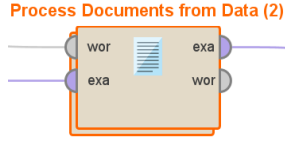
Retrive 2021\_test1: Alt sıra 1.sırada yer alan bu operatör programa aktarılan “2021\_test1” adındaki 1.test setini çağırmaya yarar. 1.test seti yukarıda belirtildiği üzere içinde sınıflandırılmamış tweet’lerin

yer aldığı veri setidir.

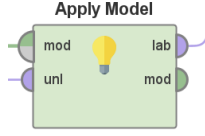


Nominal to Text – Remove Duplicates – Select Attributes: Burada kullanılan her üç operatör de eğitim seti için yaptıkları işin

aynısını yapmak için kullanılmışlardır. “Nominal to Text” operatörü verileri nominal tipinden text (metin) tipine çevirmek için, “Remove Duplicates” operatörü tweet’ler arasında olması muhtemel birebir aynı tweet’leri silmek için, “Select Attributes” operatörü de her ihtimale karşı boş ya da sınıflandırılmamış satırları silmek için kullanılmıştır.

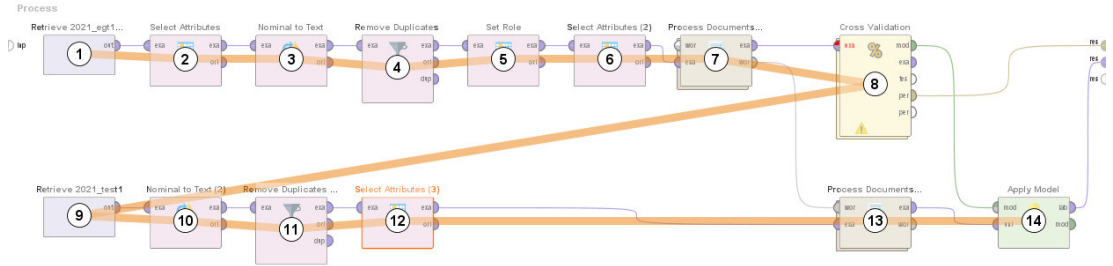


Process Documents from Data (2): Test seti için kullanılan bu operatör eğitim seti için yaptığı işin aynısını yapar, alt prosesi de birebir aynıdır. Tek farkı, “wor” girişinden eğitim seti için kullanılan “Process Documents from Data” operatöründen eğitim setine ait özelliklikleri ve ağırlıklandırma yöntemi bilgisini de içeri aktarmasıdır.



Apply Model: Duygu analizi sürecinde son olarak uygulanan “Apply Model (Modeli Uygula)” operatörü “mod” girişinden “Cross Validation” operatöründen kendisine aktarılan modeli, “unl (unlabeled)” girişinden de test verisi için uygulanan “Process Documents from Data” operatöründen kendisine aktarılan sınıflandırılmamış özelliklikleri alır. Sınıflandırılmamış test verisine modeli uygular ve “lab” çıkışından sınıflandırılmış veriyi çıkarır.

Genel algı analizi için kurulmuş 1.modelin çalışma sırası aşağıdaki gibidir:



Şekil 4.5. Genel algı analizi 1.model çalışma sırası

Şekil 4.5’te görüldüğü üzere, genel algı analizi 1.model ve bundan sonraki modellerde de olacağı üzere önce eğitim seti bir dizi ön işlemden geçirilir ve modelin öğrenmesi için uygun hale getirilir. Bu kapsamda tweet halindeki metin önce “token” adı verilen dizgecikler haline, daha sonra bir özelliklik çıkarma yöntemi kullanılarak özelliklikler haline getirilir. “Öznitelik” tweet’e ait, sınıflandırma modelinin uygulanabileceği en sade birimdir. Genel algı analizi 1.model için özelliklik çıkarma yöntemi olarak karakter tabanlı 4-Gram yöntemi kullanıldığı için özellikliklerin herhangi bir anlam ifade etmesi beklenmez. Zira 2.Bölümde detaylı olarak anlatıldığı üzere n-Gram yöntemleri olasılık hesaplama yöntemlerine göre özelliklik oluşturmaktadır. Daha sonraki modellerde kullanılan “stem” yönteminde ise bu değişiklik gösterecektir. Doğal dil işleme metodu olan stem (köke inme) methodunda

sözcüklerin köküne inme amaçlanmakta olduğu için bu yöntemle çıkarılan özniteliklerin anlamlı olması beklenir.

Eğitim seti veri madenciliği algoritmalarının uygulanabilmesi için uygun hale getirildikten sonra seçilen bir sınıflandırıcı vasıtasıyla modelin öğrenmesi sağlanır. Öğrenen modelin uygulanabilmesi için benzer ön işlemler test setine de uygulanır. Son olarak model test setine uygulanarak sınıflandırılmamış verilerin sınıflandırılması sağlanır. Sonuç, performans verileri ile birlikte sunulur.

Genel algı analizi 1.veri seti için uygulanan tüm sınıflandırıcı ve öznitelik çıkarma yöntemlerinin doğruluk yüzdesi performanslarını gösterir tablo aşağıdaki gibidir:

Tablo 4.7. Genel algı analizi 1.veri seti sınıflandırıcı – öznitelik çıkarma yöntemleri performansları

1. Veri seti	4-Gram (Karakter)	3-Gram (Karakter)	2-Gram (Terim)	3-Gram (Terim)	Stem (Snowball)
K-En Yakın Komşu	72,24	72,31	71,31	71,84	71,81
Naive Bayes	65,18	49,48	79,81	78,94	63,25
Karar Ağacı	70,21	70,54	70,88	70,94	70,44
Derin Öğrenme	70,24	69,14	79,54	79,04	68,77
Destek Vektör Makineleri	70,18	70,18	70,18	70,18	70,18

Tablo 4.7’de görüldüğü gibi genel algı analizi 1.veri seti için en yüksek performansı, 79,81 doğruluk oranı yüzdesi ile Naive Bayes sınıflandırıcısı ve terim tabanlı 2-Gram öznitelik çıkarma yöntemi göstermiştir. Onu, 79,54 doğruluk yüzdesi oranı ile Derin Öğrenme sınıflandırıcısı ve terim tabanlı 2-Gram öznitelik çıkarma yöntemi izlemektedir.

Tahmin edileceği üzere, diğer sınıflandırıcıları uygulamak için ilk modelde “k-NN” operatörünün olduğu yere sırası ile adı geçen diğer sınıflandırıcıların operatörleri yerleştirilmiştir. Öznitelik çıkarma yöntemini karakter tabanlı 3-Gram yapmak için “Generate n-Grams (Characters)” operatörünün “Parameters” alanından “length (uzunluk)” 3 olarak değiştirilmiş; terim tabanlı öznitelik çıkarma yöntemlerini uygulamak için de “Generate n-Grams (Terms)” operatörü kullanılmıştır. Stem yöntemi kullanılarak öznitelik çıkarmak için ise “Stem

(Snowball)” operatörü kullanılarak “Parameters” alanındaki “language (dil)” kısmından “Turkish (Türkçe)” seçeneği seçilmiştir.

1. Veri seti için uygulanan süreç, - analize katkısı olmadığına kanaat getirilerek terim tabanlı 3-Gram öznitelik çıkarma yöntemi terk edilmek suretiyle - 2. ve 3. veri setleri için de uygulanarak aşağıdaki tablolar elde edilmiştir.

Tablo 4.8. Genel algı analizi 2.veri seti sınıflandırıcı – öznitelik çıkarma yöntemleri performansları

2. Veri seti	4-Gram (Karakter)	3-Gram (Karakter)	2-Gram (Terim)	Stem (Snowball)
K-En Yakın Komşu	67,36	67,39	65,66	66,89
Naive Bayes	59,67	45,67	75,8	56,9
Karar Ağacı	61,69	62,01	62,33	62,01
Derin Öğrenme	67,39	67,68	74,14	66,25
Destek Vektör Makineleri	61,54	61,54	61,54	61,54

Tablo 4. 9. Genel algı analizi 3.veri seti sınıflandırıcı – öznitelik çıkarma yöntemleri performansları

3. Veri seti	4-Gram (Karakter)	3-Gram (Karakter)	2-Gram (Terim)	Stem (Snowball)
K-En Yakın Komşu	62,62	62,68	58,6	60,31
Naive Bayes	56,64	49,85	66,63	52,68
Karar Ağacı	45,67	45,39	45,6	45,69
Derin Öğrenme	64,86	64,9	67,14	64,45
Destek Vektör Makineleri	44,59	44,59	44,59	44,59

Tablo 4.8 ve 4.9’da görüldüğü üzere 2.veri seti için en yüksek performansı gösteren model 75,8 doğruluk oranı yüzdesi ile Naive Bayes sınıflandırıcısı terim tabanlı 2-Gram öznitelik çıkarma yöntemi; 3.veri seti için en yüksek performansı gösteren model ise 67,14 doğruluk oranı yüzdesi ile Derin Öğrenme sınıflandırıcısı terim tabanlı 2-Gram öznitelik çıkarma yöntemi olmuştur. Veri setleri arasında da en

yüksek doğruluk oranına ulaşılan veri seti 1.veri seti olmuş, onu sırasıyla 2. ve 3.veri setleri takip etmiştir.

3.Bölümde anlatıldığı üzere duygu analizinde performans ölçümü için en çok dikkat edilen ölçüt doğruluk oranı (accuracy) olsa da tek ölçüt değildir. Doğruluk oranı sınıfların ayrı ayrı doğruluk yüzdelerine dikkat etmediği, tüm sınıfların doğruluk yüzdelerinin bir ortalaması olduğu için tüm sınıfları aynı oranda temsil ettiği kesin olarak söylenemez. Bu noktada sınıfların ayrı ayrı performanslarını daha fazla dikkate alacak bir ölçüte ihtiyaç duyulur ki bu da çoğunlukla F-ölçütüdür. F-ölçütünün kendi içinde çeşitleri olsa da 3 sınıflı genel algı analizi modeli için sınıfları eşit derecede dikkate alacak olan ölçüt makro-F1 ölçütüdür. Makro-F1 ölçütü her sınıfın ayrı ayrı F1-ölçütünün aritmetik ortalamasıdır. Aritmetik ortalama her sınıfın eşit derecede dikkate alındığının kanıtıdır. F1-ölçütü de yine 2.Bölümde anlatıldığı üzere sınıfların kesinlik (precision) ve duyarlılık (recall) değerlerinin harmonik ortalamasıdır.

Bu noktada her 3 veri seti için en yüksek performansı gösteren sınıflandırıcılar olan Naive Bayes ve Derin Öğrenme sınıflandırıcıları ve tez kapsamında kullanılan tüm öznelilik çıkarma yöntemleri için makro-F1 ölçütleri hesaplanmıştır.

Tablo 4. 10. Naive Bayes sınıflandırıcısı makro-F1 ölçütü değerleri

Veri Setleri	4-Gram (Karakter)	3-Gram (Karakter)	2-Gram (Terim)	Stem (Snowball)
1	0,42	0,36	0,67	0,43
2	0,45	0,39	0,67	0,45
3	0,52	0,47	0,62	0,5

Tablo 4. 11. Derin Öğrenme sınıflandırıcısı makro-F1 ölçütü değerleri

Veri Setleri	4-Gram (Karakter)	3-Gram (Karakter)	2-Gram (Terim)	Stem (Snowball)
1	0,52	0,51	0,65	0,51
2	0,55	0,56	0,64	0,53
3	0,61	0,6	0,63	0,59

Tablo 4.10 ve 4.11’de görüleceği üzere her iki sınıflandırıcı için de en yüksek makro-F1 ölçütü değerlerine, doğruluk oranlarında olduğu gibi terim tabanlı 2-Gram yöntemi ile ulaşılmıştır. Onun dışında Naive Bayes sınıflandırıcısı için 1.ve 2.veri setleri 3.veri setine göre daha yüksek bir performans göstermiş; Derin Öğrenme sınıflandırıcısı için ise aralarında büyük bir fark olmamakla birlikte performans sıralaması 1., 2. ve 3.veri seti şeklinde olmuştur.

Tüm performans ölçümlerine bakılarak genel algı analizinin en yüksek performans gösteren 1.veri seti ile, Naive Bayes ve Derin Öğrenme sınıflandırıcıları ve terim tabanlı 2-Gram öznitelik çıkarma yöntemi kullanılarak yapılmasına karar verilmiştir. Sonuçlar aşağıdaki gibidir:

Tablo 4. 12. Genel algı analizi en yüksek performans gösteren modellerin tahmin sonuçları

Tahmin edilen tweet sayısı: 12854					Nötr çıkarıldığında		
Sınıflandırıcı	Doğruluk %	F-Ölç.	Olumlu %	Olumsuz %	Nötr %	Olumsuz %	Olumlu %
Naive Bayes	79,81	0,67	15,5	68,5	16	81,5	18,5
Derin Öğrenme	79,47	0,65	11,9	76,5	11,6	86,5	13,5

12854 adet tweet’in genel algı sınıfının tahmin edildiği analizde, Tablo 4.12’de görüldüğü üzere en yüksek performansı 0,34’lük bir doğruluk oranı farkı ile de olsa Naive Bayes sınıflandırıcısı göstermiştir. Buna göre Türkiye’de yaşayan Suriyeli sığınmacılar ile ilgili olarak Twitter sosyal medya platformunda bu konuda atılan ve görüş belirten tweet’lerin %81,5’i olumsuz, %18,5’i olumludur. İkinci en yüksek performansı gösteren Derin Öğrenme sınıflandırıcısına göre yapılan analizde ise bu oran %86,5 olumsuz, %13,5 olumlu şeklindedir.

