



## Hate speech framework in social media: A comprehensive review from past to present

Büşra Büyüktanır\*<sup>ID</sup>, Özlem Yakar<sup>ID</sup>, Ayşe Berna Altinel Girgin<sup>ID</sup>

<sup>1</sup>Department of Computer Engineering, Faculty of Technology, Marmara University, 34854, Maltepe, Istanbul, Turkey

### Highlights:

- A Review on hate speech in the context of traditional and digital networks
- Challenges and current status of the Turkish hate speech problem
- Use of machine learning and natural language processing techniques in Turkish hate speech studies

### Keywords:

- Hate speech
- Social media
- Cyber bullying
- Turkish hate speech
- Natural language processing

### Article Info:

Review Article  
Received: 15.07.2023  
Accepted: 02.03.2024

### DOI:

10.17341/gazimmfd.1327840

### Acknowledgement:

This work was partially supported by the Scientific and Technological Research Council of Turkey (TÜBİTAK) with grant number 120E187. The authors would like to thank TÜBİTAK for their support

### Correspondence:

Author: Büşra Büyüktanır  
e-mail: busra.buyuktanir@marmara.edu.tr  
phone: +90 216 777 3816

### Graphical/Tabular Abstract

With the introduction of technology into our lives, social networking platforms are developing rapidly and the use of these platforms is increasing day by day. In terms of the Turkish language, hate speech is frequently encountered in social media environments; but the number of studies on this subject is very few. In addition, hate speech on social media generally covers today's studies. Based on these two considerations, in this review study, the development of hate speech from past to present and hate speech in the Turkish language are discussed. The scope of our study is expressed in the hierarchical drawing of hate speech in Figure A. Here, hate speech is structured according to two main headings: application-based studies and social analysis studies. Accordingly, the studies were collected and evaluated through a literature review. Application-based studies include machine learning-based and deep learning-based applications for hate speech detection. As for social analysis studies, they are examined under four headings: subject-based attention-raising studies, descriptive analysis studies, discourse analysis studies and content analysis studies for hate speech detection and evaluated under the relevant headings.

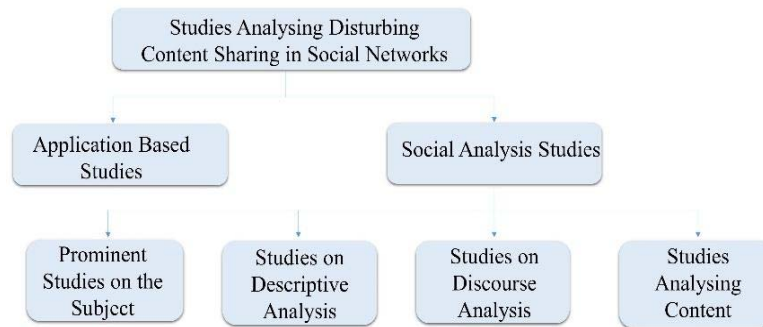


Figure A. Visual presentation of practice-based studies and social analysis studies in the field of hate speech

**Purpose:** The aim of the study is to raise awareness about the increase in Turkish hate speech in social media and the precautions taken in this regard with the developing technology.

**Theory and Methods:** The article is a compilation study. In this context, it covers traditional and current studies on hate speech on social networks. While conducting the literature review, the studies between 2019 and 2023 were examined and carried out according to two main headings: application-based studies and social analysis studies.

**Results:** In this study, studies examined under two separate headings are included in order to analyze the sharing of disturbing content on social networks. In application-based studies, algorithms such as SVM and LSTM were used for automatic hate speech detection systems, and Twitter datasets were preferred. It has also been stated that social media algorithms can encourage hateful posts. Social analysis studies, on the other hand, were examined under four sub-headings: descriptive analysis, discourse analysis, and content analysis. These studies dealt with the effects of social media on individuals, hate speech, and social problems in detail.

**Conclusion:** As a result of the study, hate speech in Turkish; detection difficulties, the inadequacy of open-access datasets, and sample limitations were observed. As long as all social media tools continue to create viral content with polarizing posts, it can be said that the production of hate speech concepts will continue and it will risk spreading rapidly in the future.



## Sosyal medyada nefret söylemi çerçevesi: Geçmişten günümüze kapsamlı bir derleme

Büşra Büyüktanır\*<sup>ID</sup>, Özlem Yakar<sup>ID</sup>, Ayşe Berna Altınel Girgin<sup>ID</sup>

<sup>1</sup>Marmara Üniversitesi, Teknoloji Fakültesi, Bilgisayar Mühendisliği Bölümü, 34854, Maltepe, İstanbul, Türkiye

### ÖNEÇIKANLAR

- Geleneksel ve dijital ağlar bağlamında nefret söylemi üzerine bir derleme
- Türkçe nefret söylemi probleminin zorlukları ve mevcuttaki durumu
- Makine öğrenmesi ve doğal dil işleme tekniklerinin Türkçe nefret söylemi çalışmalarında kullanılması

### Makale Bilgileri

Derleme Makaleleri  
Geliş: 15.07.2023  
Kabul: 02.03.2024

### DOI:

10.17341/gazimmfd.1327840

### Anahtar Kelimeler:

Nefret söylemi,  
sosyal medya,  
siber zorbalık,  
Türkçe nefret söylemi,  
doğal dil işleme

### ÖZ

Nefret söylemi, bireye veya gruba yönelik nefret duygusu içeren ve her türlü olumsuz uygulama ya da uygulamaya teşvik etmek amacıyla kullanılan ifadeler bütünüdür. Nefret söylemleri kapsamında üretilen olumsuz ifadelerin tümü ve bunların kullanıma yayıldığı ortamların başında sosyal medya ortamları gelmektedir. Türkçe dilinde nefret söylemlerine sosyal medya ortamlarında sıkça rastlanmaktadır; fakat bu konuda yapılan çalışma sayısı örneğin İngilizce dilinde yapılan çalışma sayısına göre oldukça azdır. Bununla birlikte, Türkçe dilindeki nefret söylemi çalışmalarını içeren kapsamlı bir literatür taraması bizim bildiğimiz kadarıyla henüz bulunmamaktadır. Bu çalışma, bahsedilen bu boşluğu doldurarak bu konuda araştırmalar yapmayı amaçlayan hem ticari hem de akademik platformların faydalanabilmesi için hazırlanmıştır. Bu literatür taraması içerisinde Türkçe dilindeki nefret söylemleri çalışmaları, uygulama tabanlı çalışmalar ve sosyal analiz tabanlı çalışmalar kategorileri bazında detaylı şekilde incelenmiştir. Bununla birlikte, bu konuda çalışmalar yapmayı planlayan tüm diğer araştırmacıların faydalanması için konuyla ilgili temel kavramlar, kullanılan öğrenme algoritmaları ve doğal dil işleme teknikleri de anlatılmıştır. Ayrıca yine bu çalışma kapsamında, Türkçe nefret söylemi problemi özelindeki zorluklar, gelecekte yapılabilecek olası çalışmalar ve diğer dillerdeki nefret söylemi çalışmalarının geldiği durum gibi okuyuculara faydalı olabilecek ek bölümlere de yer verilmiştir. Bu literatür taraması, Türkçe nefret söylemi problemiyle alakalı bahsedilen bütün bu bölümleri içeren detayda hazırlanmış bildiğimiz kadarıyla ilk çalışmadır.

## Hate speech framework in social media: A comprehensive review from past to present

### HIGHLIGHTS

- A Review on hate speech in the context of traditional and digital networks
- Challenges and current status of the Turkish hate speech problem
- Use of machine learning and natural language processing techniques in Turkish hate speech studies

### Article Info

Review Article  
Received: 15.07.2023  
Accepted: 02.03.2024

### DOI:

10.17341/gazimmfd.1327840

### Keywords:

Hate speech,  
social media,  
cyber bullying,  
Turkish hate speech,  
natural language processing

### ABSTRACT

Hate speech is a set of expressions that contain a sense of hatred towards an individual or a group and are used to encourage all kinds of negative practices or practices. All of the negative expressions produced within the scope of hate speech and the environments in which they are disseminated are social media environments. Hate speech in the Turkish language is frequently encountered in social media environments; however, the number of studies on this subject is quite low compared to the number of studies conducted in the English language. In addition, to our knowledge, there is not yet a comprehensive literature review on hate speech in the Turkish language. This study has been prepared to fill this gap and to benefit both commercial and academic platforms that aim to conduct research on this subject. In this literature review, hate speech studies in the Turkish language are analysed in detail based on the categories of application-based studies and social analysis-based studies. In addition, the basic concepts, learning algorithms and natural language processing techniques are also explained for the benefit of all other researchers who plan to conduct studies on this topic. In addition, this study also includes additional sections that may be useful for the readers, such as the challenges specific to the Turkish hate speech problem, possible future studies, and the current state of hate speech studies in other languages. To the best of our knowledge, this literature review is the first study on the Turkish hate speech problem that includes all these sections in detail.

\*Sorumlu Yazar/Yazarlar / Corresponding Author/Authors : \*busra.buyuktanir@marmara.edu.tr, ozlemyakar@marun.edu.tr, berna.altinel@marmara.edu.tr / Tel: +90 216 777 3816

## 1. Giriş (Introduction)

Nefret söylemi, kişilere, insanlara ve olaylara karşı hakaret içeren, ayrımcı ifadeler kullanarak toplumsal hoşgörüyü ve saygıyı zedeleyen olumsuz bir söylem şeklidir. Türkçe’de kullanılan “Nefret” kelimesinin, Türk Dil Kurumu’na göre tanımı, “bir kimsenin mutsuzluğunu, kötülüğünü istemeye dair his” veya “tiksinme/tiksinti” anlamlarında ifade edilmektedir [1]. Amerikan Anayasası Ansiklopedisi’nde hukuk profesörü olan Nockleby, nefret söylemini “bir kişiyi ya da grubu renk, ırk, etnik köken, cinsiyet, cinsel yönelim, din, milliyet ya da diğer nitelikler gibi ayrımcı her türlü ifade” biçiminde açıklamıştır [2]. Genellikle cinsiyetçilik, yabancı düşmanlığı, ırkçılık, islamofobi, homofobi ve diğer ayrımcılık biçimlerini içeren nefret duygularından kaynaklanan nefret söylemi, sözlü veya yazılı olarak, görsellerle ve diğer iletişim araçlarıyla iletilir. Gazete, televizyon, radyo, sosyal medya, bloglar, forumlar, haber siteleri ve diğer internet platformları nefret söyleminin iletilmesi için yaygın kullanılan alanlardır. Şekil 1’de nefret söyleminin kullanıldığı alanlar görselleştirilmiştir.

### 1.1. Sosyal Medyada Nefret Söyleminin Tanımı ve Önemi (Definition and Importance of Hate Speech in Social Networks)

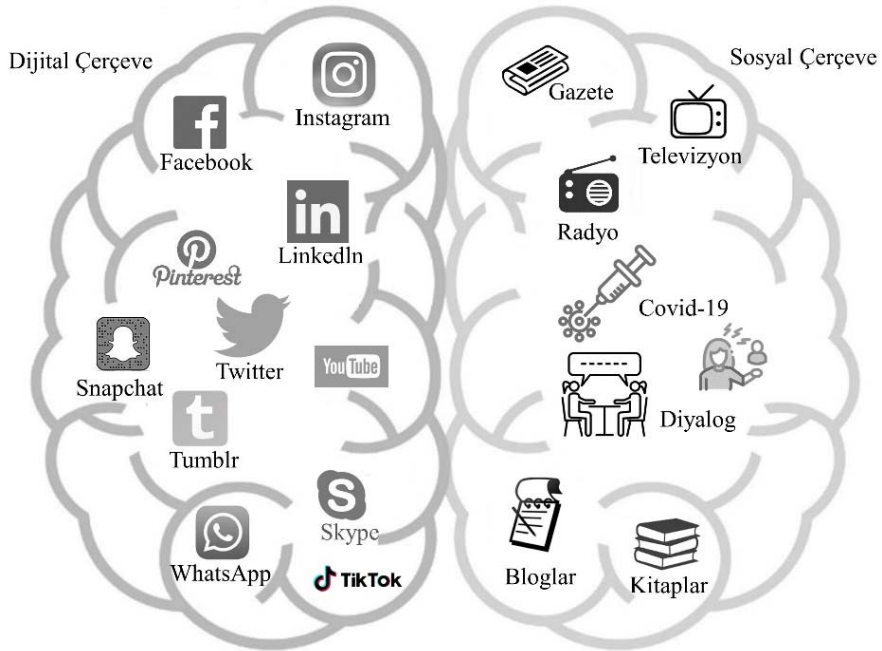
Sosyal medya, insanların birbirleriyle etkileşimde bulunarak, iletişim kurmasını ve bilgi paylaşımını sağlayan, çeşitli ve farklı uygulamaların olduğu çevrimiçi platformlardır [3]. Bu platformların kullanımının artması, sosyal medyada sürekli bir içerik üretimine ve veri paylaşımına neden olmaktadır. Üretilen içerikler arasında nefret söylemi içeren ifadeler de yer almaktadır. Sosyal medyada nefret söylemi, cinsiyet, cinsel yönelim, milliyet, din, dil, ırk gibi konularda kişi ya da gruplara olumsuz söz veya hakaret içeren ifadelerin çeşitli platformlarda kullanılmasındır. Bu tür söylemler, toplumda psikolojik şiddete, nefret duygularının uyanmasına, ayrımcılığa, önyargılara ve insan hakları ihlallerine yol açabilmektedir. İnsanların geniş kitlelere ulaşabildiği sosyal medya içerisinde kullanılan nefret söylemi, bireylerin haklarını ve özgürlüklerini ciddi şekilde etkilediği için insan ve toplum üzerinde zararlıdır. Bu tür söylemler, hedeflenen

kişilerin kendilerini güvende hissetmemelerine, toplum tarafından haksız muameleye, şiddete veya dışlanmaya maruz kalmalarına neden olabilmektedir. Bu nedenle toplumsal barış, adalet ve insan hakları açısından nefret söyleminin önlenmesi önemlidir.

