
GAZİ ÜNİVERSİTESİ
Gazi
1926

TWITTER ÜZERİNDEKİ İSLAMOFOBİK TWEETLERİN DUYGU ANALİZİ İLE TESPİTİ

Buğra AYAN

YÜKSEK LİSANS TEZİ
BİLİŞİM SİSTEMLERİ ANA BİLİM DALI

GAZİ ÜNİVERSİTESİ
BİLİŞİM ENSTİTÜSÜ

OCAK 2020

Buğra AYAN tarafından hazırlanan "TWITTER ÜZERİNDEKİ İSLAMOFOBİK TWEETLERİN DUYGU ANALİZİ İLE TESPİTİ" adlı tez çalışması aşağıdaki jüri tarafından OY BİRLİĞİ ile Gazi Üniversitesi Bilişim Enstitüsü Bilişim Sistemleri Ana Bilim Dalında Yüksek Lisans Tezi olarak kabul edilmiştir.

Danışman: Doç. Dr. Bünyamin CİYLAN
Teknoloji Fakültesi, Gazi Üniversitesi
Bu tezin, kapsam ve kalite olarak Yüksek Lisans Tezi olduğunu onaylıyorum/onaylamıyorum. ...……

Başkan: Prof. Dr. Şeref SAĞIROĞLU

Teknoloji Fakültesi, Gazi Üniversitesi

Bu tezin,kapsam ve kalite olarak Yüksek Lisans Tezi olduğunu onaylıyorum/onaylamıyorum. ...……

Üye: Doç. Dr. Harun ARTUNER

Teknoloji Fakültesi, Gazi Üniversitesi

Bu tezin,kapsam ve kalite olarak Yüksek Lisans Tezi olduğunu onaylıyorum/onaylamıyorum. ...……

Tez Savunma Tarihi: 03/01/2020

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

...……..…….

Doç. Dr. Aslıhan TÜFEKCİ

Bilişim Enstitüsü Müdürü

ETİK BEYAN

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

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

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

İmza
Buğra AYAN
.. /../….

TWITTER ÜZERİNDEKİ İSLAMOFOBİK TWETLERİN DUYGU ANALİZİ İLE TESPİTİ
(Yüksek Lisans Tezi)

Buğra AYAN

GAZİ ÜNİVERSİTESİ
BİLİŞİM ENSTİTÜSÜ

Ocak 2020

ÖZET

Özellikle sosyal medya ağlarındaki büyük yükseliş, internet kullanıcıları tarafından oluşturulan web içeriğinin analiz edilmesi üzerine yapılan araştırmalarında artmasına neden olmuştur. Araştırmalar, insanların nefret içerikli, rahatsız edici yorumlar yapmak için popüler sosyal ağları, özellikle Twitter platformunu kullanabildiğini ve bu tür içeriklerin belli bir toplumsal grubu hedef alabildiğini göstermektedir. Özellikle son yıllarda İslamofobia konusunda yapılan çalışmalarda, islamofobik tweetlerin sınıflandırması için tweet içerisinde "hateislam" gibi anahtar kelimelerin geçip geçmediğine bakıldığı görülmektedir. Bununla birlikte, İslamofobi söyleminin çok yönlü doğası, kavramsal arka planı düşünüldüğünde, sadece önceden belirlenen anahtar kelimelerin taranmasına göre yapılacak bir sınıflandırmanın yeterli doğrulukta sonuç vermeyeceği düşünülmüştür. Bu tez çalışmasında islamofobik tweetlerin sınıflandırılması için yapay zeka tekniklerinden gözetimli makine öğrenmesi yöntemi kullanılmıştır. Yapay sinir ağının eğitimi ve test işlemlerinde kullanılmak üzere bir veri seti oluşturulmuştur. Bu amaçla twitterin arama kısmı üzerinden "Islam", "Muslim" gibi kelimelerin geçtiği İngilizce dilinde yazılmış 290.000 tweet, Twitter API kullanılarak geliştirilen yazılımla elde edilmiştir. Çeşitli ön işleme adımlarının ardından kalan 162.000 tweet, beş kişilik bir ekip tarafından İslamofobik ve İslamofobik değil şeklinde işaretlenmiştir. İşaretlemenin ardından tweetlerin %80'i eğitim ve %20'si test amaçlı olarak ayrılmıştır. Elde edilen veri seti Bayes Regresyonu, Ridge Regresyonu ve derin öğrenme modeli olmak üzere 3 farklı modele uygulanmıştır. Testlerde her üç model içinde %95'in üzerinde doğruluğa ulaşılmış olmasına rağmen veri seti içerisinde olmayan 100 yeni tweet ile yapılan deneysel çalışmada bu oranların nispeten düştüğü gözlemlenmiştir.

Bilim Kodu : 92429

Anahtar Kelimeler: Twitter, islamofobi, makine öğrenmesi

Sayfa Adedi : 81

Danışman : Doç. Dr. Bünyamin CİYLAN

DETECTION OF ISLAMOPHOBIC TWEETS ON TWITTER USING SENTIMENT ANALYSIS
(M.Sc. Thesis)

BUĞRA AYAN

GAZI UNIVERSITY
INFORMATICS INSTITUTE
January 2020

ABSTRACT

Especially, the great rise in social networks has caused an increase in the reseaches on the analysis of web content created by internet users. Research shows that people can use popular social networks, especially the Twitter platform, in order to make hateful, offensive comments, and it is shown that such content can target a particular social group. Especially, in recent years, it is observed whether keywords such as 'hateislam' are included in an tweet in order to classify the Islamophobic tweets in studies on Islamophobia. Besides, consedering the versatile nature and conceptual background of Islamophobia, it is thought that a classification based on the search of predetermined keywords will not yield sufficient accuracy. In this thesis study, machine learning method, which is one of the artificial intelligence techniques, is used to classify islamophobic tweets. A data set has been created to be used in training and testing operations of the artificial neural network. For this purpose, 290,000 tweets written in the English language in which words such as “Islam” and “Muslim” pass through the search section of the Twitter, are obtained with the software developed using the Twitter API. The remaining 162,000 tweets after various preprocessing steps have been marked by a team of five as Islamophobic and non-Islamophobic. 80% of the tweets are reserved for training and 20% for testing purposes after marking. The data set obtained was applied to 3 different models: Bayes Regression, Ridge Regression and deep learning model. Although more than 95% accuracy was achieved in all three models in the tests, it was observed that these rates decreased relatively in the experimental study conducted with 100 new tweets that are not included in the data set.

Science Code: 92429

Key Words : Twitter, islamophobia , machine learning

Page Number : 81

Supervisor : Assoc. Prof. Dr. Bünyamin CİYLAN

TEŞEKKÜR

Çalışmalarım boyunca değerli yardım ve katkılarıyla beni yönlendiren saygıdeğer hocam Doç. Dr. Bünyamin CİYLAN'a, kıymetli tecrübelerinden yararlandığım Cumhurbaşkanı Özel Kalem Müdürü Dr. Hasan DOĞAN'a, Cumhurbaşkanı Başdanışmanı Dr. M. Mücahit KÜÇÜKYILMAZ'a, Cumhurbaşkanlığı Sağlık Politika Kurulu Üyesi Prof. Dr. Necdet ÜNÜVAR'a, Diyanet İşleri Başkan Yardımcısı Selim ARGUN'a, değerli bilgilerini esirgemeyen Dr. Fatma ÖZTÜRK'e ve her zaman her durumda yanımda olan aileme en içten dileklerimle teşekkür ederim.

İÇİNDEKİLER

Sayfa

Sayfa

ÇİZELGELERİN LİSTESİ

Çizelge **Sayfa**

ŞEKİLLERİN LİSTESİ

Şekil **Sayfa**

RESİMLERİN LİSTESİ

Resim **Sayfa**

SİMGELER VE KISALTMALAR

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

Kısatmalar	Açıklama
AI	Yapay zeka (Artificial intelligence)
API	Uygulama programlama arayüzü
FN	Yanlış negatif (False negative)
FP	Yanlış pozitif (False positive)
F1	F1-skoru (F1-score)
GİF	Grafik değişim biçimi **(**Graphics interchange format**)**
IDF	Ters terim frekansı (Inverse document frequency)
ML	Makine öğrenmesi (Machine learning)
NLP	Doğal dil işleme (Natural language processing)
NLTK	Doğal dil işleme araç seti (Natural language toolkit)
TF	Terim frekansı (Term frequency)
TN	Doğru negatif (True negative)
TP	Doğru pozitif (True positive)

1. GİRİŞ

İnternet kullanımının 2000'li yılların başından itibaren gün geçtikçe artmasının en önemli sonuçlarından biri web 2.0 araçları olarak tanımlanan uygulamaların ortaya çıkmasını sağlamasıdır [1]. İkinci nesil internet hizmetleri olarak tanımlanan bu sitelerde kullanıcılara kendi içeriklerini sağlama imkânı tanınmıştır [2]. Bu durum kullanıcıların etkileşimlerinin artması, paylaşım kültürünün ortaya çıkması gibi bazı olumlu neticeler doğururken; dijital nefret söylemi, siber terörizm, siber zorbalık gibi bazı olumsuz neticeleri de ortaya çıkarmıştır [3]. İnternet ortamındaki nefret söylemini de içinde barındıran bu durumlar ile ilk etapta makine öğrenmesi yaklaşımı kullanmayan geleneksel yollar incelenmiştir. Fakat Twitter'da bulunan verinin hacminin devasa boyutlarda olması, nefretin kavramsal arkaplanıın karmaşıklığı bu nefret söyleminin elle tespitini imkânsız kılmıştır. Teknoloji altyapısının gelişmesi, veri biliminde yaşanan gelişmeler ile veri analizlerine dayalı yöntemler bu sorunların çözümünde gündeme gelmeye başlamıştır [4]. Örneğin Thomas Davidson ve arkadaşları tarafından 2017 yılında yapılan bir çalışmada Twitter üzerindeki 33 bin kullanıcıdan toplanan 85 milyon Tweet ile nefret söylemi ve saldırgan dil otomatik olarak %90 doğrulukla tespit edilmiştir [5]. Kelly Reynolds (2011) tarafından yapılan bir çalışmada ise internetteki bir başka problem olan siber zorbalığı makine öğrenmesi algoritmalarıyla %78,5'lik bir doğrulukla tespit edilmiştir [6].

Nefret söylemi, siber zorbalık gibi konuların yanında; terörist grupların dijitaldeki hareketlerinin tespitinde de makine öğrenmesi kullanılabilmektedir. Michael Ashcroft ve arkadaşları tarafından yapılan bir çalışmada, IŞID gibi radikal terör gruplarından yayılan mesajlar makine öğrenmesi kullanılarak %98 doğrulukla tespit edilmiştir [7].

Bu problemlerin yanında İnternette yer alan bir başka sorun da Müslümanlara karşı olan kaygı ve nefret söylemi olan İslamofobidir. İslamofobi Türkiye'de gerek kamuda gerek sivil toplum tarafındaki birçok kurumun gündemindeki önemli başlıklardan biridir. İnternetteki İslam veya Müslümanlara yönelik nefretin hızlı bir şekilde tespiti fiziksel vakaları önleme noktasında da faydalı olabilmektedir. Guardian tarafından 2018 yılında yayınlanan bir rapor, İngiltere'de yaşayan Müslümanlara karşı yapılan fiziksel saldırılar ile dijital ortamdaki saldırılar arasında bir paralellik olduğunu göstermektedir. 2016 yılında fiziksel ortamda gerçekleşen saldırı sayısı 642 iken 2017 yılında bu sayı 839'a çıkmıştır. Benzer şekilde dijital ortamda tespit edilen vaka sayısı da 311'den 362'ye yükselmiştir [8].

İngiltere'nin yanında ABD'de de güncel vakalar yaşanmaktadır. ABD Kongresi'ne giren ilk iki Müslüman kadından biri olan İlhan Omar'a karşı Twitter üzerinden yürütülen Müslüman karşıtı kampanyalar, Foreign Policy dergisinin gündemine gelmiş ve 2018 yılında yapılan araştırmaya göre Amerika Birleşik Devletleri'ndeki orta sınıfta İslamofobi'nin arttığını gündeme getirmiştir [9].

Benzer şekilde İslamofobi konusu her geçen gün dünyanın daha çok gündemine aldığı bir problem olarak karşımıza çıkmaktadır. Bu konuda Google Trends hizmetinin verdiği istatistikler ilgideki artışı göstermektedir [10].

Makine öğrenmesi konusunda olduğu gibi İslamofobi konusunda da akademik çalışmalar yapılmaktadır. Örneğin Imran Awan tarafından 2014 yılında yapılan çalışmada sosyal medyadaki İslamofobi verisi makine öğrenmesi yöntemleri kullanılmadan, elle incelenmiştir [11]. 2015 yılında Paris'te yaşanan terör saldırısı sonrasında Walid Magdy ve arkadaşları Twitter üzerindeki İslam karşıtlığını ölçümlemek için bir çalışma yapmışlardır. Bu çalışmada saldırıyı takip eden 50 saat içerisinde atılan 900 bin tweet üzerinden yapılan analiz makine öğrenmesi kullanılmadan geleneksel yolla yapılmıştır [12].

Gerek İslamofobi konusunda gerek makine öğrenmesi üzerine birçok çalışma olmasına rağmen, İslamofobi konusunu ele alan, özellikle de bu konudaki tabloyu bir skorlama sistemi ile ortaya koyulan çalışmalara henüz rastlanmamıştır. Çalışma kapsamında farklı makine öğrenmesi algoritmalarıyla, Twitter üzerindeki İslamofobi konulu nefret söyleminin tespiti üzerine bir çalışma gerçekleştirilmiştir.

Bu çalışmada Twitter uygulamasındaki verilerin incelenmesi yoluyla makine öğrenmesi algoritmalarının İslamofobik tweetlerin tespiti konusundaki başarısı araştırılmıştır ve birbiri ile kıyaslanmıştır. Çalışmada;

- İslamofobi ile ilgili tweeetlerin genel yapısı nasıldır?
- İslamofobi konusundaki tweetlerle ilgili olarak bir makine öğrenmesi çalışması yapılabilir mi?
- İslamofobi konusunda hangi makine öğrenme algoritmaları daha başarılı sonuçlar vermektedir?

sorularına cevap bulunmaya çalışılmıştır.

Tez; özet, giriş, literatür, metod materyali, İslamofobinin Makine Öğrenmesi ile Tespiti, Deneysel çalışma ile Sonuç ve Öneriler kısımlarından oluşmaktadır.

Literatür; nefret söylemi üzerine yapılan makine öğrenmesi çalışmalarını ve Islamofobi konusunda makine öğrenmesi kullanılmadan yapılan çalışmaları içermektedir.

Metod materyali kısmında ilk olarak İslamofobi kavramının nasıl geliştiğine değinilmiştir. Kavramın ilk olarak ne zaman gündeme geldiği, hangi olaylarla kullanımında hızlı bir artış yaşandığı, kapsadığı alanlar ve kapsamadığı alanlar da bu bölümde ele alınmıştır. Ardından sosyal ağ platformu Twitter'in gelişim süreci, sahip olduğu ve olmadığı teknik kabiliyetlere yer verilmiştir. Twitter'da çalışma yaparken diğer platformlara karşı sahip olunan kolaylıklar ve zorluklara da değinilmiştir. Sonrasında Nefret Söylemi konusunun gerek fiziksel ortamda gerek dijital ortamdaki gelişiminden bahsedilmiştir. Bu bölüm konunun sosyolojik derinliğine inmeden daha pratik örnekler ve akademik çalışmalar üzerinden ele alınmıştır. Nefret söyleminin ardından duygu analizi çalışmalarının nasıl ortaya çıktığı, gelinen nokta akademik örneklerle incelenmiş ve son olarak ise doğal dil işlemenin tarihsel süreci, kullanılan algoritmalar, geliştirme sırasında yaşanan zorluklara değinilecek ve bu parantez içerisinde metin madenciliğine de değinilmiştir.

İslamofobinin makine öğrenmesi ile tespiti kısmında ilk olarak problemin tanımı yapılmış ardından bu problemi çözmek için kullanılabilecek sistem açıklanmıştır. Sistemin detaylarının belirtilmesinin ardından Metodlar ve Araçlar kısmında sistemi inşa etmek için kullanılabilecek metodlar, algoritmalar, geliştirme ortamları ve diğer araçlar tanıtılmıştır. Bu bölümü verinin temini ve işaretlenmesinin nasıl olacağı, bahsedilen metodlar ve araçlar ile sistem mimarisinin nasıl yapılacağı takip etmiştir.

Deneysel çalışma iki kısımdan oluşmaktadır. İlk kısımda veri seti ile ilgili detaylar verilmiş ve veri önişleme adımları detaylıca anlatılmıştır. Ön işleme sonrasında verinin vektör haline getirilmesi ve eğitilmesinden bahsedilmiştir. Burada Ridge modeli, Naive Bayes modeli ve Derin Öğrenme modelinin ne olduğu anlatılmış ve eğitim sonuçları da karşılaştırılmalı olarak tablo olarak sunulmuştur. Deneysel çalışmanın ikinci bölümünde ise veri setinde olmayan 100 tweet üzerinden modeller test edilmiş ve ulaşılan oranları tabloda verilmiştir. Bölümün nihayetinde ise bu sonuçlar üzerinde tartışma yapılmıştır.

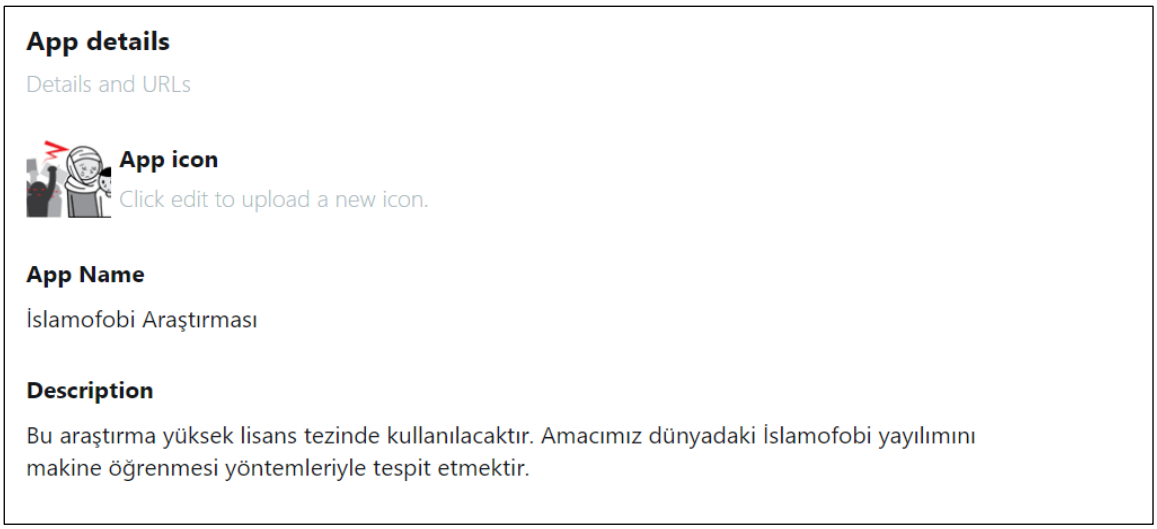
2. TWITTER

Twitter 2006 yılında kurulmuş anlık ve hızlı paylaşım yapmayı sağlayan bir mikroblog hizmetidir. Twitter insanların ortak ilgi alanları ile diğer insanlarla etkileşim kurduğu sosyal bir ağ sitesidir. Twitter' daki ilk tweet kurucuları kabul edilen Jack Dorsey ve Biz Stone tarafından atılmıştır.

Twitter'ın asıl amacı kullanıcılarına kısa ve anlamlı mesajlar kullanarak kendilerini ifade etmeleridir. 2017 yılından önce 140 karakter kullanımına olanak sağlayan twitter 2017'den sonra 280 karaktere olanak sağlamıştır. 2011 yılının Nisan ayının 25' den itibaren Türkçe dil seçeneğini eklemiştir.

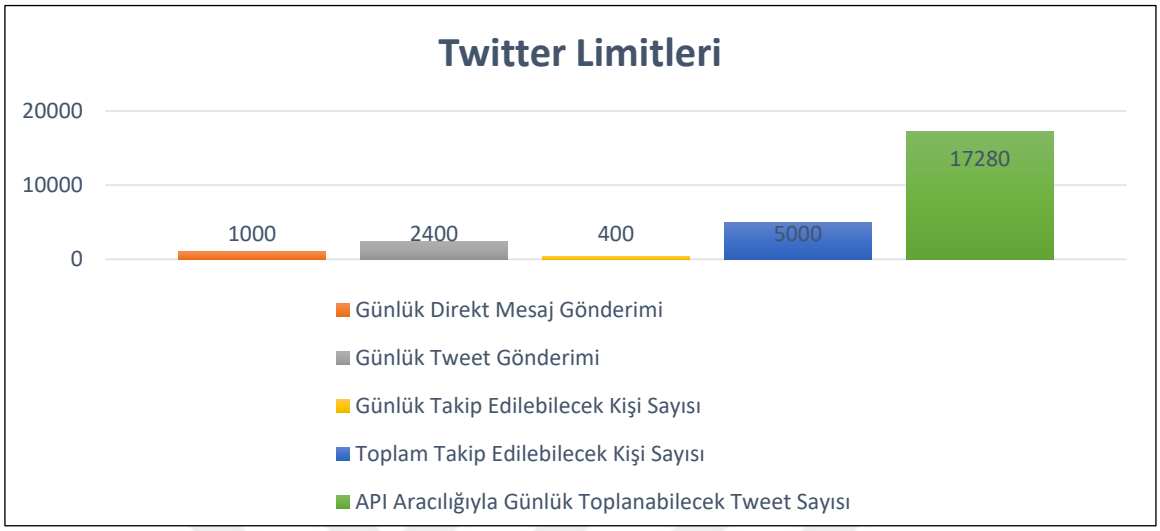
Twitter; sohbet etmek, arkadaş bulmak, ilgi duyulan müzisyen, yazar ya da diğer grup ve kişileri takip etmek ilişki kurmak için kullanılan sosyal bir ağ platformudur.

Twitter'ı gelişim sürecinde öne çıkaran özelliklerden biri platformun geliştiriciler için sunduğu "Uygulama Programlama Ara yüzü" olarak tanımlanan API hizmetidir. Geliştiriciler bu ara yüzü kullanarak Twitter üzerinde bulunan ve herkese açık olan tweetleri belirlenen limitler dâhilinde alabilmekte ve Twitter Kuralları ve politikalarına uymak koşuluyla kullanabilmektedir [13]. API vasıtasıyla konuma, dile, tarihe, kişiye göre özelleşmiş şekilde tweetler temin edilebilmektedir. Araştırma sırasında oluşturulan uygulamanın görüntüsü Resim 2.1.'de verilmiştir.



Resim 2.1. Twitter üzerinde oluşturulan uygulama

Belirlenen günlük limitin aşılması durumunda ise bir sonraki güne kadar hizmet kullanılamamaktadır. Twitter'ın geliştiriciler için sunduğu limitler Şekil 2.1'de verilmiştir [14].



Şekil 2.1. Twitter'ın geliştiriciler için belirlediği limitler

Çalışmada Python Programlama Dili kullanılarak, Twitter API vasıtasıyla İngilizce dilinde 290 bin tweet elde edilmiş ve ön işleme adımları uygulanarak 162.000 tweet kalacak şekilde veri seti oluşturulmuştur.

2.1. Twitter Kullanıcı Kitlesi

Ülkemizde ve Dünyada Twitter sosyal ağına üyelik sayısı gittikçe artmaktadır. 2019'un 4. Çeyrek raporlarına göre toplam 260 milyon kullanıcı sayısı bulunmakta ve bu kitlenin büyük bir bölümü erkek kullanıcılardan oluşmaktadır.

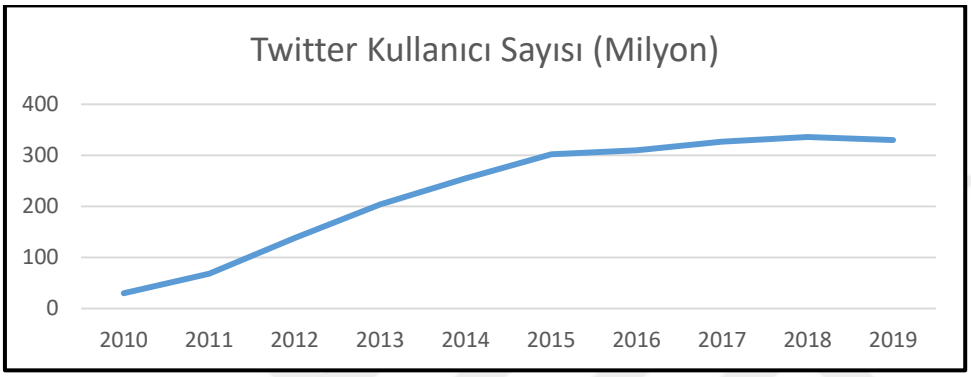
Twitter yaş dağılımına göre;

Çizelge 2.1. 2019'un 4. Çeyrek Dünya Twitter Kullanıcı Yaş Dağılımları

	13-17	18-24	25-34	35-49	50+
Kadın	4.3%	10%	8.6%	6.7%	5.2%
Erkek	5.2%	14%	21%	15%	10%

Twitter kullanıcıları 25-34 yaş arası erkeklerde, 18-24 yaş arası kadınlarda yoğunluk vardır.Twitter kullanıcısı bireysel olabileceği gibi kurum-kuruluş da olabilir, ayrıca şarkıcılar, ünlüler, sporcular, ulusal liderler ve politikacılar da twitter kullanıcılarıdır.