#### 4.5. Algının Niteliği Analizi

Daha önce belirtildiği üzere Türkiye’de yaşayan Suriyeli sığınmacılar hakkındaki algıyı ölçmek üzere gerçekleştirilen duygu analizinin ikincisi algının niteliği analizidir. Bu analizde Suriyeli sığınmacılar hakkındaki algının nitelikleri ortaya konmaya çalışılmıştır. Analiz, olumsuz ve olumlu nitelik analizi olmak üzere ikiye ayrılmaktadır.

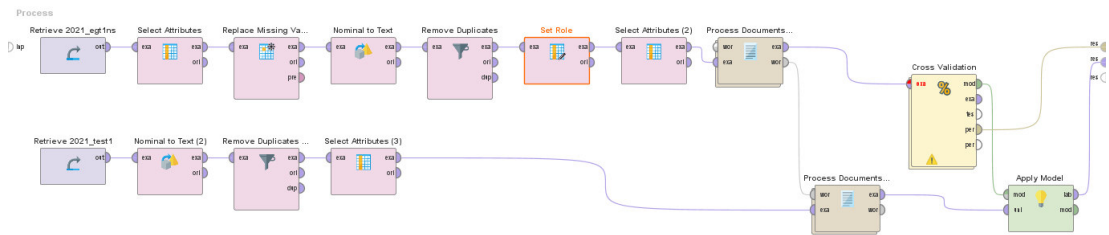
#### 4.5.1. Olumsuz Algının Niteliği Analizi

Bu analizde Türkiye’de yaşayan Suriyeli sığınmacılar hakkındaki olumsuz algının nitelikleri ortaya konmaya çalışılmıştır. Genel algı analizi için kullanılan ve 3024 adet tweet’ten oluşan 1.eğitim setindeki nitelik belirten tweet’ler bu analiz için kullanılmıştır. 1.eğitim setinde nitelik belirten 2477 adet tweet tespit edilmiştir. 2.ve 3.eğitim setlerinde olumsuz algı belirten daha fazla tweet yer almadığı için analize dahil edilmemiştir. Olumsuz algının niteliği analizi için oluşturulan eğitim setinin nitelik sınıflandırmaları aşağıdaki gibidir:

Tablo 4. 13. Olumsuz algının niteliği analizi eğitim seti sınıflandırma verileri

	Nitelik	Sınıflandırma sayısı	Yüzde	Sıra
Olumsuz nitelikler	Hoşnutsuzluk	1839	74,2	1
	Nefret	132	5,3	3
	Kızgınlık	93	3,8	4
	Alay etme	21	0,8	5
	Aşağılama	19	0,8	6
	Üzgün olma	9	0,4	7
	Olumlu nitelikler	364	14,7	2
	Toplam	2477	100	

Oluşturulan eğitim setine karşılık, genel algı analizinde kullanılan ve 975 adet tweet’ten oluşan 1. test seti tercih edilmiştir. Bu şekilde eğitim seti - test seti oranları yaklaşık olarak %72 – %28 şeklinde oluşmuştur. Buna göre oluşturulan olumsuz algının niteliği analizi 1.modeli aşağıdaki gibidir:



Şekil 4.6. Olumsuz algının niteliği analizi 1.modeli

Olumsuz algının niteliği analizi görüntü olarak büyük oranda genel algı analizi modeline benzese de kullanılan operatör ve içerik olarak bazı farklılıklara sahiptir. Öncelikle ve doğal olarak yukarıda belirtildiği üzere eğitim seti değişmiştir. “Select Attributes” operatöründen bu sefer “tweet” ve “nitelik” kolonları seçilmiştir. Bu, bundan sonraki işlemlerin sadece bu iki kolona uygulanacağı anlamına gelmektedir. Genel algı analizi modelinde kullanılmayan “Replace Missing Values (Kayıp Değerleri Değiştir) operatörü kullanılmıştır. Bu operatör, nitelik belirtmeyen tweet’lerin “nötr” olarak sınıflandırılması için kullanılmıştır. “Set Role” operatörü ile tahmini yapılacak olan kolon “nitelik” olarak seçilmiştir. Onun dışındaki diğer operatör ve değişkenler genel algı analizi ile aynıdır. Buna göre olumsuz algının niteliği için uygulanan modellerin doğruluk oranı performansları aşağıdaki gibidir:

Tablo 4. 14. Olumsuz nitelik analizi sınıflandırıcı – öznelik çıkarma yöntemleri performansları

1. Veri Seti	4-Gram (Karakter)	3-Gram (Karakter)	2-Gram (Terim)	Stem (Snowball)
K-En Yakın Komşu	64,08	64,31	63,25	63,35
Naive Bayes	56,65	42,95	35,15	54,28
Karar Ağacı	61,15	61,11	61,21	61,08
Derin Öğrenme	62,18	61,45	67,14	61,05
Destek Vektör Makineleri	61,01	61,01	61,01	61,01

Tablo 4.14’te görüldüğü üzere, algının niteliği analizi için en yüksek performans gösteren model 67,14’lük doğruluk oranı ile Derin Öğrenme sınıflandırıcısı ve terim tabanlı 2-Gram öznelik çıkarma yöntemi olmuştur. 2021 yılına ait toplanmış 12854 adet tweet’e bu model uygulanmış ve aşağıdaki sonuçlara ulaşılmıştır.

Tablo 4. 15. Olumsuz algının niteliği analizi tahmin sonuçları

Tahmin edilen tweet sayısı: 12854		Doğruluk oranı: 66,84		
Nitelik	Sınıflandırma sayısı	Yüzde	Sıra	Olumlu ve nötr’ler çıkarıldığında yüzde
Hoşnutsuzluk	9600	74,7	1	91,2

Nefret	588	4,6	4	5,59
Kızgınlık	247	1,9	5	2,3
Alay etme	42	0,3	6	0,3
Aşağılama	29	0,2	7	0,2
Üzgün olma	9	0,1	8	0,08
Olumlu nitelikler	1087	8,5	3	
Nötr	1252	9,7	2	

Tablo 4.15’te görüldüğü üzere, olumsuz algının niteliği analizinde olumsuz nitelik belirten tweet’lerin %91,2’si “hoşnutsuzluk” niteliğindedir. Onu, %5,5 ile “nefret” ve %2,3 ile “kızgınlık” nitelikleri takip etmektedir. Bu sonuç, Türkiye’de yaşayan Suriyeli sığınmacılar ile ilgili olarak olumsuz görüş belirten tweet’lerin büyük bir çoğunluğunun “hoşnutsuzluk” niteliğinde olduğunu göstermektedir.

#### 4.5.2. Olumlu Algının Niteliği Analizi

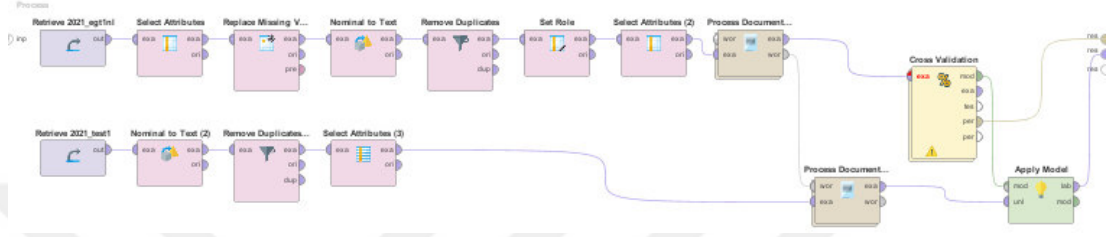
Olumsuz algının niteliği için oluşturulmuş olan 2477 adet tweet içeren eğitim seti, olumlu algının niteliği analizi 1.eğitim seti olarak üzerinde bazı düzenlemeler yapılmak suretiyle kullanılmıştır. Bu düzenlemeler, olumsuz niteliklerin hepsinin tek bir başlık altında toplanması ve olumlu niteliklerin kendi içinde ayrı ayrı sınıflandırılmasıdır. Buna göre oluşturulan 1.eğitim seti aşağıdaki gibidir:

Tablo 4. 16. Olumlu algının niteliği analizi 1. eğitim seti sınıflandırma verileri

Nitelik	Sınıflandırma sayısı	Yüzde	Sıra	
Olumsuz nitelikler	2113	85,3	1	
Olumlu nitelikler	Sempati besleme	112	4,5	2
	Kabul etme / kabul görme	109	4,4	3
	Sevme	44	1,8	4
	Hak verme	33	1,3	5
	Acıma / üzülme	31	1,3	6
	Beğenme	27	1,1	7
	Saygı duyma	5	0,2	8

Değer verme	3	0,1	9
Toplam	2477	100	

1.eğitim setine karşılık genel algı analizi için kullanılmış olan 1.test seti kullanılmış, buna göre olumsuz nitelik analizinde olduğu gibi eğitim seti – test seti oranı yaklaşık olarak %72 - %28 olarak oluşmuştur. Bu şekilde oluşturulan 1.veri seti ile olumlu algının niteliği analizi için kurulmuş olan 1.model aşağıdaki gibidir:



Şekil 4.7. Olumlu algının niteliği analizi 1.veri seti 1.modeli

Şekil 4.7’de görüldüğü üzere modelin omurgası olumlu algının niteliği analizi ile aynıdır. Farklı olan tek şey eğitim setinde yapılmış olan değişikliktir. Buna göre sınıflandırıcı ve öznitelik çıkarma yöntemlerinin değiştirilmesi suretiyle uygulanan modellerin doğruluk performansları aşağıdaki gibidir:

Tablo 4. 17. Olumlu nitelik analizi 1.veri seti sınıflandırıcı – öznitelik çıkarma yöntemleri performansları

1.Veri Seti	4-Gram (Karakter)	3-Gram (Karakter)	2-Gram (Terim)	Stem (Snowball)
K-En Yakın Komşu	71,28	71,31	71,38	71,24
Naive Bayes	66,44	54,35	43,82	64,41
Karar Ağacı	70,08	70,11	70,11	70,11
Derin Öğrenme	68,34	68,54	72,81	69,34
Destek Vektör Makineleri	70,11	70,11	70,11	70,11

Tablo 4.17’de görüleceği üzere 1.veri seti için olumlu nitelik analizi en yüksek performansına Derin Öğrenme sınıflandırıcısı ve terim tabanlı 2-Gram öznitelik çıkarma yöntemi ile ulaşılmıştır ki bu da 72,81’lik doğruluk oranıdır. Bu noktada hatırlanacağı üzere genel algı analizi için veri setinin daha dengeli bir hal alması için

oluşturulmuş olan 2. ve 3.veri setleri de analize dahil edilmiş ve daha dengeli veri setlerinin sonuca etkisi irdelenmiştir. Buna göre 2.eğitim seti sınıflandırma verileri aşağıdaki gibidir:

Tablo 4. 18. Olumlu algının niteliği analizi 2. eğitim seti sınıflandırma verileri

Nitelik	Sınıflandırma sayısı	Yüzde	Sıra	
Olumsuz nitelikler	2113	80,9	1	
Olumlu nitelikler	Kabul etme / kabul görme	146	5,6	2
	Sempati besleme	138	5,3	3
	Hak verme	55	2,1	4
	Sevme	54	2,1	5
	Beğenme	44	1,7	6
	Acıma / üzülme	41	1,6	7
	Değer verme	13	0,5	8
	Saygı duyma	7	0,3	9
	Toplam	2611	100	

2.eğitim setine karşılık genel algı analizinde kullanılan 2.test seti kullanılmıştır. Olumlu nitelik analizi 1.veri seti ile kurulan modellerden farklı olarak, analize katkı sağlamadığına kanaat getirilerek karakter tabanlı 4-Gram öznitelik çıkarma yöntemi analiz dışında bırakılmış ve diğer sınıflandırıcı ve öznitelik çıkarma yöntemleri kullanılarak aşağıdaki performans sonuçlarına ulaşılmıştır.

Tablo 4. 19. Olumlu nitelik analizi 2.veri seti sınıflandırıcı – öznitelik çıkarma yöntemleri performansları

2.Veri Seti	3-Gram (Karakter)	2-Gram (Terim)	Stem (Snowball)
K-En Yakın Komşu	66,01	63,88	66,13
Naive Bayes	50,82	37,64	57,36
Karar Ağacı	61,48	61,48	61,48
Derin Öğrenme	63,33	67,1	64,82

Destek Vektör Makineleri	61,48	61,48	61,48
--------------------------	-------	-------	-------

Tablo 4.19’da görüleceği üzere, 2.veri seti ile yapılan analizlerde en yüksek performansa yine Derin Öğrenme sınıflandırıcısı ve terim tabanlı 2-Gram öznitelik çıkarma yöntemi ile ancak 67,1’lik daha düşük bir doğruluk oranı ile ulaşılmıştır. Sınıflar arası daha dengeli bir dağılım daha düşük bir performans ile sonuçlanmıştır. Olumlu algının niteliği analizi için oluşturulan 3. eğitim seti verileri aşağıdaki gibidir:

Tablo 4. 20. Olumlu algının niteliği analizi 3. eğitim seti sınıflandırma verileri

Nitelik	Sınıflandırma sayısı	Yüzde	Sıra
Olumsuz nitelikler	2117	67,8	1
Kabul etme / kabul görme	275	8,8	2
Sempati besleme	185	5,9	3
Değer verme	147	4,7	4
Beğenme	135	4,3	5
Hak verme	100	3,2	6
Sevme	87	2,8	7
Acıma / üzülme	68	2,2	8
Saygı duyma	8	0,3	9
Toplam	3122	100	

3. eğitim setine karşılık olarak genel algı analizinde kullanılmış olan 3. test seti kullanılmıştır. Olumlu nitelik analizi 2. veri seti için uygulanan tüm sınıflandırıcı ve öznitelik çıkarma yöntemleri olumlu nitelik analizi 3. veri seti için de uygulanmış ve aşağıdaki sonuçlara ulaşılmıştır.