### 1.2. Nefret Söylemi Kavramlarına Genel Bakış (Overview of Hate Speech Concepts)

Günümüzde sosyal medyanın da etkisiyle nefret söylemlerinde sayıca artış görülmektedir. Özellikle sosyal medyanın interaktif yapısı, nefret söylemlerine dair çeşitli kavramların günlük hayatta ve dijital alanda kolayca yayılmasına zemin oluşturmaktadır. Nefret söylemlerinin yaygınlaşmasının nedeni, öncelikle kişilerin toplum içerisinde diğerleri için birtakım önyargılar oluşturması, zihinlerde farklı düşüncelerin varlığının getirdiği kabullenememe durumları, kişilerin kendisinden farklı bireyleri “öteki, diğeri, şu, bu” şeklinde kategorilendirerek meydana getirdikleri ruh halinin oluşturduğu önyargılar ve ötekileştirmeden kaynaklı tüm olumsuz duygu durumlarından (korku, huzursuz hissetme, mobbing, öfke, dışlanma, nefret, şiddet vb.) kaynaklanmaktadır. Bunun yanı sıra, dolaylı yoldan da olsa nefret söylemlerinin kinaye, imâ, çarpıtma, abartma şeklindeki ifadeleri, kişilerde söylenen sözlere olumsuz anlamlar yükleme, hakaret, küfür, aşağılama, nefret hissi gibi duygulara dönüşerek, hızla paylaşma ve yayılmasına imkân tanımaktadır. Kimi çevrelerce ifade özgürlüğünün bir parçası veya bir “iç dökme eylemi” olarak kabul edilen bu ifadeler [4], sosyal medyanın doğal bir uzantısı durumuna getirilmiştir.

Nefret söyleminin genelde olumsuz duygu durumlarından kaynaklı ifadeler olduğu söylene de sosyal medyada sıkça kullanılan nefret söylemi kavramının, evrensel nitelikte kabul edilen bir tanımı bulunmamaktadır [5]. Birçok ülke, nefret söylemine karşılık gelen ifadeleri yasaklayan kanunları kabul etseler de yasaklanan olgunun ne olduğu belirlenirken tanımlar değişiklik göstermektedir [6]. Avrupa Konseyi Bakanlar Komitesi tarafından 1997 yılında yayınlanan R (97)20 sayılı Tavsiye Kararı ile “nefret söylemi” kavramı şu şekilde ifade edilmiştir: “Nefret söylemi kavramı, yabancı düşmanlığı, ırkçı



Şekil 1. Nefret söyleminin genel çerçevesi. (General framework of hate speech.)

nefreti, Yahudi düşmanlığını veya azınlıklara, göçmenlere ve göçmen kökenli insanlara yönelik etnik merkezlik ve saldırgan ulusalcılık, düşmanlık ve ayrımcılık şeklinde ifade edilen, dinsel açıdan hoşgörüsüzlük dahil olmak üzere her türlü hoşgörüsüzlüğe dayalı farklı nefret biçimlerini teşvik eden, kışkırtan, yayan veya meşrulaştıran her türlü ifade biçimini içerecek şekilde anlaşılacaktır [7]. Bu anlamda nefret söylemi, mutlaka belirli bir kişiye ya da gruba yönlendirilmiş yorumları içermektedir.” Komite tarafından nefret söylemi çerçevesinin tanımı, bu şekilde genel açıdan belirlenmiş olsa da dünyada olduğu kadar ülkemizde de herkese göre olumsuzluk içeren ifadelerin tümü, nefret söylemi çatısı altında toplanabilir. Tüm bu olumsuz ifadelerin dünya çapında yayılması kadar ülkemizde de Türkçe dilinde olumsuz söylemlerin denetimsiz biçimde yayılımı, farklı bireyler ve gruplara yönelik nefret söylemlerine kolayca erişim ve ulaşım olanağı oluşturmaktadır. Aşağıda genel olarak, nefret söylemi kavramlarının ifadeleri incelenmiştir.

### 1.2.1. Ayrımcılık (Discrimination)

Ayrımcılık, bireye veya gruba sahip olduğu nitelikler nedeniyle eşit şekilde davranmama ya da özensif davranma şekli olarak tanımlanır [8]. “Adil olmayan davranış şekli” olarak da ele alınan “Ayrımcılık” kavramı, kişi/kişilere veya bir gruba karşı takınılan negatif tutumlar nedeniyle, meydana gelen olumsuz davranışlardır. Ayrımcılık, nefret söylemlerinde esas alınan ilk kavramdır ve sosyal medyada nefret söylemleri arasında sıkça kullanılmaktadır. Bu kavram, mantık çerçevesinde olmadan sebepsiz bir yere birey/bireylere farklı davranışlar göstermek şeklinde ortaya çıkmaktadır.

Genelde ayrımcılık ile ilgili yapılmış tanımlamalarda vurgulanan, bireylerin ya da toplumdaki grupların, geçerli olmayan sebeplerle olumsuz şekilde kategorilendirilmesi ve buna göre söylem ve eylemde bulunulması mantığı yer almaktadır [9]. Sosyal medya ifadelerinde bu mantık, üstü kapalı şekilde “ötekiler, biz” olarak ifade edilmektedir. Bu durum, günlük hayatta olduğu gibi sosyal medya alanında da kullanıcılar arasında ayrımcı sözlere ve ayrımcılık oluşturan özel ifadelere dönüşmektedir. Farklı insan grupları tarafından bu durumlar, eşitsizlik ve dışlanma şeklinde algılanarak, bireyler veya gruplar arasında nefret söylemleri şeklinde yorumlandığı görülmüştür. Tüm bu durumlar göz önüne alındığında, farklı özelliklere sahip bireylere yapılan ayrımcılık ifadeleri ve davranışları, geçmişte olduğu gibi, günümüz teknoloji çağında da sosyal medyada devam etmektedir.

### 1.2.2. Ötekileştirme (Marginalization)

Günlük hayatta, toplumsal bağlamda ve ülkemiz içindeki edindiğimiz bilgilerin önemli ve büyük bir kısmı medya araçları aracılığıyla temin edilmektedir. Teknolojinin hızla gelişimi, bilgi edinme açısından medya araçlarına sosyal medya kavramını da katmıştır. Dünyada ve ülkemizde gerçekleşen farklı bilgiler ve düşüncelere dair sosyal medyanın bu denli giderek önem kazanmış olması, geniş kitlelere ileteceği olumlu ifadeler kadar, olumsuz ifadeleri de beraberinde getirmektedir. Sosyal medyanın insanlar üzerindeki gücünün bu denli artması da nefret söylemlerinden “ötekileştirme” kavramını ortaya koymaktadır.

Her ülkede olduğu gibi, ülkemizde de farklı insan grupları, toplum içinde bir arada yaşamaktadır. Bu durum, sosyal medya araçları vasıtasıyla toplumsal normların yeniden üretilmesine yol açabilecek şekilde, farklı grupların ötekileştirilmesine neden olabilmektedir. Var olan bu eğilim ile sosyal medyanın ilgi ve beğenisini ayakta tutacak şekilde toplumsal bireylerin gruplandırılarak oluşturulduğu görülmüştür. Bu yüzden, örneğin, Türk toplumu içerisindeki “biz kavramı”, sosyal medya tarafında da Türk ifadesini belirginleştirmek için “biz” kavramı şeklinde ele alınmaktadır. Buradaki “biz” kavramı, “Türk, Sünnî, Müslüman, Muhafazakâr, Erkek, Kadın,

Heteroseksüel” gibi sözcük kalıpları ile ifade edilmektedir [10, 11]. Toplumda bu tür kalıplar yer aldığından, sosyal medya bu türdeki sözcük kalıplarına göre ötekileştirmeyi sürekli şekilde yenileyerek, bunların dışında kalan tüm grupları da dışlamaktadır [12].

### 1.2.3. Önyargı (Prejudice)

Önyargı, günlük hayattaki sadeleştirilmiş kullanımının yanı sıra ontogenetik, evrimsel, ekonomik, tarihsel ve sosyo-politik alanlardan etkilenen çok kapsamlı bir olgudur [13]. Bu olgu, bireylerin genelde farkında olmadan karar verdiği olumsuz bir fenomen veya kavram olarak karşımıza çıkmaktadır. Bu kavramın insanlar arasında ve toplumsal ilişkiler açısından negatif etkileri bulunmaktadır. Bireyler, karar alma süreçlerinden seçimlerine kadar toplumsal yaşamın çeşitli alanlarında bilinçsiz bir şekilde önyargılara sahip olabilirler veya bu önyargıları dışa vurabilirler. Önyargılar, yabancılara karşı (bizden olmayanlar) olumsuz tutum ve düşmanlığın oluşmasında önemli bir role sahiptir. Diğer taraftan, bireysel veya geniş grupların korunmasında da önyargılar önemli bir göreve sahiptir.

Önyargı kavramı, gelişen teknolojiyle birlikte sosyal medya alanlarında sıkça görülmektedir. Günümüzde, insanlar aynı anda birden çok sayıda bilgiye maruz kalmaktadır. Bu nedenle, karar verme aşamasında zihnin kısa yollarına başvurmaktadır [14]. Teknolojinin gelişimi ile birlikte insanların günlük hayattaki önyargılarını kısa yol olarak sosyal medyaya taşıdığı görülmüştür. Özellikle sosyal medyanın görsel etkileri, bireyler üzerinde daha çok önyargı oluşturmaya neden olmaktadır. Hem bireylerin hem de büyük grupların sergilediği ayrımcı tutumlar ve davranışlar, önyargıları besleyen durumlardandır. Sosyal medya platformlarında yaşanan bu durum, bireyler ve toplumlar arasında stresli ve belirsiz durumları arttırarak, nefret söylemlerini teşvik etmektedir. Önyargı da bir çeşit ayrımcı tutumdur ve nefret söylemleri arasında sıkça karşılaşılmaktadır. Geçmişte olduğu gibi, günümüzde de önyargı kavramının doğası araştırılmakta, ayrımcı tutumlar üzerindeki artışın neden yaşandığı ve bunun nasıl azaltılabileceği sorularına cevap aranmaktadır [15].

### 1.2.4. Kalıp yargı (Stereotype)

Geçmişten günümüze kişilerin ve toplumların birbirleriyle olan etkileşimi sürerek devam etmiştir. Bu etkileşimler, günümüze kadar işitsel, görsel ve yazınsal kaynaklarla aktararak, kişiler ve toplumlar arasında olumlu/olumsuz düşünce, tutum ve davranışların gelişmesine de yardımcı olmuştur. Bu noktada, iletişim süreçlerinde bireyler ve toplumlar arasında, hedef kitleye yönelik düşünsel ve davranışsal açıdan “kalıp yargı (stereotip)” içeren imgesel yaklaşım ve bakış açılarıyla sürekli karşılaşılmaktadır [16]. Kalıp yargı, “diğer bireyleri ya da grupları” içine yerleştirdiğimiz “kategorileri” ifade etme şeklidir; bunun yanında, “diğer bir bireyi veya bireysel grupları” ifadesini tanımlayabilmek için de kullanılan basitleştirilmiş betimsel kategorileri göstermektedir [17]. Genellikle yaş, cinsiyet, milliyet, azınlık grupları ve meslek grupları gibi kategorilere odaklanan kalıp yargılar, bu kesimlerde yer alan bireyleri, sahip oldukları betimsel özelliklerle değerlendirerek, çevrelerindeki diğer kesimlerdeki bireyler hakkında kalıp yargılara varmalarına neden olmaktadır. Bireylerin, farklı kişileri değerlendirdiklerinde hangi kalıp yargılara sahip olacakları, değerlendirdikleri bireylerin çeşitli niteliklere sahip olmasına bağlı olarak oluşmaktadır [18]. Bu niteliklerin, günümüz teknolojisinin de gelişimiyle birlikte sosyal medya alanında yer bulduğu söylenebilir. Sosyal medyada farklı kalıp yargılarla hareket ederek birbirlerini destekleyen bireyler, nefret söylemleri üretmektedir. Bunun yanında, sosyal medyada manipülatif metinlerin oluşturulması da insanları kalıp yargılarla hareket etmeye sevk etmektedir. Tüm bu durumlar birleştirildiğinde, kalıp yargıların önyargılardan oluşturulan kalıplaşmış ifadeler olduğu söylenebilir.

### 1.2.5. Çokkültürlülük (Multiculturalism)

Çokkültürlülük, bir toplumdaki kültürel, etnik, ırksal, dilsel, dinsel ve/veya bireysel tüm farklılıkların, kamusal ve siyasal alanda kendini tanıması ve özgürce ifade etmesi anlamına gelir [19]. Farklı sosyal grupları ötekileştiren (işçiler, kadınlar, ateistler, geysler, engelliler vs.) bir kullanımının yanında çokkültürlülük, ulusal ve etnik bağlamdan kaynaklanan dinamikleri de ifade etmektedir. Bu ifadeler, çokkültürlü gruplar üzerinde kendinden olmayı dışlama, aşağılama ya da hakaret şeklinde olup, üretilen ifadelerin tümü sosyal medya aracılığıyla dünyaya duyurulmaktadır. Dünyada bir yerden bir başka yere yapılan büyük göçlerin yaşanması teknolojik açıdan sosyal medya araçlarının gelişimini ve yaygınlaşmasını sağlarken, çokkültürlülük kavramında da etnik çapta bir dönüşüme yardımcı olmaktadır. Bu açılarından bakıldığında, çokkültürlülük kavramı farklı kültürlerin, grupların, dinlerin ve milletlerin değerlerini içeren ve bu değerleri bir bütün olarak tanımlayan kavram olarak karşımıza çıkmaktadır.

Türkiye’de çokkültürlülük kavramı incelendiğinde, konumu itibarıyla jeopolitik açıdan farklı göç hareketlerinin odak noktası halinde olmuştur [20]. Bu durum, günümüzdeki teknoloji çağını da düşününce, sosyal medya “küresel teknolojik çağ” olarak adlandırılabilir. Sosyal medyanın iletişimden ziyade, genelde nefret söylemleri üreten bir ağ olduğu düşünüldüğünde, ülkemizde ve dünyada göçler sonrası çokkültürlü yapıya sahip olan tüm gruplara yönelik, özel nefret söylemleri üretildiği ve çokkültürlü toplumlarda bir dezavantaj oluşturduğu söylenebilir. Tüm bu durumlar, toplumun bir arada barış ve hoşgörü içinde yaşamasını engelleyen faktörleri oluşturur.