Twitter'daki aylık aktif kullanıcı sayısının yıllara göre değişimi Şekil 3.3'de verilmiştir [15]. Çizelgede görüleceği üzere 2015 yılından itibaren hesap artışı hızı eskiye nazaran azalmaktadır. Bunda 2016 yılındaki ABD seçimlerinde 10 binlerce bot hesabın Rusya tarafından seçimleri etkilemek için kullanıldığının Twitter tarafından kabul edilmesi ve bunun üzerine Twitter'ın sahte hesapları sıklıkla kapatması da etkili olmuştur [16]. Sahte hesaplar gerçek bir kişiyi temsil etmeyen ve belli bir algı oluşturmak için genellikle yazılımlar tarafından kullanılan hesap türüdür.



Şekil 2.2. Twitter kullanıcı sayısının yıllara göre değişim [17]

2.2. Twitter'a Ait Temel Kullanım Özellikleri

Twitter kullanıcıların sohbet etmek amacıyla istenilen konular hakkında konuşabilmek için toplandıkları ve paylaşım yaptıkları bir mikroblog ağıdır. Twitter ne kadar mikroblog özelliklerini barındırsa da kendine ait birçok özelliği vardır. Bunlar; Tweet, retweet, etiket(hashtag), profil, yanıt, beğenme, takip etme, bahsetme, direkt mesaj, anket gibi özelliklerdir.

Tweet: Twitter paylaşımlarına verilen addır. 280 karakter uzunluğuna sahiptir. Metin, fotoğraf, anket, canlı yayın, GIF gibi içerikleri vardır. Atılan tweet hesap sahibinin kendi sayfasında tüm kişilere görünür, gizlilik ayarları sayesinde sadece takipçilerine açık kalabilir.

Retweet: Bir kullanıcı tarafından paylaşılan bir tweetin başka bir kullanıcı tarafından orijinal halinin kendi hesabı üzerinden aynen paylaşmasıdır. Retweetlenen tweet gerçek sahibinin kullanıcı adı ve orijinal tweet ile birlikte retweet yapan hesabın sayfasında görüntülenir. Retweet yapılması için kullanıcı profili herkese açık olması gerekir.

Etiket(hashtag): "#" işaretiyle bir tweetin konusunu ya da kategorisini belli etmek için kullanılır.

Profil: Kullanıcıların, kullanıcı adıyla oluşturdukları diğer kullanıcıları takip etmek ve diğer kullanıcılar tarafından takip edilmek için oluşturulan, isteğe göre kişisel bilgilerine de yer verildiği hesaptır.

Yanıt: Tweet'e verilen cevaptır. Tweetlerin altında bulunan yanıtla iconuna basarak yapılır.

Beğenme: Tweet'in takipçiler tarafından beğenilmesidir. Tweetin altında yer alan kalp iconuna basarak gerçekleştirilir.

Takip Etme: Twitter kullanıcısının diğre twitter hesaplarını takibe almasıdır. Genellikle aynı fikirde olan hesapları ya da ilgi duyulan hesaplar takip edilir.

Bahsetme: "@" simgesini kullanarak bit tweet içinde bahsedeceğimiz kullanıcının hesap adını yazarak yapılır. Örneğin @kullanıcıadı

Direkt Mesaj: Bir twitter kullanıcısının diğer kullanıcıya attığı mesajdır.

Anket: Belirlenen bir içeriğe göre takipçilerin katılmasıyla oluşan bir özelliktir.

2.3. Duygu Kaynağı Olarak Twitter

Twitter insanların duygu ve düşüncelerini, görüşlerini paylaştıkları sosyal bir ağdır. Araştırmacı ve uygulayıcılar için bu paylaşımlar çok önemli bir kaynaktır. Twitter'daki bu paylaşımlar büyük bir veri kaynağı olduğundan kriz yönetiminde, güncel olayları belirleme gibi daha bir çok alanda kullanılmaktadır.

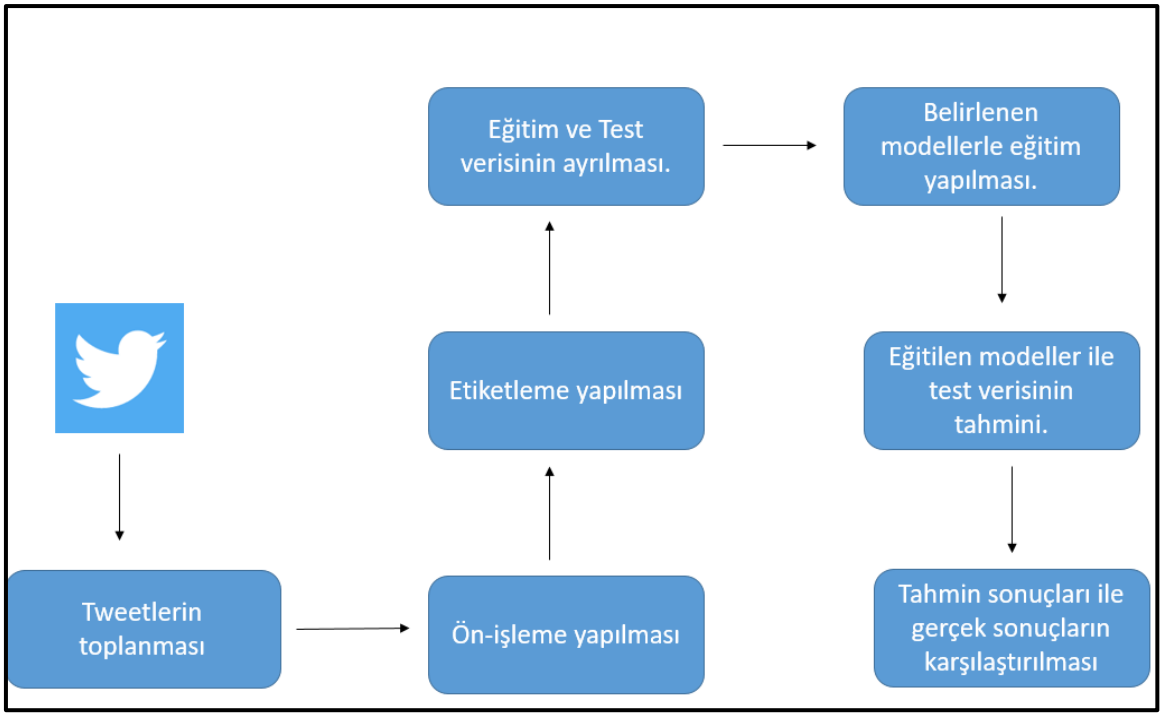
Kullanıcıların duygu, tutum ve görüşlerini belirlemede, duygu analizi, doğal dil işleme gibi yöntemleri kullanarak öznel bilgilerin toplanmasında çok büyük bir kaynaktır. Bu tezin amacı da twitter kullanıcılarının İslamofobi tweetlerinin duygu analizi ile tespitidir.

3. LİTERATÜR

Nefret söylemi ile ilgili literatürde pek çok çalışma yapılmıştır [18]. Bu çalışmalarda genel olarak politik, cinsiyetçi, belli bir dini hedef alan nefret söylemleri araştırılmıştır. Verilerin toplandığı platform olarak çalışmaların tamamına yakınında Twitter seçilmiştir. Twitter'in geliştiricilere sağladığı izinlerin Facebook, Instagram, Linkedin gibi diğer sosyal ağlara göre çok olması, içeriğinin politize olmaya yatkın olması bunda etkili olmuştur.

3.1. Twitter Üzerindeki Nefret Söyleminin Makine Öğrenmesinde Kullanımı Konulu Çalışmalar

Makine öğrenmesi yöntemleriyle çeşitli nefretlerin tespiti 2010 yılından itibaren aktif bir araştırma alanı olarak karşımıza çıkmaktadır. Bunda Twitter uygulaması üzerinde biriken verinin etkisi, Twitter'ın bu verinin kullanımı konusunda diğer sosyal ağlara göre tanıdığı kolaylıklar, verinin genel olarak metin tabanlı olmasının etkisi büyüktür. Bu sebeple halihazırda bu konudaki birçok çalışmada Twitter üzerindeki veriler kullanılmaktadır. Twitter uygulamasından elde edilen verilerle yapılan çalışmalarda ağırlıklı olarak kullanılan yöntem çalışma kapsamında geliştirilen Şekil 2.3'de verilmiştir.



Şekil 2.3. Twitter ile duygu analizi yöntemi

Bu konuda yapılan çalışmalara daha yakından bakılacak olursa; Theo Lynn ve arkadaşları çalışmasında kadın nefreti içeren dilin tespiti için yapay sinir ağı geliştirmiştir. Çalışmada 1999 ile 2016 yıllarını arasında atılan 2 milyon 600 bin tweetten oluşan veri seti kullanılmıştır. Sonuçlarda Bi-GRU ve Bi-LSTM algoritmalarında kesinlik ve hassasiyette %92 doğru tahmine ulaşılmış, F-1 skorunda(F1-skoru harmonik ortalamayı kullanır.) ise Random Forest algoritmasında %96 doğruluğa ulaşılmıştır [19].

Carlos Perollo ve arkadaşları tarafından yapılan çalışmada, göçmenlere ve kadınlara karşı olan nefret söylemi Twitter verisi kullanılarak incelenmiştir. 14 bin İngilizce ve 7 bin 600 İspanyolca tweetin kullanıldığı çalışmada Support Vector Machine modelinde F-1 ölçütünde %62,5 doğruluğa ulaşılmıştır [20].

Unnathi Bhandary tarafından yapılan çalışmada ise diğer çalışmalardan farklı olarak videolardaki nefret söylemi makine öğrenmesiyle tespit edilmeye çalışılmıştır. Videoların öncelikle ses dosyaları ayrıştırılarak metne dönüştürülmüş ve bu metinler 4 farklı makine öğrenmesi algoritması modeliyle eğitilmiştir. Eğitim neticesinde Random Forest Sınıflandırıcının en yüksek doğruluğa sahip sonuçları verdiği görülmüştür [21].

Poonam Sahibani ve arkadaşları tarafından yapılan çalışmada, sosyal ağlardaki radikal dilin makine öğrenmesiyle tespiti çalışılmıştır. 18 000 tweetin kullanıldığı çalışmada Support Vector Machine ve Random Forest algoritmaları %97,82 doğru tahmine ulaşılmış ve Naive Bayes onları %86,95 doğrulukla takip etmiştir [22].

Ilham Maulana Ahmad Niam ve arkadaşları tarafından yapılan çalışmada Twitter üzerindeki görsellerdeki nefret söyleminin makine öğrenmesi yöntemlerinden *latent semantic* analiz kullanılarak tespiti çalışılmıştır. 1,223 videonun bulunduğu veri setinde 645 adet nefret söylemi içeren, 578 adet ise nefret söylemi içermeyen video kullanılmıştır. Sonuç olarak ise *Latent Semantic* Analysis modeliyle F1 ölçütünde %57. 9 doğruluk yakalanmıştır [23].

Mohini S. Dadhe ve arkadaşları tarafından yapılan çalışmada Twitter üzerindeki küfürlü dilin makine öğrenmesiyle tespiti çalışılmıştır. Çalışma neticesinde Support Vector Machine kullanılarak kesinlik ölçütüne %98,24 ve duyarlılık ölçütünde %94,24 doğruluğa erişilmiştir [24].

Maarten Sap ve arkadaşları tarafından yapılan çalışmada Amerika'da yaşayan Twitter kullanıcıları arasındaki saldırgan dilin, makine öğrenmesiyle tespiti üzerine çalışılmıştır.

Çalışma sırasında ilk olarak Afroamerikan İngilizcesi olarak kabul edilen jargonu kullanan kişilerin diğer kullanıcılara göre daha fazla saldırgan dil kullandığı görülmüştür.

Günlük konuşmalarda dahi küfürlü, saldırgan dilin Afroamerikan kullanıcılarda daha fazla olduğunun tespit edildiği çalışmada bir tanesi 56 milyon diğeri ise 5,4 milyon tweet içeren veri setleri kullanılmıştır. Eğitim için model olarak BiLSTM kullanılan çalışmada saldırgan olmayan Tweetleri doğru işaretleme oranı %50'de kalmıştır [25].

Halka Mulki ve arkadaşları tarafından yapılan Twitter platformundaki nefret söyleminin tespiti için N-gram yöntemi kullanılmıştır. İngilizce ve İspanyolca tweetler üzerinden yapılan çalışmada İngilizce'de daha başarılı sonuçlar alınmıştır. Çalışmanın ileriki safhalarında Arapça ve Türkçe üzerinde benzer denemelerin yapılması amaçlanmıştır [26].

Nabiila Adani Setyadi ve arkadaşları tarafından yapılan çalışmada Twitter üzerindeki nefret söyleminin geri yayılımlı çok katmanlı sinir ağları ile tespiti çalışılmıştır. Toplam 1235 tweetin olduğu veri seti ilk olarak %10 eğitim %90 test seti olacak şekilde ardındansa %20 eğitim %80 test seti olacak şekilde ayrılmıştır. İlk durumda F-1 ölçütünde %80,92 doğruluğa ulaşılırken ikinci durumda %81,26 doğru tahmine ulaşılmıştır [27].

Alexandra Siegel ve arkadaşları tarafından yapılan çalışmada 2016 Amerika Birleşik Devletleri seçimi ve sonrasını esas alan ABD başkanı Trump taraftarlarının Twitter üzerindeki nefret söylemine etkileri çalışılmıştır. Seçim ile ilgili 750 milyon ve rastgele alınmış 400 milyon Tweet'ten oluşan veri seti ile yapılan çalışmada *Bölünmüş Zaman Serisi Regresyonu* ile eğitim gerçekleştirilmiş ve çalışma neticesinde söylenenlerin aksine Trump taraftarlarının kullandığı nefret söyleminin Clinton taraftarlarından çok farklı olmadığı ve regresyon sonuçlarının birbirine benzer olduğu görülmüştür [28].

Erryan Sazany ve Indra Budi tarafından yapılan çalışmada Endonezya'ca dilindeki tweetlerdeki nefret söyleminin makine öğrenmesi yaklaşımlarından derin öğrenmeyle tespiti çalışılmıştır. Uzun / Kısa Süreli Bellek (LSTM) kullanılan çalışmada katman sayısı 100, epoch değeri 1 ve batch size 32 olarak belirlenmiş, sonuç olarak ise F-1 ölçütünde %94,5 doğruluğa ulaşılmıştır [29].

Kelvin Kiema Kiilu ve arkadaşları tarafından yapılan çalışmada Twitter üzerindeki nefret söyleminin tespiti Naiive Bayes algoritması ile yapılmıştır. 68 bin Tweet'ten oluşan veri setinde Naive Bayes sınıflandırıcıda unigram tekrar oranı için %64,47, bigram tekrar oranı için %70 doğruluğa ulaşılmıştır. Support Vector Machine için bu değerler %53,59-%59 olurken, Benroulli Naive Bayes için %56,25'e %66 olarak gerçekleşmiştir [30].

Pushkar Mishra ve arkadaşları tarafından yapılan çalışmada ise Twitter üzerindeki nefret söylemi yazar profilleme için kullanılmıştır. Bu çalışmada tweet odağı yerine kullanıcı odağı seçilmiş ve makine öğrenmesi yöntemleriyle kullanıcıların ırkçı veya cinsiyetçi olup olmadığı bulunmaya çalışılmıştır. Irkçı profilleme için Lojistik Regresyon için %72,28 doğruluğa erişilirken c için bu değer %73,24 olmuştur. Cinsiyetçi profillemedeki tahminlerde ise lojistik regresyon F-1 ölçütünde %72,09 değerine ulaşırken tekrarlayansinir ağı modeli %72,24 doğrulukla tahminde bulunmuştur[31].

Çizelge 3.1. Logistik regresyon, tekrarlayan sinir ağı, ırkçı ve cinsiyetçi profilleme

	Logistik Regresyon	Tekrarlayan sinir ağı modeli
Irkçı Profilleme	%72,28	%73,24
Cinsiyetçi Profilleme	%72,09	%72,24

Irene Kow ve Yuzhou Wang tarafından yapılan çalışmada Twitter üzerindeki siyahilere karşı olan nefret çalışılmıştır. 24 582 tweetten oluşan veri setinde bag-of-words modeli kullanılmış ve %76 doğrulukla tahminde bulunan sistem geliştirilmiştir [32].

Rijul Magu ve arkadaşları tarafından yapılan çalışmada sosyal medyadaki nefret söylemi çalışılmıştır. 1 milyonun üstünde Tweet kullanılan veri seti Support Vector modeli ile eğitilmiş ve 10 katlı doğrulama yapılmıştır. Sonuç olarak precision ölçütünde %79,5 recall ölçütünde ise %79,4 doğruluğa ulaşılmıştır [33].

Özetle; bu çalışmalar kapsamında Twitter uygulamasından elde edilen verilerle hangi nefret söylemlerinin incelendiği incelenmiş ve birbirinden farklı makine öğrenme yöntemlerinin kulanıldığı, tek bir yönteme bağlı kalınmadığı görülmüştür.

3.2. Sosyal Ağlar Üzerindeki İslamofobik Verinin İncelenmesi Konulu Çalışmalar

Imran Awan tarafından yapılan çalışmada Müslümanlara karşı olan nefretin Twitter üzerindeki topolojisi çıkarılmıştır. 100 farklı kullanıcıdan elde edilen 500 tweet esas alınan çalışmada tweetlerin %72'sinin erkekler tarafından gönderildiği, Müslüman karşıtlığının, muzrad gibi bazı özel terimler ve çocuk istismarı, terörizm gibi kavramlar ile ilişkilendirilerek yapıldığı tespit edilmiştir [34].

Awan tarafından yapılan bir başka çalışmada Facebook üzerindeki Müslüman karşıtı etkileşimler incelenmiştir. Bu çalışmada ise Facebook üzerinde Müslüman karşıtlığını artırmak için kullanılan 191 görsel incelenmiştir. Görsellerde en çok vurgulanan Müslümanların bir güvenlik sorunu olarak görülmesidir. Bu durumu; Müslümanları kültürel tehdit, ekonomik tehdit olarak gösteren görseller izlemiştir. Ardından ise doğrudan Müslümanları şeytan olarak gösteren görseller gelmektedir. Doğrudan Müslümanlara soykırım yapılmasını savunan görseller ise daha alt sıralarda yer almaktadır [35].

Bünyamin Ayhan ve Muhammet Emin Çifçi tarafından yapılan Işid, Propaganda ve İslamofobi isimli bir çalışma yapılmıştır. Çalışmada IŞİD gibi terör örgütlerinin sosyal medya platformları üzerinden mesajlarını yaydıkları, bu mesajların da yükselen İslamofobiye katkı sağladıkları belirtilmiştir [36].

Çilem Tuğba Koç tarafından "Yeni Medyada İslamofobik Söylemin Üretimi: Facebook Örneği" isimli çalışmada 35 adet Facebook grubu üzerinden analiz yapılmıştır.

Çalışma sonucunda Müslüman karşıtlığının barbar, şeytani, sabit fikirli, eğitimsiz, terörist, pedofoliyi destekleyen gibi olumsuz kavramlarla beraber anıldığı görülmüştür [37].

Çalışmadan elde edilen bulgular İslamofobi'nin en çok terörist, pedofili, cahil gibi kavramlarla birlikte ele alındığını göstermektedir.

Çizelge 3.2. İslamofobi çalışma tablosu

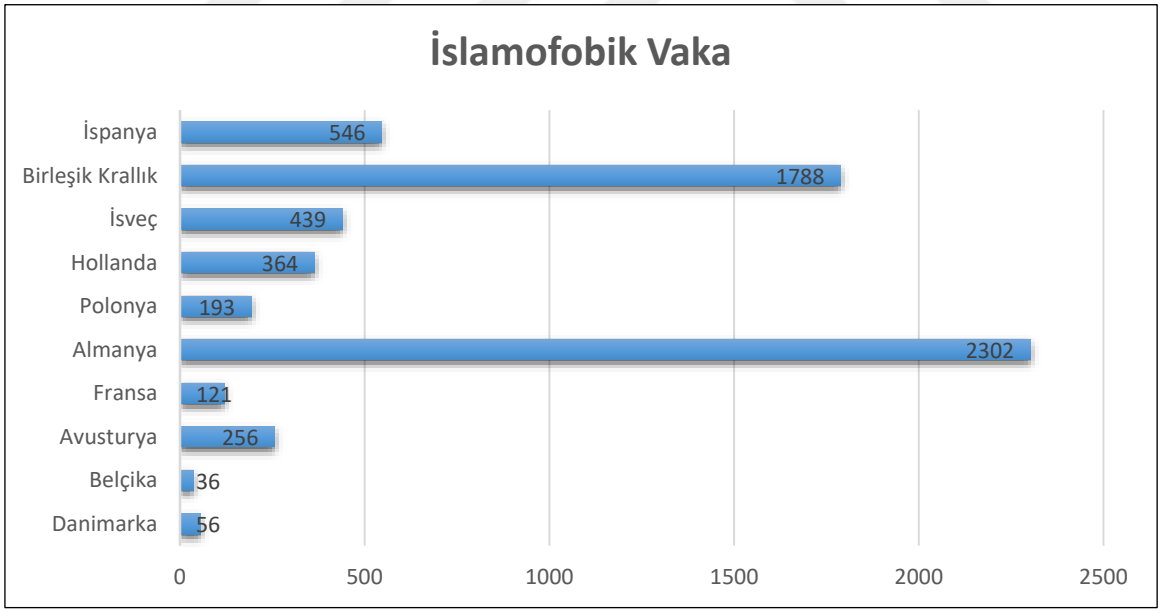
Çalışan Grup	Konu	Dil	Model	Başarı	Sonuç
Imran Awan	Müslüman Nefretinin Twitter Topolojisi	Python	(RNNs,CNNs)	%78,9	Müslüman karşıtlığının çocuk istismarı ile ilişkilendirilmiştir
Bünyamin Ayhan ve Muhammed Emin Çiftçi	Işid, Propaganda ve İslamofobi	Python	Multilayer Perceptrons	%75,2	Işid destekli islamofobi çalışması yapılmıştır.
Çilem Tuğba Koç	Yeni Medyada İslamofobik Söylemin Üretimi	Python	Multilayer Perceptrons(RNNs)	%82,7	Müslümanların barbar,şeytani fikirli eğitimsiz olarak yansıtılmıştır

4. METOD MATERYAL

Bu bölümde çalışmada kullanılan metod materyal hakkında bilgi verilmiştir. Gerek İslamofobinin gerke makine öğrenmesi ve duygu analizinin gelişimine değinilmiştir.

4.1. İslamofobi

İslamofobi, son yıllarda oldukça popüler olan terimlerden birisidir. Bir kesim bu terimin anlamının doğrudan İslam'a veya Müslümanlara karşı duyulan korku, endişe olarak çevrilebileceğini savunurken bir kesim ise bunun aslında olmayan bir fenomen olduğu konusunda ısrarcı olmaktadır [37-38]. 11 Eylül 2001 tarihinde Amerika Birleşik Devletleri'nde gerçekleşen terör saldırıları, Temmuz 2005 tarihinde Londra'da yapılan bombalama eylemleri Müslüman olduğunu iddia eden El-Kaide terör örgütü tarafından sahiplenilmiştir [39-40]. Bu durum da tüm dünyada Müslümanlara karşı olan endişe ve korkunun artmasını tetiklemiştir [41]. Siyaset, Ekonomi ve Toplum Araştırmaları (SETA) Vakfı tarafından 2017 yılında yayınlanan rapora göre 2016 yılında Avrupa'nın farklı ülkelerinde yaşanan fiziki İslamofobik saldırı sayıları Şekil 4.1'de verilmiştir [42].



Şekil 4.1. Avrupa'daki Islamofobik vaka sayıları

Yeni Zelanda'nın Christchurch şehrindeki Nur Camii ve Linwood İslam Merkezine 15 Mart 2019 tarihinde gerçekleşen saldırılarda 49 kişi hayatını kaybetmiştir [43]. Bu saldırı, Yeni Zelanda'da son 20 yılda yapılan ilk toplu katliam olarak tarihe geçmiştir. Terörist Brenton Tarrant, saldırı anındaki 16 dakikalık videoyu Facebook üzerinden canlı olarak yayınlamıştır.,

Ayrıca saldırgan, saldırı öncesinde yine internet ortamını kullanarak "Büyük Yenilenme" isimli bir manifesto yayınlamış, askıya alınan Twitter hesabında da Neo-Nazi paylaşımlarda bulunmuştur [44].

Kanada'da 29 Ocak 2017 tarihinde Quebec City cami saldırısında, akşam namazı kılan 6 Müslüman hayatını kaybetmiş, 19 Müslüman ise yaralanmıştır [45]. Fransa'da ortalama üç haftada bir camiye karşı Vandalizm davranışları olmaktadır [46]. Hollanda'da İslam karşıtı Geert Wilder yönetimindeki Özgürlük Partisi son seçim olan 2017 yılı seçiminde 5 koltuk daha alarak parlamentoda 20 koltuğa ulaşmıştır [47]. Almanya'da Mısır asıllı Müslüman Marwa El-Sherbani mahkeme sırasında bir Neonazi tarafından vahşice öldürülmüştür [48].

Avustralya Viktorya eyaletinde yapılan bir araştırmada okuldaki öğrencilerin yarısından fazlasının Müslümanları terörist olarak gördüğü, 5'te 2'sinin ise Müslümanlığı temiz olmamak, kirli olmakla ilişkilendirdiği sonucuna ulaşılmıştır [49]. İsveç'te "İslam ikinci dünya savaşından beri karşılaştığımız en büyük dış tehlike" diyen İsveç Demokrat Parti başkanı Jimmie Akesson 2018 yılında yapılan seçimde oyunu %17,53 e çıkararak 13 sandalye daha almış ve parlamentoda 62 sandalyeye ulaşmıştır [50].