Tablo 4. 21. Olumlu nitelik analizi 3. veri seti sınıflandırıcı – öznitelik çıkarma yöntemleri performansları

3. Veri Seti	3-Gram (Karakter)	2-Gram (Terim)	Stem (Snowball)
K-En Yakın Komşu	59,26	57,12	58,86

Naive Bayes	45,1	37,17	52,05
Karar Ağacı	44,55	44,55	44,55
Derin Öğrenme	59,95	60,29	59,57
Destek Vektör Makineleri	44,55	44,55	44,55

Tablo 4.21’de görüleceği üzere en yüksek performansı yine Derin Öğrenme sınıflandırıcısı ve terim tabanlı 2-Gram öznitelik çıkarma yöntemi ve yine performansı daha düşük olmak suretiyle 60,29 doğruluk oranı olarak göstermiştir. Daha dengeli bir eğitim seti yine daha düşük bir performans ile sonuçlanmıştır. Buna göre en yüksek performansı gösteren 1.veri seti ile analize gidilmesine karar verilmiş, en yüksek performansı gösteren Derin Öğrenme sınıflandırıcısı ve terim tabanlı 2-Gram öznitelik çıkarma yönteminin kullanıldığı model uygulanarak aşağıdaki sonuçlara ulaşılmıştır.

Tablo 4. 22. Olumlu algının niteliği analizi tahmin sonuçları

Tahmin edilen tweet sayısı: 12854			Doğruluk: 73,34	
Nitelik	Sınıflandırma sayısı	Yüzde	Sıra	Olumlu ve nötr’ler çıkarıldığında yüzde
Sempati besleme	400	3,1	3	32,6
Kabul etme / kabul görme	258	2	4	21
Sevme	183	1,4	5	14,9
Hak verme	139	1,1	6	11,3
Acıma / üzülme	121	0,9	7	9,8
Beğenme	108	0,8	8	8,8
Saygı duyma	12	0,1	9	0,9
Değer verme	4	0,03	10	0,3
Olumsuz nitelikler	10197	79,3	1	
Nötr	1432	11,1	2	
Toplam		100		100

Tablo 4.22’de görüleceği üzere Türkiye’de yaşayan Suriyeli sığınmacılar ile ilgili olarak olumlu nitelik belirten tweet’lerin %32,6’sı “sempati besleme” niteliğindedir. Bunu, %21 ile “kabul etme/kabul görme”; %14,9 ile “sevme”; %11,3 ile “hak verme”; %9,8 ile “acıma/üzülme”; %8,8 ile “beğenme” nitelikleri takip etmektedir.

#### 4.6. Algının Sebep Analizi

Türkiye’de yaşayan Suriyeli sığınmacılar hakkındaki algı üzerine gerçekleştirilen son duygu analizinde algının sebepleri irdelenmiştir. Sebep analizleri olumlu ve olumsuz olmak üzere ikiye ayrılmıştır.

##### 4.6.1. Olumsuz Algının Sebep Analizi

Olumsuz algının sebeplerini ortaya koymak için her duygu analizi sürecinin başında olduğu gibi öncelikle eğitim seti oluşturulmuştur. Eğitim seti için genel algı analizi 1.eğitim setindeki olumsuz sebep belirten tweet’ler kullanılmış ve 1888 adet tweet’ten oluşan bir eğitim seti oluşturulmuştur. Eğitim setine dair sınıflandırma verileri aşağıdaki gibidir:

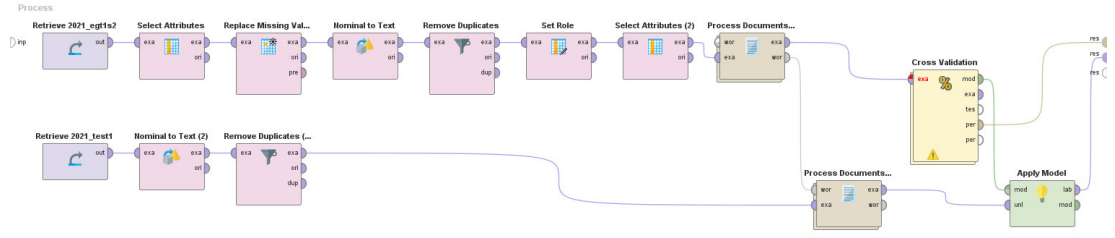
Tablo 4. 23. Olumsuz algının sebep analizi eğitim seti sınıflandırma verileri

Olumsuz Algı Sebep Başlıkları	Sebep	Sınıflandırma sayısı	Yüzde	Sıra
Suriyeliler bize zarar veriyor	Ekonomik yük / zarar	500	26,5	1
	Güvenlik sorunu			
	Suç işliyorlar			
	Rahatsız ediyorlar			
	İşsizliğin sebeplerinden biri Suriyeliler			
	Toplumsal huzurumuzu bozuyorlar			
	Daha düşük sağlık hizmeti almamıza sebep oluyorlar			
Daha düşük eğitim hizmeti almamıza sebep oluyorlar				
	Kurallara uymuyorlar			

	Sığınma / ikamet hakkı			
	Vatandaşlık hakkı			
Suriyelilerin Türkiye'deki hakları	Eğitim hakkı	452	23,9	2
	T. C. Vatandaşlarına karşı hukuki yollara başvurma			
	İstihdam hakkı			
Suriyeliler bizden daha iyi durumdadır / mevcut durumları	Bizden daha çok hakka / imtiyaza sahipler / daha iyi durumdadır / Suriyelilere daha çok değer veriliyor	406	21,5	3
	Rahat içindeler			
Suriyelilerin nüfusu	Sayınca çok fazlalar	315	16,7	4
	Çok fazla çocuk yapıyorlar			
Suriyelilerin kötü özellikleri var	Kötü özellikleri var	122	6,5	5
Kişisel düşünceler / Diğer	Kişisel düşünceler	93	4,9	6
	Diğer			
Toplam			100	

Tablo 4.23'de görüleceği üzere, sınıflandırma sürecinde ortaya çıkan sebepler, daha önce belirtildiği üzere gruplandırmalara tabi tutulmuşlar, ortak noktaları bulunan sebepler aynı başlık altında toplanmışlardır. Böyle bir uygulama yapmanın sebebi daha az sınıf ile yapılan duygu analizinin daha yüksek performans göstermesidir. Yapılan testlerde de bu teyit edilmiştir. Sonuç olarak duygu analizindeki sınıflar olarak sebep başlıkları kullanılmıştır.

Oluşturulan eğitim setine karşılık genel algı analizinde kullanılan 1.test seti kullanılmıştır. Buna göre oluşturulan 1.model aşağıdaki gibidir:



Şekil 4.8. Olumsuz algının sebep analizi 1.modeli

Şekil 4.8’de görüleceği üzere, olumsuz algının sebep analizi 1.modeli daha önceki duygu analizleri modelleri ile büyük oranda benzerlik göstermektedir. Farklılık olarak doğal olarak eğitim seti değişmiştir. Onun dışında daha önceki analizlerde test seti için kayıp değerleri ve boş satırları temizlemek için kullanılan “Select Attributes” operatörü bu modelde kullanılmamış, onun yerine “Remove Duplicates” operatöründen “treat missing values as duplicates (kayıp değerlere kopya gibi davran) seçeneği işaretlenerek tek bir operatörle hem kayıp değerleri hem de birebir aynı tweet’leri temizleme yoluna gidilmiştir.

Eğitim seti için kullanılan “Select Attributes” operatöründe bu kez “tweet” ve “olumsuz sebep” kolonları seçilmiştir ki bu, bundan sonraki işlemlerin sadece bu iki kolon için yapılacağı anlamına gelir. Eğitim seti için kullanılan “Replace Missing Values” operatöründen “olumsuz sebep” kolonu seçilerek bu kolondaki sınıflandırılmamış tweet’ler temizlenmiştir. “Set Role” operatöründen tahmini yapılacak olan kolon olarak “olumsuz sebep” kolonu seçilmiştir. Onun dışındaki operatör ve ayarları önceki analizler ile aynıdır. Buna göre uygulanan modeller ve doğruluk oranlarını gösterir performans sonuçları aşağıdaki gibidir:

Tablo 4. 24. Olumsuz algının sebep analizi sınıflandırıcı – öznelitik çıkarma yöntemleri performansları

1. Veri Seti	4-Gram (Karakter)	3-Gram (Karakter)	2-Gram (Terim)	Stem (Snowball)
K-En Yakın Komşu	47,72	48,02	45,95	47,78
Naive Bayes	40,05	34,62	41,49	38,22
Karar Ağacı	37,32	37,25	37,65	37,52
Derin Öğrenme	48,35	49,25	56,82	49,45
Destek Vektör Makineleri	37,39	37,39	37,39	37,39

Tablo 4.24'te görüleceği üzere, olumsuz algının sebep analizi için en yüksek performansı Derin Öğrenme sınıflandırıcısı ve terim tabanlı 2-Gram öznitelik çıkarma yöntemi kullanılarak uygulanan model, 56,82'lik doğruluk oranı şeklinde göstermiştir. Buna göre en yüksek performansı gösteren model kullanılarak eldeki tüm tweet'ler analizi tabi tutulmuş ve aşağıdaki sonuçlar elde edilmiştir.

Tablo 4. 25. Olumsuz algının sebep analizi tahmin sonuçları

Tahmin edilen tweet sayısı: 12854		Doğruluk: 57,38			
Olumsuz Algı Sebepler Başlıkları	Sebepler	Tahmin sayısı	Yüzde	Sıra	Nötr'ler çıkarıldığı da yüzde
Suriyeliler bize zarar veriyor	Ekonomik yük / zarar				
	Güvenlik sorunu				
	Suç işliyorlar				
	Rahatsız ediyorlar				
	İşsizliğin sebeplerinden biri Suriyeliler				
	Toplumsal huzurumuzu bozuyorlar	1816	14,1	2	28,2
	Daha düşük sağlık hizmeti almamıza sebep oluyorlar				
Suriyelilerin nüfusu	Daha düşük eğitim hizmeti almamıza sebep oluyorlar				
	Kurallara uymuyorlar				
	Sayınca çok fazlalar				
Suriyelilerin Türkiye'deki hakları	Çok fazla çocuk yapıyorlar	1635	12,7	3	25,4
	Sığınma / ikamet hakkı				
	Vatandaşlık hakkı	1307	10,2	4	20,3
	Eğitim hakkı				

		T. C. Vatandaşlarına karşı hukuki yollara başvurma			
		İstihdam hakkı			
Suriyeliler bizden daha iyi durumdadır / mevcut durumları	Bizden daha çok hakka / imtiyaza sahipler / daha iyi durumdadır / Suriyelilere daha çok değer veriliyor	1156	9	5	18
	Rahat içindeler				
Suriyelilerin kötü özellikleri var	Kötü özellikleri var	255	2	6	3,9
Kişisel düşünceler / Diğer	Kişisel düşünceler	253	2	7	3,9
	Diğer				
	Nötr	6432	50	1	
Toplam			100		100

Tablo 4.25'te görüleceği üzere, test analizinden daha yüksek bir doğruluk oranı olan 57,38'lik bir doğruluk oranı ile olumsuz algının sebep analizi tamamlanmış ve sebep başlıkları diğer analizlere nazaran daha dengeli bir dağılım göstermiştir. Olumsuz sebep belirten tweet'lerin %28,2'si "Suriyeliler bize zarar veriyor" başlığı altında toplanmaktadır. Onu %25,4 ile "Suriyelilerin nüfusu"; %20,3 ile "Suriyelilerin Türkiye'deki hakları" ve %18 ile "Suriyeliler bizden daha iyi durumdadır / mevcut durumları" başlıkları takip etmektedir.