### 1.3. Sosyal Medyada Nefret Söylemi Olarak “Siber Zorbalık” ve “Siber Mobbing” Terimleri

*(The Terms "Cyberbullying" and "Cyber Mobbing" as Hate Speech on Social Media)*

#### 1.3.1. Siber zorbalık (Cyberbullying)

İçinde bulunduğumuz teknoloji çağının en önemli sorunlarından birisi, dijital teknolojiler ve sosyal medya kullanılarak insanlara zarar vermeye amaç haline getiren “siber zorbalık” teriminin göstermiş olduğu davranışlardır. Siber zorbalık, başkalarına verilen rahatsızlık veya zarar vermek amacıyla bireylerin ya da grupların çevrimiçi ortamlarda gösterdikleri tutum ve davranışlar olarak adlandırılır [21]. Bu terim, nefret söylemi ile benzer işleve sahiptir. Aradaki temel fark, nefret söyleminde ırk, din, dil, milliyet, cinsiyet, cinsel yönelim nedeniyle bir kişi ya da grup hedef alınırken, siber zorbalıkta çevrimiçi ortamlar kullanılarak genellikle bireylere veya gruplara yönelik saldırılar gerçekleştirilmektedir [22]. Siber zorbalık, çevrimiçi ortamlarda bir kişinin, başka bir kişiye bilerek ve düzenli şekilde rahatsızlık vermesi, zarar verme amacı taşıyan eylemlerde bulunması ve olumsuz olarak tanımlanabilecek her türlü nefret söylemi ifadelerini karşı tarafa yansıtmasıdır. Siber zorbalıkta genellikle kısa mesajlar, e-postalar, tüm mesajlaşma uygulamaları, forumlar, sosyal medya uygulamaları, oyun platformları gibi bilgi paylaşımına izin veren çevrimiçi ortamlar kullanılmaktadır. Bir kişi hakkında kötü niyetli metinlerin üretimi, olumsuz içerikli mesajların gönderimi, bireyler veya gruplar hakkında çevrimiçi yanlış bilgi yayılması, bireyi/hedefi/grupları aşağılama, taciz etme veya tehdit etme, utanç verici fotoğrafların/videoların paylaşımı siber zorbalık türünün örneklerindedir. Siber zorbalık, teknolojinin kötüye kullanımı ile giderek büyüyen ve müdahale edilmesi gereken toplumsal bir sorundur [23]. Çevrimiçi farklı platformlar kullanılarak gerçekleştirilebilen bu olgu, bireylerde psikolojik, duygusal ve ruhsal sorunlara yol açmaktadır.

### 1.3.2. Siber mobbing (Cyber mobbing)

Gelişen teknolojinin insanlar arasındaki iletişimi şekillendirme biçimi, iş hayatında da etkisini göstermektedir. İş hayatında sanal ve yüz yüze iletişim iç içe geçmiştir. Bu durum, çalışma hayatını ve iş ilişkilerini iki katmanlı bir yapıya doğru dönüştürmektedir. Yüz yüze ve sanal iletişimin iç içe geçtiği bu süreçte, iş yerlerinin önemli sorunlarından birisi olan “mobbing” kavramı, “çevrimiçi” düzlemde varlık göstermeye başlamıştır. Siber mobbing, teknolojik yıldırma ile aynı anlama gelen, iş yerinde olumsuz ayrımcılık türü olan mobbing davranışların bilişim teknolojileri ve iletişim kullanılarak çevrimiçi ortamda sergilendiği bir terimdir [24]. Bu terim, maruz kalınan çevrimiçi saldırıların nefret söylemleri kullanılarak ifade edilmesidir. Siber mobbing ile siber zorbalık kavramları birbirine yakın kavramlardır. Arasındaki temel fark şudur: siber zorbalık için hedef kitle “herkese karşı yapılan saldırılar” şeklindedir [25]; etki alanı daha geniş, etki gücü daha yüksek ve etki süresi daha uzun olabilir. Siber mobbing için ise, iş yerine özgü yüz yüze iletişimdeki mobbing sürecin çevrimiçi uzantısıdır; etki alanı daha kısıtlı, etki gücü daha düşük ve etki süresi daha kısa bir mobbing türü olarak söylenebilir [24].

### 1.4. Sosyal Medyada Türkçe ve Diğer Diller Üzerinde Nefret Söylemi Tespitinde Durum Değerlendirmesi

*(Situation Assessment in Detecting Hate Speech on Social Media in Turkish and Other Languages)*

Günümüzde, teknolojinin gelişimi ve internetin yaygınlaşmasıyla insanlar kendilerini ifade etme yolu olarak sosyal medyayı seçmiştir. Bu tercihler sosyal medyanın da gelişimini hızlandırarak, kullanıcıların önemli ölçüde farklı içerikler üretmelerine yardımcı olmaktadır. Sosyal medyanın bu derece yoğun biçimde kullanımını, nefret söylemi gibi çeşitli sosyal medya problemlerini de beraberinde getirmektedir [17]. Nefret söylemi içeriklerinin önlenmesine yönelik sosyal medya çalışanları belirli politikalarla bu durumun önüne geçmeye çalışsa da içeriklerin boyutunun hacimce artışı, bu durumun önüne geçmeyi zorlaştırmaktadır. Bu yüzden, sosyal medya çalışanları da nefret söylemini önlemek adına kendi politikalarını belirlemektedirler.

Sosyal medyada nefret söylemi tespitinde, son yıllarda yapılan çalışmalar incelendiğinde, literatürde en çok İngilizce-Hintçe dillerindeki [26-29] çalışmalar göze çarpmaktadır. Bunun yanında, İngilizce, Almanca ve İtalyanca dilleri [30], Hintçe’nin Bengali [31], Sinhali [32] ve Dravid dilleri [33], Bangladeş’in Bangla dili [34], Arapça [35, 36], Kazakça [37], Roman-Urduca [38] ve Endonezce [39] gibi farklı dillerde çalışmalar mevcuttur. Bu çalışmalar incelendiğinde, İngilizce-Hintçe dili hariç diğer dillerde daha az sayıda çalışma olduğu görülmüştür. Bunun nedeni, farklı dillere ait kamuya açık veri kümelerinin yetersiz oluşu ve farklı dil yapılarındaki yapısal zorluklar olduğu düşünülmektedir. Türkçe dilindeki nefret söylemleri çalışmalarına [40, 83, 91] bakıldığında ise, dilin yapısal farklılıkları ve veri kümelerindeki yetersizlikler daha çok sayıda çalışma yapılmasını gerektirmektedir.

Makalenin geri kalanı şu şekilde organize edilmiştir: Bölüm 2’de, Türkçe nefret söylemi probleminin otomatik olarak uygulamalar tarafından tespit edilmesi için, model eğitiminde kullanılan makine öğrenmesi algoritmalarına, doğal dil işleme tekniklerine ve modelin performans değerlendirme metriklerine yer verilmiştir. Ardından Bölüm 3’te, literatürde yer alan konuyla ilgili uygulama geliştirilmiş çalışmalar ve sosyal içerikli çalışmalar ayrıntılı olarak açıklanmıştır. Bölüm 4’te ise, yapılan literatür taramalarında kullanılan teknik yöntemlerin kıyaslanması ve bu literatür çalışmalarında paylaşılan açık erişimli veri kümeleri ele alınmıştır. Son olarak Bölüm 5’te ise incelenen makalelerden hareketle gözlemlenen güncel zorluklara,

gelecekte yapılmasına dair açıklanan yönlendirmelere ve çalışma değerlendirme sonucuna yer verilmiştir.

## 2. Türkçe Nefret Söylemi Problemi Üzerinde Kullanılan Makine Öğrenmesi Algoritmaları, Doğal Dil İşleme Teknikleri ve Performans Değerlendirme Metrikleri Hakkında Temel Kavramlar

(Basic Concepts on Machine Learning Algorithms, Natural Language Processing Techniques and Performance Evaluation Metrics Used on the Turkish Hate Speech Problem)

Teknolojinin gelişmesiyle birlikte, iletişim konusunda sosyal medya araçlarının önemi giderek artmaktadır. Bu durum, sosyal medya araçlarının sayısına bağlı olarak sosyal medya kullanıcı sayısının ve bu platformlar üzerindeki kullanıcı etkileşimleri ile nefret söylemi içerikli ifadelerin artmasına neden olmaktadır. Bunun yanında, kullanıcılar tarafından sürekli üretilen veri hacminin büyümesi nedeniyle nefret söylemi içeriklerinin tespitinde otomatik sistemlere ihtiyaç duyulmaktadır. Bu sebeple, Türkçe nefret söylemi tespiti için farklı makine öğrenmesi algoritmaları ve doğal dil işleme teknikleri kullanılarak otomatik tespit sistemleri geliştirilmektedir. Çalışmanın bu bölümünde, bu kapsamda kullanılan farklı algoritmaların ve tekniklerin neler olduğuna yer verilmiştir. Ayrıca, bu algoritmalar ile eğitilen modellerin performansını ölçmek için kullanılan performans metrikleri de açıklanmıştır.

### 2.1. Makine Öğrenmesi Yöntemleri (Machine Learning Methods)

Makine öğrenmesi, bilgisayar sistemlerinin verileri kullanarak öğrenme yeteneği kazanmasını sağlayan yapay zekanın bir dalıdır. Makine öğrenmesi yöntemi için çeşitli algoritmalar geliştirilmiştir. Algoritmalar ile öğrenme yöntemi iki temel kategoride incelenebilir: denetimli öğrenme ve denetimsiz öğrenme. Denetimli öğrenme, etiketli veri kümeleri üzerinden öğrenme yaparak girdi ve çıktı arasındaki ilişkiyi modellemeyi amaçlar. Denetimsiz öğrenme ise, etiketlenmemiş veri kümeleri üzerinden desenler ve ilişkiler çıkarmayı hedefler. Bu öğrenme yöntemleri dışında takviyeli öğrenme ve yarı denetimli öğrenme yöntemleri de bulunmaktadır. Makine öğrenmesi, birçok uygulama alanında kullanılarak karmaşık problemlerin çözümüne ve veri analizi süreçlerine katkı sağlar [41].

#### 2.1.1. Destek vektör makineleri (Support vector machines - SVM)

Destek Vektör Makineleri (Support Vector Machines, SVM), denetimli öğrenme yöntemlerinden birisi olup, sınıflama ve regresyon problemlerinde sıkça kullanılan bir algoritma türüdür. SVM'nin temel amacı, veri kümesindeki örnekleri sınıflar arasında en geniş marjı sağlayacak şekilde ayırmaktır. Algoritma önce veri kümesindeki örnekleri iki veya daha fazla sınıfa ayırır ve ardından bu sınıflar arasında en geniş boşluğu sağlayan bir karar sınırı belirler. Bu karar sınırına "marjın" adı verilir. Marjın ne kadar geniş olursa sınıflar o kadar iyi ayrıştırılır. Öte yandan, SVM boyut indirgeme yöntemlerini kullanarak özneliklerin boyutunu azaltır. Veri özneliklerinin doğrusal olarak ayrılabilir olmadığı durumlarda ise "kernel trick" olarak bilinen bir teknik kullanılır. Bu teknikte, veri özellikleri başka bir boyuta yansıtılır (mapping) ve doğrusal olarak ayrılabilir hale getirilir. SVM, büyük ve karmaşık veri kümeleri üzerinde performansıyla öne çıkan bir makine öğrenmesi yöntemidir [41-43].

#### 2.1.2. Karar ağaçları (Decision tree - DT)

Karar Ağaçları, veri kümelerinin yapısını görsel bir şekilde temsil etmek için ağaç yapısını kullanan sınıflandırma algoritmaları arasındadır. Bu ağaçlar, bir dizi karar kuralını takip ederek bir girdiyi sınıflandırır veya bir değeri tahmin eder. Ağaç yapısındaki her bir düğüm veri kümesindeki özelliği, karar noktasını; her bir kenar ise

özellik değerlerini, karar sonuçlarına götüren yolları temsil eder. Hedef değışkeni tahmin etmek için veri kümesindeki özelliklerin değerleri alınır. Yaprak düğümler hedef değışkenin tahmin edilen değerlerini içerir. Kökten hedef yaprağına kadar olan yol, hedefi sınıflandıran bir "kural" olarak adlandırılır. Bu kurallar "eğer-sonra" (if-then) yapısıyla ifade edilir. Karar ağaçları, açıklanabilirlik, basitlik ve yüksek performans gibi avantajları sayesinde en çok tercih edilen bir denetimli öğrenme algoritmasıdır [41, 44]. En çok kullanılan karar ağacı algoritmaları: ID3, C4.5, CHAID ve CART şeklinde sıralanabilir [45]. Bu algoritmalar arasındaki seçim, uygulama senaryosuna, veri tipine ve öğrenme hedeflerine bağlı olarak değışebilir.

#### 2.1.3. Rastgele ormanlar (Random forest - RF)

Rastgele Orman algoritması, sınıflandırma ve regresyon problemleri için sık kullanılan bir denetimli öğrenme algoritmasıdır. Birden fazla karar ağacının bir araya gelmesiyle oluşur. Algoritmanın temel çalışma prensibi, birbirinden bağımsız ve rastgele seçilmiş alt veri kümeleri üzerinde ayrı ayrı karar ağaçları oluşturup ardından bu ağaçların tahminlerini birleştirmektir. Alt kümeler bagging (bootstrap aggregating) yöntemi ile ana veri kümesinden bir veya birden fazla kez seçilerek üretilir. Ardından her alt kümede birbirinden bağımsız olarak karar ağaçları oluşturulmaktadır. Her bir alt kümede oluşturulan bu karar ağaçları kendi alt kümesi üzerinde tahminler yapmaktadır. RF, bu tahminleri bir araya getirerek genel bir tahminde bulunmaktadır. Rastgele Orman algoritması, kategorik/sürekli veri türlerinde ve farklı boyutlardaki veri setlerinde başarılı bir şekilde performans göstermektedir [41, 46].

#### 2.1.4. Naive Bayes sınıflandırıcıları (Naive Bayes classifier - NB)

Naive Bayes sınıflandırıcı, istatistik tabanlı bir denetimli öğrenme algoritmasıdır. Kolay uygulanabilir olması ve başarılı sonuçlar elde edebilme potansiyeline sahip olması sebebiyle model eğitiminde sıkça tercih edilmektedir. Temel çalışma mantığı, Bayes Teoremine dayanır.

$$P(A \setminus B) = P(B \setminus A) * P(A) \setminus P(B) \quad (1)$$

Yukarıda yer alan denklemde (Eş. 1), Bayes Teoremini ifade eden formüldür. Formüle göre A bir sınıfın gerçekleşmesini, B ise gözlemlenen öznelikleri temsil eder. Naive Bayes sınıflandırıcı algoritması ile her bir etiket/sınıf/hedef öznelik için Bayes Teoremi uygulanır. Örnekler sınıflandırılırken en yüksek olasılığa sahip olan sonuç seçilir [41, 47].