Polonya'da ana akım medya ve bazı politikacılar Müslümanları tanımlarken "insan görünümlü şeytan" anlamına gelen "folk devil" ifadesini kullanmıştır [51]. Danimarka'da Jullands-Posten isimli gazete Hz. Muhammed'in sözde karikatürlerini yayınlayarak onu terörist olarak sembolize etmiştir. Ayrıca ülkede Kur'an-ı Kerim'i aşağılayan Fitna isimli film yapılmıştır [52].

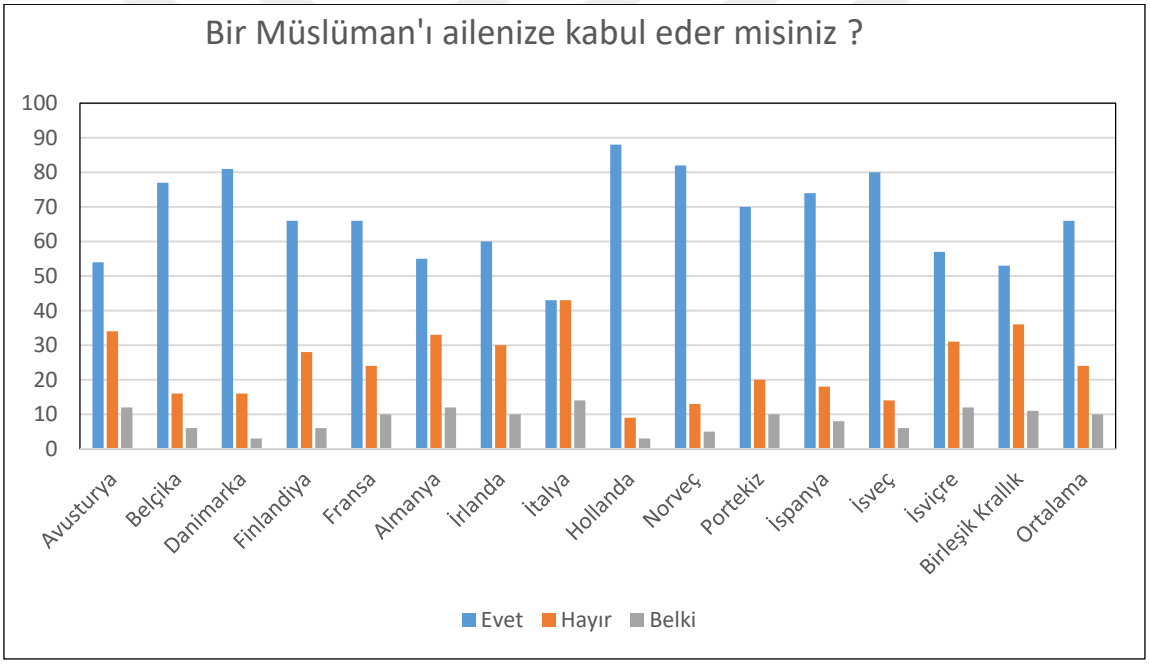
Hindistan'da iktidarda olan Hindu Milliyetçisi Bharatiya Janata Partisi ve lideri Narendra Modi Ayodya isimli bölgede Babri Mescidinin yıkılması için propaganda yapmaya devam etmektedir [53]. Fransa ve Belçika kamuya açık yerlerde burka ve peçe giyilmesini yasaklayan kanunları yapılırken İsviçre yeni yapılacak camilerde minare olmasına karşı çıkmaktadır [54-55-56].

Japonya'da Müslümanların ibadet yerlerinin, helal restoranların, İslam ile ilişikli organizasyonların düzenli olarak izlenmesi mahkeme tarafından onaylanmıştır. 2010 yılında internete sızan bazı polis araştırmaları, Müslümanların bir potansiyel suçlu gibi doğrudan izlendiğini ortaya koymaktadır [57].

Sri Lanka'da camilere ve Müslümanların sahibi olduğu marketlere kundaklama saldırıları gerçekleşmiştir. [58].

Amerika Birleşik Devletleri'nde Cumhuriyetçi Partinin Fort Worth Bölgesi Başkanvekili olan Pakistan asıllı Shadid Shafi'nin camiye düzenli olarak gitmesinden dolayı, "İslami şeriat düzeninin getirilmesine yardımcı olmakla" itham edilerek görevden alınması istenmiştir [59]. Yine Amerika Birleşik Devletleri Ordusu tarafından yayınlanan bir yönergede tesettür, "pasif terörizm" olarak tanımlanmış ve başörtüsünün terör örgütlerinin şiddetine bir tür kapalı destek olduğu savunulmuştur [60].

Pew Research tarafından 2017 yılında 15 ülkede bir araştırma yapılmış ve katılımcılara "Bir Müslüman'ı ailenize kabul eder misiniz?" sorusu yöneltilmiştir. Sonuçlar ortalama her 4 Avrupalıdan birinin Müslümanları ailelerine kabul etmek istemediğini belirtmiştir. İtalya'da ise bu oran %43'e çıkmıştır. İtalya'yı, %36 ile Birleşik Krallık ve %33 ile Almanya takip etmektedir [61]. Sonuçlar Şekil 4.2'de verilmiştir.



Şekil 4.2. Pew Research tarafından yapılan islamofobi anketi

Daha da genişletilebilecek bu liste göstermektedir ki, bugün dünyanın birçok bölgesinde Müslümanlara karşı olan endişe ve korku artmaktadır. Bu endişe ve korkuyu önceden tespit etmek çok önemlidir. Tespit sonrasında ilgili paydaş kurumların tespit edilen bölgelerde yapacağı etkinlikler ve faaliyetler ile endişe ve korkunun yerini iletişim alabilecektir. Türkiye'ye gelen Amerikalı felsefeci Grand Valley State Üniversitesi Öğretim Üyesi Prof. Dr. Kelly James Clark Anadolu Ajansına verdiği röportajda Müslümanları tanıdıkça İslamofobi'den uzaklaştığını belirtmiştir [62].

4.2. İslamofobinin Temel Sebepleri

İslamofobinin ortaya çıkmasının Temel Sebeplerinden biri, İslam'ın hayat sunan ilke ve prensiplerinden insanların etkilenmesi gösterilebilir. Çünkü İslam'ın hayat sunan, yön veren, yol gösteren ve ilkeleri anlaşıldıkça, İslam karşıtı kalmayıp islama yönelmeler artmaktadır. Bu nedenle İslam karşıtı oluşumların İslamofobiyi kullanarak islamdan uzaklaştırma yolunu kullanmışlardır. İslamofobinin yayılmasında en önemli kaynaklardan biri de sosyal medyadır.

İslamofobinin ettkisinin sürmesinde Avrupalı siyasetçilerin birkaç oy fazla alma uğruna nefret söylemleri kullanarak politik dengeleri bozmak için

İslamofobiyanın etkisini sürdürmesinin bir diğer nedeni de Batı ülkelerinde İslam dininin resmi din olarak kabul edilmemesidir. Bunun etkisiyle islamofobi giderek arttırmıştır ve Batı ülkelerinin islam'a uzak durmalarına ve aşırıların da saldırı ve hakaretler yapmasına neden olmaktadır.

Diğer ülkelerdeki müslümanların en temel haklarından mahrum kalmalarının nedenlerinden biri de ülke yöneticileri, İslamofobiya karşısında daima sessiz ve tepkisiz kalmalarıdır [62].

4.3. Nefret Söylemi

Geçtiğimiz yıllarda kullanımındaki yoğun artışa rağmen nefret söyleminin evrensel anlamda kabul gören bir tanımı bulunmamaktadır. Bu konuda genel olarak kabul edilen tanımlardan biri Avrupa Konseyi'nin Bakanlar Komitesi tarafından yayınlanan "nefret söylemi" konulu 97(20) sayılı kararında geçen ifadedir [63]. Bu kararda nefret söylemi şu şekilde tanımlanmıştır:

"Nefret söylemi belirli bir grubu veya din, ırk, ulus, mezhep, cinsel yönelim, göçmenler gibi konularda ayrımcılık ve düşmanlık şeklinde ifade bulan, bir inanca karşı hoşgörüsüzlük de dahil olacak şekilde hoşgörüsüzlüğe dayalı nefreti yayan, yayılmasını teşvik eden veya meşrulaştıran her türlü ifade biçimini kapsamaktadır."

Nefret söyleminin yaptırımları ülkelere göre hatta ülke içerisindeki eyaletlere göre dahi değişiklik gösterebilmektedir.

Örneğin Avustralya'da, Avustralya Başkent Bölgesi cinsel yönelimi sebebiyle bir kişiye karşı işlenen nefret suçlarına karşı tazminat ödenmesini isterken Victoria, Queensland eyaletleri bu konuda tazminat ödenmesini gerekli görmemektedir.

Benzer şekilde İslamofobi'yi de içine alan ve bir dine karşı olan nefret söyleminde ise Tasmania, Queensland ve Avustralya Başkent Bölgesi tazminat ödemesini kabul ederken diğer bölgeler bunu tazminat kapsamına almamaktadır [64-65].

Brezilya'da 1997 yılında alınan bir karara göre bir dine yönelik nefret söylemi suç kapsamındadır [66].

Kanada'da nefret söylemine karşı yaptırımlar bazı konularda çok ağır olmaktadır. Örneğin herhangi bir gruba karşı yapılan soykırımı olumlu gören kişilere en az 5 yıl olmak üzere hapis cezası verilmektedir [67]. Fakat Müslümanlara karşı yapılan nefret söyleminde aynı hassasiyette bulunulmamaktadır. Örneğin Mark Steyn isimli yazar "Müslümanlar sinek gibi ürüyor." gibi birçok İslamofobik ibareyi barındıran Maclean dergisini yayınlamış fakat açılan davalarda şikayetler hem Ontario İnsan Hakları Komisyonu hem de Kanada İnsan Hakları Komisyonu tarafından reddedilmiştir [68].

Şili'de bir dini gruba karşı yapılan nefret söylemi para cezası sonucu doğurmaktadır. Buna internet yoluyla gerçekleşen nefret söylemi de dahildir [69]. Hırvatistan'da bir dini gruba karşı nefret söylemi anayasal bir suç olarak görülmektedir [70].

Danimarka'da bir dini gruba karşı nefret söyleminin suç olarak görüldüğü yasalarda vurgulanmamıştır, bunun yerine ifade özgürlüğü ön plana çıkarılmıştır. Özellikle Hz. Muhammed karikatürlerinde bu konu gündeme gelmiş ve Danimarka Müslüman kamuoyunun tüm tepkilerine rağmen konuyu ifade özgürlüğü kapsamında değerlendirmiştir [71].

Finlandiya, Fransa, İzlanda, İrlanda, Malta, Hollanda, Norveç, Rusya, Sırbistan, Singapur, Güney Afrika, Birleşik Krallık yasalarında da nefret söyleminde bir dine karşı nefret söylemi için ayrı bir parantez açılmış ve suç olarak nitelendirilmiştir. Bunların dışında bazı ülkeler uluslararası medeni ve siyasi haklar sözleşmesine imza atmakta fakat nefret söylemi suçlarına karşı kapsamlı bir yasa çıkartmamaktadır. Örneğin Endonezya'da Balinese Hinduları ve İslami olduğunu iddia eden bir örgüt arasındaki sözlü nefret söylemlerine hukuksal bir süreç başlamamış Ulusal Polis Şefi General Badrodin Haiti, bu nefret söyleminin kötü sonuçlar doğurabileceğine dair sadece tarafları uyarmış ve herhangi bir yaptırım uygulamamıştır [72].

Amerika Birleşik Devletleri nefret söylemine karşı herhangi bir yaptırım uygulamamakta ve bunun ABD anayasasındaki düşünce özgürlüğüne aykırı olduğunu savunmaktadır [73].

4.4. Makine Öğrenmesi Tarihçesi

Bilgisayarlardaki problemleri çözmek için algoritmalara ihtiyaç duyulur. Algoritma, verilen girişi istenen çıktıya dönüştürmek için yapılması gereken talimatlar dizisine verilen isimdir [74]. Örneğin verilen hayvan isimlerini alfabetik sıraya göre listeleyen bir algoritma olabilmektedir. Böyle bir durumda girişler kelime kümesi, çıktı ise sıralı bir listedir.

Fakat bazı durumlar için algoritma geliştirmek daha karmaşık bir iş olabilmektedir. Bu duruma örnek olarak bir cep telefonuna gelen bir mesajın dolandırıcılık amaçlı olup olmadığını öğrenilmesi düşünülebilir. Bu durumda giriş, karakter bazlı bir metin dosyasıdır. Çıkış ise bu mesajın dolandırıcılık amaçlı atılıp atılmadığına dair bilgidir. Fakat girdinin çıktıya nasıl dönüştüreceği bilinmemektedir. Çünkü dolandırıcıların yöntemler zaman içinde ve kültürden kültüre değişebilmektedir. Buradaki bilgi eksikliğini veri bilimci elinde bulunan veri ile telafi edilebilmektedir. Bu noktada yapılacak olan önceden dolandırıcılık mesajı olduğunu kabul ettiğimiz binlerce örneği ve dolandırıcılık mesajı olmayan binlerce örneği yazılıma göstererek bazı matematiksel veya istatistiksel yaklaşımlar ile algoritmayı otomatik olarak çıkartmasını istemektir [75].

Makine öğrenmesinin temelleri ilk olarak 1795 yılında Carl Friedrich Gauss tarafından geliştirilen ve 1806 yılında Adrien-Marie Legendre tarafından geliştirilerek yayınlanan en küçük kareler yöntemine, 1812 yılında tanımlanan Bayes teoremine ve 1813 yılında Andrey Markov tarafından geliştirilen Markov Zincirine dayanmaktadır. Bu üç teknik modern makine öğrenmesi algoritmalarının temelini oluşturmaktadır [76].

Bu tekniklere yakından bakılacak olursa, en küçük kareler yöntemi birbirine bağlı olarak değişen iki büyüklük arasındaki matematiksel bağlantıyı, olabildiğince gerçeğe yakın bir denklem olarak yazmaya çalışır. Bunun için gerçeğe mümkün olduğunda yakın bir fonksiyon oluşturur.

Benzer şekilde farklı yoğunluklarda gözlemler için farklı eğimlerde veya parabolik eğriler de oluşabilir. Bayes teoremi ise bir olayla ilgili olabilecek durumun önceki bilinen bilgilere dayanarak olma olasılığını bulmaktadır [77].

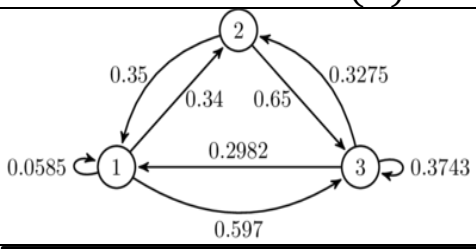
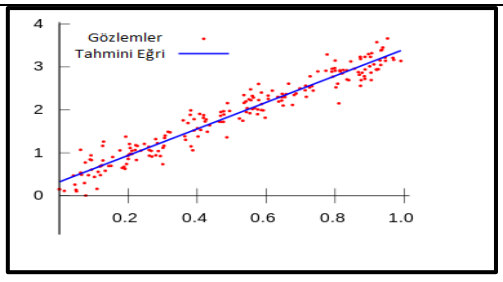
$$P\,(A|B) = \frac{P(B|A)P(A)}{P(B)} \tag{3.1}$$

Eş. 3.1'de ifade edilen Bayes teoremi P(B)>0 olma koşulu ile özetle B'nin gerçekleştiği durumda A'nın gerçekleşme olasılığını açıklamaktadır.

Örneğin bir tenis maçının günlük hava durumuna göre oynanıp oynanamayacağını bulmak için şöyle bir yol izlenir. Daha önceki günlerdeki dışarıdaki hava durumu, sıcaklık, nem oranına göre maçların yapılıp yapılmadığına bakılır. Formüle göre olasılıkları yerine koyarak maçın yapılabilme ihtimalini hesaplanır.

Üç temel teknikten sonuncusu olan Markov Zinciri ise mevcut durumun açıklamasının, sürecin gelecekteki evrimini etkileyebilecek tüm bilgiyi kapsadığını söylemektedir. Gelecekteki durumlara belirlenen şekilde değil zaman veya mekâna göre değişen/evrilen olasılıksal bir süreçle ulaşılacağını iddia etmektedir [78].

Örneğin bir bebeğin davranışlarını Markov zinciriyle açıklamak isterseniz oynama, yemek yeme uyuma ve ağlamaları durumlar olarak tanımlar ve bu durumlardan bir "durum alanı" oluşturulur. Markov zinciri önceki yaşanan durumlara bakarak, size şu an oynayan bir bebeğin önümüzdeki beş dakika içerisinde ağlamadan uykuya dalma olasılığını söylemektedir. Özellikle protein dizilimleri, DNA yapıları gibi alanlarda Markov zincirlerinden faydalanılmaktadır.

Kullanılan Yöntemler	
Bayes Teoremi	$$P(A\|B) = \frac{P(B\|A)P(A)}{P(B)}$$
Markov Zinciri	
Lineer regresyon	

Şekil 4.3. Üç teknik modern makine öğrenmesi

19. Yüzyılın başında yaşanan bu gelişmelerin ardından 1940'lı yıllara kadar henüz adı tam olarak konulmamış makine öğrenmesi alanında önemli bir gelişme yaşanmamıştır. 1940'lı yılların sonunda ise önceden girilen verileri hafızasında tutabilen bilgisayarlar için geliştirmelere başlanmıştır. Bu türden ilk bilgisayarlar 1948 yılında Manchester Küçük-Ölçekli Deneyim Makinesi adı altında geliştirilmiş ve sonrasında ismi doğasına uygun olarak bebek anlamına gelen "Baby" olarak konulmuştur. Bu çalışmayı 1949 yılında Cambridge laboratuvarlarındaki EDSAC ve 1951 yılında Pensilvanya Üniversitesinde geliştirilen EDVAC isimli bilgisayarlar izlemiştir [79-80-81].

1950 yılına gelindiğinde Alan Turing "Makinalar düşünebilir mi?" isimli makalesini yayınlamıştır [82]. Bu makalede "Turing testi" isimli bir testten bahsetmiştir. Bu test ile Turing bir bilgisayarın, tıpkı bir insan gibi davranabilmesinin çerçevesini çizmiştir. Turing testin detaylarında şu örneği vermektedir. Bir duvarın arkasında iki bilgisayar olduğunu düşünelim. Soru soran kişi duvarın arkasındaki bilgisayarlara aynı soruyu gönderir. Cevapların birini bilgisayarda oturan bir kişi diğerini ise diğer bilgisayardan bir yazılım verir. Soruyu soran kişi cevaplardan hangisini yazılımın hangisini gerçek kişinin verebildiğini bulmalıdır. Bu test, birçok yapay zekâ probleminin çatısını teşkil etmektedir.

1951 yılında Marvin Minsky ve Dean Edmonds tarafından ilk yapay sinir ağı geliştirilmiştir. Burada sinir ağı ibaresinin kullanılma sebebi gerçek beyni taklit eden bir simülasyon olmasıdır [83]. SNARC ismi verilen bu bilgisayarın yaptığı labirentte duvarlara kendini vura vura çıkış bulan bir farenin psikolojik deneyimlerine benzetilmektedir [84].

1959 yılına gelindiğinde "makine öğrenmesi" kavramı ilk olarak Arthur Samuel tarafından kullanılmıştır. Samuel makine öğrenimini "makinelerin dış bir programlama olmadan sonuçları öğrenebilme kabiliyeti" olarak tanımlamıştır. Samuel bu kapsamda kendi hatalarını görerek kendini geliştiren bir dama oyunu da yapmıştır [85].

1960'lı yıllarda yapay zekâ ve makine öğrenmesi konusuna yoğun bir ilgi olmasına rağmen çalışmaların başarısızlıkla sonuçlanması hayal kırıklıklarını beraberinde getirmiştir. 1973 yılında İngiltere Bilim Araştırma Konseyi tarafından düzenlenen sempozyumda "Lighthill Raporu" olarak kayıtlara geçen bir sunum yapılmıştır. James Lighthill tarafından gerçekleştirilen bu sunumda yapay zekânın görkemli hedefler sunamadığına dikkat çekilmiştir [86].

Fakat 1980'li yıllara gelindiğinde karamsarlık yerini yavaş yavaş heyecana bırakmaya başlamıştı. 1982 yılında John Hopfield'in, Hopfield Ağı olarak bilinen yapay sinir ağını geliştirmesi, yapay sinir ağlarını tekrar gündeme getirdi [87]. Aynı günlerde Japonya- ABD arasında yapılan ortak konferansta Japonya yapay sinir ağları konusundaki gelişmelerde geride kalmanın onları endişelendireceğini dile getirerek bu konuya fon ayıracaklarını belirtti [88].

1985 yılında itibaren Amerikan Fizik Enstitüsü yapay sinir ağları ile ilgili yıllık toplantılar düzenlemeye başlamıştır [89].

1987 yılında yapay sinir ağları ile ilgili IEEE tarafından düzenlenen konferansa bin 800 kişi katılmıştır [90].

1996 yılına gelindiğinde yapılan yapay zekâ çalışmalarını sunmak için önemli bir fırsat doğmuştur. Dünya satranç şampiyonu Garry Kasparov, Deep Blue isimli yapay zekâ ile 1996-1997 yıllarında maçlar yaptı ve 1997 yılının mayıs ayında yapılan maçta Kasparov, yapay zekâya yenilmiştir [91]. Bu maç insanlık tarihini değiştiren maç olarak kayıtlara geçerken yapay zekâ konusundaki çalışmalara olan ilgiyi de artırmıştır.

1997 yılında Sepp Hochreiter ve Jürgen Schmidhuber tarafından uzun-kısa süreli hafıza (LSTM) önerilmiştir [92].

2006 yılında makine öğrenmesinin en önemli alanlarından biri olan derin öğrenme konusunda önemli bir adım atılmıştır. Temeli 1960'lı yıllardaki kontrol teorisine dayanan geri yayılmalı yapay sinir ağları Geoff Hinton ve arkadaşları tarafından yapılan çalışmalarla başarısını göstermiştir [93-94].

2010 yılında video oyunları oynamayı öğrenebilen DeepMind isimli bir ağ geliştirilmiştir [95]. 2016 yılında DeepMind'ın İngiltere'deki hasta verilerini alarak tıbbi çözümler üretmesi gündeme geldi fakat bu konuda İngiltere'de verileri korumakla ilgilenen kurum olan UK Information Commissioners Office tarafından veri koruma yasasını ihlal ettiğine karar verilerek süreç durdurulmuştur [96].

2016 yılında DeepMind araştırmacıları tarafından geliştirilen AlphaGo isimli yazılım Go oyuncularına karşı yarışmış ve dünyanın en iyi oyuncularını yenmiştir [97].

Bu yıllarda akıllı telefonlar ve sosyal ağlarda birçok makine öğrenmesi uygulaması geliştirilmiştir.

Örneğin Facebook üzerindeki fotoğraflarda yüzlerin otomatik kutu içerisine alınması, fotoğraflara otomatik metin oluşturulması, yabancı dillerdeki paylaşımlar için otomatik çeviri yapılması bunlardan birkaçı olmuştur [98].

Benzer şekilde akıllı telefon kullanıcılarının mesaj yazarken karşılaştıkları otomatik kelime tamamlama da bir makine öğrenmesi uygulaması olarak her gün yüz milyonlarca kişi tarafından kullanılmaktadır.

Bugün ise internet ile bağlantılı; arama motorları, tavsiye motorları, spam algılama filtreleri, yetkisiz erişim algılama sistemlerinin gelişmesi makine öğrenme uygulamalarının kullanımında artışa yol açmıştır.

4.5. Makine Öğrenmesi Türleri

Makine Öğrenmesi kendi içerisinde Makine öğrenmesi temelde gözetimli, gözetimsiz ve pekiştirmeli olmak üzere 3 ana gruba ayrılmaktadır [99]. Gözetimli öğrenmede eğitim verisindeki girdilerin hangi çıktıyı vereceği işaretlenmektedir. Örnek olarak, araştırmacının elinde bulunan kedi ve köpek resimlerini ayrıştırmak istemesi düşünülebilir. Bu durumda ilk olarak eğitim seti ve test setini ayırdıktan sonra, eğitim setindeki kedileri kedi, köpekleri ise köpek olarak işaretleyerek modeli eğitilir. Bu işaretlenmiş verilerden makine öğrenme yaparak test setindeki resimlerin kedi mi yoksa köpek mi olduğunu tahmin eder [100].

Gözetimli öğrenme konusunda sıklıkla kullanılan algoritmalar; doğrusal regresyon, lojistik regresyon, en yakın komşular (KNN), yapay sinir ağları, destek vektör makinaları ve karar ağaçları şeklindedir.

Gözetimsiz öğrenmede ise veriler özel olarak etkilenmemektedir. İşaretlenmemiş veri üzerinden yapı tahmin edilmeye çalışılır. Örnek olarak bir markette süt alan kişilerin başka hangi ürünü alabileceğini tahmin etmek istenmesi düşünülebilir. Bunun için verideki kümelere bakarak süt alan diğer müşterilerin hangi ürünlerini aldıklarını öğrenilir ve beraber alınma yoğunluğuna göre bu ürünlere ağırlık çıkartılır. Artık işaretlenmemiş verimizdeki yapıyı tahmin edilebilmektedir [101]. Gözetimsiz öğrenme konusunda yoğun olarak kullanılan algoritmalar kümeleme, birliktelik kuralı ve temel bileşen analizidir.