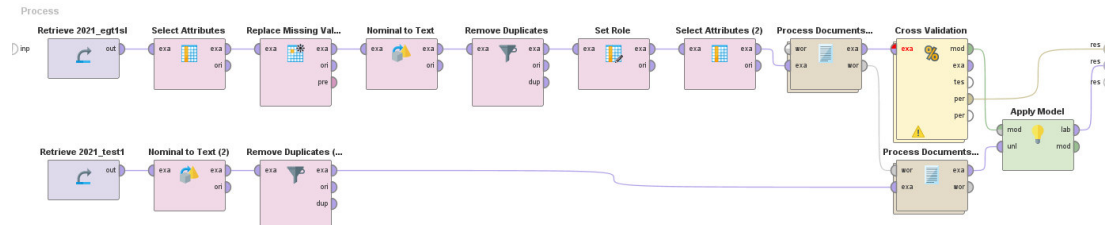
#### 4.6.2. Olumlu Algının Sebep Analizi

Türkiye'de yaşayan Suriyeli sığınmacılar hakkındaki algı ile ilgili olarak yapılan son analiz olumlu algının sebep analizidir. Her analiz öncesinde olduğu gibi ilk olarak eğitim seti oluşturulmuştur. Eğitim seti olarak genel algı analizi için kullanılan her üç eğitim seti de kullanılmıştır. Farklılık olarak her üç eğitim seti için de "olumlu algı" kolonu seçilmiştir. Bu şekilde oluşturulan 1.eğitim seti sınıflandırma verileri aşağıdaki gibidir:

Tablo 4. 26. Olumlu algının sebep analizi 1.eğitim seti sınıflandırma verileri

Sebepler	Sebepler	Sınıflandırma sayısı	Yüzde	Sıra
Olumlu sebepler	Olumsuz sebepler	1894	87,9	1
	Yardıma ihtiyaçları olma / mağdur / mazlum olma durumları	114	5,3	2
	Suriyelilerin iyi özellikleri	48	2,2	3
	Ortak yanlarımız	41	1,9	4
	Suriyelilerin faydalı olma durumu	32	1,5	5
	Suriyelilerin zararlarının olmaması	11	0,5	6
	Suriyeliler ile arkadaş olma durumu	9	0,4	7
	Diğer	6	0,3	8
Toplam		2155	100	

Bu şekilde oluşturulan 1.eğitim setine karşılık genel algı analizi için kullanılan 1.test seti kullanılmıştır. Bu şekilde oluşturulan 1.veri seti ile uygulanan 1.model aşağıdaki gibidir:



Şekil 4.9. Olumlu algının sebep analizi 1.modeli

Şekil 4.9’da görüldüğü üzere olumlu algının sebep analizi modeli olumsuz algı sebep analizi modeli ile büyük oranda benzerlik göstermektedir. Farklılık olarak eğitim seti değiştirilmiştir; “Select Attributes” operatöründen analize dahil edilecek kolon olarak “”tweet” ve “olumlu sebep” kolonları seçilmiştir; “Replace Missing Values” operatöründen sınıflandırılmamış tweet’lerin temizlenmesi için “olumlu sebep” kolonu seçilmiştir ve “Set Role” operatöründen tahmini yapılacak olan kolon olarak “olumlu sebep” kolonu seçilmiştir. Sınıflandırıcı ve öznitelik çıkarma yöntemlerinin değiştirilmesi suretiyle uygulanan modellerin doğruluk oranlarını gösterir tablo aşağıdaki gibidir:

Tablo 4. 27. Olumlu algının sebep analizi 1.veri seti sınıflandırıcı – öznitelik çıkarma yöntemleri performansları

1.Veri Seti	4-Gram (Karakter)	3-Gram (Karakter)	2-Gram (Terim)	Stem (Snowball)
K-En Yakın Komşu	66,68	67,41	65,84	67,41
Naive Bayes	59,45	52,52	43,05	55,71
Karar Ağacı	62,81	62,81	62,75	62,81
Derin Öğrenme	66,11	66,28	67,88	64,91
Destek Vektör Makineleri	62,81	62,81	62,81	62,81

Tablo 4.27’de görüleceği üzere, 1.veri seti kullanılarak gerçekleştirilen olumlu algının sebep analizi modellerinden en yüksek performansı Derin Öğrenme sınıflandırıcısı ve terim tabanlı 2-Gram öznitelik çıkarma yöntemi kullanılarak uygulanan model, 67,88’lik bir doğruluk oranı ile göstermiştir.

Genel algı analizinde kullanılan 2. ve 3.eğitim setlerinin “olumlu sebep” kolonları kullanılarak olumlu sebep analizi için 2. ve 3.eğitim setleri oluşturulmuştur. 2. eğitim setine dair sınıflandırma verileri aşağıdaki gibidir:

Tablo 4. 28. Olumlu algının sebep analizi 2.eğitim seti sınıflandırma verileri

Sebepler	Sebepler başlığı	Sınıflandırma sayısı	Yüzde	Sıra
	Olumsuz sebepler	1894	83,3	1
Olumlu sebepler	Yardıma ihtiyaçları olma / mağdur / mazlum olma durumları	167	7,3	2
	Suriyelilerin iyi özellikleri	75	3,3	3
	Ortak yanlarımız	53	2,3	4
	Suriyelilerin faydalı olma durumu	49	2,2	5
	Suriyelilerin zararlarının olmaması	15	0,7	6
	Suriyeliler ile arkadaş olma durumu	12	0,5	7
	Diğer	10	0,4	8
	Toplam		2275	100

2.eđitim setine karřılık genel algı analizinde kullanılan 2.test seti kullanılmıř ve bu řekilde 2.veri seti oluřturulmuřtur. Oluřturulan veri seti kullanılarak uygulanan modellerin dođruluk oranlarını gsterir tablo ařađıdaki gibidir:

Tablo 4. 29. Olumlu algının sebep analizi 2.veri seti sınıflandırıcı – znitelik ıkarma yntemleri performansları

2.Verdi Seti	3-Gram (Karakter)	2-Gram (Terim)	Stem (Snowball)
K-En Yakın Komřu	64,17	62,19	63,68
Naive Bayes	52,95	44,45	53,3
Karar Ađacı	55,08	55,08	55,08
Derin đrenme	64,5	64,44	62,74
Destek Vektr Makineleri	55,08	55,08	55,08

Tablo 4.29’da grleceđi zere, en yksek performansı Derin đrenme sınıflandırıcısı ve karakter tabanlı 3-Gram znitelik ıkarma ynteminin kullanıldıđı model, 64,5’lik dođruluk oranı ile gstermiřtir. Olumlu nitelik analizinde olduđu gibi, veri setinin daha dengeli bir hal alması yine performansta dřle sonulanmıřtır.

3.eđitim setine dair sınıflandırma verileri ařađıdaki gibidir:

Tablo 4. 30. Olumlu algının sebep analizi 3.eđitim seti sınıflandırma verileri

Sebeb bařlıđı	Sınıflandırma sayısı	Yzde	Sıra	
Olumsuz sebepler	1894	69,7	1	
Yardıma ihtiyaları olma / mađdur / mazlum olma durumları	359	13,2	2	
Olumlu sebepler	Suriyelilerin iyi zellikleri	213	7,8	3
	Suriyelilerin faydalı olma durumu	105	3,9	4
	Ortak yanlarımız	82	3	5
	Suriyeliler ile arkadař olma durumu	25	0,9	6
	Diđer	20	0,7	7

Suriyelilerin zararlarının olmaması	19	0,7	8
Toplam	2717	100	

Oluşturulan 3.eğitim setine karşılık olarak genel algı analizinde kullanılan 3.test seti kullanılmış ve 3.veri seti oluşturulmuştur. Oluşturulan veri seti ile gerçekleştirilen modellerin doğruluk oranlarını gösterir tablosu aşağıdaki gibidir:

Tablo 4. 31. Olumlu algının sebep analizi 3.veri seti sınıflandırıcı – öznitelik çıkarma yöntemleri performansları

3.Verdi Seti	3-Gram (Karakter)	2-Gram (Terim)	Stem (Snowball)
K-En Yakın Komşu	61,88	57,82	61,07
Naive Bayes	44,17	44,72	51,16
Karar Ağacı	42,9	43,05	43,09
Derin Öğrenme	61,56	63,08	62,15
Destek Vektör Makineleri	42,96	42,96	42,96

Tablo 4.31’de görüleceği üzere en yüksek performansı gösteren model, Derin Öğrenme sınıflandırıcısı ve terim tabanlı 2-Gram öznitelik çıkarma yönteminin kullanıldığı model olmuş ve 63,08’lik bir doğruluk oranına ulaşmıştır.

Olumlu sebep analizlerinin tümüne bakılarak en yüksek performans gösteren 1.veri setinin kullanıldığı analiz tercih edilmiş ve eldeki tüm tweet’lerin kullanıldığı olumlu sebep analizine bu analizde en yüksek performansı gösteren – Derin Öğrenme sınıflandırıcısı ve terim tabanlı 2-Gram öznitelik çıkarma yönteminin kullanıldığı – model ile gidilmiştir. Buna göre gerçekleştirilen analiz sonucu aşağıdaki gibidir:

Tablo 4. 32. Olumlu algının sebep analizi tahmin sonuçları

Tahmin edilen tweet sayısı: 12854		Doğruluk: 68,48			
Sebepler	Sebepler	Kodlama sayısı	Yüzde	Sıra	Olumsuz ve nötr'ler çıkarıldığında yüzde
Olumlu sebepler	Yardıma ihtiyaçları olma / mağdur / mazlum olma durumları	520	4	3	51,6
	Suriyelilerin iyi özellikleri	259	2	4	25,7
	Ortak yanlarımız	142	1,1	5	14,1
	Suriyelilerin faydalı olma durumu	39	0,3	6	3,8
	Suriyelilerin zararlarının olmaması	19	0,1	7	1,8
	Diğer	15	0,1	8	1,4
	Suriyeliler ile arkadaş olma durumu	12	0,1	9	1,1
Olumsuz sebepler		9928	77,2	1	
Nötr		1920	14,9	2	
Toplam			100		100

Tablo 4.32’de görüleceği üzere, olumlu sebep analizi test analizinden daha yüksek bir yüzde olan %68,48’lik bir doğruluk oranı ile tamamlanmıştır. Buna göre olumlu sebep belirten tweet’ler arasında en yüksek yüzde %51,6 ile “Yardıma ihtiyaçları olma / mağdur / mazlum olma durumları” sebep başlığına aittir. Onu %25,7 ile “Suriyelilerin iyi özellikleri” ve %14,1 ile “Ortak yanlarımız” sebep başlıkları takip etmektedir.

#### 4.7. Sonuç ve Tartışma

Türkiye’de yaşayan Suriyeli sığınmacılar hakkındaki algıyı Twitter’da 2021 yılı boyunca bu konuda atılan tweet’ler üzerinden ölçmeyi amaçlayan araştırma kapsamında duygu analizi metodu kullanılarak toplam 283 farklı model kurgulanmış ve çalıştırılmıştır. Konu üç farklı boyutta incelenmiştir. Bunlardan ilki, genel algı analizidir. Bu analiz kapsamında Suriyeli sığınmacılar ile ilgili olarak Twitter’daki genel algının ortaya konması amaçlanmıştır. Genel algı analizi için gerçekleştirilen

test analizlerinde iki modelin performans olarak diğerlerinden ayrıştığı görülmüştür. Bunlardan ilki ve en yüksek performans göstereni Naive Bayes sınıflandırıcısının kullanıldığı ve %79,81'lik doğruluk oranına ulaşılan model; diğeri ise Derin Öğrenme sınıflandırıcısının kullanıldığı ve %79,54'lük doğruluk oranına ulaşılan modeldir. Modeller arasındaki 0,34'lük doğruluk oranı farkının görece küçük olması ve Derin Öğrenme sınıflandırıcısının genel algı analizi dışında kalan diğer tüm test analizlerinde en yüksek performansı göstermesi sebebiyle genel algı analizi kapsamında Naive Bayes'in yanı sıra Derin Öğrenme sınıflandırıcısı ile de genel algı analizi gerçekleştirilmiş ve şu sonuçlara ulaşılmıştır:

Naive Bayes sınıflandırıcısının kullanıldığı modele göre, Türkiye'de yaşayan Suriyeli sığınmacılar ile ilgili olarak 2021 yılı boyunca Twitter'da atılan ve görüş belirten tweet'lerin %81,5'i olumsuz, %18,5'i ise olumludur. Derin Öğrenme sınıflandırıcısının kullanıldığı modele göre ise bu oran %86,5 olumsuz, %13,5 olumlu şeklindedir. Buna göre, Türkiye'de yaşayan Suriyeli sığınmacılar ile ilgili olarak en iyi ihtimalle algının %81,5 olumsuz, en kötü ihtimalle ise %86,5 olumsuz olduğu söylenebilir. Her iki sonuç da, Suriyeli sığınmacılar ile ilgili olarak Twitter'daki algının büyük oranda olumsuz olduğunu göstermektedir. Twitter'ın insanların herhangi bir konu hakkındaki görüşlerini belirttiği Türkiye'deki en büyük platform olduğu göz önünde bulundurulacak olursa, bu sonucun Türk halkının Suriyeli sığınmacılara karşı genel algısına işaret etme ihtimalinin gözlerden kaçırılmaması gereken bir ihtimal olduğu ifade edilebilir.

Araştırma kapsamında gerçekleştirilen ikinci analiz, algının niteliği analizidir. Algının niteliği analizi de kendi içerisinde olumlu ve olumsuz algının niteliği şeklinde ikiye ayrılmaktadır. Olumsuz nitelik belirten tweet'lerin ezici bir çoğunluğunun "hoşnutsuzluk" niteliğinde olduğu görülmektedir. Olumlu nitelik belirten tweet'lerin ise 3'te 1'i "sempati besleme" niteliğindedir. Onu, sırasıyla "kabul etme/kabul görme", "sevme", "hak verme", "acıma/üzülme" ve "beğenme" nitelikleri takip etmektedir. Görüldüğü üzere Suriyeli sığınmacılara karşı olumsuz algıya sahip tweet'lerin niteliği büyük oranda "hoşnutsuzluk" şeklinde tarif edilebilecekken, olumlu nitelikler daha dengeli bir dağılım göstermektedir.