#### 2.1.5. Derin Öğrenme yaklaşımları (Deep Learning approaches)

Derin öğrenme, insan beyni gibi çalışacak şekilde modellenmiş olan ve yapay sinir ağları olarak isimlendirilen algoritmaların, çok büyük miktardaki veriler ile öğrenme eylemini gerçekleştirdiği makine öğrenmesinin bir alt dalıdır. İnsan beynindeki nöronlar düşünülerek geliştirilmiş olan yapay sinir ağlarında, tüm nöronlar birbirlerine bağlıdır ve çıktısı etkilemektedir. Nöronlar, giriş katmanı, gizli katmanlar ve çıkış katmanı olmak üzere üç farklı katmandan oluşmaktadır. Her katmandaki sinir ağları, önceki katmanlardan gelen bilgileri kullanır ve hiyerarşik şekilde sürekli olarak öğrenmeye devam eder. En sık kullanılan derin öğrenme algoritmaları arasında Yapay Sinir Ağları (Artificial Neural Networks, ANN) [48], Evrişimli Sinir Ağları (Convolutional Neural Networks, CNN) [49], Uzun-Kısa Vadeli Hafıza (Long Short Term Memory, LSTM) [50] ve Transformer tabanlı modeller (GPT, BERT vb.) [51] bulunmaktadır. Bu algoritmalar, büyük veri kümelerinde karmaşık desenleri öğrenme yeteneğiyle öne çıksalar da farklı görevlere uyum sağlamak üzere tasarlanmıştır [52].

## 2.2. Doğal Dil İşleme Teknikleri (Natural Language Processing Techniques)

İnternette sosyal medyanın yaygınlaşması ile nefret söylemlerinde artış olmaktadır. Sosyal medya madenciliği alanında yer alan nefret söylemi kavramının oluşturduğu içeriklerden oluşan birçok verinin analiz edilmesinde Doğal Dil İşleme teknikleri kullanılmaktadır. Nefret söylemi, son yıllarda aktif bir araştırma konusu halindedir. Dünyada farklı dillerde olduğu kadar ülkemizde de Türkçe dilinde nefret söylemi problemi üzerinde çalışmalara ağırlık verilmiştir ve farklı Doğal Dil İşleme teknikleri [53] kullanılarak çalışmalar yürütülmektedir.

### 2.2.1. Kelimeleri vektör uzayına dönüştürme (Converting words to vector space)

Bilgisayarın doğal dili anlayabilmesi için, metinlerin en küçük birimi olan kelimeleri anlaması gerekir. Makinelerin anlayacağı formatta kelimeler sembolize edilerek sayısal birime çevrilmelidir. Kelimelerin sayısal temsili ile her türlü matematiksel işlem yapılabilmektedir. Matematiksel olarak kelimelerin sayısal temsili vektörlerdir. Bu sayede makineler, kelimeleri anlayarak kodlara dönüştürüp gerekli işlemleri gerçekleştirebilir.

Doğal dil işleme alanına bakıldığında, dil modellemede farklı tekniklerden yararlanılmaktadır. Bu teknikler arasında, kelimelerin vektör uzayında temsil edildiği yaklaşımlardan Word2Vec ve GloVe modelleri yer almaktadır. Word2Vec modeli [54, 55], dil modellemede en çok kullanılan yöntemlerden birisidir ve kelime temsil vektörü oluşturmada kullanılır. Bu yöntemde, her kelime belirli uzunluğa sahip bir vektörle temsil edilir. Bu sayede, benzer kelimelerin vektörel açıdan birbirine yakın olması sağlanarak, kelimeler arasında anlamsal açıdan bir benzerlik ilişkisi sağlanır. Kelimeleri vektör uzayında temsil eden diğer bir yaklaşım da GloVe modelidir. GloVe (Global Vectors for Word Representation), Word2Vec'ten sonra en çok kullanılan kelime temsil yöntemlerinden birisidir. Bu yöntemde eğitim şekli, bir corpus üzerinde yakın anlamlı kelimelerin tespiti global kelime-kelime şeklinde yapılmaktadır [56]. GloVe yöntemine göre, optimize edilecek hata fonksiyonu Eş. 2'de [57] verilmiştir.

$$E_g = \frac{1}{2} \sum_{i,j=1}^M f(W_{ij}) (a_i^T b_j - \log W_{ij})^2 \quad (2)$$

Eş. 2'ye göre,  $a$  ve  $b$  simgeleri global kelime-kelime sayıları ile oluşturulan matrisin satır ve sütun değerlerini;  $f(W_{ij})$ , ağırlık fonksiyonunu;  $M$  ise, oluşturulan sözlük boyutunu ifade etmektedir.

### 2.2.2. Duygu analizi (Sentiment analysis)

Duygu analizi, belirli bir metindeki anlamsal yönelime dair görüşlerin gücünü ve kutupluluğunu tanımlayan duygudurumsal analizlerdir [58]. Yani, metnin belirli bir bölümüne ait duygu polarite durumunu çıkarmaktadır. Duygu polaritesi olarak metindeki görüşlere dair pozitif, negatif ya da nötr şekilde veya kanonik açıdan belirli bir ölçekteki görüşü temsil edecek şekilde sayısal olarak sınıflandırılmaktadır.

Son yıllarda özellikle sosyal medyanın popüler olmasıyla birçok sosyal medya sitelerinin (örneğin Twitter, Facebook, Instagram, YouTube vb.) ve alışveriş sitelerinin (örneğin AliExpress, Amazon, Hepsiburada vb.) kullanıcı faaliyetlerinin artışı ile büyük miktarda veri elde edilmektedir. Tüm bu verileri elle okuma ve yorumlama işlemi zor olacağından, bu süreci otomatik hale getirmek için duygu analizi tekniklerine ihtiyaç duyulmaktadır. Film yorumlarının analizi, sosyal medyada üretilen içeriklerin analizi, alışveriş sitelerinde

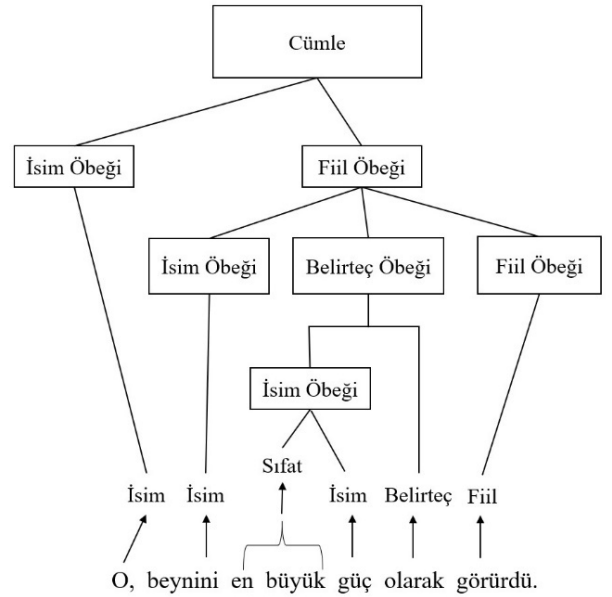
yapılan yorumlara göre müşteri duygudurum analizi gibi örnekler duygu analizi için verilebilecek örnekler arasındadır.

### 2.2.3. Morfolojik analiz (Morphological analysis)

Morfolojik analiz, kelimelerin tek başına ele alınarak dilin kural yapısına uygun biçimde inceleme şeklidir [61]. Bu inceleme ile her sözcüğe ait her bir parçasının analizi yapılmaktadır. Sözcüğe ait yapıların kökleri, ekleri, gövde yapılarının analizi ve bunlara ait kuralların sınıflandırılması morfolojik analiz çatısı altında ele alınır. Her dilin biçimbilimsel olarak yapısı farklı olmakla beraber, Türkçe dili için kelimelerin çekimlenmesi morfolojik analize örnek olarak verilebilir.

### 2.2.4. Sözdizimi analizi (Syntax analysis)

Sözdizimi, bir dilin cümle yapısını, bir araya gelen sözcüklerin nasıl kurullandığını ve bu sözcüklerin cümle içerisindeki sıralanmasını gramatik olarak ele alan bir dil bilimi dalıdır. Bir cümle içerisindeki kelimelerin anlamlı bir bütün oluşturabilmesi için birbirine uygun şekilde bir araya getirilmesi gerekir. Oluşturulacak bu bütünde, anlamlı bir cümlede varlığını koruyan her bir kelime "dizim" olarak adlandırılır. Bir cümle farklı şekillerde ayrıştırıldığında ya da çözümlendiğinde sözdizimsel olarak analizi yapılabilir demektir [59]. Analiz işlemi cümle içerisindeki gramatik yapıyı anlamak için, kelimeler arasındaki ilişkinin belirlenmesi gerekir. Bu kelimelerin bir araya gelmesi durumuna ait kurullarda, sözcüklerin birbirleriyle olan bağlantısı ve birbirlerini nasıl/ne kadar etkiledikleri sözdizim biliminin konusudur ve sadece sözdizimsel analiz yöntemleriyle açıklanabilir. En temelde bir cümlenin sözdizimi analizi Şekil 2'deki [59] gibi verilebilir.



Şekil 2. Cümlenin ağaç yapısını oluşturma örneği [59].  
(An example of creating tree structure of a sentence [59].)

### 2.2.5. Adlandırılmış varlık tanıma (Named entity recognition - NER)

Bilgisayar tarafından okunabilen büyük miktardaki metinsel verilerin ve bu verileri işleyebilen donanımların varlığı, bilgi projelerinde odağı Doğal Dil İşleme tekniklerine doğru kaydırmıştır. Doğal Dil İşleme tekniklerinden özellikle Adlandırılmış Varlık Tanıma (NER), bilgi çıkarmanın (IE) bir alt görevidir ve büyük metinlerden bilgi çıkarma görevi için kullanılmaktadır.

Metindeki önemli nesnelere (kişi, konum, kurum) belirleme, metindeki isimleri/yerleri tanıma, özel isimleri sınıflandırma, zaman ifadeleri/parasal değerler/tıbbi kodları belirleme gibi yapılandırılmamış farklı metinlerdeki semantik türler için önceden tanımlanan kategorilerin sınıflandırılması işlemini gerçekleştirir [60]. Genellikle makine çevirisi (Machine Translation, MT), metin özetleme, metin anlama ve soru yanıtama (Question Answering, QA) gibi birçok doğal dil uygulamalarında kullanılan NER, araştırmacılara farklı bir öğrenme yöntemini sunar. NER ile ilgili yapılan literatür çalışmaları incelendiğinde, dil alanındaki çalışmalarda genellikle zengin kaynak dillerinden İngilizce ve Çince dillerinde [61-67] kullanıldığı görülmüştür [68]. Son yıllarda NER sistemlerini oluşturan yaklaşımlar, araştırmacıların giderek ilgisini çektiği için, düşük kaynaklı diller arasında sayılan Türkçe dili [69-72] üzerinde de kullanılarak geliştirilmektedir.

Genel olarak, NER metinlerdeki varlık isimlerini (insanlar, konumlar, değerler vb.) tespit ederek, önceden belirlenen kategorilere göre sınıflandırma işlemi yapmaktadır. Örnek vermek gerekirse, “Ayşe, Eylül 2020’den beri Microsoft, Washington’da çalışıyor.” cümlesinde, Kişi: Ayşe, Konum: Washington, Kurum: Microsoft ve Tarih: Eylül 2020 şeklinde bilgi çıkarımı ile varlık isimleri sınıflandırılır. NER yönteminin genel çalışma prensibi Şekil 3’te [73] gösterilmiştir.

### 2.2.6. N-gram modeli (N-gram model)

N-gram, bir metin içerisindeki sözcüklerin n tane karakter tabanlı gösterim yöntemidir [74]. Bu yöntemde, bir metinde/dokümanda yer alan her bir kelimenin n-boyutlu karakterlere ayrılması işlemi yapılır. Burada, “N” karakterlere ayırma işlemi ve “Gram” ise karakterlerin ağırlığını ifade etmektedir. Bu işlemler için metinde her bir kelimenin tek başına gösterimi şeklinde Uni-gram (1-gram), ardışık iki kelimenin gösterimi şeklinde Bi-gram/Di-gram (2-gram), ardışık üç kelimenin gösterimi şeklinde Tri-gram (3-gram) ve bunun gibi gösterim sayısı artırılarak ardışık N kelimenin gösterimi de N-gram ( $N \geq 1$ ) şeklinde listelenebilir.

N-gramlar için bir örnek olarak “Akademik Çalışma” metnini ele alalım [74]. Burada boşluk karakterinin gösterimi “\_” (alt çizgi)

sembolü ile gösterilebilir. Bu metindeki kelimelerin N-gram’ları şöyledir:

Bi-gram/Di-gram (2-gram): “Ak”, “ka”, “ad”, “de”, “em”, “mi”, “ik”, “\_k”, “\_Ç”, “Ça”, “al”, “lı”, “ış”, “şm”, “ma”

Tri-gram (3-gram): “Aka”, “kad”, “ade”, “dem”, “emi”, “mik”, “ik\_”, “k\_Ç”, “\_Ça”, “Çal”, “alı”, “lış”, “ışm”, “şma”

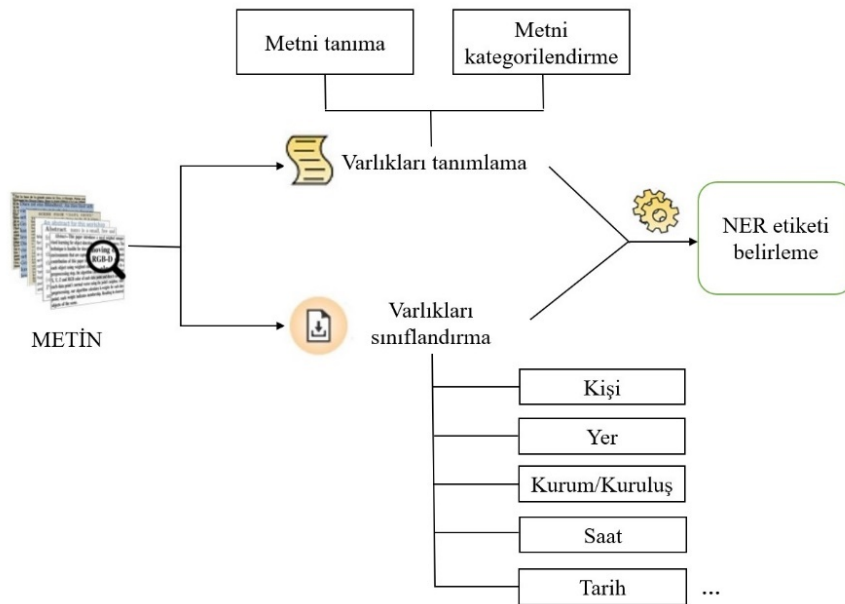
### 2.2.7. Metin sınıflandırma (Text classification)

Bir corpus içerisinde metinlerin nasıl temsil edileceği metin sınıflandırma çalışmalarının en önemli kısımlarından birisidir. En kısa tanımıyla, metin sınıflandırma, karmaşık metinlerin ve dokümanların önceden belirlenmiş sınıflardan hangisine/hangilerine ait olduğunun belirlenmesi işlemidir [75]. Metin sınıflandırma problemi  $T=\{t1,t2, \dots, tn\}$  kümesindeki her bir dokümana (metne) ait önceden tanımlanan  $C=\{c1, c2, \dots, cm\}$  kümesindeki sınıflara ait olup olmadığının belirlenmesidir. Yani her  $(tj, ci) \in T \times C$  çifti için doğru ya da yanlış şeklinde bir mantıksal düşünceyle değerlendirilmesi gerekir. Örneğin, bir  $f$  fonksiyonunun  $j$ . dokümanı  $i$ . sınıfa ait ise doğru, değilse yanlış değerlerini üretsin. O halde,  $f: D \times C \rightarrow \{doğru, yanlış\}$  şeklinde gösterilebilir. Bu durumda, benzer bir mantıkla bir  $g$  fonksiyonu,  $g: D \times C \rightarrow \{doğru, yanlış\}$  makine öğrenmesi yöntemleri ile oluşturulabilir [75]. Makine öğrenmesinde “denetimli öğrenme (supervised classification)” yöntemi ile metin sınıflandırma problemleri çözülebilir. Şekil 4’te [75] çalışma prensibi gösterilmiştir.