Pekiştirmeli öğrenme ile bulunduğu ortamı algılayabilen ve yalnız başına kararlar alabilen sistemler geliştirmektedir. Bu sistemlerde ödül-ceza olarak tabir edilen fonksiyonlarla yapay zekâ aldığı kararları iyileştirebilmektedir [102]. Pekiştirmeli öğrenme konusunda en çok kullanılan algoritmalar Q-öğrenme yöntemi ve TD öğrenme yöntemidir [102].

Çizelge 4.1. Gözetimli Öğrenme Algoritmaları ve Gözetimsiz Öğrenme Algoritmaları

Gözetimli Öğrenme Algoritmaları	**Gözetimsiz Öğrenme Algoritmaları**
En Yakın Komşuluk → k-Nearest Neighbors (KNN)	Kümeleme → Clustering
Yapay Sinir Ağları → Artificial Neural Network (ANN)	Birliktelik Kuralları → Association Rules
Destek Vektör Makinaları → Support Vector Machine (SVM)	Temel Bileşen Analizi → Principal Component Analysis (PCA)
Karar Ağaçları → Decision Trees (DTs)	
Doğrusal Regresyon → Linear Regression	

4.6. Doğal Dil İşleme

İnsanlık tarihinde yaşanan bazı gelişmeler, insanlığın ilerlemesinde önemli değişimlere neden olmuştur. Dilin icadı da insanlığın ilerlemesindeki önemli sıçrayışlardan biridir. İnsanlar dil vasıtasıyla bir başkasıyla iletişim kurabilmiş ve derdini anlatabilmiştir. Yine toplumdaki kültür, bilim, sanat gibi birçok alandaki ilerlemelerin sağlamasında dilin önemli bir etkisi olmuştur. Dil, tabiri caizse "canlı"dır. Geçen yüzyıllar içerisinde insan ilişkilerinde değişimlere bağlı olarak dile farklı sözcükler eklenmekte veya bazı sözcükler kullanılmayarak unutulmaktadır. Benzer şekilde insanların dili kullanırken cümle kurma şekilleri de değişebilmektedir. Bu gibi sebeplerle önceki yüzyıllarda kaleme alınan metinler günümüzde okunduğunda, anlatmak istediği anlaşılamayabilmektedir. İletişim teknolojilerinin gelişmesiyle de dilin kullanımı çeşitli değişikliklere uğramıştır. Cep telefonlarının kullanımında SMS ismi verilen kısa mesajlarda karakter sınırlaması olması insanların daha çok kısaltmalar kullanarak mesajlarını göndermelerine sebep olmuştur. Bu süreçte "slm, nbr" gibi bugün çok kullanılan kısaltma ibareleri ortaya çıkmıştır.

Benzer şekilde Twitter'da tweet gönderirken uyulması gereken karakter limiti kullanıcıların bazen kısaltmalar kullanması, bazen noktalama işaretlerini kullanmaması, bazense sesli harfleri kullanmaması sonucunu doğurmuştur. Bu çalışmada da olduğu gibi Twitter üzerinde yapılan doğal dil işleme çalışmalarında bu bir zorluk olarak ortaya çıkmaktadır [103].

Doğal dil işleme çalışmaları bilişim teknolojilerindeki gelişmeler ile dil biliminin yolunun kesişmesiyle ortaya çıkmıştır. Doğal dil işleme olarak tanımlanan bu yeni alan ile bilgisayarların insanların konuşmaları ve yazdıklarının anlaşılması, benzer şekilde insanlar ile konuşarak veya yazarak iletişim kurmasını hedeflemektedir. Fakat bunun sağlanması için bilgisayarın, dilin tüm özelliklerini bilmesi gerekmektedir. Şöyle bir örnek düşünülebilir; akıllı telefonlar üzerinden bir dijital asistan ile konuşan kişi selam vermek için; "Merhaba" yerine "Naber", "Naptın", "Napıyon" gibi ifadeler kullanabilir veya bu ifadeleri şiveli bir şekilde söyleyebilmektedir. Bu durumların tamamında yapay zekanın söylenen ifadeyi anlaması ve ona göre cevap vermesi gerekmektedir. İşte bu yüzden dil bilimcilerin biçim bilimi, ses bilimi, söz dizimi, anlam bilimi gibi dil biliminin alt alanları üzerine de çalışmaları gerekmektedir.

Doğal dil bilimi çalışmalarının başlangıcı olarak 1950'li yıllar kabul edilmektedir. 1950 yılında Alan Turing ünlü Turing Testini yayınlamış ve makinanın da insan gibi davranabileceğine atıfta bulunulmuştur [104]. 1954 yılında 60'dan fazla Rusça kelime otomatik olarak İngilizce'ye çevrilmiştir [105]. Bu çalışma sonrasında teknoloji yazarlarının birkaç yıl içinde makine çevirisinin tüm sorunları çözeceğini iddia etse de süreç daha yavaş ilerlemiştir [106].

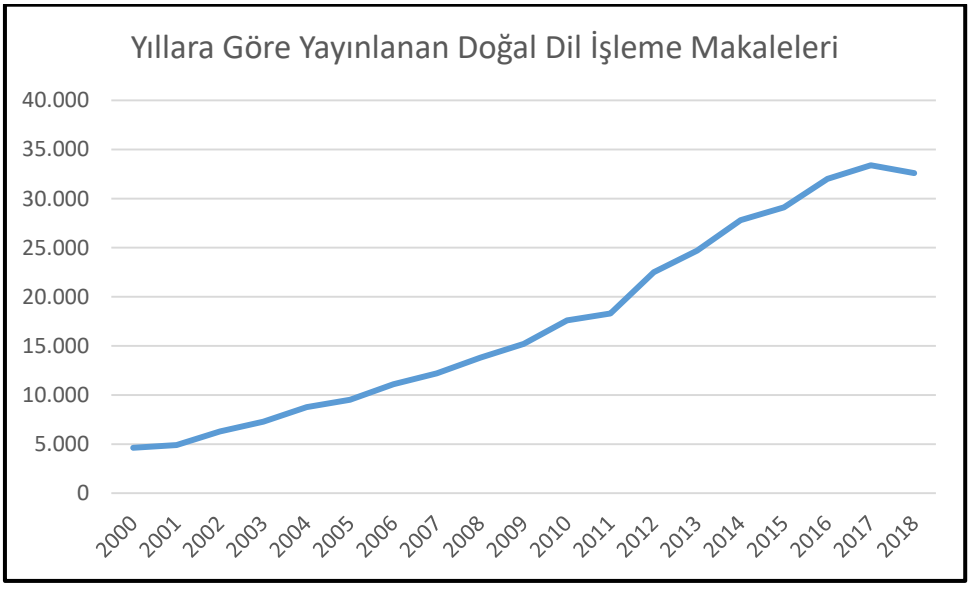
1966 yılı doğal dil işleme çalışmaları için can sıkıcı bir tarih olmuştur. ABD hükümeti tarafından genel olarak hesaplamalı dilbilim alanındaki ilerlemeyi ve özellikle makine çevirisini değerlendirmek amacıyla kurulan bir komisyon olan, ALPAC tarafından yayınlanan rapor ile on yıllık çalışmaların başarısızlığa uğradığı dile getirilmiş ve makine çevirisi konusundaki çalışmaları popülerliğini nispeten kaybetmiştir [107]. 1970'li yıllarda çeşitli sohbet botları geliştirilmiş fakat bunlar beklenen ilgiyi görmemiştir. 1980'li yıllarda ise makine öğrenmesi algoritmalarının doğal dil işleme süreçlerinde kullanılmasıyla devrim niteliğinde gelişmeler yaşanmıştır. Bu tarihten itibaren çeşitli doğal dil işleme yazılımları geliştirilmiştir. Ön plana çıkan bazı yazılımlar Çizelge 4.2'de gösterilmiştir.

Çizelge 4.2. Doğal dil işleme çalışmaları

Projenin İsmi	Yılı	Geliştirici
Plot Units	1981	Lehnert
Jabberwacky	1982	Rollo Carpenter
Mumble	1982	McDonald
Racter	1983	William Chamberlain ve Thomas Etter
Moptrans	1984	Lytinen
Kodiak	1986	Wilensky
Absity	1987	Hirst

2006 yılına gelindiğinde ise IBM tarafından geliştirilen Watson isimli yapay zekâ ABD televizyonlarında 1964 yılında yayınlanmaya başlayan ve popüler bir kelime yarışması olan Jeopardy'de en iyi oyuncuları yenmeyi başarmıştır [108]. Gelişen teknoloji altyapısı, internetteki veri hacminin artması gibi sebeplerle doğal dil işleme çalışmaları her geçen gün daha da gelişmiştir.

Google Scholar veri tabanına göre; doğal dil işleme alanını içeren makalelerin sayısı 2000 yılında 4 630 iken, bu sayı 2018 yılında 30 000'in üzerine çıkmıştır [109]. Google Akademik veri tabanındaki doğal dil işleme makalelerinin yıllara göre değişimi çalışma kapsamında oluşturulan Şekil 4.4.'da verilmiştir.



Şekil 4.4. Yıllara göre yayınlanan doğal dil işleme makaleleri

Süreç içerisinde farklı kullanım alanları da doğan doğal dil işleme, yazım yanlışlarının düzeltilmesi, dijital ortamdaki metinlerin sesli olarak okunması, metinlerin özetlerinin çıkarılması, metinlerden istenilen konudaki kanaatin çıkarılması, konuşmanın metne dönüştürülmesi, soru yanıtlama, çeviri yapma alanlarında kullanılabilmektedir.

Fakat bu projelerin geliştirilmesinde elbette çeşitli zorluklar vardır. Özellikle kuralsız ve anlaşılmaz konuşmalar veya yazılar bu problemlerin başında gelmektedir. Twitter gibi bazı platformlarda ise sitenin konseptinden dolayı oluşan konuşma jargonu bir zorluk olarak geliştiricinin karşısına çıkmaktadır.

4.7. Duygu Analizi

Duygu analizi veya diğer adıyla fikir madenciliği bir metindeki duygusal durumları ve öznel bilgileri tanımlamak, ölçümlemek ve analiz etmek için doğal dil işleme, metin analizi gibi metodların kullanılması anlamına gelmektedir. Örneğin doğru çalışan bir duygu analizi yazılımı "Uçaklardan nefret ederim." cümlesini negatif "Uçaklara bayılırım." cümlesini pozitif, "Uçaklar yolcu taşır." cümlesini ise nötr olarak yorumlamaktadır.

Duygu analizindeki yaklaşımlar üç ana kategoride toplanmaktadır. Bunlar bilgi bazlı teknikler, istatistik metodlar ve hibrid yaklaşımlardır. Bilgi bazlı tekniklerde mutluyum, üzgünüm, korktum, sıkıldım gibi bazı durum bildiren kelimelere göre kategorilere ayrılır. Bazı bilgi bazlı tekniklerde sadece bu listeler kullanılmaz, rastgele olarak seçilen kelimelere de belli duygusal yakınlıklar atanır [110].

İstatistiksel yöntemler ise gizlenmiş anlamsal analiz bulma yoluna gider. Bunun için "sözcük torbası" olarak çevrilebilecek "bag of words" veya derin öğrenme gibi metodlar kullanmaktadır [111].

Hybrid yaklaşım ise hem makine öğrenmesi hem de semantik ağlar gibi gibi bilgi sunumu (knowledge representation) yöntemlerini beraber kullanarak duyguları çıkarmaya çalışmaktadır [112].

Duygu analizi çalışmaları birçok çalışmada kullanılmaktadır. Örneğin, Ali Reza Alaei ve arkadaşları tarafından yapılan çalışmada turistlerin duygu, düşüncelerini ölçümlemek, Salud MaríaJiménez-Zafra ve arkadaşları tarafından yapılan çalışmada doktorlar ve ilaçlar hakkındaki hastaların düşüncelerini öğrenebilmek amacıyla kullanılmıştır [113-114].

Benzer şekilde ekonomiden marka yönetimine, siyasetten pazarlamaya kadar birçok alanda kullanılabilmektedir. Bu çalışmada metinlerin vektörlere dönüştürülmesi için Tfidf Vectorizer kullanılmıştır.

4.8. Metin Madenciliği

Metin madenciliği, veri kaynağı olarak metni esas alan veri madenciliği çalışmalarına verilen isimdir. Metin madenciliğinde, veri madenciliğine benzer şekilde, ilginç kalıpların tanımlanması ve araştırılması yoluyla veri kaynaklarından faydalı bilgiler çıkarma amaçlanmaktadır.

Farklı olarak ise görseller, ses kayıtları, videolar gibi farklı veri kaynaklarında değil belge koleksiyonlarında çalışılır. Bundan dolayı da veri madenciliği alanındaki gelişmeler metin madenciliğini de dolaylı olarak etkilemektedir.

Metin madenciliğinin kullanıldığı alanlara metinlerden konu başlıklarının elde edilmesi, farklı alanlardaki metinlerin sınıflandırılması, metnin özetinin çıkarılması örnek verilebilir. Bunların dışında çalışmamıza konu olan metinlerden görüş analizi yapılması da metin madenciliğinin çalışma alanıdır.

Yüksek miktarda metin verisi içeren Twitter üzerinde de metin madenciliği çalışmaları sıklıkla yapılmaktadır [115].

Bunun yanında metin madenciliği çalışmaları güvenlik amacıyla [116], biyomedikal çalışmalarda [117], dijital suçların tespitinde [118], sosyal ağ uygulamalarında, iş ve pazarlama çözümlerinde [119], duygu analizinde [120] ve bilimsel literatür madenciliğinde [121] kullanılmaktadır.

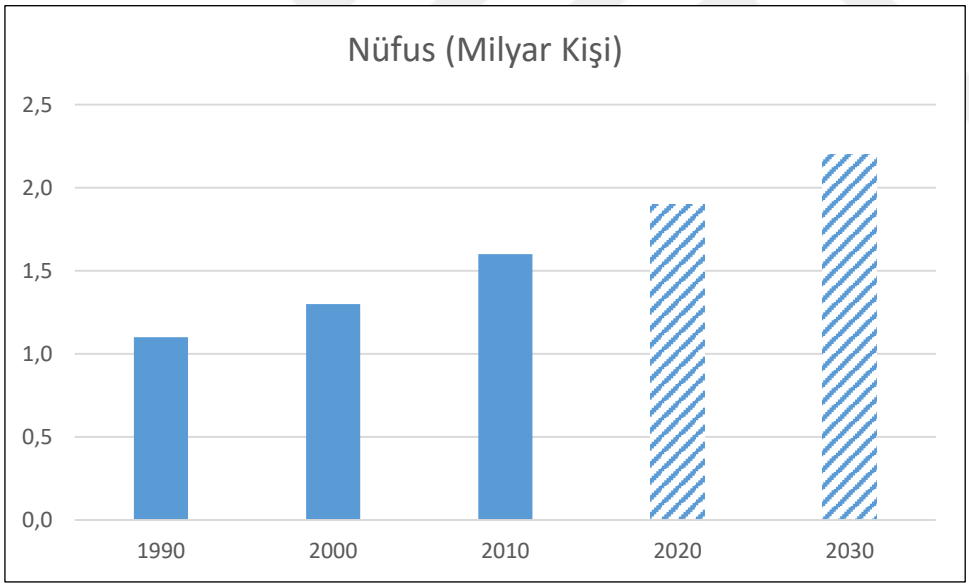
Bu çalışmada ise metin madenciliği ve doğal dil işlemenin ortak çalışma alanı olan metinler üzerinden istatistiksel inceleme yapmak için TF-IDF yöntemi kullanılmıştır. İngilizcedeki Term Frequency – Inverse Document Frequency (Terim frekansı – ters metin frekansı) olarak geçen sözcüklerin ilk harflerinden oluşan yöntem, bir metinde geçen terimlerin çıkarılması ve bu terimlerin geçtiği miktara göre çeşitli hesapların yapılması üzerine kurulmuştur [122]. Bir terim diğer dokümanlarda ne kadar çok olursa TF-IDF değeri düşmektedir. Aynı şekilde bir terim sadece kendi dokümanında ne kadar çok olursa TF-IDF değeri artmaktadır.

5. DENEYSEL ÇALIŞMA

Bu bölümde çalışma kapsamında yapılan deneysel çalışma ele alınmıştır. Deney kapsamında kurulan hipotez Twitter üzerindeki Tweetlerin İslamofobik olup olmadığının makine öğrenmesi metodlarıyla belirlenebileceğidir. Deneysel çalışma kısmında sunulan ve daha önceden görülmemiş veride Bayes Regresyonu ile ulaşılan %89'luk doğruluk bu hipotezi ispatlamıştır.

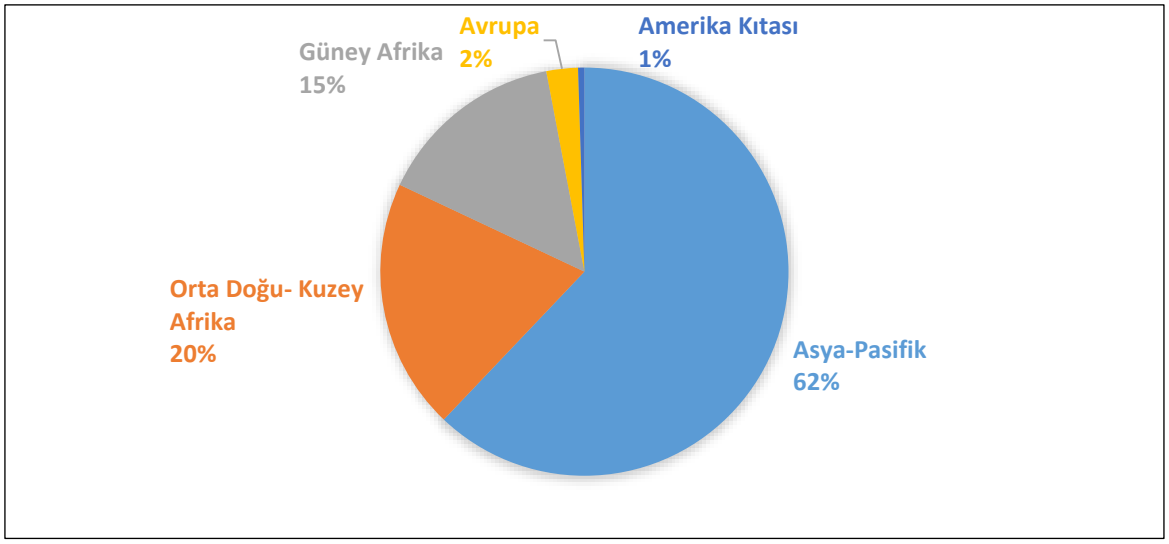
5.1. Problemin Tanımı

İslam, dünyada en çok inananı olan dinlerden biridir. Pew Research tarafından yapılan araştırmaya göre 2020 yılında 1,9 milyar, 2030 yılında ise 2,2 milyar Müslüman olması beklenmektedir. Bu da dünya nüfusunun %26,4'ünün Müslüman olacağı anlamına gelmektedir [123]. Dünyadaki Müslüman nüfusunun bugüne kadar olan sayıları ile 2020 ve 2030 yıllarındaki tahmini sayılar Şekil 5.1.'de verilmiştir.



Şekil 5.1. Dünya'daki Müslüman nüfusu ve gelecek tahminleri [123]

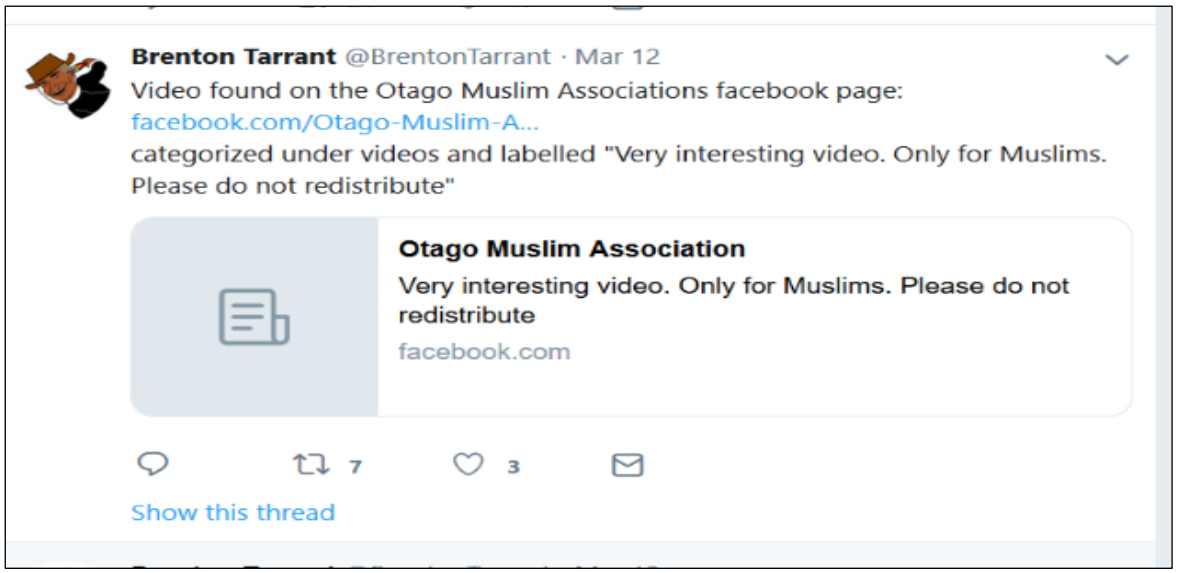
Müslümanlar ağırlıklı olarak Asya-Pasifik ülkelerinde yaşamaktayken bunu sırasıyla Orta Doğu- Kuzey Afrika ülkeleri, Güney Afrika ülkeler, Avrupa ve Amerika kıtası izlemektedir. Yüzdesel olarak karşılaştırma Şekil 5.2.'de verilmiştir [124].



Şekil 5.2. Müslümanların yaşadığı bölgeler

Bu rakamlar göstermektedir ki Müslümanlar her geçen yılda nüfuslarını artırmakta ve tüm kıtalarda varlıklarını sürdürmektedirler. Fakat Müslümanların her alanda olması diğer inanç topluluklarında İslamofobi olarak tanımlanan bir İslam korkusu veya endişesini gündeme getirmektedir. Bu kaygılar bazen ekonomik bazen ise kültürel olurken, kaygının nefrete dönüşmesi durumunda Müslümanlar terörist veya sapkın olarak görülebilmektedir. Çalışmanın İslamofobi bölümünde ele alınan dünyanın her yanından örnekler İslamofobi'nin birçok farklı şekilde ortaya çıkabileceğini göstermektedir. Bu şekillerden biri de çevrimiçi ortamda oluşan nefret söylemidir. Kullanıcılar bazen anonim kalabileceği inancıyla bazen de anlık bir tepkiyle Müslüman karşıtı söylemlerini veya Müslümanlara karşı olan endişelerini bu platformlardan gündeme getirebilmektedirler.

Yeni Zelanda'da onlarca Müslüman'ın hayatını kaybetmesiyle sonuçlanan saldırıda teröristin saldırı öncesi tarihlerde İslam karşıtı Tweetler attığı bilinmektedir. Örneğin; Resim 5.1.'de saldırgan, Müslüman nüfusunun arttığını ve bunun ileride diğer insanlar için bir tehdit olacağını anlatan videoyu paylaşmıştır [125].



Resim 5.1. Yeni Zelanda saldırganın saldırı öncesi islamofobik paylaşımı

Sadece bu olay bile sosyal ağlardaki İslam karşıtı söylemin önceden tespiti ve analizinin ne kadar önemli olduğunu göstermektedir. Fakat Twitter'da bulunan verinin hacminin devasa boyutlarda olması bu nefret söyleminin elle tespitini imkânsız kılmaktadır. Bunun yerine metinden anlam çıkarılabilecek metin madenciliği ve doğal dil işleme çalışmaları yapılarak, makine öğrenmesi algoritmalarından faydalanılarak otomatik bir sistem kurulması nefretin önceden tespiti ve yeni saldırıların olmaması, can kayıplarının yaşanmaması için zaruri hale gelmiştir. Çalışma da esas olarak çevrimiçi ortamdaki İslamofobik söylem problemini ele almakta ve buna makine öğrenmesine dayalı çözüm önerisi sunmaktadır.

5.2. Sistemin Tanımı

Çalışmada dört uygulama katmanı olan bir sistem kullanılmıştır. Bu katmanlar sosyal medya arama katmanı, veri tabanı yönetim katmanı, doğal dil işleme modülü ve sınıflandırma modülüdür.

Resim 5.2. Arama Modülü

5.2.1. Sosyal medya arama katmanı

Sosyal medya platformlarında geliştiriciler için API (Application Programming Interface) ismi verilen özel araçlar kullanılmaktadır. Bu yolla sosyal medya platformları geliştiricilerin hangi verileri kullandığını görebilmekte ve veri aktarımına çeşitli sınırlamalar getirerek kullanıcılara verilen hizmetin aksamasının önüne geçebilmektedir. Bu çalışma esnasında da yukarıda değinildiği gibi Twitter API'si kullanılmıştır. Oluşturulan uygulamada Twitter içeriklerini okuma izni alınmıştır. Bu izin içeriği herkese açık hesapları kapsamaktadır. Kilitli profillerin içeriklerine ulaşılamamaktadır. API izinleri alındıktan sonra, PHP tabanlı olarak geliştirilen yazılım yardımıyla belirlenen anahtar sözcüklere göre Twitter verileri toplanmıştır. Bu toplama işleminde zamana dayalı bir iş zamanlayıcısı olan Cron Job fonksiyonu kullanılarak, periyodik olarak veriler toplanmakta ve gelen sonuçlar veri tabanına yazılmaktadır. Kullanım sırasında Twitter'ın verdiği limitler doğrultusunda işlem yapılabilmektedir. Bu sebeple belirli sürelerin sonunda Cron Job sayesinde yazılım tekrar çalışmakta ve limitlere takılmadan veri toplanabilmektedir.