Araştırma kapsamında gerçekleştirilen son analiz ise algının sebepleri analizidir. Algının sebepleri analizi de kendi içerisinde olumlu ve olumsuz sebep analizleri şeklinde ikiye ayrılmaktadır. Olumsuz sebep belirten tweet'ler içinde en

yüksek yüzdenin %28,2 ile “Suriyeliler bize zarar veriyor” başlığı olduğu görülmektedir. Bu başlık, “Ekonomik yük/zarar”, “güvenlik sorunu”, “suç işliyorlar”, “rahatsız ediyorlar”, “işsizliğin sebeplerinden biri Suriyeliler” gibi, bu görüşleri dile getirenlerin kendileri için bir zarar olarak algıladıkları alt sebepler içermektedir. Bu başlığı %25,4 ile “Suriyelilerin nüfusu” başlığı takip etmektedir. Bu başlık da “sayıca çok fazlalar” ve “çok fazla çocuk yapıyorlar” alt sebeplerini içermektedir. Bu başlığı %20,3 ile “Suriyelilerin Türkiye’deki hakları” başlığı takip etmektedir. Bu başlık da “sığınma/ikamet hakkı”, “vatandaşlık hakkı” gibi Suriyeli sığınmacılara Türkiye’de tanındığı düşünülen haklardan oluşmaktadır. Bu başlığı da %18 ile “Suriyeliler bizden daha iyi durumdalar / mevcut durumları” başlığı takip etmektedir. Bu başlık da “bizden daha çok hakka / imtiyaza sahipler / daha iyi durumdalar / Suriyelilere daha çok değer veriliyor”, “rahat içindeler” gibi alt sebeplerden oluşmaktadır. Görüldüğü gibi olumsuz sebepler çeşitlilik göstermekle birlikte, en fazla dile getirilen olumsuz sebeplerin Suriyeli sığınmacıların ekonomik yük ya da zarar olarak görülmesi, güvenlik sorunu olarak algılanmaları, sayıca fazla olmaları ve çok çocuk yaptıklarına dair algılar, sığınma, ikamet etme ve vatandaşlık haklarına dair algılar, T.C. vatandaşlarından daha fazla hakka ve imtiyaza sahip olduklarına, daha iyi durumda olduklarına ve Suriyeli sığınmacılara daha çok değer verildiğine dair algılar olduğu görülmektedir.

Olumlu sebeplere bakıldığında ise en büyük yüzdenin %51,6 ile “yardıma ihtiyaçları olma / mağdur / mazlum olma durumları” başlığına ait olduğu görülmektedir. Onu %25,7 ile “Suriyelilerin iyi özellikleri” ve %14,1 ile “ortak yanlarımız” başlıkları takip etmektedir. Görüldüğü üzere Suriyeli sığınmacılar ile ilgili olarak olumlu görüş ve sebep belirten tweet’lerin yarıya yakını sığınmacıların yardıma ihtiyacı olduğu ya da sığınmacılar mağdur ya da mazlum durumda oldukları için olumlu görüş belirtmektedir. Suriyelilerin sahip oldukları iyi özellikler sebebi ile olumlu görüş belirten tweet’ler ise olumlu görüş belirten tweet’lerin 4’te 1’ini oluşturmaktadır. Tweet’lerin içeriğine bakıldığında ise bu iyi özelliklerin çoğu “cinsel olarak beğenme” başlığı altında değerlendirilebilir. Son olarak dile getirilen “ortak yanlarımız” sebebi ise yine tweet’lerin içeriğine bakıldığında daha çok “aynı dine mensup olma”, “insan olma” gibi ortak olduğu düşünülen unsurlara vurgu yapmaktadır.

Türkiye’de yaşayan Suriyeli sığınmacılar hakkındaki algıyı ortaya koyma amacı ile gerçekleştirilen araştırmayı özetlemek gerekirse, algının büyük oranda olumsuz ve “hoşnutsuzluk” niteliğinde olduğu söylenebilir. Ayrıca bu algının başlıca sebeplerini, Suriyeli sığınmacıların kendilerine verdiklerini düşündükleri zarar, sayıca fazla olmaları, Türkiye’de sahip oldukları başta sığınma ve ikamet hakları ile kendilerinden daha iyi durumda olduklarına dair sahip olunan algılar oluşturmaktadır. Olumsuz algıya göre çok daha düşük orandaki olumlu algının niteliği ise çoğunlukla “sempati besleme” ve “kabul etme” niteliğinde olup sebebini çoğunlukla Suriyeli sığınmacıların yardıma ihtiyacı olduğuna dair sahip olunan algı oluşturmaktadır.

Daha önce belirtildiği üzere araştırma, Türkiye’de yaşayan insanların gündemdeki herhangi bir konu hakkında görüş belirtmek için en çok kullandıkları platform olan Twitter kapsamında gerçekleştirilmiştir. Görüşlerin 2021 yılının tamamını yansıtabilecek şekilde dengeli bir şekilde ve olabildiğince çok sayıda toplanmasına dikkat edilmiştir. Tüm görüşler çeşitli ön işlemlerden geçirilerek herhangi bir kampanya veya manipülasyon ürünü olmalarının mümkün olduğunca önüne geçilmiştir. Metot olarak, büyük veri olarak nitelenebilecek bu denli çok sayıda görüşü günümüzde işleyebilecek en verimli yöntem olan duygu analizi kullanılmıştır. Tüm bu sebeplerden ötürü, araştırma sonuçlarının Türkiye’nin genel kanısına dair işaretler taşıma ihtimali gözlerden kaçırılmamalıdır.

#### **4.8. Öneriler**

Aynı veya çalışılmak istenen bir başka konu, yine duygu analizi yöntemi kullanılarak, ancak bu sefer daha fazla doğal dil işleme metotlarına başvurularak gerçekleştirilebilir. Araştırma kapsamında doğal dil işleme yöntemi olarak uygulanan tek işlem Rapidminer programında hali hazırda var olan Türkçe Stem (köke inme) işlemidir. Bunun dışında örneğin cümlenin öğelerinin bulunması, öğelerin anlama katkısının saptanması ve ona göre değerlendirilmesi gibi işlemler gerçekleştirilebilir. Farklı köke inme metotları uygulanarak aralarındaki performanslar kıyaslanabilir. Duygu analizinde Türkçe ifadelerin anlamını daha iyi ortaya çıkarabilecek yöntemler üzerinde çalışılabilir. Analize dahil edilebilecek bir başka husus emoji ve noktalama işaretleri olabilir. Emoji ve noktalama işaretleri anlama anlamlı katkılar sunabilmektedir.

## KAYNAKLAR

- Acun, C. ve Keskin, B. (2016). *PKK'nın Kuzey Suriye Örgütlenmesi PYD-YPG*. İstanbul: SETA Yayınları.
- AFAD (2013). *Türkiye'deki Suriyeli Sığınmacılar, 2013 Saha Araştırması Sonuçları*. Türkiye Cumhuriyeti Başbakanlık, Afet ve Acil Durum Yönetimi Başkanlığı. [https://www.afad.gov.tr/kurumlar/afad.gov.tr/25336/xfiles/1a-Turkiye deki Suriyeli Siginmacilar 2013 1.pdf](https://www.afad.gov.tr/kurumlar/afad.gov.tr/25336/xfiles/1a-Turkiye%20deki%20Suriyeli%20Siginmacilar%202013%201.pdf) adresinden 08.05.2022 tarihinde erişilmiştir.
- AFAD (2017). Türkiye'deki Suriyelilerin Demografik Görünümü, Yaşam Koşulları ve Gelecek Beklentilerine Yönelik Saha Araştırması. Türkiye Cumhuriyeti Başbakanlık, Afet ve Acil Durum Yönetimi Başkanlığı [https://www.afad.gov.tr/kurumlar/afad.gov.tr/25337/xfiles/17a-Turkiye deki Suriyelilerin Demografik Gorunumu Yasam Kosullari ve Gelecek Beklentilerine Yonelik Saha Arastirmasi 2017.pdf](https://www.afad.gov.tr/kurumlar/afad.gov.tr/25337/xfiles/17a-Turkiye%20deki%20Suriyelilerin%20Demografik%20Gorunumu%20Yasam%20Kosullari%20ve%20Gelecek%20Beklentilerine%20Yonelik%20Saha%20Arastirmasi%202017.pdf) adresinden 08.05.2022 tarihinde erişilmiştir.
- Ağdeniz, Ş. (2017). Finansal Raporların Analizinde Metin Madenciliğinin Kullanımı: Borsa İstanbul Şirketlerinin Kurumsal Yönetim Niteliklerinin Tahmini. (*Yayımlanmamış Doktora Tezi*). Eskişehir Osmangazi Üniversitesi/Sosyal Bilimler Enstitüsü, Eskişehir.
- Akpınar, H. (2000). Veri Tabanlarında Bilgi Keşfi ve Veri Madenciliği. *İstanbul Üniversitesi İşletme Fakültesi Dergisi*, 29 (1), 1-22.
- Akpınar, H. (2014). *Data*. İstanbul: Papatya Yayıncılık Eğitim.
- Aksaya, H. (2021). Sözlük Tabanlı Yaklaşım ve Makine Öğrenmesi ile Türkçe Doğal Dil İşleme: Eğitim Kurumlarında Duygu Analizi. (*Yayımlanmamış Doktora Tezi*). İstanbul Üniversitesi/Fen Bilimleri Enstitüsü, İstanbul.
- Aksoy, A. (2016). Türkçe Stopwords. *Github*. <https://github.com/ahmetax/trstop/blob/master/dosyalar/turkce-stop-words> adresinden 23.10.2021 tarihinde erişilmiştir.
- Aksoy, T., Çelik, S., ve Gülseçen, S. (2020). "Data Pre-Processing in Text Mining". In S. Gülseçen, S. Sharma, and E. Akadal (Eds.), *Who Runs the World: Data* (pp.123-145). İstanbul: İstanbul University Press .
- Aksu, M. Ç. (2021). Turizm için Sosyal Medya Temelli Mekansal Karar Destek Sisteminin Geliştirilmesi ve Uygulanması: Artvin İli Örneği. (*Yayımlanmamış Doktora Tezi*). Atatürk Üniversitesi/Sosyal Bilimler Enstitüsü, Erzurum.
- Aktan, E. (2018). Büyük Veri: Uygulama Alanları, Analitiği ve Güvenlik Boyutu. *Bilgi Yönetimi Dergisi*, 1 (1). 1-22.
- Akyol, H. (2019). Demokratikleşme Yaklaşımları Açısından "Arap Baharı" ve Geleceği: Mısır Örneği. (*Yayımlanmamış Doktora Tezi*). Gazi Üniversitesi/Sosyal Bilimler Enstitüsü.
- Albayrak, A. S. (2006). *Uygulamalı Çok Değişkenli İstatistik Teknikleri, 1.Baskı*. Ankara: Baran Ofset.
- Albayrak, M., Topal, K. ve Altıntaş, V. (2017). Sosyal Medya Üzerinde Veri Analizi: Twitter. *Süleyman Demirel Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi*, 22 (Kayfor 15 Özel Sayısı). 1991-1998.
- Altuncu, M. A. (2021). Makine Öğrenmesi ve Derin Öğrenme Yöntemleri Kullanılarak Saldırı Tespit ve Önleme Sistemi Geliştirilmesi. (*Yayımlanmamış Doktora Tezi*). Kocaeli Üniversitesi/Fen Bilimleri Enstitüsü, Kocaeli.