### 2.3. Performans Değerlendirme Metrikleri (Performance Evaluation Metrics)

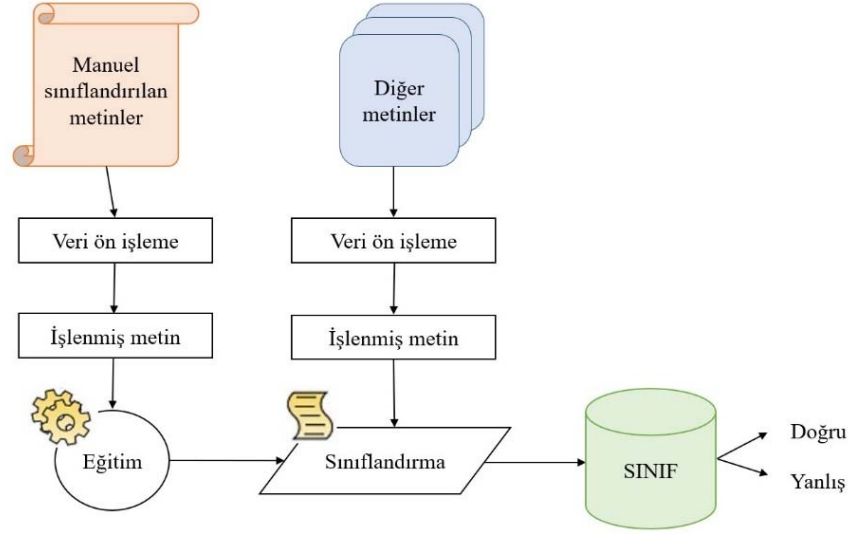
#### 2.3.1. Karışıklık Matrisi (Confusion Matrix)

Karışıklık matrisi, makine öğrenme algoritmalarının ve ürettikleri sınıflandırma modelinin performans değerlendirmesi gösterimi olarak tanımlanır [76]. Sınıflandırma kalitesine dair ölçümler, her sınıf için doğru ve yanlış etiketlenen örnekleri içeren bir karışıklık matrisinden oluşturulur. Bu matris, doğru pozitiflerin (DP), yanlış pozitiflerin (YP), doğru negatiflerin (DN) ve yanlış negatiflerin (YN) sayısını görüntüler. Bu sayede, bu matris ile model performans analizi, yanlış sınıflandırmaların belirlenmesi ve tahmin doğruluğunun



Şekil 3. Adlandırılmış Varlık Tanıma yönteminin çalışma prensibi [73]. (Working principle of the Named Entity Recognition method [73].)





Şekil 4. Makine öğrenmesi ile metin sınıflandırmanın genel çalışma prensibi [75].  
(General working principle of text classification with machine learning [75].)

iyileştirilmesi hesaplanır. Tablo 1, modelde ikili sınıflandırma problemi [76] için toplamda dört değere sahip olan  $2 \times 2$  değerindeki karışıklık matrisini göstermektedir. Çok sınıflı sınıflandırma modelleri için matris şekli  $n \times n$  değerindeki karışıklık matrisi olarak gösterilebilir.

**Tablo 1.** İkili sınıflandırma problemi için örnek karışıklık matris formu [76].  
(Example confusion matrix form for binary classification problem [76].)

Gerçek \ Tahmin	Pozitif	Negatif
Pozitif	DP	YN
Negatif	YP	DN

Tablo 1'de Doğru Pozitif (DP): Ait olduğu gözlem sınıfı ve tahmin edilen sınıfının değerinin pozitif olma durumu, Yanlış Pozitif (YP): Ait olduğu gözlem sınıfı negatif olup tahmin edilen sınıfının değerinin pozitif olma durumu, Doğru Negatif (DN): Ait olduğu gözlem sınıfı ve tahmin edilen sınıfının değerinin negatif olma durumu, Yanlış Negatif (YN): Ait olduğu gözlem sınıfı pozitif olup tahmin edilen sınıfının değerinin negatif olma durumu şeklinde ifade edilmektedir.

### 2.3.2. Doğruluk (Accuracy)

Algoritmanın genel etkinliğini değerlendirmek için, sınıf etiketinin gerçek değerinin olasılığını hesaplayarak algoritmanın performansını yaklaşık olarak tahmin eden metrik türüdür [76]. Doğruluk, ampirik açıdan en çok kullanılan performans ölçüm metriğidir ve algoritmanın genel etkinliğini değerlendirir. Eş. 3'te [76] doğruluk metriği gösterilmiştir:

$$\text{Doğruluk} = (DP + DN) / (DP + YP + DN + YN) \quad (3)$$

### 2.3.3. Kesinlik/Hassasiyet (Precision)

Kesinlik, sınıfa bağlı olarak hesaplanan bir etiketin pozitif veya negatif olma durumunu tahmin eden bir metrik türüdür [76]. Bu metrik, algoritmanın tahmin gücünü değerlendirmektedir ve Eş. 4'te [77] gösterilmiştir:

$$\text{Kesinlik} = DP / (DP + YP) \quad (4)$$

### 2.3.4. Duyarlılık (Recall)

Duyarlılık, bir sınıfa ait pozitif veya negatif etiketin yaklaşık olarak doğru olma olasılığını tahmin eden bir metriktir [76]. Bu metrik, algoritmanın tek bir sınıf üzerindeki etkinliğini değerlendirerek sınıflandırmayı gerçekleştirir. Duyarlılık metriği Eş. 5'te [77] gösterilmiştir:

$$\text{Duyarlılık} = DP / (DP + YN) \quad (5)$$

### 2.3.5. F-puanı (F-score)

F-puanı, yüksek kesinlik (Precision) ve duyarlılık (Recall) değerine sahip olan algoritmaları tercih eden birleşik bir ölçme türüdür [76]. Bu ölçüm değeri, kesinlik ve duyarlılık değerlerinin harmonik ortalaması alınarak hesaplanır. Eş. 6'da [76] F-puanı için hesaplama metriği gösterilmiştir:

$$\text{F-puanı} = 2 * (\text{Kesinlik} * \text{Duyarlılık}) / (\text{Kesinlik} + \text{Duyarlılık}) = 2 * (DP / (DP + YP + YN)) \quad (6)$$

Genel olarak, model başarısının ölçümü için karmaşıklık metrikleri kullanılarak performans ölçüm metriklerinin (doğruluk, kesinlik, hassasiyet ve F-puanı) değerleri hesaplanmaktadır. Performans ölçümleri ve karmaşıklık metriklerinin gösterimi Tablo 2'de [97] gösterilmiştir.

## 3. LİTERATÜR İNCELEMESİ (LITERATURE REVIEW)

Nefret söylemi geçmişten günümüze çeşitli formlarda üretilerek süregelmiştir. Başlangıçta bu faaliyetler için radyo, televizyon, gazeteler veya mesajlar gibi geleneksel medya araçları kullanılırken, günümüzde bu durum yerini sosyal medya ağlarına (Twitter, Instagram, Facebook, YouTube, Reddit vb.) bırakmıştır. İncelenen çalışmalar, nefret söylemi tespitinde, internet başta olmak üzere sosyal medya çerçevesinde toplanmaktadır. Nefret söylemleri değerlendirilirken, olumsuz açıdan genel bir fikir birliği oluşturacak şekilde ırk, din, dil, cinsiyet, etnik köken veya siyasi görüş gibi farklı açılardan kişileri veya gruplara bölmeye odaklanan bir yaklaşım ile hareket edildiği görülür. Tüm bu durumlar, nefret söyleminin tespiti ve uygulamaları şeklinde literatürde yerini almaktadır.

Bu bölüm, nefret söylemlerinin geleneksel ve günümüzdeki sosyal medyada yapılan çalışmalarını kapsamaktadır. Bu çalışmada, literatür incelemesi yapılırken 2019-2023 yılları arasındaki çalışmalar dikkate alınmıştır. Literatürde inceleme yapılırken, çalışmamızda ele alınan araştırma konusunun literatürdeki çalışmalarla bağlantısı, literatürdeki çalışmaların güncelliği ve yayınlanma tarihleri göz önüne alınmıştır. Çalışmamızın kapsamı, Şekil 5'te yer alan nefret söyleminin hiyerarşik çizimi şeklinde ele alınmaktadır. Burada nefret söylemleri, uygulama tabanlı çalışmalar ve sosyal analiz çalışmaları şeklinde iki ana başlığa göre yapılandırılmıştır. Uygulama tabanlı çalışmalarda, nefret söylemi tespiti için, makine öğrenme tabanlı ve derin öğrenme tabanlı yapılan uygulamalar yer almaktadır. Sosyal analiz çalışmalarında ise, yine nefret söylemi tespiti için, konu temelli dikkat uyandıran çalışmalar, betimsel analiz yapan çalışmalar, söylem analizi yapan çalışmalar ve içerik analizi yapan çalışmalar şeklinde ele alınarak, dört başlık altında incelenmiş ve ilgili başlıklar altında değerlendirilmiştir.

### 3.1. Sosyal Medyada Nefret Söylemleri İçin Uygulama-Tabanlı Çalışmalar (Application-Based Studies for Hate Speech in Social Networks)

Günümüzde yapay zekâ, makine öğrenmesi ve derin öğrenme yöntemleri her alanda kendini göstermektedir. Bu yöntemler, büyük veri kavramının içerisinde yer alan sosyal medya konusunda da sıkça kullanılmaktadır. Sosyal medyada nefret söyleminin otomatik olarak tespit edilmesi için de bu yöntemler kullanılarak uygulamalar geliştirilmektedir. Şekil 6'da uygulama geliştirme adımları görselleştirilmiştir.

Şekil 6'da, geliştirilecek uygulama için veri toplama işlemleri, kullanılan sosyal medya araçlarının belirlenmesi ile başlamaktadır. Ardından, bu ağlarda kullanılan tüm nefret söylemi içerikli paylaşımlar, uygun anahtar kelimelerle aratılarak veri kümesi oluşturulmaktadır. Oluşturulan veri kümesi üzerinde, veri ön işleme

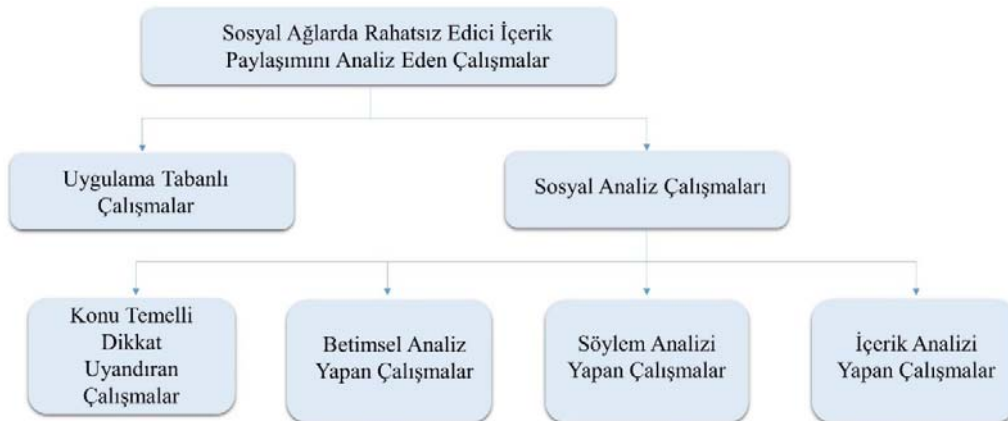
adımları uygulayarak veri işlenebilir hale getirilmektedir. Son olarak, veriler üzerinde uygun yapay zekâ yöntemleri kullanılarak model eğitimi gerçekleştirilmektedir. Sosyal medya platformlarında güvenli ortam sağlamak, nefret söylemi ile mücadele etmek ve bu alanda farkındalık uyandırarak kullanıcıların bilinçlendirilmesi için geliştirilen uygulamalara yönelik makale örnekleri bu başlık altında toplanmıştır.

Sosyal medyada nefret söylemi içerikli paylaşımların artmasına çözüm olarak, Baydoğan ve Alataş'ın [78] hazırladığı çalışma kapsamında, otomatik nefret söylemi tespit sistemi geliştirilmiştir. Açık kaynak olarak Kaggle platformunda [79] paylaşılan 40.623 İngilizce dilinde tweetten oluşan veri kümesi üzerinde deneyler gerçekleştirilmiştir. Bu veri kümesi 28.540 adet 'hate-nefret' ve 12.083 adet 'nohate-nefret değil' şeklinde etiketli dengesiz bir veri kümesidir. Veri kümesi içerisinde noktalama işaretlerinin temizlenmesi, sayısal ifadelerin kaldırılması, büyük-küçük harf dönüşümünün yapılması, dört karakterden az kelimelerin silinmesi ile veri ön işleme adımları tamamlanmıştır. Ardından Snowball-Stemmer algoritması [80] ile kelime köklerine ayrılmıştır. Terim frekans ağırlığı ile köklerin kaç kez kullanıldığı hesaplanmış, en çok geçen kelimeler için Kelime Çantası (BoW) oluşturulmuştur. Elde edilen terim doküman matrisi üzerinde Lib-Destek Vektör Makinesi (Lib-SVM), Multinomial Naive Bayes (MNB), Çok Katmanlı Algılayıcı (MLP) ve iki farklı Karar Ağacı (DT-Part, CVF-DT) algoritmaları ile model eğitimi gerçekleştirilmiştir. Deneysel sonuçlardan elde edilen en yüksek değerler sırasıyla %80 doğruluk değeri ile MLP ve Karar Ağaçları, %78 F-skor değeri ile MLP, %81 kesinlik değeri ile CVF-DT ve %80 duyarlılık değeri ile MLP, DT-Part ve CVF-DT algoritmaları şeklindedir. Geliştirilen nefret söylemi tespit çalışmasının tüm adımlarını özetleyen akış diyagramı Şekil 7'de paylaşılmıştır.