5.2.2. Veri tabanı yönetim katmanı

Veri tabanı yönetim katmanı MYSQL üzerine iki farklı tablo üzerine veriler eklenmiştir. Mysql seçilirken PHP ile sorunsuz çalışması göz önünde bulundurulmuştur [126]. İlk olarak anahtar kelime yapısına göre İslamofobik olma ihtimali yüksek Tweetler bir tabloya, İslamofobik olmama ihtimali yüksek Tweetler ise başka bir tabloya eklenmiştir.

Buradaki tablo seçimi esnasında "muzrad" gibi Müslümanlara hakaret etmek için kullanılan kelimeleri içeren Tweetler İslamofobik olma ihtimali yüksek tabloya eklenirken, "mashallah" benzeri Müslüman jargonundaki sözcükler İslamofobik olma ihtimali düşük tabloya eklenmiştir. Bu şekilde işaretleme sırasında 5 gönüllü kişinin işlemleri daha pratik bir şekilde yapabilmesi sağlanmıştır.

5.2.3. Doğal dil işleme modülü

Doğal dil işleme modülünde Python dili kullanılmıştır. Python modülünün Mysql ile bağlantısı web servis kullanarak sağlanmıştır. İlk olarak metin ve etiket olarak veriler birleştirilmiştir.

Ardından metni sayıya dönüştürmek için 'kelime' bazlı ve hesaplamak için Scikit-learn kütüphanesinin içerisinde bulunan TfidfVectorizer sınıfı kullanılmıştır. Scikit-learn yapay zekâ alanında en yaygın olarak kullanılan ve birçok regresyon ile sınıflandırma yöntemini içeren bir kütüphanedir.

Bu aşamada eğitim kümesinde geçen kelimeler TfidfVectorizer metodu ile terim frekansı (tf), ters belge sıklığı (idf) özellik matrisine dönüştürülmüştür.

5.2.4. Sınıflandırma modülü

Metod seçimi ardından %80-%20 oranıyla veri eğitim ve test olarak ikiye ayrılmıştır ve gözetimli öğrenme kullanılarak eğitilmiştir. %80-%20 seçiminde Scikit-learn tarafından önerilen bölme oranları esas alınmıştır [127]. Bu işlemin ardından eğitim seti 'Negatif' ve 'Pozitif' durumları için küçük varyanslı tahminler için öncelikle uygun olan iki ayrı lineer Ridge Modeli eğitilmiştir. Eğitilen bu modeller 'Negatif.pkl' ve 'Pozitif.pkl' isimleri ile kaydedilmiştir. Ridge Regresyonu veya bir diğer ismiyle Tikhonov normalizasyonu regresyon katsayılarının tahmininde, bağımsız değişkenlerin birbirleri üzerindeki etkilerini minimum yapmayı hedefleyen ve kararlı katsayı tahminleri elde edebilmeyi sağlayan bir yöntemdir. Ridge regresyonunun ardından aynı eğitim ve test seti Naive Bayes Sınıflandırıcıyla ve son olarak derin öğrenme yöntemiyle eğitilmiştir. Naive Bayes sınıflandırıcı Twitter üzerindeki duygu analizlerinde çeşitli araştırmalarda kullanılan olasılıkçı bir yaklaşımdır. Derin öğrenme ise çoklu soyutlama yapısı ile verinin temsillerini öğrenmek için bir araya getirilmiş çoklu işleme katmanlarında oluşan bir yaklaşımdır.

Sınıflandırma modülünde eğitilen modeller ile test verisi tahmin edilmeye çalışılmıştır.

Tahmin sonuçları ile gerçek sonuçların karşılaştırma matrisi hesaplanmış ve çeşitli ölçütlerle karşılaştırmalar yapılmıştır. Bu ölçütler kesinlik, duyarlılık, f1-score türleridir. Kesinlik ölçütü Pozitif olarak tahmin edilen bir durumdaki başarıyı gösterir.Duyarlılık pozitif durumların ne kadar başarılı tahmin edildiğini gösterirken, F1 ölçütü ise Duyarlılık ve Kesinlik ölçütlerinin çarpma işlemine göre terslerinin, iki değerin toplamına bölümü ile hesaplanan harmonik ortalamasıdır.

5.3. Verilerin Toplanması ve İşaretlenmesi

Veri toplama işleminde, toplanan veriler olgusal ve yargısal olamak üzere iki ana gruba ayrılmaktadır. Twitter üzerindeki kişinin kullanıcı adı, doğum tarihi, tam ismi, kullandığı cihazlar, bir tweeti beğenme veya retweet yapma durumu gibi bilgiler kişisel yargılardan bağımsız olması nedeniyle olgusal veriler olarak değerlendirilebilir. Diğer yandan twitter kullanıcısının paylaştığı Tweetlerin içeriği ise yorum gerektirdiğinden yargısal veriler grubuna girmektedir.

Çalışma sırasında hem olgusal hem de yargısal veriler toplanmıştır. Olgusal veriler kelime bulutlarının çıkarılması, listelerin oluşturulması gibi geleneksel programlama işlemlerinde kullanılırken, yargısal veri olan tweet içeriği makine öğrenmesi algoritmalarında değerlendirilmiştir.

Daha detaylıca bakılacak olursa Twitter'ın sağladığı programlama arayüzü vasıtasıyla Tweetler toplanırken her Tweet için Tweet'in içeriğinin yanı sıra o tweetin id numarası, diyalog id numarası, oluşturulma tarihi, oluşturulma zamanı, kullanıcının bu tweeti atarken kullandığı zaman dilimi, kullanıcı adı, tam ismi, konumu, içerisinde bir başka kişiden bahsedilip bahsedilmediği, bir web sayfasına bağlantı bulundurup bulundurmadığı, içerdiği fotoğraf, kaç cevap verildiği, kaç retweet aldığı, içerisinde bir etiket kullanılıp kullanılmadığı, içerdiği video bilgileri de otomatik olarak elde edilmektedir.

```
In [3]: def etiket_analizi(tweetler):
            import pandas as pd
            id_list = [tweet.id for tweet  in tweetler]
            tablom = pd.DataFrame(id_list, columns = ["id"])

            tablom["metin"] = [tweet.text for tweet in tweetler]
            tablom["gonderim_zamani"] = [tweet.created_at for tweet in tweetler]
            tablom["rt_edilmismi"] = [tweet.retweeted for tweet in tweetler]
            tablom["rt_sayisi"] = [tweet.retweet_count for tweet in tweetler]
            tablom["ekran_ismi"] = [tweet.author.screen_name for tweet in tweetler]
            tablom["kullanici_takipci_sayisi"] = [tweet.author.followers_count for tweet in tweetler]
            tablom["kullanici_konumu"] = [tweet.author.location for tweet in tweetler]
            tablom["etiket_kullanimi"] = [tweet.entities.get('hashtags') for tweet in tweetler]

            return tablom
```

Resim 5.3. Tweet fonksiyonu örneği

Burada bahsi geçen ID numarası tekil bir numaradır ve kullanıcı ismini, kullanıcı adını, fotoğrafını değişse dahi bu ID numarası kullanılarak Tweet'in önceden hangi hesaptan atıldığı tespit edilebilmektedir. Çalışmada da tweetlerin çekilmesi sırasında ID numarası aynı gelen tüm tweetlerin kopyaları silinerek 1 adet veri tabanında bırakılmıştır.

Oluşturulma tarihi Tweetlerin atıldığı tarihi gösterir. *Since* ve *For* parametreleri kullanılarak belli tarihler arasındaki veriler toplanabilmektedir.

Örneğin “Since 2016-01-01 islam” araması bize 1 Ocak 2016’dan itibaren gönderilen ve İslam kelimesini içeren Tweetleri verirken, “For 2016-01-01 İslam” ise 1 Ocak 2016 tarihine kadar gönderilen Tweetleri verecektir. Veri setinin homojen oluşması için çalışma esnasında bu parametreler kullanılarak farklı zaman dilimlerinden Tweetler alınmıştır.

Tweet’in herhangi bir kişiden bahsedip bahsetmediği bilgisi ise çalışma sırasında nefret söyleminin doğrudan bir kişiyi hedef alıp almadığı veya bir kişiye atıf yapılarak oluşturulup oluşturulmadığının analizi sırasında kullanılmıştır. Bu yolla oluşturulan sistemde #StopIslam etiketine gönderilen Tweeetlerde en çok @ABC kişisinden bahsediliyorsa ABC isimli kullanıcının durumu özel olarak incelenebilecektir.

Tweet’in herhangi bir etiket içerip içermediği bilgisi çalışma sırasında bir kelime bulutu oluşturulurken kullanılmıştır. Kelime bulutları basitleştirilmiş bir grafik olduğundan nefret söyleminin yoğunlaştığı kişi, kavram, konum gibi bilgileri pratik olarak görmeyi sağlamaktadır.

Tweet’in herhangi bir fotoğraf veya video içerip içermediği bilgisi bir sıralama fonksiyonu içerisinde kullanılmıştır. Bu nefret söyleminin yayılırken en çok kullandığı fotoğraf ve video içerikler görülebilecek ve uzman kişi tarafından bu içerik incelenebilecektir. Burada herhangi bir görüntü işleme çalışması yapılmamaktadır. Fakat aynı görsel veya video kaynağının yayılması URL (Uniform Resource Locator) adresi üzerinden takip edilmektedir. İlerleyen aşamalarda görüntü işleme de değerlendirme kapsamına alınabilecektir.

Tweet’in kaç retweet veya favori aldığı bilgisi yine bir sıralama fonksiyonu içerisinde kullanılarak hızlı bir şekilde yayılan nefret söyleminin aynı hızda tespiti hedeflenmiştir. Sosyal medyada bazen dezenformasyon amaçlı bir içerik hızlı bir şekilde yayılarak algının gerçeğin önüne geçmesine sebep olabilmektedir. Bu tarz durumlarda erken tespit ile yanlış bilginin yayılması engellenebilmektedir.

Verinin toplanmasının ardından ilk olarak ön işleme adımları uygulanmıştır.

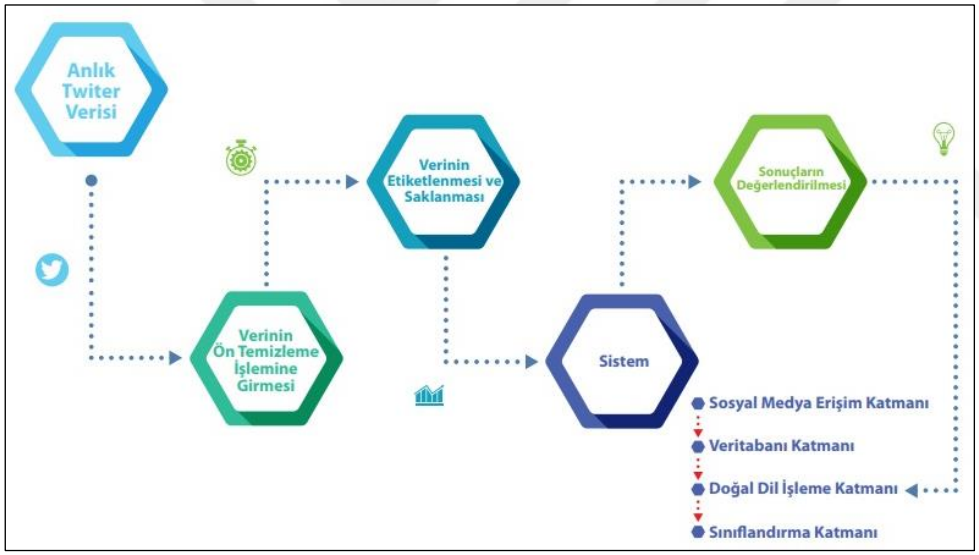
Bu ön işleme adımları, bütün kelimelerin küçük harfe dönüştürülmesi, özel karakterlerin, emojilerin, bağlantı adreslerinin silinmesidir. Ayrıca İngilizce istenmesine rağmen Twitter API’den kaynaklanan sorunlar sebebiyle farklı dillerde gelen Tweetler de veri setinden çıkarılmıştır.

Çıkarılma işleminin ardından 5 kişi gönüllü olarak Tweetleri İslamofobik veya İslamofobik Değil şeklinde işaretlemiştir. İşaretlenme sonrasında bir Tweet için hangi görüş 3 veya üstü kanaatte ise bu görüş kabul edilerek eğitim setine bu şekilde eklenmiştir.

5.4. Genel Sistem Mimarisi

Sistemin genel mimarisinde Twitter üzerinden Twitter programlama ara yüzü ile belirtilen sınırlara riayet ederek verilerin PHP programlama dili kullanılarak alınır ve veriler SQL veri tabanı üzerinde saklanmaktadır. Twitter'dan alınan Tweet'leri içeren bu verilere web ara yüzü üzerinden Python programlama dili yardımıyla ön işleme işlemleri uygulanmaktadır. Ön işleme işlemi uygulanan veriler okunma kolaylığı açısından bir Microsoft Excel dosyasına aktarılmaktadır.

Gönüllü kişilerce burada etiketleme işlemi yapıldıktan sonra veri seti eğitim ve test veri seti olmak üzere ikiye ayrılmaktadır. Eğitim seti belirlenen makine öğrenmesi algoritmalarıyla eğitilir ve test seti üzerinde bu eğitimin sonuçları test edilebilmektedir. Sistemin genel gösterimi Şekil 5.3.'te verilmiştir.



Şekil 5.3. Sistemin gösterimi

5.5. Kullanılan Araçlar ve Yazılımlar

Bu bölümde deneysel çalışmanın yapılması sırasında Python, Numpy, Pandas, Matplotlib, Jupyter notebook, PHP, MySQL gibi yazılım dili, kütüphaneler ve geliştirme ortamları hakkında bilgiler ele alınmıştır.

5.5.1. Python

Python etkileşimli ve nesne yönelimli yüksek seviyeli bir programlama dilidir. Farklı veri türleri ile işlem yapabilmesi, geliştiriciler tarafından sunulan modülleri desteklemesi ve yazım dilinin sade olması tercih edilirliğini artırmaktadır. Geliştirilen Python kodları Windows, Mac ve Unix türevlerinde çalışabilirken, C C++ gibi programlama dilleri ile etkileşime giren ara yüzler de barındırmaktadır [128]. Python programlama dili ismini genelde düşünülenin aksine bir yılandan değil İngiliz mizah grubu Monty Python'ın, Monty Python'ın Uçan Sirki isimli televizyon dizisinden almıştır [129]. İlk stabil versiyonu 1994 yılının ocak ayında yayınlanan programlama dili iki farklı seride geliştirilmeye devam etmektedir. 2019 yılının mart ayında 2 serisinin 2.7.16 numaralı versiyonu geliştirilirken 3 serisinin de 3.7.3 versiyonu dağıtıma sunulmuştur. Çalışmada Python programlama dili ile makine öğrenmesi modelleri inşaa edilmiştir.

5.5.2. Numpy

Python da veri odaklı etkili olarak çalışabilmek için veriyi tutma ve üzerinde işlem yapmayı iyi anlamak gerekmektedir. Bu konuda öne çıkan kütüphanelerden biri Numpy kütüphanesidir. Numpy, numerical python ifadesinin kısaltmasıdır. Bilimsel hesaplamalar için kullanılır ve Python'daki veri saklama maliyetlerini düşürür. Diziler veya çok boyutlu matrisler üzerinde yüksek performanslı çalışma imkânı sağlamaktadır [130]. Kütüphanenin temelleri 1995 yılında Python programlama dilinin de kurucusu olan Guido van Rossum'un içinde bulunduğu bir topluluk tarafından atılmış ve nihai sürümü 2005 yılında Travis Oliphant tarafından hayata geçirilmiştir [131]. Çalışmada Numpy ile tweetlerden oluşan veri setinin tutulması ve üzerinde işlem yapılması sağlanmıştır.

5.5.3. Pandas

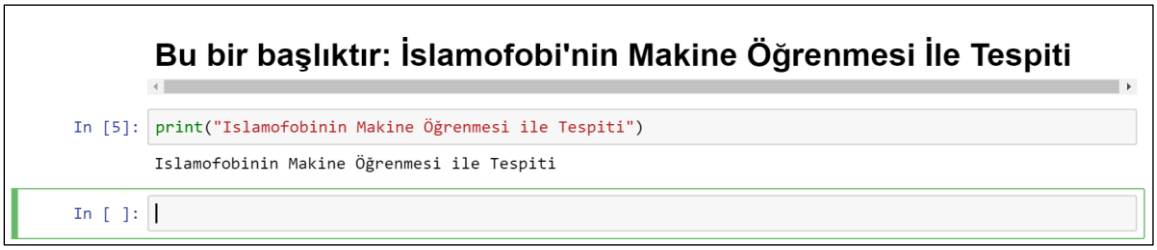
Numpy kütüphanesi array işlemlerinde son derece başarılı olmasına rağmen yapısal anlamda satırlar ve sütunların olduğu veri analitiği işlemlerinde zayıf kalabilmektedir. Pandas Numpy'nin eksik kaldığı bu yönlerin geliştirildiği bir kütüphanedir [132]. Çalışmada veriseti üzerindeki işlemlerde bu kütüphaneden destek alınmıştır.

5.5.4. Matplotlib

Numpy ve Pandas gibi kütüphaneler ile yapılan veri işlemlerinin nihai sonucunun grafiksel olarak ekrana yansıtılması için bir kütüphaneye ihtiyaç duyulmaktadır.

Matplotlib bu konuda geliştiricilere yardımcı olan önemli kütüphanelerden birisidir. Matplotlib sayesinde veriler görselleştirilerek hem iki boyutlu hem de üç boyutlu grafikler üretilebilmektedir [133].

5.5.5. Jupyter notebook



Resim 5.4. Jupyter notebook kullanım örneği

Tez çalışmasında Jupyter notebook ile Tweetlerin elde edilmesi ve analiz edilme işlemleri yapılmıştır. Örnek bir kullanım aşağıda verilmiştir. İlk olarak Tweepy isimli Python kütüphanesi vasıtasıyla aşağıdaki görselde görüldüğü üzere bir sorgulama yapılmıştır. Tweepy uygulaması Resim 5.5.'de verilmiştir.

```
In [2]: sonuclar = api.search(q = "islam",
                             lang = "en",
                             result_type = "popular",
                            count = 1000)
```

Resim 5.5. Tweepy kütüphanesi ile tweetlerin elde edilmesi

Tweetler üzerinde inceleme kolaylığının yapılması açısından bu işlemin ardından pandas kütüphanesi yardımıyla veriler bir DataFrame üzerine aktarma fonksiyonu yazılmıştır. Fonksiyon Resim 5.6.'de verilmiştir.

```
In [3]: def etiket_analizi(tweetler):
            import pandas as pd
            id_list = [tweet.id for tweet  in tweetler]
            tablom = pd.DataFrame(id_list, columns = ["id"])

            tablom["metin"] = [tweet.text for tweet in tweetler]
            tablom["gonderim_zamani"] = [tweet.created_at for tweet in tweetler]
            tablom["rt_edilmismi"] = [tweet.retweeted for tweet in tweetler]
            tablom["rt_sayisi"] = [tweet.retweet_count for tweet in tweetler]
            tablom["ekran_ismi"] = [tweet.author.screen_name for tweet in tweetler]
            tablom["kullanici_takipci_sayisi"] = [tweet.author.followers_count for tweet in tweetler]
            tablom["kullanici_konumu"] = [tweet.author.location for tweet in tweetler]
            tablom["etiket_kullanimi"] = [tweet.entities.get('hashtags') for tweet in tweetler]

            return tablom
```

Resim 5.6. Tweet fonksiyonu örneği

Son olarak fonksiyonumuz çağrılarak tablomuz Jupyter Notebook içerisine yansıtılmıştır. Sonuçların örneği aşağıdaki görseldeki şekildedir. Tweetlerin tablo şeklinde gösterimi Resim 5.7.'de verilmiştir.

Out[5]:

	id	metin	gonderim_zamani	rt_edilmismi	rt_sayisi	ekran_ismi	kullanici_takipci_sayisi	kullanici_konumu
0	1144941855517188096	Deoband has issued a fatwa against TMC MP Nusr...	2019-06-29 12:13:05	False	2123	ippatel	69796	फतेहपुर,उत्तर प्रदेश,भारत
1	1145189600421224448	Pakistan doesn't care about Islam, they only c...	2019-06-30 04:37:31	False	701	Imamofpeace	460819	Australia
2	1145243458099191808	'Dangal' star Zaira Wasim quits Bollywood. Cla...	2019-06-30 08:11:32	False	486	TarekFatah	518098	Toronto, Ontario
3	1144985275535245312	We had suspected Congress manifesto was drafte...	2019-06-29 15:05:37	False	188	madhukishwar	2064744	New Delhi, India
4	1145174646108540929	Darul Uloom Deoband is most respected institut...	2019-06-30 03:38:06	False	213	DivyaSoti	17031	India

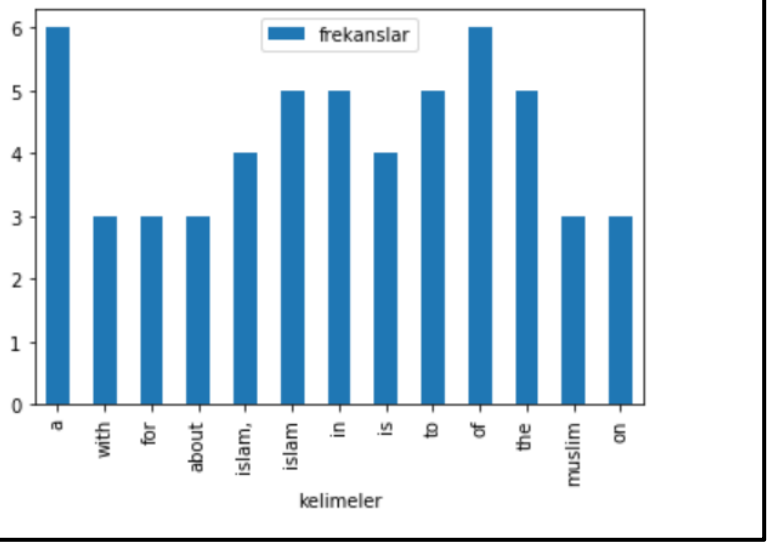
Resim 5.7. Tweetlerin tablo şeklinde gösterimi

Bu tablo üzerinden sıralama, yoğunluk ölçümü, kelime bulutu oluşturma gibi işlemler yapılabilmektedir.

Tweetlerde geçen kelime frekansları üzerinden Matplotlib kütüphanesiyle bir grafik aşağıdaki şekilde oluşturulabilir. Kelime frekanslarına göre örnek gösterim Şekil 5.4.'de verilmiştir.

Şekil 5.4. Yoğunluklarına göre tweetlerde geçen kelimeler

```
In [15]: cizim.plot.bar(x = "kelimeler",y = "frekanslar");
```



Bu kullanıcılar takipçi sayılarına göre sıralanabilmektedir. Takipçi sayısına göre paylaşım yapan kişilerin sıralaması Resim 5.8.'de verilmiştir.

Out[41]:

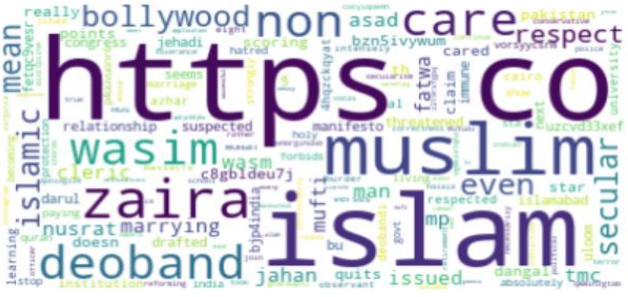
	ekran_ismi	kullanici_takipci_sayisi
3	madhukishwar	2064744
2	TarekFatah	518098
1	Imamofpeace	460819
6	ZaidZamanHamid	258906
11	PamelaGeller	200315
13	OpIndia_com	152893
5	ARanganathan72	148096
7	jihadwatchRS	123216
10	DVATW	89280
0	ippatel	69796

Resim 5.8. Takipçi sayısına göre paylaşım yapan kişilerin sıralaması

Tweetlerde geçen kelimelerden kelime bulutu oluşturulabilmektedir. Kelime bulutu gösterimi Resim 5.9.'de verilmiştir.

```
In [54]: kelimebulutu = WordCloud(background_color = "white").generate(metin)
         #Kelimebulutu oluştu. Şimdi çizelim.

         plt.imshow(kelimebulutu, interpolation = "bilinear")
         plt.axis("off")
         plt.tight_layout(pad = 0)
         plt.show()
```



Resim 5.9. Kelime bulutunun oluşturulması

5.5.6. PHP

PHP programlama dili çalışmanın ilk aşamasında Tweetlerin, Twitter programlama ara yüzü kullanılarak elde edilmesi aşamasında kullanılmıştır. Geliştirilmeye 1994 yılında başlanan Rasmus Lerdorf isimli mühendisin kendi kişisel web sayfasını yönetmek için Perl betiklerinin temel alarak geliştirmesiyle başlamıştır. Diğer kullanıcılara duyurulması ise 1995 yılının haziran ayında hata ayıklama süreçlerinin hızlandırılması ve dilin iyileştirilmesi ile olmuştur [134].