- Amanet, H. (2017). Türkçe Sosyal Medya Metinlerinde Duygu Analizi. (*Yayımlanmamış Yüksek Lisans Tezi*). Karadeniz Teknik Üniversitesi/Fen Bilimleri Enstitüsü, Trabzon.
- Archenaa, J. ve Anita, E. A. M. (2015). A Survey of Big Data Analytics in Healthcare and Government. *Procedia Computer Science*, 50, 408-413.
- Ataman, F. (2020). Büyük Veri Tabanlı Sistemlerde Duygu Analizine Yönelik Bir Uygulama: Avrupa'daki Twitter Kullanıcılarının Mülteci Algısı. (*Yayımlanmamış Doktora Tezi*). Van Yüzüncü Yıl Üniversitesi/Fen Bilimleri Enstitüsü, Van.
- Atılgan, K. Ö. ve Yoğurtcu, H. (2021). Kargo Firması Müşterilerinin Twitter Gönderilerinin Duygu Analizi. *Çağ Üniversitesi Sosyal Bilimler Dergisi*, 18 (1), 31-39.
- Ayan, B., Kuyumcu, B. ve Ceylan, B. (2019). Twitter Üzerindeki İslamofobik Twitlerin Duygu Analizi ile Tespiti. *Gazi Üniversitesi Fen Bilimleri Dergisi*, Part C: Tasarım ve Teknoloji, 7 (2), 495-502.
- Aydın, C. R., Erkan, A., Güngör, T., ve Takçı, H. (2013, Kasım). Sözlük Tabanlı Kavram Madenciliği: Türkçe İçin Bir Uygulama, 30. *Ulusal Bilişim Kurultayı*, Ankara.
- Aykırı (2022). İçişleri Bakanı Süleyman Soylu: "Türkiye'de kendi vatandaşlarımızın suç oranı %2.2, Suriyelilerin 1.3, hemen hemen yarısı kadar. <https://twitter.com/aykirimtr/status/1506019692263055360> adresinden 01.06.2022 tarihinde erişilmiştir.
- Balaban, M. E. ve Kartal, E. (2015). *Veri Madenciliği ve Makine Öğrenmesi: Temel Algoritmaları ve R Dili ile Uygulamaları*, İstanbul: Çağlayan Kitabevi.
- Bagadiya, J. (2022). 430+ Social Media Statistics You Must Know in 2022. *SocialPilot*. <https://www.socialpilot.co/blog/social-media-statistics#fb-usage-stats> adresinden 03.02.2022 tarihinde erişilmiştir.
- Baykal, A. ve Coşku, C. (2009, Şubat). Web Madenciliği Teknikleri. *Akademik Bilişim '09 – XI. Akademik Bilişim Konferansı Bildirileri*, Şanlıurfa.
- BBC (2021). "Türkiye'deki Suriyeliler hakkında güncel bilgiler neler?" haberi. <https://www.bbc.com/turkce/haberler-turkiye-58329307> adresinden 21.02.2022 tarihinde erişilmiştir.
- Budak, İ. (2021). Veri ve Metin Madenciliği ile Hava Yolu İşletmelerinin Sosyal Medya Yorum ve Skorlarının Değerlendirilmesi. (*Yayımlanmamış Doktora Tezi*). Pamukkale Üniversitesi/Sosyal Bilimler Enstitüsü, Denizli.
- Bulur, S. (2018). Suriyelilerin karıştığı suç oranı %1,46'ya düştü. Anadolu Ajansı. <https://www.aa.com.tr/tr/turkiye/suriyelilerin-karistigi-suc-orani-yuzde-1-46ya-dustu/1289461> adresinden 01.06.2022 tarihinde erişilmiştir.
- Burun, E. ve Kanlı, M. (2019). Fuat Oktay: Suriyeli sığınmacılar için harcamamız 40 milyar doları aştı. *Hürriyet*. <https://www.hurriyet.com.tr/gundem/son-dakika-cumhurbaskani-yardimcisi-oktaydan-guvenli-bolge-aciklamasi-41382767> adresinden 21.05.2022 tarihinde erişilmiştir.
- Caro, L. P. (2020). Türk İşgücü Piyasasında Suriyeli Mülteciler. *ILO Türkiye Ofisi*. [https://www.ilo.org/wcmsp5/groups/public/---europe/---ro-geneva/---ilo-ankara/documents/publication/wcms\\_739463.pdf](https://www.ilo.org/wcmsp5/groups/public/---europe/---ro-geneva/---ilo-ankara/documents/publication/wcms_739463.pdf) adresinden 11.05.2022 tarihinde erişilmiştir.
- Cemaloğlu, N. ve Duykuluoğlu, A. (2020). *Sosyal Bilimlerde Veri Madenciliği*. Ankara: Pegem Akademi.

- Ceritoğlu, E., Gürcihan-Yüncüler, H.B., Torun, H. ve Tümen, S. (2017). The Impact of Syrian Refugees on Natives' Labor Market Outcomes in Turkey: Evidence from Quasi-Experimental Design. *T.C. Merkez Bankası Working Paper No: 17/05*.
- Chowdhury, G. G. (2003). Natural Language Processing, *Annual Review of Information Science and Technology*, 37(1), 51-89.
- CNN Türk (2019). *Bakan Pekcan: 15 bin 159 Suriyeli şirket var* <https://www.cnnturk.com/ekonomi/bakan-pekcan-15-bin-159-suriyeli-sirket-var> adresinden 07.06.2021 tarihinde erişilmiştir.
- Çabuk, H., Yüksel, Ç., Mocan, Z., Diri, B. ve Amasyalı, F. (2003, Haziran). Metin Analizi ve Sorgulama. *11.Sinyal İşleme ve İletişim Uygulamaları Kurultayı*, İstanbul, 679-681.
- Çalış, A., Kayapınar, S. ve Çetinyokuş, T. (2014). Veri Madenciliğinde Karar Ağacı Algoritmaları ile Bilgisayar ve İnternet Güvenliği Üzerine Bir Uygulama, *Endüstri Mühendisliği Dergisi*, 25 (3), 2-19.
- Çelik, A. H. (2015). Buazizi'den Rabia'ya, Trablus'tan Şam'a Arap Baharının Sonuçları Üzerine Bir Değerlendirme. *Akademik Yaklaşımlar Dergisi*, 6 (2), 34-53.
- Çelikyay, E. K. (2010). Metin Madenciliği Yöntemiyle Türkçe'de En Sık Kullanılan ve Birbirini Takip Eden Harflerin Analizi ve Birliktelik Kuralları. (*Yayımlanmamış Yüksek Lisans Tezi*). Beykent Üniversitesi/Fen Bilimleri Enstitüsü, İstanbul.
- Çölkesen, R. (2014). *Veri Yapıları ve Algoritmalar, 9.Basım*. İstanbul: Papatya Yayıncılık Eğitim.
- ÇSGB (2020). *Çalışma ve Sosyal Güvenlik Bakanlığı Yabancıların Çalışma İzinleri 2020 Raporu*. <https://www.cs.gb.gov.tr/media/87487/yabanciizin2020.pdf> adresinden 01.12.2021 tarihinde erişilmiştir.
- Dean, B. (2022). WhatsApp 2022 User Statistics: How Many People Use WhatsApp?. *Backlinko*. <https://backlinko.com/whatsapp-users> adresinden 09.02.2022 tarihinde erişilmiştir.
- Değer, N. S. (2017). Sosyal Medya Mesajlarında Veri Madenciliği ile Duygu Analizi. (*Yayımlanmamış Doktora Tezi*). İstanbul Üniversitesi/Sosyal Bilimler Enstitüsü, İstanbul.
- Del Carpio, X. V. ve Wagner, M. (2015). The impact of Syrian refugees on the Turkish labor market. *Policy Research Working Paper 7402*, World Bank Group, Washington, DC.
- Delen, D. ve Crossland, M. D. (2008). Seeding the Survey and Analysis of Research Literature with Text Mining. *Expert Systems with Applications*, 34 (3), 1707-1720.
- Demir, S. (2013). Lingo Algoritmasının Kümelerle İlişkili Dokümanların Belirlenmesi ve Küme Etiketlerinin Çıkarılması Aşamalarının İyileştirilmesi. (*Yayımlanmamış Yüksek Lisans Tezi*). Hacettepe Üniversitesi/Fen Bilimleri Enstitüsü, Ankara.
- Deniz, T. (2013). Arap Baharı ve Türkiye: Siyasi Coğrafya Açısından Bir Değerlendirme. *Doğu Coğrafya Dergisi*, 18 (29), 65-78.
- Deng, S., Sinha, A. P. ve Zhao, H. (2017). Adapting Sentiment Lexicons to Domain-Specific Social Media Texts. *Decision Support Systems*, 94, 65-76.
- DHA (2022). Bakan Soylu, sınır dışı edilen Suriyeli sayısını açıkladı. *Habertürk*. <https://www.haberturk.com/son-dakika-bakan-soylu-sinir-disi-edilen-suriyeli-sayisini-acikladi-3407829> adresinden 01.06.2022 tarihinde erişilmiştir.
- Doğan, R. Ö. (2021). Pankreas Kanseri Dokularının Segmentasyonunda Farklı Derin Öğrenme Tekniklerinin Performanslarının İncelenmesi. (*Yayımlanmamış Doktora Tezi*). Karadeniz Teknik Üniversitesi/Fen Bilimleri Enstitüsü, Trabzon.

- Dondurmacı, G. (2011). Veri Madenciliği'nde Regresyon Ağaçları ile Sınıflandırma ve Bir Uygulama. (*Yayımlanmamış Doktora Tezi*). Mimar Sinan Güzel Sanatlar Üniversitesi/Fen Bilimleri Enstitüsü, İstanbul.
- Durmaz, O. (2011). Metin Sınıflandırmada Boyut Azaltmanın Etkisi. (*Yayımlanmamış Yüksek Lisans Tezi*). Gazi Üniversitesi/Fen Bilimleri Enstitüsü, Ankara.
- Emel, G.G. ve Taşkın, Ç. (2005), Veri Madenciliğinde Karar Ağaçları ve bir Satış Analizi Uygulaması, *Eskişehir Osmangazi Üniversitesi Sosyal Bilimler Dergisi*, 6 (2), 221-239.
- Elkin, S. L., Topal, K. ve Bebek, G. (2017). Network Based Model of Social Media Big Data Predicts Contagious Disease Diffusion. *Information Discovery and Delivery*, 45 (3), 110-120.
- Erdoğan, M. M. (2019). *Türkiye'deki Suriyeli Mülteciler*. Konrad Adenauer Stiftung.
- Erdoğan, M. M. (2020). *Suriyeliler Barometresi 2019: Suriyelilerle Uyum İçerisinde Yaşamın Çerçevesi*. Ankara: Orion Kitabevi.
- Ersöz, F. (2019). *Veri Madenciliği Teknikleri ve Uygulamaları*. Ankara: Seçkin Yayıncılık.
- Esuli, A. ve Sebastiani, F. (2006, Mayıs). Sentiwordnet: A Publicly Available Lexical Resource for Opinion Mining. *Proceedings of the 5th Conference on Language Resources and Evaluation (LREC'06)* (s. 417-422). Genoa, Italy: European Language Resources Association (ELRA).
- Feinerer, I., Hornik, K. ve Meyer, D. (2008), Text Mining Infrastructure in R, *Journal of Statistical Software*, 25 (5), 1-54.
- Gaikwad, S.V, Chaugule, A. ve Pramod P. (2014), Text Mining Methods and Techniques, *International Journal of Computer Applications*, 85 (17), 42-45.
- Gandomi, A. ve Haider, M. (2015). Beyond the Hype: Big Data Concepts, Methods, and Analytics. *International Journal of Information Management*, 35 (2), 137-144.
- Gençler, A. (2020). Türkiye'deki Suriyeli Sığınmacıların Sosyo-Ekonomik Yaşama Etkileri: Fayda Maliyet Ekseninde Bir Bakış. *Sosyal Siyaset Konferansları Dergisi*, 78, 111-145.
- GİB (2022a). Geçici Koruma İstatistikleri. *T.C. İçişleri Bakanlığı Göç İdaresi Başkanlığı*. <https://www.goc.gov.tr/gecici-koruma5638> adresinden 03.05.2022 tarihinde erişilmiştir.
- GİB (2022b). İkamet İzinleri İstatistikleri. *T.C. İçişleri Bakanlığı Göç İdaresi Başkanlığı*. <https://www.goc.gov.tr/ikamet-izinleri> adresinden 03.05.2022 tarihinde erişilmiştir.
- GİB (2022c). Uluslararası Koruma Başvurusu İstatistikleri. *T.C. İçişleri Bakanlığı Göç İdaresi Başkanlığı*. <https://www.goc.gov.tr/uluslararasi-koruma-istatistikler> adresinden 03.05.2022 tarihinde erişilmiştir.
- Goes, P. B. (2014). Big Data and IS Research. *MIS Quarterly*, 38 (3), 3-13.
- Gokte, S. A. (2020). Most Popular Distance Metrics Used in KNN and When to Use Them. *KDnuggets*. <https://www.kdnuggets.com/2020/11/most-popular-distance-metrics-knn.html> adresinden 06.04.2022 tarihinde erişilmiştir.
- Güneş, B. (2018). Libya İç Savaşı ve Kriz Yönetimi. *Güvenlik Bilimleri Dergisi*, 7 (2), 263-291.
- Güran, A., Uysal, M. ve Doğrusöz, Ö. (2014). Destek Vektör Makineleri Parametre Optimizasyonunun Duygu Analizi Üzerindeki Etkisi. *Dokuz Eylül Üniversitesi Mühendislik Fakültesi Fen ve Mühendislik Dergisi*, 16 (48), 86-93.