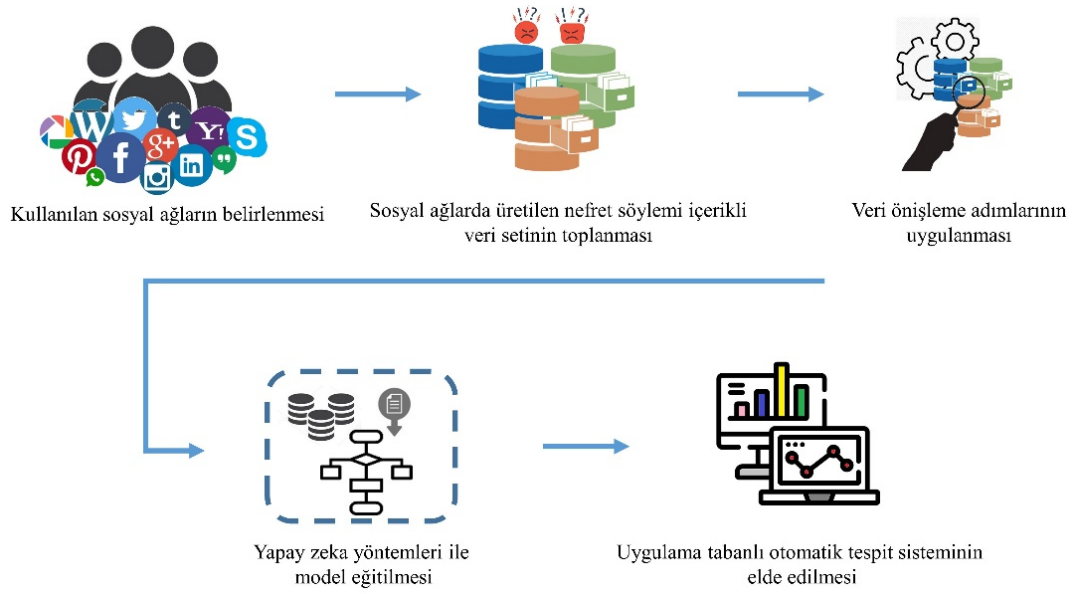
EğİN ve Bulut [81] tarafından yapılmış olan bu çalışmada amaç, mülttecilere yönelik yapılan hakaret, ayrımcılık, şiddet ve diğer olumsuz davranışları tespit etmektir. Bu doğrultuda, nefret söylemi

**Tablo 2.** Performans ölçümleri ve karmaşıklık metrikleri [97]. (Performance measurements and Complexity metrics [97].)

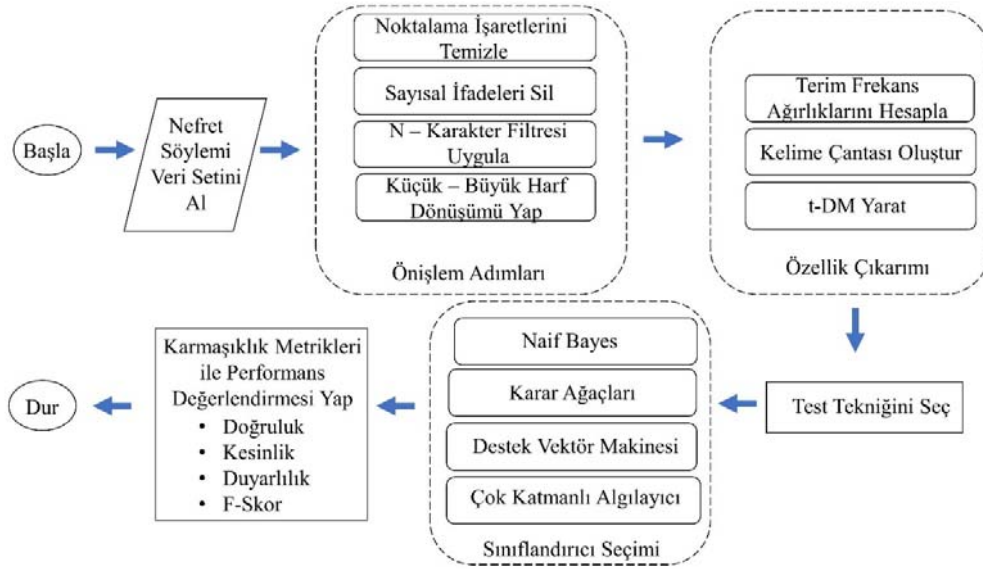
Performans Ölçümü	Denklem	Karmaşıklık Metriği
Doğruluk	$\frac{DP + DN}{DP + DN + YP + YN}$	DP: Doğru Pozitif DN: Doğru Negatif
Kesinlik	$\frac{DP}{DP + YP}$	YP: Yanlış Pozitif YN: Yanlış Negatif
Duyarlılık	$\frac{DP}{DP + YN}$	
F-puanı	$2 * \frac{DP}{DP + YP + YN}$	



**Şekil 5.** Nefret söyleminin hiyerarşik çizimi. (Hierarchical plot of hate speech.)



**Şekil 6.** Otomatik nefret söylemi tespiti için uygulama geliştirme adımları.  
(Application development steps for automatic hate speech detection.)



**Şekil 7.** Çalışmanın akış diyagramı [78]. (Flow diagram of the study [78].)

tespiti için makine öğrenmesi modeli geliştirmişlerdir. Öncelikle Twitter'da mülteciler hakkında atılan tweetleri içeren bir veri kümesi oluşturulmuştur. SNScrape [82] ile çekilen 9.778 adet tweet verisine "Nefret söylemi" ve "Nefret söylemi değil" şeklinde etiketlemeler yapılmıştır. Daha sonra bu veri kümesi, doğal dil işleme teknikleriyle Türkçe dolgu kelimelerden arındırılarak işlenmiştir. Veri önileme adımları tamamlandıktan sonra çeşitli makine öğrenmesi algoritmaları kullanılarak, mültecilere yönelik nefret söylemi tespiti yapılmaya çalışılmıştır. Model geliştirmek için Lojistik Regresyon (LR), SVM, DT, ANN ve RF algoritmaları kullanılarak sınıflandırma yapılmıştır. Çalışmada, beş farklı model arasında yapılan karşılaştırmalı analiz sonucu LR, RF ve ANN modellerinin doğruluk değerleri 0.68 ile eşit olduğu, SVM'nin 0.64 ve DT'nin 0.65 doğruluk değerleri ile daha düşük olduğu görülmüştür. Bu sonuçlar, ile mültecilere yönelik nefret söylemi tespitinde LR, RF ve ANN modellerini ön plana çıkarmaktadır.

Yılmaz vd.'nin [83] gerçekleştirdiği bu çalışma, sosyal medyada gittikçe artan saldırgan dil kullanımına çözüm sağlamak amacıyla oluşturulmuştur. Veri kümesi Twitter platformundaki 14.752 Türkçe tweetten oluşmaktadır. Python dili kullanılarak Twitter'dan çekilen 1 milyon 860 bin tweet ile genişletilmiş derlem oluşturulmuş, bu derlem kullanılarak etiketleme performansında artış olup olmadığı gözlemlenmiştir. Bu derlemede yer alan 150'den fazla saldırgan içerikli kelime temsilleri veri kümesine dahil edilerek etiketleme yapılmıştır. Kelime temsillerinin sınıflandırma performansına katkısının ölçümü için Word2Vec-CBOW yöntemi kullanılmıştır. Mikolov vd. [84, 85] tarafından kullanılan Word2Vec-CBOW yöntemi [86], kelime yerleştirme için oluşturulmuş bir mimaridir ve anlamlı ilişkilere sahip kelime vektörleri üretebilmektedir. Word2Vec ile modelin eğitiminde, veriler kayan bir pencereyle (sliding window) iterasyonu yapılır. Pencerenin ortasındaki kelime hedef kelime ve diğer kelimeler de bağlam

(context) kelimeler olarak adlandırılır. Her iterasyon sonrası bağlam ve hedef kelimeleri Tek-sıcak gösterimle (One-hot representation) girdi ve çıktı şeklinde sınır ağına verilir. Word2Vec modelinin diğer bir çeşidi olan Sürekli Kelime Çantası (CBOW)'nda ise, bağlam kelimeler modele girdi olarak verilir; çıktı ise hedef kelimenin Tek-sıcak vektörüne (One-hot vector) yaklaşması beklenir. Yani diğer bir deyişle, çıktı hedef kelimenin olasılığını 1'e ve diğer tüm kelimelerin olasılıklarını 0'a yaklaştırdığı her kelime için bir olasılıklar vektörü oluşturmaktadır [87].

CBOW'un amaç fonksiyonu [85] Eş.7'deki gibidir.

$$J_0 = \frac{1}{T} \sum_{t=1}^T \log p(w_t | w_{t-n}, \dots, w_{t-1}, w_{t+1}, \dots, w_{t+n}) \quad (7)$$

Burada  $T$ , eğitim verisindeki kelime boyutudur;  $t$ , hedef kelimenin her iki taraftaki bağlam boyutu, yani pencere boyutunun yarısıdır;  $w$ , kelime temsilidir;  $p$ , softmax fonksiyonunu ifade eder;  $n$  ise bir kelime yerleştirme oluşturmak için  $n$  adet gelecekteki ve  $n$  adet geçmişteki kelimeleri temsil eder. Kelime temsilinin sınıflandırılması sonrası, Twitter'dan çekilen verilere "saldırgan değil, hedefsiz, birey, grup, diğer" şeklinde beş farklı etiket ile çoklu sınıflandırma yapılmıştır. Sınıflandırma yapılırken LSTM ve GRU (Gated Recurrent Units) algoritmaları kullanılmıştır. İkili sınıflandırma için, LSTM F1-makro değeri %85.19 ve GRU F1-makro değeri ise %87.99; genişletilmiş derlem kullanımı için, LSTM F1-makro değeri %94.21 ve GRU F1-makro değeri %94.49'dur. Çoklu sınıflandırma için, GRU modelinin F1-makro değeri %54.10 ve genişletilmiş derlem kullanımı için ise GRU F1-makro değeri %71.97'dir. Deneysel sonuçlara göre, genişletilmiş derlemin sınıflandırma performanslarına iyileştirici etkisi olduğu sonucuna ulaşılmaktadır. Çalışma bu yönüyle, Türkçe dilinde çoklu sınıflandırma yapan ilk çalışma olarak adlandırılmaktadır.

Günümüz dünyasında gereken önemin verilmediği mobbing ve siber mobbing konuları, Atatanır'ın [88] bu çalışmasıyla ele alınarak ayrıntılı şekilde tanıtılmıştır. Çalışmada, iş yaşamında karşılaşılan mobbing durumunun aynı zamanda dijitalleşme tarafı olan siber mobbing kavramına da değinecek şekilde bir yaklaşım benimsenmiştir. Bu konuya yaklaşım birkaç açıdan ele alınarak detayları incelenmiştir. Çalışmada veri kümesi, mobbing ve siber mobbing olmak üzere ikiye ayrılmıştır. Siber mobbing veri kümesinde, siber mobbinge maruz kalan kişi hakkında e-posta, mesaj, sosyal medya gibi dijital alanda elde edilen veriler yer almaktadır. Mobbing veri kümesinde de, mobbinge maruz kalan kişilerin fotoğrafları, habersiz video kayıtları, ses dosyaları ve ifşa oluşturabilecek paylaşımları (toplumsal statü, fiziksel üstünlük, köken gibi hususlar) yer almaktadır. Çalışma, siber mobbingin etkilerini incelemek için nitel araştırma yöntemlerinden biri olan vaka çalışması üzerine kuruludur. Verileri sınıflandırmak için JEL Sınıflandırması (J53, J71, J83) tekniğinden yararlanılmıştır. Araştırmacılar, siber mobbing kurbanı olan kişilerin görüşlerini yorumlamaya çalışmışlardır. Görüşmenin kayıt altına alınması ve deşifre edilmesinin ardından verilere içerik analizi yapılmıştır. Yapılan analizlerin ardından mobbing ve siber mobbingin yol açtığı psikolojik sorunlar ve ekonomik zararlar gibi konular aydınlatılarak çalışma tamamlanmıştır.

Siber zorbalık probleminin tespiti için, Yazgılı ve Baykara [89] çalışmasında, sosyal medyada yer alan siber zorbalık çeşitlerini ele almıştır. Ardından, siber zorbalık tespiti için gerekli olan makine öğrenme metodlarındaki çalışmalar genel açıdan incelenmiştir. Literatüre uygulamalı katkılar sunulması açısından, makine öğrenmesi metodlarında kullanılan uygulamalar da incelenerek, dijital medya alanında kullanıcıların bilinçlendirilmesi hedeflenmiştir. Uygulamalar için Beacon-Less Routing Algorithm (BLR) [90], RF, MLP, Karar Ağacı (J48), SVM, MNB, NB, KNN (K-Nearest Neighbors), ANN, AvOID algoritması [91], GRU ve LSTM makine

öğrenmesi metodları kullanılmıştır. Çalışmada literatür kısmı 2009-2018 yıllarını kapsamaktadır. Twitter, Facebook, Wikipedia ve Fromspring.me uygulamalarından alınan veri kümeleriyle İngilizce, Arapça, Litvanca ve Bengalce dillerinde yapılan çalışmalar incelenmiştir. Çalışmada incelenen veri kümesi, kullanılan dil aile yapısı ve uygulamalarda kullanılan makine öğrenmesi algoritmalarına genel açıdan bakıldığında, sınıflandırmada başarı açısından en iyi performansı SVM'nin gösterdiği sonucuna varılır.