Bu tarihten itibaren çeşitli güncelleştirmeler ile yeni versiyonunu yayınlayan dil son versiyonu 7.3 sürümü olarak 2018 yılının aralık ayında yayınlanmıştır. Bir sonraki sürümün ise 2019 yılının Kasım ayında yayınlanması planlanmaktadır [135]. Bu çalışmada PHP'nin son sürümü olan 7.3 versiyonu kullanılmıştır.

PHP, HTML işaretleme dili içerisinde kolaylıkla kullanılabilmektedir. Örneğin aşağıda örneği verilen, HTML etiketleri arasında yer alan <?php echo '<p>Bu bir cümledir. </p>'; ?> ifadesiyle ekrana "Bu bir cümledir." cümlesi yazdırılabilmektedir.

```
<body>
   <?php echo '<p>Bu bir cümledir. </p>'; ?>
</body>
```

Çalışmada yapılan PHP programlama dili ile yapılan Twitter bağlantısı yapılmış ve MYSQL veri tabanı yapısı kullanılarak kaydedilmektedir

5.5.7. MySQL

Çalışmada verilerin toplanması sırasında kullanılan PHP programlama dili ile veriler Mysql veri tabanı kullanılarak kaydedilmiştir. Mysql çok kullanıcılı bir veri tabanı yönetim sistemidir. Sistem, çalışmada tercih edilen Windows işletim sistemi için ücretsiz olarak kullanılabilmektedir. C ve C++ dilleri kullanılarak 1994 yılında geliştirilmeye başlanan Mysql'ın ilk sürümü İsveç şirketi MySQL AB tarafından 1995 yılında yayınlanmıştır [136]. Bu tarihten itibaren çeşitli versiyonları yayınlanan veri tabanı sisteminin 2018 yılının Nisan ayında 8.0 sürümü kullanıcılarla paylaşılmıştır. Çalışmada bu sürüm tercih edilmiştir [137].

Mysql üzerinde işlemlerin daha hızlı yapılabilmesi için phpMyAdmin isimli ücretsiz ve açık kaynak kodlu bir yazılım mevcuttur.

Bu araçla CSV ve SQL formatındaki dosyalar kolayca içe ve dışarı aktarılabilmekte, veri tabanı tablosu ve kullanıcıları oluşturulabilmektedir. Çalışmada phpMyAdmin aracılığıyla öncelikle tablo ve kullanıcı oluşturulmuş, kullanıcıya yazma, okuma, değiştirme, silme gibi tüm veri tabanı izinleri verilmiştir.

5.6. Veri Seti ve Önişleme

Tez kapsamında yapılan deneysel çalışmada ele alınan veri seti 290 bin tweetin ön işleme adımlarından geçirilerek azaltılması sonucu elde edilen 162 bin tweetten oluşmaktadır. Tweetlerin elde edilmesi aşamasında ekonomi, sağlık gibi özel bir konu belirlenmemiş aksine "Islam", "Muslim" gibi kelimeleri farklı konulardan Tweetler alınmıştır. Bu seçim sırasında İslamofobi'nin sadece belirli bir alan üzerinden değil sosyal hayatın tüm evrelerinde oluşabilmesi düşünülmüştür. Bir kişi ekonomik kaygılarla Müslüman karşıtı olabilmekteyken, başka bir kişi kültürel kaygıları sebebiyle İslamofobik söylem kullanabilmektedir.

Veriler ilk etapta PHP programlama dili yardımıyla ile temin edildiğinde ilk olarak Tweetler'de bazı istenmeyen durumlar ile karşılaşılmıştır:

- İngilizce talep edilmiş olmasına rağmen farklı dillerde gelen Tweetler
- Aynı ID'ye sahip Tweetler
- Farklı ID'ye fakat aynı metinsel içeriğe sahip Tweetler
- Pornografi veya kumar içeriği olan fakat sadece "islam" kelimesini bir tür çok görüntülenme taktiği olarak kullanan Tweetler

Bu Tweetlerin de dahil olduğu veri seti yaklaşık 290 bin Tweet içermektedir. Fakat daha doğru bir çalışma olması için bu Tweetler kümeden çıkarılmıştır. Kalan 170 bin Tweet içerisinde; emojilerin ve özel karakterlerin silinmesi, internet sayfası bağlantılarının temizlenmesi gibi bazı veri önişleme işlemleri yapılmıştır.

Bu işlemin ardından 170.000 Tweet Islamfobi konusunda çalışma yapan, alanında uzman 5 kişi tarafından işaretlenmiştir. Bu İşaretleme sonucunda 3 veya daha üstü görüş hangisi olarak verilmişse bu esas kabul edilerek veri seti işaretlenmesi tamamlanmıştır. Fakat bu aşamada veri setinin %53 olumlu %47 olumsuz olduğu görülmüş ve dengeli bir şekilde oluşması için rastgele bir seçimle 8 bine yakın olumlu içerikli Tweet çıkarılarak veri seti 162 bin Tweet'e indirilmiştir. Bu durumda veri setinin %51,25 olumlu %48,75 olumsuz tweetlerden oluşur hale gelmiştir.

İşaretlenen bu Tweetler'de doğal dil işleme adımlarının atılması öncesinde; bütün kelimelerin küçük harfe dönüştürülmesi, noktalama işaretlerinin silinmesi benzeri bazı veri ön işleme işlemleri yapılmıştır. Bu aşamada Jupyter notebook'tan faydalanılmıştır. Resim 5.10'de örnek ön işleme adımlarının nasıl yapıldığına dair kod yapısı verilmiştir. Bu işlemler sırasında Python programlama dili ve Jupyter Notebook ara yüzünden faydalanılmıştır.

```
In [ ]: #buyuk-kucuk donusumu
        df['metin'] = df['metin'].apply(lambda x: " ".join(x.lower() for x in x.split()))

        #noktalama işaretleri
        df['metin'] = df['metin'].str.replace('[^\w\s]','')

        #sayılar
        df['metin'] = df['metin'].str.replace('\d','')
```

Resim 5.10. Tweet ön işleme fonksiyonları

Metnin istenilen şekle gelmesinin ardından stopwords olarak adlandırılan ve Türkçe karşılığına "dolgu sözcük" denilebilecek bazı kelimelerin silinmesi aşamasına gelinmiştir. Dolgu sözcükler tek başına bir anlam ifade etmeyen, doğal dil işleme çalışmalarında çalışmanın doğruluğunu etkilemeyen fakat donanımı güç ve performans harcama yönünden yavaşlatabilen unsfardır. Kavramın daha da iyi anlaşılması açısından bu kelimelere Türkçeden örnek olarak "acaba, ama, aslında, az, bazı, belki, biri, birkaç, bir şey" kelimeleri verilebilir.

Bilgisayar bilimi öncülerinden Hans Peter Luhn tarafından 1959 yılında bir sunum sırasında "stoplist" ismiyle kullanılmış geçen yıllar içerisinde isim değiştirerek "stopwords" olarak kullanımı sürmektedir [138].

Geçen yıllarda stopwords kullanımı için birçok kütüphane geliştirilmiştir. Python programlama dilinde en popüler olan kütüphanelerden biri, diğer doğal dil işleme fonksiyonlarını da barındıran NLTK kütüphanesidir [139]. NLTK kütüphanesi 1991 yılında geliştirilmesine Python üzerinde başlanan, sınıflandırma, etiketleme, köklendirme, ayrıştırma, anlamsal akıl yürütme gibi birçok yürütme işlevini destekleyen bir kütüphanedir. Farklı diller için "stopwords" listeleri barındıran kütüphanenin çalışma sırasında İngilizce listesi kullanılmıştır.

5.7. Algoritmaların Veri Seti Kullanılarak Eğitimi

Makine öğrenmesinde model eğitimi bir makine öğrenme algoritmasının üzerinde çalıştığı verileri öğrenerek yeni tahminlerde bulunmasıdır. Modelin eğitimi makine öğrenmesinin özüdür. Modelin verilere tam olarak uymaması durumunda ürettiği sonuçlar doğru olmayacaktır. Benzer şekilde eğitimin fazlalaşması overfitting ismi verilen soruna sebep olacak ve tabiri caizse makinenin eğitim setini ezberlemesine sebep olacaktır.

Modelin eğitimi sırasında uygulanan yapay öğrenme, bilgisayar algoritmalarının verileri kullanarak karar verme yeteneğini kazanması olarak tanımlanabilir ve model oluşturmada istatistik kuramını kullanmaktadır [140]. Yapay öğrenmede amaç eldeki verilerden faydalanarak kendi kararını verebilen algoritmaları oluşturmadır. Yapay öğrenme genel olarak 3 ana gruba ayrılmaktadır. Bunlar gözetimli, gözetimsiz ve pekiştirmeli öğrenmedir.

Örnek olarak, Twitter'ın site üzerindeki Tweetleri spam veya spam değil diye ayırmaya çalışan bir algoritma geliştirmek istediğini düşünülürse, bunun için ilk olarak elindeki spam olarak etiketlediği Tweetler ile spam olmayan Tweetleri makinaya öğretecektir.

Ardından da yeni gelen bir Tweet'in spam olup olmadığını tahmin edebilecek ve eğer spam ise kullanıcıların zaman akışından gizleyerek sitede daha kaliteli vakit geçirmelerini sağlayabileceklerdir. Bahsi geçen örnek makine öğrenimi bölümünde ele alınan gözetimli öğrenme kapsamındadır.

Twitter'ın elindeki milyonlarca Tweet'i kendi içerisindeki gizli örüntüleri kullanarak bunları haber, spor, magazin, bilim gibi konulara otomatik olarak ayırması ise gözetimsiz öğrenmeye örnek olacaktır. Bu örneklerin ötesinde pekiştirmeli öğrenme için ise şöyle bir örnek verilebilir. Facebook'a 3 fotoğraf yüklendiğinde Facebook'un bu fotoğrafların her birinde Facebook hesabı olmamasına rağmen annenizin yüzünü kare içine alması, bir fotoğrafa isim yazınca diğerlerine otomatik olarak bu ismi tanımlaması bir pekiştirmeli öğrenme örneğidir.

Bu çalışma kapsamında toplanan veriler, eğitim ve test seti olarak ayrılmıştır. Eğitim setindeki veriler gözetimli öğrenme uygulanmak üzere işaretlenmiştir. Gözetimli öğrenme elde bulunan bağımsız değişkenleri kullanarak bağımlı bir değişkenin bir özelliğini tahmin etmek için kullanılabilmektedir. Bu kapsamda veriler ilk olarak Ridge Regresyonu ardından Bayes Ridge regresyonu ve son olarak da derin öğrenme yöntemi modeli ile eğitilmiştir.

5.7.1. Ridge regresyonu

Tahminler için sıklıkla kullanılan ve çalışmada kullanılan Ridge Regresyonu'nun temellerini oluşturan yöntem "En Küçük Kareler Yöntemi" dir [141]. Matematikte bir bağımlı değişken olan x ile bağımsız değişken olan y arasındaki ilişki sunan temel denklem Eş. 4.1.'de görülen doğru denklemidir.

$$y = \beta_0 + \beta_1 X_i + \varepsilon_i \quad (4.1)$$

Formülde;

y= Bağımlı değişken

β_0 , β_1=Tahmin edilecek parametrelerdir.

X_i= Bağımsız değişken

ε_i = Rastgele hata

Buradaki doğru denklemini bulurken veri kümesindeki noktlar ile doğru arasındaki uzaklıkların kareleri toplamını en aza indirecek doğru denklemini bulmaya çalışılır. Bu yöntem en küçük kareler yöntemi olarak isimlendirilmektedir. Eş. 4.2.'de görülen temel denklem doğru denklemindeki ε_i hatasını denklemde yalnız bırakıp her hata için kareler toplamını aldığımızda aşağıdaki formüle ulaşılmaktadır.

$$q = \sum_{i=1}^{n} [y - (a + bx)] = \sum_{i=1}^{n} [y - a - bx]^2 \quad (4.2)$$

Ridge regresyonunda X formülünde belirtilen bu temel regresyonuna ek olarak en küçük kareler farkı terimi olarak tanımlanan q'ya yeni bir terim eklenmektedir. Eş. 4.3.'de görülen bu yolla bağıntıya katkısı olmayan değişkenlerin beta katsayıları küçültülmeye çalışılır.

$$q + a \sum \beta_j{}^2 \quad (4.3)$$

Bu yeni denklemde X numaralı denklemin sol tarafında yer alan q'ya $a \sum \beta_j{}^2$ eklenen ifadesindeki a katsayısını doğruluğu yükseltmek için ayarlama yapan bir ceza terimi olarak düşünülebilmektedir. Eğer bu terim 0'a yaklaşırsa $a \sum \beta_j{}^2$ kısmı da 0 olacağından sonuç en küçük kareler yöntemi ile aynı olacaktır.

Aksi durumda ise a katsayısı çok büyük değerlere ulaştığında ise beta katsayıları sıfıra yaklaşacağı için hatalı bir sonuç bulunabilecektir. Dolayısıyla Ridge Regresyonunda a değerinin doğru seçimi önemlidir.

Ridge modeli Python'da *from sklearn.linear_model import Ridge* ifadesiyle içe aktarılmaktadır. Ardından Tfidf ile vektör haline getirilen veri seti, %80'i eğitim %20'si test için kullanılacak şekilde ayrılmıştır. Eğitim seti kullanılarak model eğitilmiş ve test kısmındaki tweet sonuçları tahmin edilmiştir. Çalışma sırasında sınıflandırıcı tanımlanırken Resim 5.11'da görülen kod çalıştırılmıştır.

```
for i,class_name in enumerate(class_names):
    train_target = [ get_score(x == class_name) for x in train_labels ]
    classifier = Ridge(alpha=1.0, copy_X=True, fit_intercept=True, solver='auto',
                       max_iter=120,   normalize=False, random_state=0,  tol=0.0025)
    classifier.fit(train_features, train_target)
    prediction[:,i] = classifier.predict(test_features)
    joblib.dump(classifier, class_name+'.pkl')
```

Resim 5.11 Ridge regresyonu ile eğitim

Buradaki parametrelere yakından bakılacak olursa;

Alpha: Teorik kısımda bahsi geçen alfa değeridir. Eğer 0 seçilirse Ridge Regresyonu en küçük kareler farkına eşit olacaktır. Çok büyük bir değerde ise beta katsayıları anlamsız kalabilecektir. Bu nedenle dikkatli seçilmelidir. Eğitim sırasında 1.0 değeri seçilmiştir.

Copy_x: X kopyalanasın mı yoksa üzerine mi yazılsın sorusuna cevap vermektedir. Çalışmada True diyerek kopyalanması istenmiştir.

Fit_intercept: Bu model için kesişim noktası hesaplanıp hesaplanmayacağı durumu bildirilir. False olarak seçildiğinden hesaplamada hiçbir kesişim noktası kullanılmayacaktır.

Solver: Veri türüne bağlı olarak çözücünün nasıl seçileceğini bildirir. Auto parametresinin yanında svd, cholesky, lsqr, sparse_cg, sag ve saga şeklinde alternatifleri vardır. Bu alternatiflerin her biri farklı matematiksel yaklaşımlar içerir. Auto seçimi ise verinin tipine bakarak otomatik bir seçim yapmasını sağlar.

Max_iter: Gradyan izdüşürme için maksimum yineleme sayısını verir. Varsayılan değeri 1000'dir. Eğitim sırasında 120 alınmıştır.

Normalize: Bu parametre, fit_intercept parametresi False olarak seçildiğinden göz ardı edilir. Fakat True olarak seçilme durumlarında regresyonda normalleştirme işlemi yapılmasını sağlar.

Random_state: Bu değer rastgele sayı üretecini başlatmak için kullanılmaktadır. Eğer bu değer belirtilmezse, rastgele sayı üreteci her çalışmasında rastgele bir değer kullanacaktır. Sabit bir değer belirleyerek yazılımın her çalıştırılmasında aynı sayı dizisinin belirlenmesi sağlanır.

Eğitim sırasında verilen 0 değeri de bu kapsamda düşünülebilir. 0 yerine 34,42,123 gibi başka rastgele bir sayı da verilebilir.

Tol: Çözümün hassasiyet değerini belirtir. Eğitim sırasında 0,0025 olarak seçilmiştir.

5.7.2. Bayes regresyonu

Bayesçi regresyon yöntemi, normal dağılımda yer alan rastlantı değişkenleri için Bayes teoreminden yararlanılarak oluşturulmuştur. Bayes teoremi 18. Yüzyılda ismini aldığı Thomas Bayes tarafından bulunan ve olasılık kuramı içerisinde incelenen koşullu olasılık hesaplama formülüdür [142]. Teorem bir rassal değişken için olasılık dağılımı içinde koşullu olasılıklar ve marjinal olasıklar arasındaki ilişkiyi göstermektedir. Bayes formülü Eş. 4.4 de verildiği şekildedir.

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)} \tag{4.4}$$

Burada bahsi geçen terimlere şu şekilde bakılabilir:

P (A | B) = B olayı gerçekleştiğinde A olayının gerçekleşme olasılığı

P (A) = A olayının gerçekleşme olasılığı

P (B | A) = A olayı gerçekleştiğinde B olayının gerçekleşme olasılığı

P (B) = B olayının gerçekleşme olasılığı

Bayes teoreminin makine öğrenmesi modeli olarak kullanılması Naive Bayes sınıflandırıcı ile olmaktadır. Bu model gözetimli öğrenme alanında bir yaklaşımdır.

Bayes lineer regresyonu, lineer regresyona bayesçi bir yaklaşımdır. Regresyon modelinin normal dağılıma sahip hataları olduğunda, Bayeçi liner regresyon ile modelin parametrelerinin sonsal olasılık dağılımları için iyi sonuçlar elde edilebilmektedir. Bayesçi bakış açısında nokta tahminlerinden ziyade olasılık dağılımları kullanılarak lineer regresyon formüle edilir. Bağımsız değişken sonucu tek bir değeri tahmin etmez fakat olasılık dağılımından bir değer bulmaya çalışılır.

Bayesçi lineer regresyon kullanımı için ilk olarak Sklearn.linear_model kütüphanesi içerisinden BayesianRidge sınıfı içe aktarılmıştır. Ardından Tfidf ile vektör haline getirilen veri seti, %80'i eğitim %20'si test için kullanılacak şekilde ayrılmıştır. Eğitim seti kullanılarak model eğitilmiş ve test kısmındaki Tweet sonuçları tahmin edilmiştir. Modelin eğitimi Resim 5.12.'da verilmiştir.

```
prediction = np.zeros((len(test_labels),len(class_names)))
for i,class_name in enumerate(class_names):
    train_target = [ get_score(x == class_name) for x in train_labels ]
    classifier = BayesianRidge()
    classifier.fit(train_features.toarray(), train_target)
    prediction[:,i] = classifier.predict(test_features)
    joblib.dump(classifier, class_name+'_B.pkl')
```

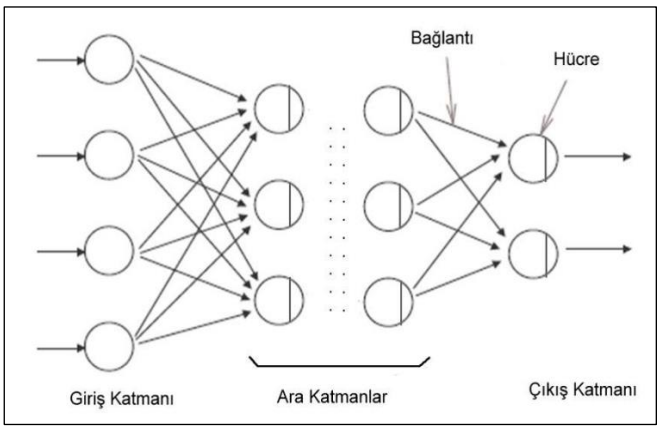
Resim 5.12. Bayes lineeri ile eğitim

Bayesçi yaklaşımda eğitimin uzun sürmesi sebebiyle bu kısımda bir makine değişikliğine gidilmiş ve daha yüksek geçici belleğe sahip bir makine ile eğitim tamamlanmıştır.

5.7.3. Derin öğrenme

Derin öğrenme veriden bilgi çıkarmak için birden çok katmanlı bir yapı kullanarak yüksek kararlılıkta bilgi, kavram ve nesne tespiti yapabilen bir makine öğrenme sınıfıdır. Derin öğrenme çalışmalarında hem gözetimli hem de gözetimsiz öğrenme uygulanabilmektedir [143].

Yapay Sinir Ağları derin öğrenmenin temelini oluşturmaktadır. YSA insan beyninden esinlenerek tasarlanmıştır, insan beyninde olduğu gibi nöronlardan oluşur. Bu nöronların tümü birbirine bağlıdır ve çıktıyı etkilemektedir. Örnek bir yapay sinir ağı modeli Şekil 5.5.'de verilmiştir.



Şekil 5.5. Örnek bir yapay sinir ağı

Nöronlar üç farklı katmana ayrılır:

- Giriş Katmanı
- Gizli Katman(lar)
- Çıkış Katmanı

Giriş katmanı: Fonksiyonu giriş verilerini almaktır. Bu çalışmada giriş katmanı Twitter platformu üzerinden paylaşılan tweetlerden oluşmaktadır.

Gizli katmanlar: Bu katmanda giriş katmanında alınan verilerde matematiksel hesaplamalar yapılmaktadır. Derin öğrenmedeki "derin" ifadesi birden fazla gizli katman ibaresinden gelmektedir. Kaç gizli katman olacağına geliştirici karar vermelidir. Bu çalışmada katman sayısı 64 olarak seçilmiştir.

Çıkış katmanı: Gizli katmanlarda yapılanan matematiksel hesaplamalar sonrasında oluşan çıktı verilerini göstermektedir. Bu çalışmada çıktı verisi bir Tweet'in İslamofobik söyleme sahip olup olmadığıdır.

Bir yapay sinir ağı şu çerçevede çalışmaktadır. İlk olarak sinir ağı oluşturulup eğitim verileri hazırlanmakta, ardından algoritmayı başlatmak için ağırlıklar (weight) değerlerine sıfır olmayacak şekilde rastgele değerler atanmaktadır. Eğitim verileri ile ileri besleme uygulanır. Geri yayılım algoritması ile ağdaki her bir ağırlık değeri için kısmi türevler bulunur. Meyilli azalım algoritması uygulanarak her bir ağırlık güncellenir. Bu güncellemeler sonrasında maksimum iterasyon sayısı ile bir minimum değere ulaşılmışsa eğitim tamamlanmış olur. Aksi durumda ise eğitim verileri ile ileri besleme adımına tekrar dönülür.

Derin öğrenme ağlarında ağların karmaşıklığı parametre sayısı ile belirtilmektedir. Derin öğrenme ağlarında parametre sayısı bazen 10 milyarı dahi aşabilmektedir [144].

Bu ağlarda her katmanda bulunan nöronlar önceki katmanın çıktılarına, girdi vektörüne ve ağırlıklar kümesine bakarak bir aktivasyon fonksiyonu üretir. Sigmoid, ReLU, Leaky ReLU ve Softmax gibi farklı isimlerle olabilen aktivasyon fonksiyonları önceki katmanın bütün girdilerini alarak bir çıktı değeri üretir ve sonraki katmana aktarır. Bu çalışmada softmax aktivasyon fonksiyonu kullanılmıştır.

Softmax aktivasyon fonksiyonu derin öğrenme ağı tarafından üretilen ağırlık değerlerinden faydalanarak olasılık temelli kayıp değeri üretmektedir. Aktivasyon işlemi sonucunda test girdisinin her bir sınıfa ait benzerliği için olasılık değeri üretilir. Derin öğrenme ağının çıktı

olarak verdiği skor değerler normalize edilmemiş değerlerdir. Softmax aktivasyon fonksiyonu bu değerleri normalize ederek olasılık değerlerine dönüştürür.

Bu dönüştürme işlemini en çok olabilirlik (maximum likelihood) fonksiyonuyla yapmaktadır. Softmax aktivasyon fonksiyonu, değerlerin büyük olmasından ötürü likelihood fonksiyonun logaritmasını alarak dağılımın yapısını korumaktadır.

Aktivasyon fonksiyonu seçiminin yanında, derin öğrenme çalışmalarında önemli bir parametre seçimi de hangi optimizasyon algoritmasının uygulanacağıdır. Bu algoritmalar doğrusal olmayan problemlerin çözümünde optimum değeri bulmakta kullanılmaktadır. Yaygın olarak kullanılan optimizasyon algoritmaları rmsprop, stochastic gradient descent, adagrad, adadelta, adam, adamax'tır. Bu çalışmada rmsprop isimli optimizasyon algoritması kullanılmıştır. Bu fonksiyon ilk olarak Geoffrey Hinton tarafından bir çevrimiçi kursta önerilen indirgeme tabanlı optimizasyon tekniğidir. Rmsprop, harekeretli ortalama kare indirgemelerini kullanmaktadır. Bu yolla öğrenmenin kötüleşmesinin önüne geçilmektedir.

Çalışmada eğitim sırasında Keras modülü kullanılmıştır.

Keras Python programlama dilinde geliştirilmiş açık kaynaklı bir sinir ağı kütüphanesidir. TensorFlow, Theano, PlaidML gibi farklı yapılar üzerinde çalışabilmektedir. Keras modülü, görüntü ve metin verileriyle çalışmayı kolaylaştırmak için, katmanlar, aktivasyon fonksiyonları, optimize ediciler gibi sinir ağlarının yapı taşları olan parametrelere dair çok sayıda uygulama içermektedir. Bu çalışmada da derin öğrenme ile eğitim sırasında Keras modülünden faydalanılmıştır.