- Hamde, M. A. (2018). Kurumsal Belgelere (Metin Verilerine) Metin Madenciliği Tekniği ile Erişimin Değerlendirilmesi: Türk Özel Sektörüne Yönelik Bir İnceleme. (Yayımlanmamış Doktora Tezi). İstanbul Üniversitesi/Sosyal Bilimler Enstitüsü, İstanbul.
- Han, J., Kamber, M. ve Pei, J. (2012). *Data Mining: Concepts and Techniques*. Massachusetts: Elsevier.
- Hotho, A., Andreas, N. ve Gerhard, P. (2005). A Brief Survey of Text Mining, *LDV Forum-GLDV Journal for Computational Linguistics and Language Technology*, 20 (1), 24-27.
- Hu, M. ve Liu, B. (2004, Ağustos). Mining and Summarizing Customer Reviews. *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD-2004)* (s. 168-177). Seattle, Washington, USA: ACM New York.
- Hurwitz, J., Nugent, A., Halper, F., Kaufman, M. (2013). *Big Data for Dummies*. New Jersey: John Wiley & Sons.
- HÜNEE, (2019a). *2018 Türkiye Nüfus ve Sağlık Araştırması Suriyeli Göçmen Örnelemi*. Hacettepe Üniversitesi Nüfus Etütleri Enstitüsü, T.C. Cumhurbaşkanlığı Strateji ve Bütçe Başkanlığı ve TÜBİTAK, Ankara, Türkiye.
- HÜNEE, (2019b). *2018 Türkiye Nüfus ve Sağlık Araştırması*. Hacettepe Üniversitesi Nüfus Etütleri Enstitüsü, T.C. Cumhurbaşkanlığı Strateji ve Bütçe Başkanlığı ve TÜBİTAK, Ankara, Türkiye.
- Hürriyet (2020). *Üniversitelerde 27 binden fazla Suriyeli öğrenci var.* <https://www.hurriyet.com.tr/egitim/universitelerde-27-binden-fazla-suriyeli-ogrenci-var-41550989> adresinden 22.05.2022 tarihinde erişilmiştir.
- İçişleri Bakanlığı (2017). *Basın Açıklaması*. <https://www.icisleri.gov.tr/basin-aciklamasi05072017> adresinden 01.06.2022 tarihinde erişilmiştir.
- İnan, O. (2015). Veri Madenciliği Uygulamaları için Veri İndirgeme Algoritmalarının Geliştirilmesi ve Resim Madenciliğine Uygulanması. (Yayımlanmamış Doktora Tezi). Selçuk Üniversitesi/Fen Bilimleri Enstitüsü, Konya.
- İNGEV (2017). Suriyeli Mülteci Hayatlar Monitörü Özet Değerlendirme. İnsani Gelişme Vakfı. <https://ingev.org/wp-content/uploads/2017/07/Multeci-Hayatlar-Monitor%C3%BC.pdf> adresinden 04.06.2022 tarihinde erişilmiştir.
- Johnson, J. (2021). Number of sent and received e-mails per day worldwide from 2017 to 2025. *Statista*. <https://www.statista.com/statistics/456500/daily-number-of-e-mails-worldwide/> adresinden 09.02.2022 tarihinde erişilmiştir.
- Johnson, R. A. ve Wichern, D. W. (2002). *Applied Multivariate Statistical Analysis, Fifth Edition*. New Jersey: Prentice-Hall Inc.
- Kannan, S. ve Gurusamy, V. (2014). Preprocessing Techniques for Text Mining. *International Journal of Computer Science & Communication Networks*, 5(1), 7-16.
- Kantardzic, M. (2011). *Data Mining: Concepts, Models, Methods and Algorithms*. New Jersey: John Wiley & Sons.
- Karadağ, K. (2022). İçişleri Bakanı Soylu: Bursa ve İstanbul'daki olayın faillerini tespit ettik. *Anadolu Ajansı*. <https://www.aa.com.tr/tr/gundem/icisleri-bakani-soylu-bursa-ve-istanbuldaki-olayin-faillerini-tespit-ettik/2570348> adresinden 03.05.2022 tarihinde erişilmiştir.

- Kavzođlu, T. ve ölkesen, İ. (2010). Destek Vektör Makineleri ile Uydu Görüntülerinin Sınıflandırılmasında Kernel Fonksiyonlarının Etkilerinin İncelenmesi. *Harita Dergisi*, 76 (144), 73-82.
- Kızılay (2021). *Türk Kızılay Suriye İnsani Yardım Operasyonu Eylül 2021 Raporu*. <https://www.kizilay.org.tr/Upload/Dokuman/Dosya/9-eylul-2021-suriye-krizi-insani-yardim-operasyonu-raporu-03-11-2021-51443114.pdf> adresinden 21.05.2022 tarihinde erişilmiştir.
- Kızılkaya, Y. M. (2018). Duygu Analizi ve Sosyal Medya Alanında Uygulama. (*Yayımlanmamış Doktora Tezi*). Uludağ Üniversitesi/Sosyal Bilimler Enstitüsü, Bursa.
- Köktürk, M. S. ve Dirsehan, T. (2012). *Veri Madenciliđi ile Pazarlama Etkileşimi*. Ankara: Nobel Akademik Yayıncılık Eğitim Danışmanlık.
- Kuş, O. (2019). Suriyeli Mültecilerin Sosyal Entegrasyonunda Halkla İlişkilerin Rolü. (*Yayımlanmamış Doktora Tezi*). İstanbul Üniversitesi/Sosyal Bilimler Enstitüsü, İstanbul.
- Kuyumcu, M. İ. ve Kösematođlu, H. (2017). The Impacts Of The Syrian Refugees On Turkey's Economy (2011-2016). *Journal of Turkish Social Sciences Research*, 2 (1), 77-93.
- Laney, D. (2001). 3D Data Management: Controlling Data Volume, Velocity and Variety. *META Group Research Note*, 6 (70), 1-3.
- Lantz, B. (2013). *Machine Learning With R*. Birmingham, UK: Packt Publishing.
- Liu, B. (2012). *Sentiment Analysis and Opinion Mining*. Morgan & Claypool Publishers
- Manning, C. D., Raghavan, P. ve Schütze, H. (2009). *An Introduction to Information Retrieval*. Cambridge: Cambridge University Press.
- MEB (2022a). EBA Akademik Destek. *T.C. Milli Eğitim Bakanlığı Yenilik ve Eğitim Teknolojileri Genel Müdürlüğü*. <https://yegitek.meb.gov.tr/www/eba-akademik-destek/icerik/3014#:~:text=Mill%C3%AE%20E%C4%9Fitim%20Bakanl%C4%B1%C4%9F%C4%B1n%C4%B1n%20dijital%20e%C4%9Fitim%20platformu%20EB A'da%2C%2011..ak%C4%B1l%C4%B1%20y%C3%B6nlendirme%20yapan%20bir%20sistem>. adresinden 13.02.2022 tarihinde erişilmiştir.
- MEB (2022b). Yabancı Uyruklu Öğrenci İstatistikleri. *T.C. Milli Eğitim Bakanlığı Hayat Boyu Öğrenme Genel Müdürlüğü, Göç ve Acil Durumlarda Eğitim Daire Başkanlığı*. [https://hbogm.meb.gov.tr/meb\\_iys\\_dosyalar/2022\\_01/26165737\\_goc2022sunu.pdf](https://hbogm.meb.gov.tr/meb_iys_dosyalar/2022_01/26165737_goc2022sunu.pdf) adresinden 08.05.2022 tarihinde erişilmiştir.
- Medhat, W., Hassan, A. ve Korashy, H. (2014). Sentiment Analysis Algorithms and Applications: A Survey. *Ain Shams Engineering Journal*, 5(4), 1093-1113.
- Miah, S. J., Vu, H. Q., Gammack, J. ve McGrath, M. (2017). A Big Data Analytics Method for Tourist Behaviour Analysis. *Information and Management*, 54 (6), 771-785.
- Mijwel, M. M. (2017). Yapay Sinir Ağlar Yapısı ve Fonksiyonu. *Bilgisayar Bilimleri, Bilimleri Koleji*, 1-4.
- Miner, G. Delen, D., Fast, A., Hill, T., Elder, J., ve Nisbet, B. (2012), *Practical Text Mining and Statistical Analysis for Non-Structured Text Data Applications*, Waltham, MA: Academic Press.
- Mujahid, M., Lee, E., Rustam, F., Washington, P. B., Ullah, S., Reshi, A. A. & Ashraf, I. (2021). Sentiment Analysis and Topic Modeling on Tweets about Online Education during COVID-19. *Applied Sciences 2021*, 11 (18), 8438.

- Mülteciler Derneği (2021a). *Avrupa Birliğinin Suriyeliler İçin Türkiye'ye Ödediği Para*. <https://multeciler.org.tr/avrupa-birliginin-suriyeliler-icin-turkiyeye-odedigi-para/> adresinden 21.05.2022 tarihinde erişilmiştir.
- Mülteciler Derneği (2021b). *Yabancılara Yönelik Şartlı Eğitim Yardım Programı*. <https://multeciler.org.tr/turkiye-sartli-egitim-yardimi-programi/> adresinden 22.05.2022 tarihinde erişilmiştir.
- Mülteciler Derneği (2022). *Türkiye'deki Suriyeli Sayısı Nisan 2022*. <https://multeciler.org.tr/turkiyedeki-suriyeli-sayisi/#:~:text=3.754.591&text=G%C3%B6%C3%A7%20%C4%B0daresi%20Ba%C5%9Fkanl%C4%B1%C4%9F%C4%B1'n%C4%B1n%20yay%C4%B1nlanmad%C4%B1%C4%9F%C4%B1,286%20bin%20187%20ki%C5%9Fi%20fazla>. adresinden 04.05.2022 tarihinde erişilmiştir.
- Naik, K. ve Joshi, A. (2017, Şubat). Role of Big Data in Various Sectors. *2017 International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud)*, Tirupur, India, 117-122.
- Narasimhan, R. ve Bhuvaneshwari, T. (2014). Big Data - A Brief Study. *International Journal of Scientific & Engineering Research*, 5 (9), 350-353.
- Nasukawa, T. & Yi, J. (2003, Ekim). Sentiment analysis: Capturing Favorability Using Natural Language Processing. in *Proceedings of the 2nd International Conference on Knowledge Capture (K-CAP 2003)*, Florida, USA.
- Oğuzlar, A. (2003). Veri Ön İşleme. *Erciyes Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi*, 0(21), 67-73.
- Oğuzlar, A. (2011), *Temel Metin Madenciliği*, Bursa: Dora Yayınları.
- Oğuzlar, A. ve Kızılkaya, Y. M. (2019). *Metin Madenciliğinde Duygu Analizi: R Uygulamalı*. Bursa: Dora Basım-Yayım.
- Oğuzlu, T. (2011). Arap Baharı ve Yansımaları. *Ortadoğu Analiz*, 3(36), 8-16.
- Onan, A. (2017). Twitter Mesajları Üzerinde Makine Öğrenmesi Yöntemlerine Dayalı Duygu Analizi. *Yönetim Bilişim Sistemleri Dergisi*, 3 (2), 1-14.
- Orhan, O. ve Gündoğar, S. S. (2015). *Suriyeli Sığınmacıların Türkiye'ye Etkileri - Rapor No:195*. Ankara: ORSAM Yayınları.
- Özaydın, A. ve İlgazi, A. (2019). Doğrudan Yabancı Sermaye Yatırımı Açısından Suriyeli Sığınmacılar. *Gaziantep Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi*, 1 (1), 130-160.
- Özekes, S. (2003). Veri Madenciliği Modelleri ve Uygulama Alanları. *İstanbul Ticaret Üniversitesi Dergisi*, 2 (3), 65-82.
- Özkan, Y. (2013). *Veri Madenciliği Yöntemleri*. İstanbul: Papatya Yayıncılık Eğitim.
- Paulino, A. C., Guimaraes, L. N. F. & Shiguemori, E. H. (2019). Hybrid Adaptive Computational Intelligence-based Multisensor Data Fusion applied to real-time UAV autonomous navigation. *Inteligencia Artificial*, 22 (63), 162-195.
- Pekin, S. (2020). Finansal Performans Tahmininde Metin Madenciliğinin Kullanımı: Bist İmalat Sanayi İşletmelerinde Bir Araştırma. *(Yayımlanmamış Doktora Tezi)*. Anadolu Üniversitesi/Sosyal Bilimler Enstitüsü, Eskişehir.
- Polat, H. (2021). Halkla İlişkiler 2.0 Kapsamında Hedef Belirleme ve Ölçme Sorunsalına Bakış: Alternatif Bir Yöntem Olarak Veri Madenciliğinin Kullanılmasına Yönelik Örnek Bir Uygulama. *(Yayımlanmamış Doktora Tezi)*. Atatürk Üniversitesi/Sosyal Bilimler Enstitüsü, Erzurum.