Adlandırılmış Varlık Tanıma (NER) problemi tespitine yönelik çalışan Nesiboğlu ve Gencer [92] çalışmasında, Twitter'daki Türkçe tweetleri ele alarak hakaret, küfür ve uygunsuz sözcüklerin tespitine yönelik bir çalışma yapmışlardır. Çalışmada, sözcük ve sözcük grupları IOB2 etiketleme yöntemi (IOB Tagging) [93] ile etiketlenmiştir. Tablo 3'te çalışmada IOB etiketleme yöntemi için verilen örnek paylaşılmıştır [92].

**Tablo 3.** IOB etiketleme örneği [92]. (IOB labeling example [92].)

	IO Kodlama	IOB Kodlama
Mehmet	PER	B-PER
Edvard	PER	B-PER
Munch	PER	I-PER
'un	O	O
resmini	O	O
Ahmet	PER	B-PER
'e	O	O
gösterdi	O	O

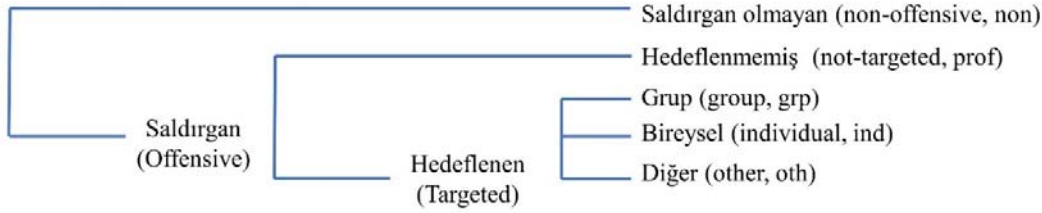
Tablo 3'e göre, etiketlenen sözcükler vektörleştirilmiştir. Vektörlerin eğitimi için derin öğrenme tabanlı Bi-LSTM (Bidirectional Long Short-Term Memory) ve BERT (Bidirectional Encoder Representations from Transformers) mimarileri kullanılarak hakaret, küfür ve uygunsuz içerikli sözcüklerin tahmini yapılmıştır. Uygulamalar için GitHub üzerinden duygu durumu analizi ile ilgili 15.110 cümlelik Türkçe tweetlerden oluşan bir veri kümesi seçilmiştir. Uygulamalarda doğru pozitif (TP), doğru negatif (TN), yanlış pozitif (FP), yanlış negatif (FN), F1-score, Accuracy, Recall ve Precision değerlendirme metrikleri ele alınmıştır. Deneysel sonuçlarına göre Bi-LSTM modeli için yaklaşık %99 ve BERT modeli için de yaklaşık %95 doğruluk oranları elde edilmiştir. Buna göre, NER problemi tespitinde çalışma hızı açısından, Bi-LSTM'nin BERT modelinden daha hızlı olduğu sonucuna varılır. Bu çalışma, Türkçe dilinde derin öğrenme yöntemleri ile uygunsuz sözcük tespitindeki ilk çalışmadır.

Çöltekin'in [94] çalışmasında, Türkçe dilinde saldırgan dil söylemi (nefret söylemi) üzerine bir derlem kullanılmıştır. Derlem, Twitter akışı API'sindeki mikro blog gönderilerinden oluşan rastgele örneklenmiş 36.232 tweet içeriğine sahiptir. Bu içeriklerin %19'u saldırgan dile sahip olup, "suçun hedefi" şeklinde alt kategorilere ayrılmıştır. Veri kümesinde saldırgan dil söylemi için dokümanlar "saldırgan" ve "saldırgan-olmayan" şeklinde sınıflandırılmıştır. Saldırgan ifadeler de kendi içerisinde "küfür, grup, bireysel ve diğer (non, prof, grp, ind and oth)" şeklinde hiyerarşik olarak alt kategorilerle etiketlenerek [95] sınıflandırılmıştır. Çalışma içerisinde yer alan etiket kümesinin hiyerarşik gösterimi Şekil 8'de yer almaktadır. Veri kümesinde [96], 10-katlı çapraz doğrulama yöntemi kullanılarak veri kümesindeki her görev için karakter ve kelime N-gramları bazında SVM sınıflandırıcı uygulamalar için kullanılmıştır. Çalışmanın deney sonuçları incelendiğinde, saldırgan tweetleri belirlemede %77.3 ile F1-puanı, belirli bir saldırgan dokümanın hedef alınıp alınmadığını belirlemede %77.9 ile F1-puanı ve hedeflenen saldırgan dokümanları üç farklı alt kategoriye ayırmada %53 ile F1-puanı ve hedeflenen saldırgan dokümanları beş farklı alt kategoriye ayırmada %45.7 ile F1-puanı elde edilmiştir.

Baydoğan ve Alataş'ın [97] çalışmasında, meta-sezgisel metodolojiye dayalı saldırgan dil söylemi ve otomatik nefret söylemi tespit sistemleri önerilmiştir. Çalışmada, Güve Alevi Optimizasyonu (Moth-Flame Optimization Algorithm, MFO) ve Karınca Aslanı Optimizasyonu (Ant Lion Optimization Algorithm, ALO) algoritmaları kullanılmıştır. Şekil 9'da MFO, Şekil 10'da ALO kaba kodu (pseudo code) yer almaktadır. Çalışma kapsamında özellik çıkarma (feature extraction), Kelime Çantası (BoW), Terim Frekansı (TF) ve Doküman vektörü (Word2Vec) yaklaşımları uygulanmıştır. Çevrimiçi web forumlarından ve Twitter'dan toplanan veriler etiketlenerek, üç farklı veri kümesi (Veri kümesi 1 [98], Veri kümesi 2 [99], Veri kümesi 3 [100]) oluşturulmuştur. Bu veri kümeleri, Veri kümesi 1 İngilizce dilinde 25K-boyutlu, Veri kümesi 2 İngilizce dilinde 4K-boyutlu ve Veri kümesi 3 İngilizce ve İspanyolca dilinde

13K-boyutlu verilerden oluşmaktadır. Anlatılan veri kümeleri üzerinde uygulanan algoritmalar, KNN ( $k=3$  ve  $k=5$ ), DT, Sıralı Minimum Optimizasyon (SMO), Çok-Sınıflı Sınıflandırma (MCC), J48, NB, RF, RIDOR algoritması (Ripple Down Rule learner Classifier algorithm), Sosyal Örümcek Optimizasyonu (SSO) ve Tunicate Swarm Algoritması (TSA) şeklindedir. Üç veri kümesi üzerinde elde edilen sonuçlar değerlendirildiğinde, ALO ve MFO algoritmalarının doğruluk değerleri sırasıyla %92.1 ve %90.7 şeklinde bulunmuştur.

Çoban vd. [101] bu çalışması, Türkçe dilinde Facebook içeriklerinin analizi ile siber zorbalık tespitine odaklanılmıştır. Çalışmada, Facebook API'si ile toplanan veriler üzerinde farklı derin öğrenme ve makine öğrenmesi yöntemleri kullanılarak siber saldırı yapan



Şekil 8. Etiket hiyerarşisi [94]. (The label hierarchy [94].)

```

Karıncaların ve karıncayiyenlerin ilk popülasyonunu rastgele başlatın
Karıncaların ve karıncayiyenlerin uygunluğunu hesaplayın
En iyi antiyonları bulun ve bunu elite (belirlenmiş optimum) olarak kabul edin
while son kriter karşılanmazken
  for her karınca için
    Rulet Çarkını kullanarak bir antlion seçin
    Operatörleri güncelleyin
    Rastgele bir yürüyüş oluşturun ve normalleştirin
    Karıncaların konumunu güncelleyin
  end
  Tüm karıncaların uygunluğunu hesaplayın
  Bir karıncayı, daha uygun hale gelirse karşılık gelen karınca ile değiştirin
  Bir antlion belirlenmiş optimumdan (elite) daha zinde hale gelirse elite güncelleyin
end
return elite
  
```

Şekil 9. ALO Sözde kod [97] (The pseudo-code of ALO [97])

```

Alev sayısını güncelleyin (FlameNumber)
Güve popülasyonunu başlatın
Hedef değerleri hesaplayın
for tüm güveler için
  for tüm parametreler için
    b ve t'yi güncelleyin
    D'yi ilgili güveye göre hesaplayın
    M matrisini ilgili güve ile ilgili olarak güncelleyin
  end
  Hedef değerleri hesaplayın
  Alevleri güncelle
end
En iyi çözümü yazdırın
  
```

Şekil 10. MFO Sözde kod [97] (The pseudo-code of MFO [97])

*M*: Küme güvelerin matris temsili, *D*: Güveler için mesafe,  
*b*: Logaritmik spiralin şeklini tanımlamak için sabit, *t*: [-1, 1] arasında rastgele bir sayı.

kullanıcıların profil tespiti yapılmıştır. Veri kümesi üzerinde 5-katlı çapraz doğrulama tekniği kullanılarak F1-makro puanı hesaplanmıştır. Deneysel sonuçlarından, kullanılan algoritmalar arasında BERT modeli 0,928'lik ortalama F1-makro değeriyle en iyi tahmin modeli olduğu söylenebilir. Literatürde incelenen çalışmalara göre, farklı sosyal medya araçlarında yer alan nefret söylemi tespiti için uygulama tabanlı çalışmaların karşılaştırılması Tablo 4 ve Tablo 5'te verilmiştir.

### 3.2. Sosyal Medyada Nefret Söylemleri İçin Sosyal Analiz Çalışmaları (Social Analysis Studies for Hate Speech in Social Networks)

Teknolojinin hayatımıza girmesi ile sosyal medya platformları hızla gelişerek, bu platformların kullanımı gün geçtikçe artmaktadır. Bunun bir sonucu olarak sosyal medya kullanıcılarının sıkça karşılaştığı sorun, nefret söylemlerine maruz kalmalarıdır. Sosyal medyada nefret söylemi, bir kişi veya gruba yönelik saldırgan, aşağılayıcı, önyargılı veya ayrımcı ifadeleri içeren sözlü, yazılı ya da görsel paylaşımlardır. Bu paylaşımlar, sosyal medya kullanıcılarına psikolojik olarak zarar verdiği gibi, aşağılama ve dışlama amacı da taşıdığı için sosyal medya kullanıcılarını olumsuz yönde etkilemektedir. Şekil 11'de bu durum görsel olarak ifade edilmiştir.

Bu çalışma kapsamında, sosyal medyada nefret söylemine maruz kalmanın sosyal analizini yapan çalışmalar da incelenmiştir. İncelenen çalışmalar konu temelli dikkat uyandıran çalışmalar, betimsel analiz yapan çalışmalar, söylem analizi yapan çalışmalar ve içerik analizi yapan çalışmalar şeklinde dört başlığa ayrılarak, bu başlıklara uygun makaleler değerlendirilmiştir.

#### 3.2.1. Konu temelli dikkat uyandıran çalışmalar (Subject-based attention-grabbing studies)

Sanal platformlarda çocukların çevrimiçi cinsel istismara maruz kalma riskini ve bunun olası sonuçlarını ele alan Bayrak'ın [102] bu çalışmasında amaç, çevrimiçi ortamda çocuk istismarlarının yaygınlaşmasına dair oluşan risk etmenlerini içeren bilgilendirmeyi yapmak ve ruh sağlığı çalışanları için farkındalık oluşturmaktır. Çalışmada, internet ortamlarının ebeveyn kontrolü dışında kullanımı ile çocuklarda ve gençlerde meydana gelebilecek olumsuz farklı durumlar ele alınmıştır. Bu durumlar için, veri kümesi olarak Türkiye İstatistik Kurumu (TÜİK) 2021 verileri kullanılmıştır. Veriler, Türkiye'de 6-15 yaş arası çocukların telefon ve internet kullanımları (düzenli cep telefonu/akıllı telefon kullanımı, düzenli internet kullanımı, sosyal medya kullanımı, oyun oynama-indirme, ödev yapma, konuşma (görüntülü-görüntüsüz), mesajlaşma) düzenli

**Tablo 4.** Nefret söylemi tespitine yönelik uygulama tabanlı çalışmaların karşılaştırılması (1).  
(Comparison of application-based studies on hate speech detection (1).)

Ref. No	Yıl	Araştırma konusu	Uygulama platformu	Veri Kümesi	Veri toplama aşaması	Performans Metriği
Çöltekin [94]	2020	Saldırgan dil söylemi	Twitter	36.232	Twitter API	F1-puanı
Baydoğan ve Alataş [78]	2021	Otomatik nefret söylemi tespit sistemi	Twitter	40.623	Kaggle	Doğruluk değeri
Baydoğan ve Alataş [97]	2021	Saldırgan dil ve otomatik nefret söylemi tespit sistemi	Twitter	Veri kümesi 1: 25.000 Veri kümesi 2: 4.000 Veri kümesi 3: 13.000	Twitter ve çevrimiçi web platformları	Doğruluk değeri
Şahiner vd. [83]	2022	Saldırgan dil derlemi	Twitter	14.752	Python dili ile geliştirilen kod	F1-makro değeri
EğİN ve Bulut [81]	2023	Mültecilere yönelik nefret söylemi tespiti	Twitter	9.778	SNScrape aracı	Doğruluk değeri
Nesiboğlu ve Gencer [92]	2023	Adlandırılmış Varlık Tanıma (NER) ile uygunsuz sözlerin sansürü	Twitter	15.110	GitHub	Doğruluk değeri
Çoban vd. [101]	2023	Siber zorbalık tespiti	Facebook		Facebook API	F1-makro değeri

**Tablo 5.** Nefret söylemi tespitine yönelik uygulama tabanlı çalışmaların sonuçları (2).  
(Results of application-based studies on hate speech detection (2).)

Ref. No	Sonuç
Çöltekin [94]	10-katlı çapraz doğrulama ve SVM için saldırgan tweetleri belirleme:%77.3 belirli bir dokümandaki saldırgan tweetleri belirleme:%77.9 saldırgan dokümanları üç alt kategoriye ayırma:%53.0 saldırgan dokümanları beş alt kategoriye ayırma:%45.7
Baydoğan ve Alataş [78]	CVFDT, DT-Part, MLP: %80
Baydoğan ve Alataş [97]	ALO: %92.1 MFO: %90.7
Şahiner vd. [83]	İkili sınıflandırma için, LSTM: %85.19 ve GRU:87.99; genişletilmiş derlem için, LSTM: %94.21 ve GRU: %94.49 Çoklu sınıflandırma için, GRU: %54.10; g genişletilmiş derlem için, GRU: %71.97
EğİN ve Bulut [81]	RF, LR, ANN: %68 DT: %65 SVM: %64
Nesiboğlu ve Gencer [92]	Bi-LSTM: %99 BERT: %95
Çoban vd. [101]	BERT: %92



Şekil 11. Sosyal medyada nefret söylemine maruz kalmanın sosyal analizi. ( Social analysis of exposure to hate speech on social networks.)