Bir keras modeli oluşturulurken gizli katman sayısı, her katmandaki düğüm sayısı, kullanılacak aktivasyon fonksiyonu, kayıp fonksiyonu ve optimizasyon yöntemi gibi parametreler belirlenerek model yapısı oluşturulmaktadır.

```
In [24]: model_checkpoint = ModelCheckpoint('best.h5', monitor='val_loss', save_best_only=True,verbose=1)
         redu = ReduceLROnPlateau(monitor='val_loss', factor=0.5, patience=4, verbose=1, mode='min', min_lr=0.00001)
         early = EarlyStopping(monitor='val_loss', min_delta=0, patience=17, verbose=1, mode='min')
         cbacks = [model_checkpoint,early,redu]

In [25]: model = Sequential()
         model.add(Bidirectional(LSTM(charlen, return_sequences=True), input_shape=(maxlen,charlen)))
         model.add(Bidirectional(LSTM(charlen)))
         model.add(Dense(outlen))
         model.add(Activation('softmax'))
         model.compile(loss='categorical_crossentropy', optimizer='rmsprop',metrics=['acc'])

In [26]: hs = model.fit(x_train, y_train, batch_size=64,
                        epochs=150,validation_data=(x_test,y_test),callbacks=cbacks)
```

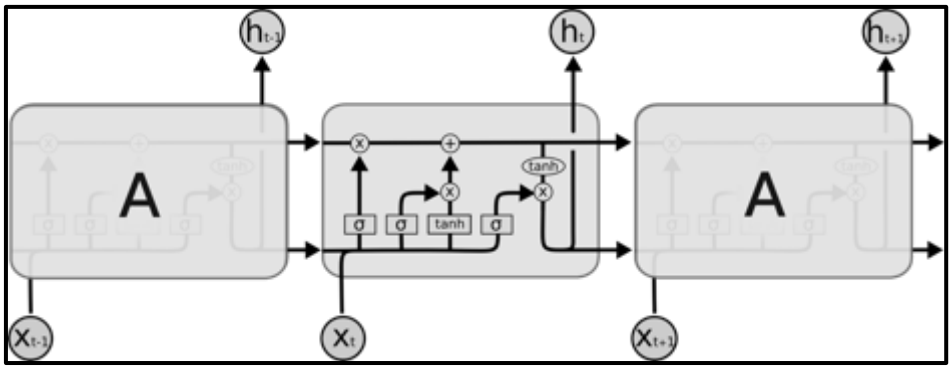
Resim 5.13. LSTM kullanımı

Çalışmada Keras ile model oluştururken Resim 4.11.'de verilen kod yapısı uygulanmıştır. Model oluşturulurken kullanılan batch_size parametresi ağırlıkların güncellenmesinin gerçekleştiği ağa verilen alt örneklerin sayısını ifade etmektedir. Genellikle 32 değeri varsayılan kabul edilmekle beraber bu sayının katları olan 64 128 ve 256 da kullanılır. Çalışmada verinin büyüklüğü de göz önünde bulundurularak 64 kullanılmıştır. Değer arttıkça daha doğru gradyan değerleri hesaplanabilir ve linerizasyon azaltılabilir. Batch_size gibi eğitim sırasında kullanılan bir başka parametre de epoch'tur. Eğitim turu anlamına gelen epoch, eğitim sırasında tüm eğitim verilerinin ağa gösterilme sayısıdır. Model eğitilirken verilerin büyüklüğünden dolayı hepsi aynı anda eğitime katılamayabilmektedir. Belirlenen sayıda parçalar halinde eğitim süreci gerçekleşmektedir. İlk belirlenen parça eğitildikten sonra elde edilen başarıma göre ağırlıklar belirlenir ve ardından geri yayılım ile başta rastgele verilen ağırlıklar güncellenmektedir. Her eğitim turunda ağırlıklar bu şekilde bir önceki ağırlıklar güncellenecek şekilde gerçekleşmektedir. Bu işlem tekrarlanarak model için en uygun ağırlık değerleri hesaplanmaktadır. Doğruluk azalmaya başlayana kadar epoch sayısı artırılır, doğruluğun en yüksek olduğu yer ideal epoch sayısını vermektedir. Bu çalışmada epoch sayısı 150 olarak belirlenmiştir[145].

Modelin eğitimi sırasında harf bazlı Uzun Kısa Vadeli Hafıza Ağları (LSTM) kullanılmıştır. LSTM'in özelliği uzun vadeli bağımlılıkları öğrenebilen bir tekrarlayan sinir ağı türü olmasıdır. İlk olarak 1997 yılında Sepp Hochreiter ve Jürgen Schmidhuber tarafından önerilmiştir. Bu tarihten itibaren birçok projede kullanılan yöntem metin sıkıştırma, el yazısı tanıma, konuşma tanıma gibi birçok alanda başarılı uygulamalar ile gündeme gelmiştir [146] [147].

LSTM yapısının tekrar eden sinir ağlarından farkı tek bir sinir ağı katmanı değil, özel bir şekilde bağlı 4 farklı katmana sahip olmasıdır. Bu katmanlar kapı olarak isimlendirilmektedir. Normal akışın dışında dışarıdan da bilgi alabilen bir yapıdır. Alınan bilgiler depolanabilir, hücreye yazılabilir, okunabilmektedir.

Hücre neyi depolayacağını, ne zaman okumasına, yazmasına veya silmesine izin vereceğini kapılar sayesinde karar verebilmektedir. Bu kapılarda bir ağ yapısı ve aktivasyon fonksiyonu içermektedir. Tıpkı nöronlarda gerçekleştiği şekilde gelen bilgi ağırlığına bakılarak geçirilmekte veya durdurulmaktadır. Bu ağırlıklar tekrarlayan ağın öğrenmesi sırasında hesaplanmaktadır. Bu yapı ile hücre veriyi kullanıp kullanmayacağı öğrenmektedir. Örnek ağ yapısı Şekil 5.7.'da verilmiştir.



Şekil 5.7. LSTM ağ yapısı örneği

5.8 Deney Sonuçları

Çalışmada Google tarafından ücretsiz olarak sağlanan ve Tesla K80 ekran kartı ile GPU desteği sağlanan Google Colab isimli arayüz kullanılarak önce veri seti içinde ardından daha önceden görülmemiş Tweetler üzerinde doğruluk tahmini yapılmıştır.

5.8.1 Veri seti için test sonuçları

Çalışmada Ridge Regresyonu, Bayesçi Ridge Regresyonu ve derin öğrenme kullanılarak model eğitilmiştir. Bu eğitimlerde veri setinin %80'lik kısmı kullanılmış ve %20'si test için ayrılmıştır. %80-%20 seçiminde Scikit-learn tarafından önerilen bölme oranları esas alınmıştır. Bu bölünme işleminde random_state değişkeni 0 değişkenine eşitlenmiştir. Bu yolla testler sırasında veri setinin farklı bölümlerinin eğitim ve test seti olarak ayrılmasının önüne geçilmiştir.

Veri seti içerisindeki tweetler için alınan sonuçlar ilk olarak her model için hata matrisinde gösterilmiştir. Hata matrisi makine öğrenimi alanında ve özellikle de istatistiksel sınıflandırma problemlerinde modelin performansının görselleştirilmesine olanak sağlayan bir matristir. Örnek bir hata matrisi Çizelge 5.1'de verildiği gibidir.

Çizelge 5.1. Hata matrisi örneği

		Gerçek Sonuç	
		False	True
Tahmini Sonuç	False	1	2
	True	3	4

Bu matris False ve True boolean değerlerini sınıflandırmakta kullanılan modelin eğitim sonrasında test sonuçlarını göstermektedir. Gerçekte False olan bir değer False olarak tahmin edilmiş, fakat False olan 3 değer True olarak tahmin edilmiştir. Benzer şekilde True olan 4 değer True olarak tahmin edilirken, 2 adet True olan değer ise False olarak tahmin edilmiştir. Bu matrisi False veya True değil de False mı yoksa False değil mi şeklinde yeniden düzenlediğinde görselde verilen tablo elde edilecektir. Bu matris örneği Çizelge 5.2'de verilmiştir.

Çizelge 5.2. False veya Not False değil sonuçlu hata matrisi örneği

		Gerçek Sonuç	
		False	Not False
Tahmini Sonuç	False	1(TP)	2(FP)
	Not False	3(FN)	4(TN)

Tablo hücrelerinde sayıların yanında yazan TP, FP, FN ve TN ifadeleri hata matrislerinde sıklıkla kullanılan True Positives, False Positives, False Negatives, True Negatives ifadelerinin kısaltmalarıdır.

True Positives = Doğru Pozitifler

False positive = Hatalı Pozitifler

True negative = Doğru Negatifler

False negative = Hatalı Negatifler anlamında kullanılmaktadır. Doğru ile başlayan ifadeler doğru tahminleri gösterirken, hatalı ile başlayan ifadeler hatalı tahminleri gösterir.

Hata matrisi ile elde edilen TP, TN, FP ve FP sayılarıyla precision, recall ve F1-score(F1-skoru harmonik ortalamayı kullanır.) ölçütlerinde modellerin başarısı test edilmiştir.

Recall: Pozitif durumların ne kadar başarılı tahmin edildiğini gösteren ölçüttür.

Recall = Doğru Pozitifler / (Doğru Pozitifler + Hatalı Negatifler)

Precision: Pozitif olarak tahmin edilen bir durumdaki başarıyı gösteren ölçüttür.

Precision = Doğru Pozitifler / (Doğru Pozitifler + Hatalı Pozitifler)

F-1 Skoru: Recall ve Precision ölçütlerinin harmonic ortalamasıdır.(Harmonik ortalama, aritmetik ortalama ve geometrik ortalama gibi pisagorik ortalamalardan biridir. Genellikle oranların ortalamasının istendiği durumlar için kullanılır.)

F-1 Skoru = 2*Precision*Recall / (Precision + Recall)

Ridge Modeli için elde edilen recall, precision ve F-1 skor değerleri Çizelge 5.3.'de verilmiştir.

Çizelge 5.3. Ridge eğitim sonuçları

	Precision	**Recall**	**F1-Score**
Negatif	0.994	0.954	0.974
Pozitif	0.934	0.002	0.962
Micro AVG	0.969	0.969	0.969
Macro AVG	0.964	0.969	0.969
Weighted AVG	0.970	0.969	0.969

Bayesçi Ridge Modeli için elde edilen recall, precision ve F-1 skor değerleri Çizelge 5.4.'de verilmiştir.

Çizelge 5.4. Bayes eğitim sonuçları

	Precision	**Recall**	**F1-Score**
Negatif	0.995	0.927	0.960
Pozitif	0.899	0.993	0.944
Ortalama	0.957	0.953	0.953

Derin Öğrenme Modeli için elde edilen recall,precision ve F-1 skor değerleri Çizelge 5.5.'de verilmiştir.

Çizelge 5.5. Derin öğrenme eğitim sonuçları

	Precision	Recall	F1-Score
Negatif	0.989	0.996	0.992
Pozitif	0.994	0.983	0.989
Ortalama	0.991	0.991	0.991

5.8.2. Harici Tweetler için Deney Sonuçları

Deneysel çalışmanın ikinci aşamasında modellerin testi için veri setinin içerisindeki Tweetler yerine, veri setinde bulunmayan 100 farklı tweet alınarak İslamofobik ve İslamofobik değil şeklinde etiketleme yapılmıştır. Ardından etiketlenen tweetler modellerde test edilerek modelin doğruluğu araştırılmıştır.

Harici Tweetler için Ridge Regresyonu Sonuçları

Çalışmada Ridge Regresyonu ile yapılan modellemede 100 tweet test edilmiştir. Bu test sonucunda 100 Tweet'in 82'i doğru bilinmiş fakat 18 tweette model hatalı cevap vermiştir. Bu sonuçlara göre oluşan hata matrisi aşağıda verilmiştir.

Çizelge 5.6. Görülmemiş veride Ridge Regresyonu eğitim sonuçları

		Gerçek Sonuç	
		False	True
Tahmini Sonuç	False	61	8
	True	10	21

Bu cevaplara birer doğru ve yanlış tahmin örneklerinin yer aldığı tablo aşağıda verilmiştir.

Çizelge 5.7. Görülmemiş veride Ridge Regresyonu test sonucu örnekleri

Test edilen Tweet	Gerçekte Olan	Tahmin
No but we should ban Islam, we are a Christian nation that separated our laws from religion centuries ago, Islam is barbaric in it's essence and isn't compatible with western culture.	İslamofobik	İslamofobik
It's something else when you are able to start a flame in a young mind and hopefully able to keep em off gang life thanks to Islam	İslamofobik Değil	İslamofobik Değil
Dont apologize for the christians plz apologize for yourself for the unjust decision with youself against God.If you have problems with people who practice islam wrong dont blame Islam.pure love you preached.	İslamofobik	İslamofobik Değil
And the fun fact is More than Half of the militia deployed in ISIS were ex con, service men from across the globe US, Bitain, Russia. They were crazy hired assasins and they did extremely wrong things in name of Islam.	İslamofobik Değil	İslamofobik

Modelin kendi veri seti içerisin yapılan testte doğruluk oranı %96 iken farklı bir veride %82 doğrulukta kalınmıştır. Bu, modelin eğitimi sırasında overfitting (aşırı öğrenme) olduğunu göstermektedir.

Harici Tweetler için Bayes Regresyonu Sonuçları

Çalışmada Bayes Regresyonu ile yapılan modellemede 100 tweet test edilmiştir. Bu test sonucunda 100 Tweet'in 89'i doğru bilinmiş fakat 11 tweette model hatalı cevap vermiştir. Bu sonuçlara göre oluşan hata matrisi aşağıda verilmiştir.

Çizelge 5.8. Görülmemiş veride Bayes Regresyonu eğitim sonuçları

		Gerçek Sonuç	
		False	True
Tahmini Sonuç	False	70	10
	True	1	19

Bu cevaplara birer doğru ve yanlış tahmin örneklerinin yer aldığı tablo aşağıda verilmiştir.

Çizelge 5.9. Görülmemiş veride Bayes Regresyonu test sonucu örnekleri

Test edilen Tweet	Gerçekte Olan	Tahmin
Because Islam is not a religion it’s an ideology to take over governments	İslamofobik	İslamofobik
“The Golden Age of Islam” that Neil is talking about was not ended by the allies in 1920, it was when 4 million (estimated) Arabs and Khwarezmian people were killed in the 1220s by mongols. Baghdad’s aqueduct system still is not fully rebuilt to this day from those events	İslamofobik Değil	İslamofobik Değil
What is wrong with the heads of this devilish people. Can we have peace & why is all this evil comes out of the Islamic world.	İslamofobik	İslamofobik Değil
I think I've read about this before. If I remember right pretty much Ishmael or his later generations began Islam because Isaac was chosen over him?	İslamofobik Değil	İslamofobik

Modelin kendi veri seti içerisin yapılan testte doğruluk oranı %95 iken farklı bir veride %89 doğrulukta kalınmıştır. Bu, modelin eğitimi sırasında nispeten de olsa overfitting (aşırı öğrenme) olduğunu göstermektedir.

Harici Tweetler için Derin Öğrenme Sonuçları

Çalışmada derin öğrenme kullanılarak yapılan modellemede 100 tweet test edilmiştir. Bu test sonucunda 100 Tweet'in 68'i doğru bilinmiş fakat 32 tweette model hatalı cevap vermiştir. Bu sonuçlara göre oluşan hata matrisi aşağıda verilmiştir.

Çizelge 5.10. Görülmemiş veride Derin Öğrenme test sonucu örnekleri

		Gerçek Sonuç	
		False	True
Tahmini Sonuç	False	43	4
	True	28	25

Bu cevaplara birer doğru ve yanlış tahmin örneklerinin yer aldığı tablo aşağıda verilmiştir.

Çizelge 5.11. Görülmemiş veride Derin Öğrenme test sonucu örnekleri

Test edilen Tweet	Gerçekte Olan	Tahmin
Pakistani jihadis r ideolocal flag holders of urself. U supported Muslim league & direct action of Jinnah.	İslamofobik	İslamofobik
My fellow muslims, unfortunately this guy has rocked my faith in Islam.	İslamofobik Değil	İslamofobik Değil
Tell me you have not learned anything from this! Meanwhile, back on the farm: What about - Halal meat, Sharia law.	İslamofobik	İslamofobik Değil
In kasmir operation... What is 370 article.	İslamofobik Değil	İslamofobik

Modelin kendi veri seti içerisin yapılan testte doğruluk oranı %99 iken farklı bir veride %68 doğrulukta kalınmıştır. Bu, modelin eğitimi sırasında diğer modellere göre çok daha fazla overfitting (aşırı öğrenme) olduğunu göstermektedir.

Harici Tweetler için Deney Sonuçlarının Karşılaştırılması

Çalışma sırasında her üç modelin testinde aynı 100 tweet kullanılarak daha doğru bir karşılaştırma yapılması amaçlanmıştır. Modellerin eğitimi sırasında en uzun süreyi derin öğrenme modeli alırken, en kısa sürede eğitim Ridge modeli ile yapılmıştır. Modellerin eğitimi sırasında harcanan süre tabloda verilmiştir.

Çizelge 5.12. Modellerin eğitim süreleri karşılaştırması

Model	Eğitim Süresi
Ridge Regresyonu	30 saniye
Bayes Regresyonu	10 dakika
Derin Öğrenme	500 dakika

Bu eğitim süreleri sonunda test edilen 100 tweet için alınan doğruluk sonuçları tabloda verilmiştir.

Çizelge 5.13. Modellerin görülmemiş verideki doğruluk yüzdeleri karşılaştırması

Model	Doğruluk Yüzdesi
Ridge Regresyonu	%82
Bayes Regresyonu	%89
Derin Öğrenme	%68

Doğruluk oranlarına ve eğitim sırasında geçen süreye bakılarak bir seçim yapıldığında Bayes Regresyonu kullanmanın daha avantajlı olduğu görülmektedir.

Deney Sonuçlarının Diğer Çalışmalarla Karşılaştırılması

Sosyal medyadaki İslamofobik nefret söyleminin tespiti konusunda Bertie Vidgen ve Taha Yasseri tarafından yapılan çalışmada [1] Naïve-Bayes, Random Forests , Lojistik Regresyon ve Karar Ağaçları algoritması kullanılmıştır.Bu algoritmalar içerisinde en yüksek doğruluk

oranına %69.13 ile Naïve-Bayes algoritmasında ulaşılmıştır. Vidgen ve Yasseri tarafından yapılan çalışmada belirlenen hesapları takip eden kullanıcıların tweetleri alınarak 140 milyonluk bir veri seti kullanılmıştır.

Tez çalışmasında ise kullanıcı bazlı bir tweet seçimi yapılmamıştır, bunun yerine kullanıcılar göz ardı edilerek doğrudan arama kısmından elde edilen tweetler üzerinden veri seti oluşturulup, bu veri setinin manuel olarak işaretlemesi yapılmıştır. Bu yaklaşım da daha küçük bir veri setiyle de olsa önceden görülmemiş veride dahi %89 gibi yüksek bir doğruluk oranına ulaşılabilmesini sağlamıştır.

Çin sosyal ağlarındaki İslamofobik nefret söyleminin tespiti için Bailey Marshech ve Kangyu Mark Wang yapılan çalışmada [2] içerisinde “muslim”, “Islam”, “halal”, “Allah” kelimelerinden biri geçen 77.462 gönderi indirilmiştir. Bu veri setinden 2.000 tweet negatif ve pozitif olarak elle işaretlenmiş, geri kalan gönderilerde gözetimli makine öğrenmesi algoritmaları kullanılarak tahminlerde bulunulmuştur. Bu sonuçlara göre Çin’in farklı eyaletlerindeki İslamofobik tweet yüzdeleri %1.2 ile %4.9 arasında bulunmuştur. Marsheck ve Wang tarafından yapılanan çalışmada test setinde ne kadarlık bir doğruluk oranına ulaşıldığı hakkında bilgi verilmemiştir. Fakat eğitim setinin, veri setinin %2,5 gibi düşük bir kısmına denk gelmesi yetersiz öğrenme (underfitting) problemi yaşanmış olabileceğini göstermektedir. Marsheck ve Wang tarafından yapılan çalışmanın eğitim sonuçları erişime açılmadığı için tez çalışması ile bir karşılaştırma yapılma şansı olmamıştır.

6. SONUÇLAR VE ÖNERİLER

Bu tez çalışması kapsamında, yapay zeka teknikleri kullanılarak, sosyal medya platformlarından twetter üzerindeki paylaşımlardan duygu analizi yapılmaya çalışılmıştır. Çalışma özelinde islamofobia duygusu araştırılmış bu amaçla içerisinde "Islam" ve "Muslim" kelimeleri geçen 162.000 ingilizce dilinde tweetten oluşan bir veri seti oluşturulmuştur. Bu tweetler beş kişi tarafından islamofobik veya değil şeklinde sınıflara ayrılmıştır. Bu veri setinin %80'i eğitim kalanlar ise test amaçlı kullanılmıştır. Ağın ezberleme yapmaması için eğitim ve test verilerinde kullanılan tweetlerin islamofobiklik oranları eşit tutulmuştur. Üç farklı sınıflandırıcı kullanılmış, her üç model içinde %95'in üzerinde doğruluğa ulaşılmıştır. Fakat araştırmanın doğruluğunu daha iyi saptayabilmek ve eğitim sırasında oluşabilecek aşırı öğrenme gibi problemleri tespit edebilmek için ikinci aşamada 100 tweetten oluşan yeni bir veri seti oluşturulmuştur. Bu yeni tweetler ile yapılan testlerde Bayes Regresyonu'nda %89, Ridge Regresyonu'nda %82, derin öğrenme modelinde ise %68 doğruluğa ulaşılmıştır. İlk veri seti ile ikinci veri seti arasındaki doğruluk oranlarındaki farkın temel sebebi olarak kelime ve cümle yapısı çok farklı tweetler ile karşılaşıldığında doğru şekilde sınıflandırma yapılamaması görülebilir. Bu problemin aşılması için test amaçlı kullanılan, görülmemiş veri setinin de çeşitli önişleme aşamalarından geçirilip sonrasında test edilmesi başarıyı artırabilir, eğitim için çok daha geniş veri setleri ele alınabilir.

Görülmemiş veride modellerin doğruluk oranları arasındaki farklılıklar incelenecek olursa; Bayes Regresyonu olasılıksal bir yaklaşım olduğu için, özellik çıkarıp ezberleme daha az olduğundan, en yüksek doğrulukta çalışan model olduğu söylenebilir. Derin öğrenme yaklaşımında ise yüksek epoch sayısı ve veri setinin yeterince büyük olmaması aşırı öğrenmeyi doğurabilmektedir. Ridge regresyonunda yaşanan nispeten düşüşün sebebi ise veri setindeki twetler arasındaki korelasyonun fazla olması durumunda özellik (feature) sayısının çok olmasının getirdiği hata oranının artması olarak değerlendirilebilir. Burada ceza algoritması değiştirilebilir. Ayrıca derin öğrenme çalışmasının içine Ridge modeli eklenerek hibrit bir modelle aşırı öğrenme azaltılabilir.

Çalışmanın başlıca kısıtlılığı verilerin kapsamıdır. Daha iyi sonuçlar elde etmek için, farklı dillerdeki İslamofobik ve İslamofobik olmayan tweetler eklenerek ve daha fazla tweet üzerinden eğitim yapılarak veri setlerinin kapsamı arttırılmalıdır. Başka bir kısıtlama yanlış yazılan kelimelerdir. Farklı bir çalışmada ek bir kütüphane geliştirilerek yanlış yazılan

kelimeler ile ilgili de otomatik düzeltme işlemi yapılabilir. Yanlış yazılmış bu kelimeler düzeltildikten sonra daha iyi sonuçlar elde edilebilir.

Tweetler içerisinde bir başkasının paylaşımına atıfta bulunurken dile getirilen küfür ifadeleri de o tweeti atan kişinin nefret söylemine sahipmiş gibi sınıflandırıcılar tarafından anlaşılabilmektedir. Gelecek çalışmalarda ifadenin cümle içindeki bağlamının da dikkate alındığı araştırmalar yapılması hedeflenmektedir. Ayrıca sınıflandırıcı tarafından oluşturulan nefret söylemine sahip tweetlerden bir sözlük oluşturularak, ilgili konuda çalışma yapan kişilerin araştırmalarına katkı sunulması planlanmaktadır.

KAYNAKLAR

1. Vossen,G. and Hagemann,S. (2007). *Unleashing Web 2.0: From Concepts to Creativity*. Amsterdam:Morgan Kaufmann Books,47.

2. Ryan, J. (2010). *A History of the Internet and the Digital Future*. London: Reaktion Books,138.

3. Assimakopoulos, S and Baider, F. (2017). *Online Hate Speech in the European Union: A Discourse-Analytic Perspective*. Manhattan: Springer International Publishing,11.

4. Russell, M. A. (2011). *Mining the Social Web: Analyzing Data from Facebook, Twitter, LinkedIn, and Other Social Media Sites*. Sebastopol: O'Reilly Media,182.

5. Thomas Davidson, D. W. (2017). Automated Hate Speech Detection and the Problem of Offensive Language. Paper presented at the Proceedings of the Eleventh International AAAI Conference on Web and Social Media, Montréal, Québec, Canada

6. Kelly Reynolds, A. K. (2011). Using Machine Learning to Detect Cyberbullying. Paper presented at the 10th International Conference on Machine Learning and Applications and Workshops , Honolulu, Hawaii,USA

7. Michael Ashcroft, A. F. (2015). Detecting Jihadist Messages on Twitter. Paper presented at the European Intelligence and Security Informatics Conference, Manchester, England.