- Rakipođlu, Z. ve Kara, M. (2022). Bakan Koca: Ayrım yapmaksızın göçmenlerin sađlık ihtiyacına iliřkin ortak bir yol izlemek hepimizin asli görevi. *Anadolu Ajansı*. <https://www.aa.com.tr/tr/saglik/bakan-koca-ayrim-yapmaksizin-gocmenlerin-saglik-ihtiyacina-iliskin-ortak-bir-yol-izlemek-hepimizin-asli-gorevi-/2537978> adresinden 07.05.2022 tarihinde eriřilmiřtir.
- Resmi Gazete (2013). 2013-2014 Eđitim-Öđretim Yılında Yükseköđretim Kurumlarında Cari Hizmet Maliyetlerine Öđrenci Katkısı Olarak Alınacak Katkı Payları ve Öđrenim Ücretlerinin Tespitine Dair Karar, *31 Ađustos 2013 Tarihli ve 28751 Sayılı Resmi Gazete*.
- Resmi Gazete (2016). Geçici Koruma Sađlanan Yabancıların Çalıřma İzinlerine Dair Yönetmelik, *15 Ocak 2016 Tarihli ve 29594 Sayılı Resmi Gazete*.
- Sarika (2022). 84 YouTube Statistics You Can't Ignore in 2022. *Invideo*. <https://invideo.io/blog/youtube-statistics/> adresinden 03.02.2022 tarihinde eriřilmiřtir.
- Savař, S., Topalođlu, N. ve Yılmaz, M. (2012). Veri Madenciliđi ve Türkiye'deki Uygulama Örneklere. *İstanbul Ticaret Üniversitesi Fen Bilimleri Dergisi*, 11 (21), 1-23.
- Sevli, O. (2016). Yapısal Olmayan Verilerin Büyük Veri Analiz Yöntemleri ile İşlenmesi ve Yapısal Olan Verilerle İliřkilendirilmesine Yönelik Bir Platform: Sosyal Medya Temelli Tavsiye Motoru Geliřtirme. (*Yayımlanmamıř Doktora Tezi*). Süleyman Demirel Üniversitesi/Fen Bilimleri Enstitüsü, Isparta.
- Silahtaröđlu, G. (2004). Veri Madenciliđinde Kümeleme Analizi ve Öđretim Başarısının Deđerlemesine İliřkin Bir Uygulama. (*Yayımlanmamıř Doktora Tezi*). İstanbul Üniversitesi/Sosyal Bilimler Enstitüsü, İstanbul.
- Silahtaröđlu, G. (2013). *Veri Madenciliđi: Kavram ve Algoritmaları*. İstanbul: Papatya Yayıncılık Eđitim.
- SOHR (2022). Syrian Revolution 11 years on - SOHR documents by names nearly 161,000 civilian deaths, including 40,500 children and women. *Suriye İnsan Hakları Gözlemevi*. <https://www.syriahr.com/en/243125/> adresinden 02.05.2022 tarihinde eriřilmiřtir.
- Srivastava, U. ve Gopalkrishnan, S. (2015). Impact of Big Data Analytics on Banking Sector: Learning for Indian Banks. *Procedia Computer Science*, 50, 643-652.
- Statista (2022). *Volume of data/information created, captured, copied, and consumed worldwide from 2010 to 2025*. <https://www.statista.com/statistics/871513/worldwide-data-created/> adresinden 04.05.2022 tarihinde eriřilmiřtir.
- Sumathy, K. L. ve Chidambaram, M. (2013). Text Mining Concepts, Applications, Tools and Issues: An Overview, *International Journal of Computer Applications*, 80 (4), 28-30.
- Suriye Gündemi (2022). *Suriye Genel Durum Haritası*. <https://www.suriyegundemi.com/suriye-genel-durum-haritasi-23-05-2022-1> adresinden 02.06.2022 tarihinde eriřilmiřtir.
- řeker, ř. E. (2013). *İř Zekası ve Veri Madenciliđi: Weka ile*. İstanbul: Cinius Yayınları.
- řeker, ř. E. (2015). Dođal Dil İşleme. *YBS Ansiklopedi*, 2 (4). 14-22.
- řeker, ř. E. (2016). Duygu Analizi (Sentiment Analysis). *YBS Ansiklopedi*, 3 (3). 21-36.
- řentürk, C. (2020). *Türkiye Şehirlerinde Suriyeli Göçmenler - ZfTI-PolicyPaper Nr. 8*. Essen: Türkiye ve Uyum Arařtırmaları Merkezi Vakfı.

- TDK (2022). *Türk Dil Kurumu Sözlükleri*. <https://sozluk.gov.tr/> adresinden 09.03.2022 tarihinde erişilmiştir.
- TEPAV & EBRD (2018). Syrian Entrepreneurship and Refugee Start-ups in Turkey: Leveraging the Turkish Experience Final Report – 2018. *Türkiye Ekonomi Politikaları Araştırma Vakfı & European Bank for Reconstruction and Development*. [https://www.tepav.org.tr/upload/files/1566830992-6.TEPAV and EBRD Syrian Entrepreneurship and Refugee Start ups in Turkey Lever....pdf](https://www.tepav.org.tr/upload/files/1566830992-6.TEPAV%20and%20EBRD%20Syrian%20Entrepreneurship%20and%20Refugee%20Start%20ups%20in%20Turkey%20Lever...pdf) adresinden 25.05.2022 tarihinde erişilmiştir.
- TEPAV (2019a). Yabancı Sermayeli Şirketler Bülteni – Suriye Sermayeli Şirketler. *Türkiye Ekonomi Politikaları Araştırma Vakfı*. [https://www.tepav.org.tr/upload/files/1551075334-6.TEPAV Suriye Sermayeli Sirketler Bulteni Aralik 2018.pdf](https://www.tepav.org.tr/upload/files/1551075334-6.TEPAV%20Suriye%20Sermayeli%20Sirketler%20Bulteni%20Aralik%202018.pdf) adresinden 25.05.2022 tarihinde erişilmiştir.
- TEPAV (2019b). *Türkiye’de Suriyeli Girişimciliği ve Mülteci İşletmeleri: Türkiye Tecrübesinden Nasıl Faydalanılabilir?* <https://www.tepav.org.tr/tr/haberler/s/10024> adresinden 25.06.2022 tarihinde erişilmiştir.
- Tınas, M. (2020). Yemen İç Savaşı’nda Husiler: Vekalet İlişkinin Karmaşık Yapısı. *Eskişehir Osmangazi Üniversitesi Sosyal Bilimler Dergisi*, 21 (1), 117-136.
- Topaçan, Ü. (2016). Sosyal Medya Paylaşımlarında Duygu Analizi: Makine Öğrenimi Yaklaşımı Üzerine Bir Araştırma. (*Yayımlanmamış Doktora Tezi*). Marmara Üniversitesi/Sosyal Bilimler Enstitüsü, İstanbul.
- Topuz, O. (2019). Suriyeli Mültecilerin Türkiye Ekonomisine Etkisi. (*Yayımlanmamış Yüksek Lisans Tezi*). Dokuz Eylül Üniversitesi/Sosyal Bilimler Enstitüsü, İzmir.
- TRT Haber (2019). *Cumhurbaşkanı Erdoğan: Başkaları harekete geçmezken Türkiye adım atıyor*. <https://www.trthaber.com/haber/gundem/cumhurbaskani-erdogan-baskalari-harekete-gecmezken-turkiye-adim-atiyor-435861.html> adresinden 28.05.2022 tarihinde erişilmiştir.
- Tunalı, V. (2011). Metin Madenciliği İçin İyileştirilmiş Bir Kümeleme Yapısının Tasarımı ve Uygulaması. (*Yayımlanmamış Doktora Tezi*). Marmara Üniversitesi/Fen Bilimleri Enstitüsü, İstanbul.
- TÜBA (2022). *Türkçe Bilim Terimleri Sözlüğü*. Türkiye Bilimler Akademisi. <http://terim.tuba.gov.tr/> adresinden 05.02.2022 tarihinde erişilmiştir.
- TÜİK (2022a). Adrese Dayalı Nüfus Kayıt Sistemi Sonuçları, 2021. *Türkiye İstatistik Kurumu*. <https://data.tuik.gov.tr/Bulten/Index?p=Adrese-Dayali-Nufus-Kayit-Sistemi-Sonuclari-2021-45500#:~:text=T%C3%9C%C4%B0K%20Kurumsal&text=T%C3%BCrkiye'de%20ikamet%20eden%20n%C3%BCfus,252%20bin%20172%20ki%C5%9Fi%20oldu>.
- TÜİK (2022b). Doğum İstatistikleri, 2021. *Türkiye İstatistik Kurumu*. <https://data.tuik.gov.tr/Bulten/Index?p=Birth-Statistics-2021-45547> adresinden 16.05.2022 tarihinde erişilmiştir.
- Tümen, S. (2016). The Economic Impact of Syrian Refugees on Host Countries: Quasi-Experimental Evidence from Turkey, *American Economic Review*, 106 (5), 456-460.
- UNHCR (2019). Update: Durable Solutions for Syrian Refugees. *United Nations High Commissioner for Refugees*. <https://data2.unhcr.org/en/documents/download/70892> adresinden 25.05.2022 tarihinde erişilmiştir.
- UNHCR (2020a). Global Trends – Forced Displacement In 2020. *United Nations High Commissioner for Refugees*. <https://www.unhcr.org/60b638e37.pdf> adresinden 02.05.2022 tarihinde erişilmiştir.

- UNHCR (2020b). Syria 3 RP Regional Strategic Overview 2021-2022. *United Nations High Commissioner for Refugees*. <https://reporting.unhcr.org/sites/default/files/SYRIA%203RP%20REGIONAL%20STRATEGIC%20OVERVIEW%202021-2022.pdf> adresinden 18.05.2022 tarihinde erişilmiştir.
- UNHCR (2020c). Türkiye Ülke Bölümü – Suriye Krizine Müdahaleye Yönelik Bölgesel Mülteci ve Dayanıklılık Planı 2020 – 2021. *Dünya Yerel Yönetim ve Demokrasi Akademisi Vakfı*. <https://www.wald.org.tr/Uploads/Pages/Module/EMW4D4KG202R1FZ595062RC6PI9O24.pdf> adresinden 22.05.2022 tarihinde erişilmiştir.
- UNHCR (2022a). UNHCR Turkey: Provincial Breakdown Syrian Refugees in Turkey - February 2022. *Birleşmiş Milletler Mülteciler Yüksek Komiserliği*. <https://data2.unhcr.org/en/documents/details/90912> adresinden 07.05.2022 tarihinde erişilmiştir.
- UNHCR (2022b). Syria 3 RP Regional Strategic Overview 2022. *United Nations High Commissioner for Refugees*. [https://www.3rpsyriacrisis.org/wp-content/uploads/2022/05/RSO\\_8thMay2022.pdf](https://www.3rpsyriacrisis.org/wp-content/uploads/2022/05/RSO_8thMay2022.pdf) adresinden 22.05.2022 tarihinde erişilmiştir.
- Wan, J., Tang, S., Li, D., Wang, S., Liu, C., Abbas, H. ve Vasilakos, A. V. (2017). A Manufacturing Big Data Solution for Active Preventive Maintenance. *IEEE Transactions on Industrial Informatics*, 13 (4), 2039-2047.
- We Are Social & Hootsuite (2022). *Digital 2022 Global Overview Report*. <https://wearesocial.com/uk/blog/2022/01/digital-2022-another-year-of-bumper-growth-2/> adresinden 03.02.2022 tarihinde erişilmiştir.
- Williams, G. (2011). *Data Mining with Rattle and R: The Art of Excavating Data for Knowledge Discovery*. New York: Springer Science + Business Media.
- Yağcı, G. (2018). *Türkiye’de işyeri açan Suriyeliler 100 bin kişiye istihdam sağlıyor*. <https://tr.euronews.com/2018/09/26/turkiye-de-isyeri-acan-suriyeliler-100-bin-kisiye-istihdam-sagliyor> adresinden 22.05.2022 tarihinde erişilmiştir.
- Yaşar, N. T. (2013). Geçmişin Gölgesinde Mısır’da Darbe. *Orsam (Ortadoğu Stratejik Araştırmalar Merkezi)*, Rapor No: 168.
- Yavuzkılıç, S. (2022). Görüntü Manipülasyonlarının Derin Öğrenme Yaklaşımı ile Belirlenmesi. *(Yayımlanmamış Doktora Tezi)*. Fırat Üniversitesi/Fen Bilimleri Enstitüsü, Elazığ.
- Yılmaz, M. E. (2012). Kaddafi Sonrası Libya’da Siyasal Dönüşüm. *Akademik Orta Doğu*, 7 (1), 1-13.
- Yu, S., Yang, D. ve Feng, X. (2017, 09-10 Ekim). A Big Data Analysis Method for Online Education. *2017 10th International Conference on Intelligent Computation Technology and Automation (ICICTA)*, Changsha, China, 291-294.
- YBYS (2022). Yükseköğretim Bilgi Yönetimi Sistemi Uyuşuna Göre Öğrenci Sayıları. *Yükseköğretim Kurulu*. <https://istatistik.yok.gov.tr/> adresinden 08.05.2022 tarihinde erişilmiştir.
- YTB (2016). Suriyeli Öğrenciler İçin Destek Bursları. *Yurtdışı Türkler ve Akraba Topluluklar Başkanlığı*. <https://www.ytb.gov.tr/duyurular/suriyeli-ogrenciler-icin-destek-burslari-2> adresinden 21.05.2022 tarihinde erişilmiştir.

## ÖZ GEÇMİŞ

İlhan PARLAK, Marmara Üniversitesi Teknik Eğitim Fakültesi Elektronik ve Haberleşme Öğretmenliği Lisans programından 2007 yılında mezun oldu. 2008 yılında Kocaeli/Darıca Denizyıldızları Mesleki ve Teknik Anadolu Lisesi'ne Radyo-TV Öğretmeni olarak atandı ve burada 5 yıl görev yaptı. 2013 yılında Giresun Fatih Mesleki ve Teknik Anadolu Lisesi'ne atandı ve burada da 5,5 yıl görev yaptı. 2014 yılında başladığı Anadolu Üniversitesi İktisat Fakültesi Uluslararası İlişkiler Lisans programından 2017 yılında mezun oldu. Halen Radyo-TV öğretmeni olarak görev yaptığı Samsun/İlkadım Piri Reis Mesleki ve Teknik Anadolu Lisesi'ne 2018 yılında atandı. 2019 yılında Ondokuz Mayıs Üniversitesi Lisansüstü Eğitim Enstitüsü Medya ve İletişim Bilimleri Yüksek Lisans programına başlayan İlhan PARLAK, iyi derecede İngilizce bilmektedir. Temel ilgi alanları, siyaset, sinema, futbol, seyahat, doğa ve teknolojidir (04.10.2022).

### İletişim Bilgileri

ORCID ID : 0000-0002-1029-4431

### Yayımlar:

1. Parlak, İ., Çakın, Ö. ve Kaya, S. (2022). Sosyal Medyada Sığınmacı Algısı: Suriyeli Sığınmacıların Türkçe Twitter Hesaplarında Görünümü, Gümüşhane Üniversitesi İletişim Fakültesi Dergisi, 10 (2), 948-983.