8. Internet: Record number of anti-Muslim attacks reported in UK last year (2018). https://www.theguardian.com/uk-news/2018/jul/20/record-number-anti-muslim-attacks-reported-uk-2017, Son Erişim Tarihi: 02.01.2020

9. Internet: Islamophobes Came for Americans on the Campaign Trail. https://foreignpolicy.com/2019/03/13/islamophobes-came-for-americans-on-the-campaign-trail, Son Erişim Tarihi: 02.01.2020

10. Internet:Islamophobia.(2019).https://trends.google.com/trends/explore?q=islamophobia&geo=US, Son Erişim Tarihi: 02.01.2020

11. Awan, I. (2014). Islamophobia and Twitter: A Typology of Online Hate Against Muslims on Social Media. *Policy and Internet*,26(4),133-151.

12. Magdy,W. Darwish,K. and Abokhodair,N.,(2015,Dec). Quantifying Public Response towards Islam on Twitter after Paris Attacks. Social and Information Networks,1-9.

13. Makice, K. (2009). Twitter API: Up and running: Learn how to build applications with the Twitter API. California :O'Reilly Media, 132-142.

14. İnternet: Twitter, About Twitter limits. URL: https://help.twitter.com/en/rules-and-policies/twitter-limits, Son Erişim Tarihi: 24.07.2019

15. Internet: Statista, Number of monthly active Twitter users worldwide from 1st quarter 2010 to 1st quarter 2019, URL: https://www.statista.com/statistics/282087/number-of-monthly-active-twitter-users/, Son Erişim Tarihi: 24.07.2019

16. Stukal, D., Sanovich, S., Bonneau, R. and Tucker, J. A. (2017). Detecting bots on Russian political Twitter. *Big Data,* 5(4), 310-324.

17. İnternet: Statista, Number of monthly active Twitter users, URL: https://www.statista.com/statistics/282087/number-of-monthly-active-twitter-users/, Son Erişim Tarihi:2019

18. Warner,W., Hirschberg.J, Detecting Hate Speech on the World Wide Web. Paper presented at the LSM '12: Proceedings of the Second Workshop on Language in Social Media 2012, Stroudsburg,USA.

19. Lynn, T., Endo, P. T., Rosati, P., Silva, I., Santos, G. L. and Ging, D. A Comparison of Machine Learning Approaches for Detecting Misogynistic Speech in Urban Dictionary. Paper presented at the IEEE Cyber Science 2019, Oxford,England.

20. Perelló, C., Tomás, D., Garcia-Garcia, A., Garcia-Rodriguez, J., & Camacho-Collados, J. (2019). Setting a Strong Linear Baseline for Hate Speech Detection. In Proceedings of the 13th International Workshop on Semantic Evaluation, Minneapolis, Minnesota, United States of America

21. Bhandary, U. (2019). Detection of Hate Speech in Videos Using Machine Learning. Master's Thesis, San Jose State University, California.

22. Sahibani, P., Rathod, K. and Padhiyar, A. Forecast Of Onlıne Radıcalızatıon Of Twıtter Text Content Usıng Enhanced Classıficatıon Methods. *Journal of Applied Science and Computations*,654-658

23. Niam, I. M. A., Irawan, B., Setianingsih, C. and Putra, B. P. (2018). Hate Speech Detection Using Latent Semantic Analysis (LSA) Method Based On Image. Paper presented at the International Conference on Control, Electronics, Renewable Energy and Communications, Bandung, Indonesia.

24. Dadhe, M., Masidkar, M. Vaidya, V. and Jalan, P. A. Detection of Abusive Language from Tweets in Social Networks. *International Journal on Recent and Innovation Trends in Computing and Communication*, 6(3), 148-151.

25. Sap, M., Card, D., Gabriel, S., Choi, Y., Smith, N. A., Bosselut, A. and Choi, Y. (2016). The Risk of Racial Bias in Hate Speech Detection. ACL,516-527.

26. Mulki, H., Ali, C. B., Haddad, H. and Babaoğlu, I. (2019). N-gram embeddings for Hate Speech Detection in Multilingual Tweets. Paper presented at the 13th International Workshop on Semantic Evaluation, Minneapolis, Minnesota, United States of America

27. Setyadi, N. A., Nasrun, M., and Setianingsih, C. (2018). Text Analysis For Hate Speech Detection Using Backpropagation Neural Network. *International Conference on Control, Electronics, Renewable Energy and Communications*, 159-165.

28. Siegel, A., Nikitin, E., Barberá, P., Sterling, J., Pullen, B., Bonneau, R. and Tucker, J. A. (2019). Trumping Hate on Twitter? Online Hate in the 2016 US Election and its Aftermath.

29. Sazany, E. and Budi, I. (2018). Deep Learning-Based Implementation of Hate Speech Identification on Texts in Indonesian: Preliminary Study. Paper presented at the International Conference on Applied Information Technology and Innovation, West Sumatra,Indonesia.

30. Kiilu, K. K., Okeyo, G., Rimiru, R. and Ogada, K. (2018). Using Naïve Bayes Algorithm in detection of Hate Tweets. *International Journal of Scientific and Research Publications*, 8(3), 99-107.

31. Mishra, P., Del Tredici, M., Yannakoudakis, H. and Shutova, E. (2018). Author Profiling for Hate Speech Detection. Paper presented at the Proceedings of the 27th International Conference on Computational Linguistics, Santa Fe, New Mexico, United States of America

32. Kwok, I. and Wang, Y. (2013). Locate the hate: Detecting tweets against blacks. In Twenty-seventh AAAI conference on artificial intelligence, Bellevue, Washington, United States of America.

33. Magu, R., Joshi, K. and Luo, J. (2017). Detecting the hate code on social media. Paper presented at the Eleventh International AAAI Conference on Web and Social Media, Montréal, Québec, Canada

34. Awan, I. (2014). Islamophobia and Twitter: A typology of online hate against Muslims on social media. *Policy & Internet*, 6(2), 133-150.

35. Awan, I. (2016). Islamophobia in Cyberspace: Hate Crimes Go Viral. New York :Routledge,172-173.

36. Ayhan, B. ve Çifçi, M. E. (2018). IŞID, Propaganda ve İslamofobi. *Medya ve Din Araştırmaları Dergisi*, 1(1), 17-32.

37. Koç, Ç. T. (2018) Yeni Medyada İslamofobik Söylemin Üretimi: Facebook Örneği. *Medya ve Din Araştırmaları Dergisi*, 1(2), 211-242.

38. Bleich, E. (2011). What is Islamophobia and how much is there? Theorizing and measuring an emerging comparative concept. *American Behavioral Scientist*, 55(12), 1581-1600.

39. Awan, I. (2016). Islamophobia in Cyberspace: Hate Crimes Go Viral. Routledge.In D. Gambetta (Ed.), *Making sense of suicide missions*.New York: Oxford University Press, 164.

40. Hoffman, B. (2009). Radicalization and subversion: Al Qaeda and the 7 July 2005 bombings and the 2006 airline bombing plot. *Studies in Conflict & Terrorism*, 32(12), 1100-1116.

41. Awan, M. S. (2010). Global terror and the rise of xenophobia/Islamophobia: An analysis of American cultural production since September 11. *Islamic Studies*, 521-537.

42. Bayrakli, E. and Hafez, F. (2016). The state of islamophobia in Europe. European Islamophobia Report, 5-10.

43. Wetherell, M. (2019). Understanding the Terror Attack: Some Initial Steps. *New Zealand Journal of Psychology*, 48(1), 6-7.

44. Besley, T. and Peters, M. A. (2019). Terrorism, trauma, tolerance: Bearing witness to white supremacist attack on Muslims in Christchurch, New Zealand. Educational Philosopy and Theory,1-11

45. Kwon, K. H., Chadha, M. and Wang, F. (2019). Proximity and Networked News Public: Structural Topic Modeling of Global Twitter Conversations about the 2017 Quebec Mosque Shooting. *International Journal of Communication*, 13, 24.

46. Cesari, J. (2002). Islam in France: The shaping of a religious minority. Yvonne Yazbeck Haddad(Ed.), Muslims in the West: From sojourners to citizens, 36-51.

47. Roth, K. (2017). The dangerous rise of populism: Global attacks on human rights values. *Journal Of International Affairs*, 79-84.

48. Carland, S. (2011). Islamophobia, fear of loss of freedom, and the Muslim woman. *Islam and Christian–Muslim Relations*, 22(4), 469-473.

49. Clyne, I. D. (1998). Cultural diversity and the curriculum: The Muslim experience in Australia. *European Journal of Intercultural Studies*, 9(3), 279-289.

50. Norocel, O. C. (2017). Åkesson at Almedalen: Intersectional Tensions and Normalization of Populist Radical Right Discourse in Sweden. *NORA-Nordic Journal of Feminist and Gender Research*, 25(2), 91-106.

51. Pędziwiatr, K. (2011). Muslims in Contemporary Poland. *Muslims in Visegrad Countries*, 10-24.

52. Chifu, I. (2013). The conflict of values in the Muslim world and its impact upon security. *Strategic Impact*, 46(1), 111-127.

53. Mehta, N. (2006). Modi and the camera: The politics of television in the 2002 Gujarat riots. *South Asia: Journal of South Asian Studies*, 29(3), 395-414.

54. Hamdan, A. (2007). The issue of hijab in France: reflections and analysis. *Muslim World Journal of Human Rights*, 4(2),1-10.

55. Coene, G. and Longman, C. (2008). Gendering the diversification of diversity: The Belgian hijab (in) question. *Ethnicities*, 8(3), 302-321.

56. Swiss ban building of minarets on mosques. (2009, November 29). *New York Times*, 29.

57. İnternet: The Independent, Japan has ruled to spy on all Muslims , URL: https://www.independent.co.uk/voices/muslims-japan-government-surveillance-im-not-surprised-islamophobia-a7113051.html, Son Erişim Tarihi: 02.01.2019

58. Singh, S. (2019). Easter Attacks in Sri Lanka. *Economic & Political Weekly*, 54(20), 13.

59. İnternet: Anadolu Agency, US Muslim Republican official faces ouster for religion, URL: https://www.aa.com.tr/en/americas/us-muslim-republican-official-faces-ouster-for-religion/1330579, Son Erişim Tarihi: 02.01.2019

60. İnternet: The Independent, Hijab wearing is 'passive terrorism', says US military publication, URL: https://www.independent.co.uk/news/world/americas/hijab-wearing-is-passive-terrorism-a6893931.html , Son Erişim Tarihi: 02.01.2019

61. Ogan, C., Willnat, L., Pennington, R. and Bashir, M. (2014). The rise of anti-Muslim prejudice: Media and Islamophobia in Europe and the United States. *International Communication Gazette*, 76(1), 27-46.

62. İnternet: Anadolu Ajansı, Amerikalı felsefeci Prof. Dr. Clark: Müslümanları tanıdıkça İslamofobi'den uzaklaştım, URL: https://www.aa.com.tr/tr/turkiye/amerikali-felsefeci-prof-dr-clark-muslumanlari-tanidikca-islamofobiden-uzaklastim/1416623, Son Erişim Tarihi: 02.01.2019

63. İnternet: The Australian Capital Territory Government, Discrimination Amendment Act 2016, URL: https://www.legislation.act.gov.au/a/2016-49/current/pdf/2016-49, Son Erişim Tarihi: 24.07.2019

64. Hernández, T. K. (2010). Hate speech and the language of racism in Latin America: a lens for reconsidering global hate speech restrictions and legislation models. *Journal of International Law*, 32, 805.

65. İnternet: Goverment of Canada, Justice Laws Website, URL: https://laws-lois.justice.gc.ca/eng/acts/C-46/section-318.html, Son Erişim Tarihi: 2019

66. İnternet: CTV News, B.C. panel rejects Muslim complaint vs. Maclean's, URL : https://www.ctvnews.ca/b-c-panel-rejects-muslim-complaint-vs-maclean-s-1.332541, Son Erişim Tarihi: 02.01.2019

67. İnternet:B.C. panel rejects Muslim complaint vs. Maclean's, URL : https://www.ctvnews.ca/b-c-panel-rejects-muslim-complaint-vs-maclean-s-1.332541, Son Erişim Tarihi: 02.01.2019

68. İnternet: URL : https://www.ctvnews.ca/b-c-panel-rejects-muslim-complaint-vs-maclean-s-1.332541, Son Erişim Tarihi: 02.01.2019

69. Paul, A. (2011). The criminalization of hate speech in chile in light of comparative case law. *Revista Chilena De Derecho*, 38, 573.

70. İnternet: The Croatian Parliament, The Constitution, URL: https://croatia.eu/article.php?lang=2&id=25, Son Erişim Tarihi: 02.01.2019

71. Weaver, S. (2010). Liquid racism and the Danish Prophet Muhammad cartoons. *Current Sociology*, 58(5), 675-692.

72. Rohde, D. (2001). Indonesia unraveling?. *Foreign Affairs*, 110-124.

73. Banks, J. (2010). Regulating hate speech online. *International Review of Law, Computers & Technology*, 24(3), 233-239.

74. Alpaydin, E. (2009). Introduction to machine learning. Cambridge :MIT press,1-2

75. Alpaydin, E. (2009). Introduction to machine learning. Cambridge :MIT press,3

76. DasGupta, A. (2011). Probability for statistics and machine learning: fundamentals and advanced topics. New York:Springer,5-20.

77. Goldstein, M. and Wooff, D. (2007). Bayes linear statistics: Theory and methods. New Jersey: John Wiley & Sons,3-10.

78. Gilks, W. R., Richardson, S. and Spiegelhalter, D. (1995). Markov chain Monte Carlo in practice. Massachusetts:Chapman and Hall/CRC,1-11.

79. Burton, C. P. (2005). Replicating the Manchester Baby: Motives, methods, and messages from the past. *IEEE Annals of the History of Computing*, 27(3), 44-60.

80. Campbell-Kelly, M. (1980). Programming the EDSAC: Early programming activity at the University of Cambridge. *Annals of the History of Computing*, 2(1), 7-36.

81. Williams, M. R. (1993). The origins, uses, and fate of the EDVAC. *IEEE Annals of the History of Computing*, 15(1), 22-38.

82. Turing, A. M. (1956). Can a machine think. *The world of mathematics*, 4, 2099-2123.

83. Benko, A. and Lányi, C. S. (2009). History of artificial intelligence. Budapest: In Encyclopedia of Information Science and Technology,1759-1762.

84. Chen, Q. and Verguts, T. (2010). Beyond the mental number line: A neural network model of number–space interactions. *Cognitive psychology*, 60(3), 218-240.

85. El Naqa, I. and Murphy, M. J. (2015). What is machine learning?. In Machine Learning in Radiation Oncology. Berlin:Springer,3-11.

86. Lighthill, J. (1973). Report to the Science Research Council on funding for artificial intelligence research. *Science Research Council*,1-43

87. Hopfield, J. J. (1982). Neural networks and physical systems with emergent collective computational abilities. *Proceedings of the National Academy of Sciences*, 79(8), 2554-2558.

88. Ishizuka, M. (1984). Japanese work in expert systems. *Expert Systems*, 1(1), 51-56.

89. Amit, D. J., Gutfreund, H. and Sompolinsky, H. (1987). Statistical mechanics of neural networks near saturation. *Annals of Physics*, 173(1), 30-67.

90. IEEE, IEEE First International Conference on Neural Networks, Sheraton Harbor Island East, San Diego, California.

91. Pandolfini, B. (1997). Kasparov and Deep Blue: The historic chess match between man and machine. New York:Simon and Schuster,13-25.

92. Hochreiter, S. and Schmidhuber, J. (1997). Long short-term memory. *Neural computation*, 9(8), 1735-1780.

93. Hinton, G. E. and Salakhutdinov, R. R. (2006). Reducing the dimensionality of data with neural networks. *Science*, 313(5786), 504-507.

94. Hinton, G. E., Osindero, S. and Teh, Y. W. (2006). A fast learning algorithm for deep belief nets. *Neural computation*, 18(7), 1527-1554.

95. Internet: Deep Mind, Solve intelligence. Use it to make the world a better place, URL: https://deepmind.com/about/ , Son Erişim Tarihi: 02.01.2019

96. Shah, H. (2017). The DeepMind debacle demands dialogue on data. *Nature News*, 547(7663), 259.

97. Gibney, E. (2016). Google AI algorithm masters ancient game of Go. *Nature News*, 529(7587), 445.

98. Machine Learning- Facebook, İnternet: URL: https://research.fb.com/category/machine-learning/, Son Erişim Tarihi: 02.01.2019

99. Marsland, S. (2009). Machine learning: an algorithmic perspective. Florida: CRC Press,

100. Azizpour, H. and Laptev, I. (2012). Object detection using strongly-supervised deformable part models. Paper presented at the 12th European Conference on Computer Vision, Firenze, Italy.

101. Hofmann, T. (2001). Unsupervised learning by probabilistic latent semantic analysis. *Machine learning*, 42(1-2), 177-196.

102. Sutton, R. S. and Barto, A. G. (1998). Introduction to reinforcement learning (Vol. 2, No. 4). Cambridge: MIT Press,144-148.

103. Gupta, N. K. (2011). Extracting descriptions of problems with product and services from twitter data. Paper presented at the 3rd Workshop on Social Web Search and Mining (SWSM2011), Beijing, China.

104. Saygin, A. P., Cicekli and I.and Akman, V. (2000). Turing test: 50 years later. *Minds and Machines*, 10(4), 463-518.

105. Jones, K. S. (1994). Natural language processing: a historical review. In Current issues in computational linguistics: in honour of Don Walker. Springer, Dordrecht, pp. 3-16.

106. İnternet: John Hutchins , The history of machine translation in a nutshell, URL: http://hutchinsweb.me.uk/Nutshell-2005.pdf, Son Erişim Tarihi: 24.07.2019

107. Hutchins, J. (2003). ALPAC: the (in) famous report. *Readings in machine translation*, 14, 131-135.

108. Ferrucci, D. A. (2012). Introduction to "this is watson". *IBM Journal of Research and Development*, 56(3.4), 1-1.

109. İnternet: Google, Natural langauge processing - google scholar, URL: https://scholar.google.com.tr/scholar?q="natural+language+processing, Son Erişim Tarihi: 24.07.2019

110. Stevenson, R. A., Mikels, J. A. and James, T. W. (2007). Characterization of the affective norms for English words by discrete emotional categories. *Behavior research methods*, 39(4), 1020-1024.

111. Dey, L. and Haque, S. M. (2009). Opinion mining from noisy text data. *International Journal on Document Analysis and Recognition*, 12(3), 205-226.

112. Cambria, E. and Hussain, A. (2012). Sentic computing. *Marketing*, 59(2), 557-577.

113. Alaei, A. R., Becken, S. and Stantic, B. (2019). Sentiment analysis in tourism: capitalizing on big data. *Journal of Travel Research*, 58(2), 175-191.

114. Jiménez-Zafra, S. M., Martín-Valdivia, M. T., Molina-González, M. D. and Ureña-López, L. A. (2019). How do we talk about doctors and drugs? Sentiment analysis in

forums expressing opinions for medical domain. *Artificial intelligence in medicine*, 93, 50-57.

115. Agarwal, A., Xie, B., Vovsha, I., Rambow, O. and Passonneau, R. (2011, June). Sentiment analysis of twitter data. Paper presented at the Proceedings of the Workshop on Language in Social Media, Portland, Oregon, United States of America.

116. Zanasi, A. (2008). Virtual weapons for real wars: Text mining for national security. Paper presented at the Proceedings of the International Workshop on Computational Intelligence in Security for Information Systems CISIS'08, Genova, Italy.

117. Badal, V. D., Kundrotas, P. J. and Vakser, I. A. (2015). Text mining for protein docking. PLoS computational biology, 11(12), 1-21.

118. Iqbal, F., Fung, B. and Debbabi, M. (2012). Mining criminal networks from chat log. Paper presented at the 2012 IEEE/WIC/ACM International Joint Conferences on Web Intelligence and Intelligent Agent Technology, Washington, DC, United States of America.

119. Coussement, K. and Van den Poel, D. (2008). Integrating the voice of customers through call center emails into a decision support system for churn prediction. *Information & Management*, 45(3), 164-174.

120. Pang, B., Lee, L. and Vaithyanathan, S. (2002). Thumbs up?: sentiment classification using machine learning techniques. Paper presented at the ACL-02 conference on Empirical methods in natural language processing, Stroudsburg, PA, United States of America.

121. Shen, J., Xiao, J., He, X., Shang, J., Sinha, S. and Han, J. (2018). *Entity set search of scientific literature: An unsupervised ranking approach*. Paper Presented at the 41st International ACM SIGIR Conference on Research & Development in Information Retrieval, Arbor, MI, United States of America.

122. Ramos, J. (2003). *Using tf-idf to determine word relevance in document queries*. Paper Presented at the Proceedings of the first instructional conference on machine learning, Alberta, Canada.

123. İnternet: Pew Research, The Future of the Global Muslim Population, https://www.pewforum.org/2011/01/27/the-future-of-the-global-muslim-population, Son Erişim Tarihi: 24.07.2019.

124. İnternet: Pew Research, The Future Global Muslim Population, URL: http://assets.pewresearch.org/wp-content/uploads/sites/11/2011/01/WebPDF-Feb10.pdf, Son Erişim Tarihi:02.01.2019.

125. İnternet: Brenton Tarrant, Brenton Tarrant Social Media Twitter Video. URL, https://heavy.com/news/2019/03/brenton-tarrant-social-media-twitter-video, Son Erişim Tarihi: 24.07.2019.

126. Greenspan, J. and Bulger, B. (2001). MySQL/PHP database applications. New Jersey: John Wiley & Sons,397-439.

127. Kramer, O. (2016). Machine learning for evolution strategies. Heidelberg: Springer,45-53.

128. Dalcín, L., Paz, R. and Storti, M. (2005). MPI for Python. *Journal of Parallel and Distributed Computing*, 65(9), 1108-1115.

129. Lutz, M. (2001). Programming python. California:O'Reilly Media,13.

130. Lee, W. M. (2019). Python Machine Learning. New Jersey: John Wiley & Sons,3-4.

131. Pine, D. J. (2019). Introduction to Python for Science and Engineering. Florida:CRC Press,1-3.

132. Numpy, İnternet: URL: https://numpy.org/, Son Erişim Tarihi: 02.01.2019

133. Hunter, J. D. (2007). Matplotlib: A 2D graphics environment. *Computing in science & Engineering*, 9(3), 90.

134. Scollo, C. and Shumann, S. (1999). Professional PHP. Birmingham: Wrox Press Ltd,15-18.

135. İnternet: Git Access, URL: https://www.php.net/git.php, Son Erişim Tarihi: 24.07.2019.

136. Kofler, M. (2001). What Is MySQL?. Berkeley:Apress,3-19.

137. İnternet: Oracle,MySQL 8.0 Release Notes, URL: https://dev.mysql.com/doc/relnotes/mysql/8.0/en/ , Son Erişim Tarihi 24.07.2019.

138. Rosenberg, D. (2014). Stop, Words. *Representations*, 127(1), 83-92.

139. Perkins, J. (2010). Python text processing with NLTK 2.0 cookbook. Birmingham:Packt Publishing,7-10.

140. Mannila, H. (1996). Data mining: machine learning, statistics, and databases. Paper Presented at the 8th International Conference on Scientific and Statistical Data Base Management, Stockholm, Sweden.

141. Willmott, C. J. and Matsuura, K. (2005). Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance. *Climate research*, 30(1), 79-82.

142. Berger, J. O. (1980). Statistical decision theory and Bayesian analysis. New York: Springer Science & Business Media,118-166.

143. LeCun, Y., Bengio, Y. and Hinton, G. (2015). Deep learning. *Nature*, 521(7553), 436.

144. Shazeer, N., Mirhoseini, A., Maziarz, K., Davis, A., Le, Q., Hinton, G. and Dean, J. (2017, January). Outrageously large neural networks: The sparsely-gated mixture-of-experts layer. Paper presented at the 5th International Conference on Learning Representations, Toulon, France.

145. Gulli, A. and Pal, S. (2017). Deep Learning with Keras. Birmingham: Packt Publishing,10-98.

146. İnternet: Matt Mahoney, Large Text Compression Benchmark, URL: http://www. mattmahoney. net/text/text. html , Son Erişim Tarihi: 24.07.2019.

147. Graves, A., Mohamed, A. R. and Hinton, G. (2013). Speech recognition with deep recurrent neural networks. Paper presented at the 2013 IEEE international conference on acoustics, speech and signal processing, Vancouver,Canada

ÖZGEÇMİŞ

Kişisel Bilgiler

Soyadı, Adı : AYAN, Buğra

Uyruğu : T.C.

Doğum tarihi ve yeri : [redacted]

Medeni hali : [redacted]

Telefon : [redacted]

E-mail : [redacted]

Eğitim

Derece	**Eğitim Birimi**	**Mezuniyet Tarihi**
Yüksek Lisans	Gazi Üniversitesi / Bilişim Enstitüsü / Bilişim Sistemleri A.B.D	Devam Ediyor
Lisans	Karadeniz Teknik Üniversitesi/ Elektrik-Elektronik Mühendisliği	2012

İş Deneyimi

Yıl	**Yer**	**Görev**
2015 - Halen	T.C Cumhurbaşkanlığı	Mühendis
2013 - 2015	Milli Savunma Bakanlığı	Mühendis

Yabancı Dil

İngilizce

Hobiler

Bilgisayar ve bilişim teknolojileri

GAZİLİ OLMAK AYRICALIKTIR…