şekilde incelenerek toplanmıştır. Denetimsiz internet kullanımının küçük yaşlardan itibaren çocukların ve gençlerin gelişimine zarar verdiği ve birçok risk faktörünü de beraberinde taşıdığı altı çizilmiştir. Özellikle çocuklarda ve gençlerde, interneti gereğinden fazla kullanmanın internet bağımlılığına, nefret söylemine veya dijital zorbalığa neden olarak gelişimlerine zarar verdiği ve bunun etkilerinin de pek çok duygu-durumsal ve düşünsel değişikliklere sebebiyet verdiği detaylıca ifade edilmiştir.

Sosyal medya etkisinin gerçek hayat üzerindeki etkileri konusunu ele alan Bülbül'ün [103] bu çalışmasında, teknoloji ile popüler kültür çerçevesinde insanların yapay dünyadaki metalaşma dönüşüm süreci ayrıntılandırılmıştır. Bu ayrıntılandırma, yapay dünyada iki çerçevede ele alınmıştır. Birinci kısımda, sosyal medyayı daha iyi anlayabilmek için sosyal medya parametreleri dikkate alınarak kültür, moda, psikoloji, etik ve hukuk kavramları ele alınmıştır. İkinci kısımda ise, sosyal medyadaki verilerin topluma imgesel, psikolojik, sosyolojik ve davranış bilimi üzerine etkileri ele alınmıştır. Zygmunt Bauman'ın düşüncelerinden yola çıkarak, sosyal medyada bireylerin ve kolektif sistemlerin dönüşümünün olumsuz etkileri bu çalışma ile ifade edilmiştir. Sosyal medyanın birey ve toplum üzerindeki etkilerini daha iyi görebilmek için nitel (teorik) çalışma yöntemi seçilmiştir. Sosyal medyada kullanıcı verilerinin toplanması için veri madenciliği yönteminden yararlanılmıştır.

Ayçiçek [104] tarafından yapılan çalışmada, çeşitli disiplinlerde önemli bir konu olan homofobi üzerinde sosyal medyanın ve sosyal medya etkileycilerinin etmenleri ele alınmaktadır. Çalışmada, sosyal medya ve sosyal medya fenomenlerinin homofobi üzerindeki etkisinin araştırılmasına yönelik, 18-35 yaş arası sosyal medya kullanıcıları arasında sosyal medya kullanımı, sosyal medya etkileycilerini takip etme ve homofobi arasındaki ilişki incelenmiştir. İnceleme aşamasında Demografik Bilgi Formu (28 soru), Eşcinsellik Tutum Ölçeği (ATHS - 56 soru) ve Hudson-Ricketts Homofobi Ölçeği (HRHS - 24 soru) gibi veri toplama araçları kullanılmıştır. Çalışmadaki veri kümesi, Ankara'da yaşayan rastgele seçilmiş ve gönüllü 200 katılımcıdan oluşmaktadır. Çalışmada kullanılan araştırma yöntemi ilişkisel tarama modelidir. Uygulama platformu olarak Twitter, Instagram, Youtube, TikTok gibi yaygın sosyal medya araçları incelenmiştir ve Blogger, Vlogger, Influencer olarak adlandırılan "fenomenlerin" sosyal medya kullanıcılarına etkisi araştırılmıştır. Bulgular, sosyal medya kullanımının katılımcılar arasında homofobi üzerinde olumlu bir etkiye sahip olduğunu göstermektedir. Sosyal medya mecrasında yapılan hakaretlerin suç unsuru olup olmadığını araştıran Acay'ın [105] çalışmasında, bu durumun ne gibi problemlere sebep olabileceği ve sosyal medyadaki

hakaretlerin Türk Ceza Kanunu'ndaki cezaî yaptırımları araştırma konusu olmuştur. Bu konuda, sosyal medyanın yanlış amaçlar uğruna kullanıldığında ciddi zararlar verebileceğinden, sosyal medya platformlarının kişilik haklarını ihlale neden olabileceğinden ve kötü amaçlı içeriklerin siber mağdurlara oluşturabileceği maddi ve manevî zararlarından bahsedilmiştir. Çalışmada, sosyal medya platformları üzerinde her türlü denetimin gerçekleştirilmesi gerektiği vurgulanmıştır. Bu çalışmada yöntem olarak, sosyal medya vasıtasıyla hakaret suçları kavramsal tartışma şeklinde ele alınmıştır. Veri kümesi olarak, hakaret suçu ve tespitinde makalelerden, kitaplardan, raporlardan, çevrimiçi ve basılı materyallerden, internet yayınları ve ilgili dergilerden elde edilen veriler kullanılmıştır. Verilerin elde edilmesi için Facebook, Twitter, YouTube, Podcasting, LinkedIn, Instagram, TikTok, Swarm, Bloglar ve Wikipedia gibi Google uygulama platformlarından yararlanılmıştır. Çalışmanın sonucunda, iletişim konusunda kolaylık sağlayan sosyal medya platformlarının, kötü amaçlı kişiler tarafından da kullanılabilirliği düşünüldükçe, kullanıcılar tarafından doğru amaçlarla kullanılması gerektiği vurgulanmaktadır.

Şiddet ve ayrımcılığın toplumsal bir sorun olarak nasıl görüldüğünü ele alan Yıldırım'ın [106] çalışması, bu sorunların nasıl toplumsallaştığını geniş açıdan ele almaktadır. Çalışmanın amacı, şiddet ve ayrımcılıkta toplumsallaşma sürecini inceleyerek, toplumsal barış ve sosyal adalet anlayışı konusundaki sorunlara çözüm aramaktadır. Çalışmada yöntem olarak, ayrımcılık ve şiddetin toplumsal yapılar üzerindeki etkileri ve bunların çözümüne dair öneriler ele alınarak Sosyal Politika, Barış Çalışmaları ve Sosyal Psikoloji gibi farklı disiplinlerin alan bilgileri ve yöntemleri kullanılmıştır. Ek olarak, tespit edilen sorunların çözümüne yönelik sosyal adalet ve toplumsal barışın önemi de vurgulanmaktadır. Sosyal adaletin ayrımcılığı önleme ve toplumsal farklılıkların kabul edilmesinde önemli bir rol oynadığına dikkat çeken bu çalışmada, toplumsal barışta şiddetin azaltılması ve çatışmaların çözümü için kritik bir önem taşıdığı belirtilmektedir. Sonuç olarak çalışmanın geneli incelendiğinde, "ayrımcılık" kavramı tartışılmış ve bu ifadeyi düzgün bir şekilde karşılayacak ifade olan "ayrıştırıcı" kavramı literatüre kazandırılmıştır.

Sosyal medya kullanıcıları arasında ortaya çıkan troll kültürü ve saldırgan mizah olgusunu inceleyen Özdemir'in [107] bu çalışmasında, özellikle Twitter'da popüler olan trollerin anlam dünyası ve deneyimleri, fenomenolojik desenler kullanılarak mülakatlar yoluyla araştırılmıştır. Araştırma, trollerin neden trollük ve saldırgan mizah yaptıklarını anlamaya yöneliktir. Araştırma sürecinde toplanan bulgular, trollerin yaptıkları eylemlerin anlamsal ortak

kodları, sistematik şekilde bir araya getirilmiştir. Özdemir'in bu çalışması, Türkçe dilinde trollerin saldırgan mizah yapma kültürü ve mizah oluşturma olgusuna dair literatürdeki boşluğu doldurmaktadır. Bunun için çalışmada, Twitter kullanıcıları 12 troll veri kümesi olarak seçilmiştir. Twitter'daki trollerin toplumsal değerlere dair dikkat çeken troll ifadeleri, nitel araştırma yöntemleri (çevrimiçi ve yüz yüze) ve örneklem seçim teknikleri (kartopu örneklem, amaçsal örneklem) kullanılarak toplanmıştır. Araştırmanın sonucunda, trollerin saldırgan mizahını bir saldırı aracı şeklinde görmedikleri için nefret söylemi ifadelerini yaydığına inanmadıkları saptanmıştır. Dolayısıyla, bu tür nefret söylemlerinin dikkat çekmek, halkı aydınlatmak ve insanların samimiyetsizliğini ortaya çıkarmak için saldırgan mizah ürettikleri sonucuna ulaşılmıştır.

Özgür ve Özkul'un [108] çalışmasında, insanların maruz kaldığı dijital şiddetin kişi üzerindeki etkileri anlatılmıştır. Araştırmada verilerin toplanması için uzman görüşlerine göre soru formu oluşturulmuştur. Oluşturulan anket formu, 18-95 yaş arası beş farklı kuşakta yer alan ayrı ayrı 100 kişiye çevrimiçi ortamda cevaplamaları için sunulmuştur. Anket cevaplarına göre internet, cep telefonları ve bilgisayarlar gibi dijital araçların yaygınlaşmasıyla birlikte, insanların çevrimiçi ortamlarda şiddet içeren davranışlar sergileyebileceği ve yaşın ilerlemesiyle dijital araçlarla ilişkinin azaldığı gözlemlenmiştir. Çalışmada, araştırma yöntemi olarak nicel yöntem araştırması, standart sapma ve ki-kare testi seçilmiştir. Çalışmanın sonucunda,

dijital araçların daha sık tercih edilmesi dijital şiddetin artmasına neden olduğu için, teknoloji kullanıcılarına yönelik bilgilendirme faaliyetlerinin düzenlenmesi, amaç dışı kullanımlar için hukuki işlemlerin uygulanması ve güvenli dijital alan yaratılması için sürdürülebilir politikalar oluşturulması gerektiği sonucuna varılabilir.

Kuş'un [109] bu çalışmasında, Metaverse ortamlarda [110, 111] çevrimiçi nefret söyleminin oluşturabileceği risklerin ve bu risklere karşı üretilebilecek çözümlerin ortaya konması amaçlanmıştır. Oluşturulacak bu çözüm, Metaverse ortamlarda olabilecek her türlü risklere karşı alınacak önlemlerin geliştirilmesi, dijital ortamdaki zararlı içeriklere dair bir denetimli kültür sisteminin oluşturulması ve oluşturulacak bu kültürün eğitilerek Metaverse ortamlarının her yerinde tüm nefret söylemi içeriklerini ortadan kaldırmaya yönelik önerileri kapsamaktadır. Bu öneriler, nefret söylemi içeriklerinde kavramsal tartışma yoluyla ele alınmaktadır. Çalışma, Metaverse ortamlarında çevrimiçi nefret söylemine dair varolan olumsuzluğun içeriğini ve sınırlarını tanımlayarak, yeni teknolojiler dahilinde bu konuya ışık tutmaktadır. Dolayısıyla tüm bu açılarından bakıldığında, bu çalışma, yeni teknolojilerle üretilen nefret söylemlerinin, dijital dünyadaki riskleri önleme konusunda çözüm önerileri sunmaktadır.

Literatürde incelenen çalışmalara göre, farklı sosyal medya araçlarında yer alan nefret söylemleri için konu temelli dikkat uyandıran çalışmalar Tablo 6 ve Tablo 7'de gösterilmiştir.

**Tablo 6.** Nefret söylemi tespitine yönelik konu temelli dikkat uyandıran çalışmalar (1).  
(Subject-based studies that draw attention to the detection of hate speech (1).)

Ref. No	Yıl	Araştırma konusu	Yöntem
Özdemir [107]	2019	Troller üzerindeki sosyolojik araştırma	Nitel araştırma yöntemi, örneklem seçim teknikleri (kartopu örneklem, amaçsal örneklem)
Bülbül [103]	2021	Sosyal medyada nefret söyleminde imaj algısı	Nitel (teorik) çalışma
Acay [105]	2021	Sosyal medyada hakaret suçlarına dair suç tespiti	Kavramsal tartışma
Yıldırım [106]	2021	Şiddet ve ayrımcılığın toplum üzerindeki olumsuz etkisi ve çözümü	Kavramsal tartışma
Bayrak [102]	2022	Çevrimiçi cinsel istismar	-
Ayçiçek [104]	2022	Sosyal medya kullanımının homofobiye etkisi	İlişkisel tarama modeli
Özgür [108]	2022	Kuşaklar arasında dijital şiddet	Nicel yöntem araştırması, standart sapma ve ki kare testi
Kuş [109]	2022	Metaverse'te dijital nefret söylemi	Kavramsal tartışma

**Tablo 7.** Nefret söylemi tespitine yönelik konu temelli dikkat uyandıran çalışmalar (2).  
(Subject-based studies that draw attention to the detection of hate speech (2).)

Ref. No	Uygulama platformu	Veri kümesi	Veri toplama aşaması
Özdemir [107]	Twitter	12 Trolle gerçekleştirilen mülakat	Çevrimiçi ve yüz yüze ortamda mülakat
Bülbül [103]	-	-	Veri toplama ağı olarak Veri Madenciliği
Acay [105]	Facebook, Twitter, YouTube, Podcasting, LinkedIn, Instagram, TikTok, Swarm, Bloglar, Wikipedia, Google uygulamaları	Makaleler, kitaplar, raporlar, çevrimiçi ve basılı materyaller, internet yayınları ve ilgili dergiler	-
Yıldırım [106]	-	-	Sosyal Politika, Barış Çalışmaları ve Sosyal Psikolojinin alan bilgileri
Bayrak [102]	Sanal platformların tümü	TÜİK 2021 verileri	Akıllı cep telefonu kullanımı, İnternet kullanımı, Sosyal medya kullanımı
Ayçiçek [104]	TikTok, Instagram, Twitter, Youtube	200 katılımcı (yaşları 18-35 yaş arası olan gönüllülere 108 soru yöneltilmiştir)	Demografik Bilgi Formu, Eşcinsellik Tutum Ölçeği (ATHS), Hudson ve Ricketts Homofobi Ölçeği (HRHS)
Özgür [108]	-	500 kişinin anket cevapları (18-95 yaş arası)	"Google formlar" anket yönetimi
Kuş [109]	Metaverse	-	-



### 3.2.2. Betimsel analiz yapan çalışmalar (Studies using descriptive analysis)

İslamofobik açıdan nefret söylemi tespitini için Çeçen [112] bu çalışmasını oluşturarak, çalışmayı iki amaç etrafında toplamıştır. İlk amacı, İslamofobik nefret söylemini kategorize etmeyi ve literatürde eksik kalan bir kategorizasyon şemasını tamamlama düşüncesidir. İkinci amacı da, internetin denetimsiz yapısının, ayrımcı dildeki artış köriklediğine ve İslamofobik söyleme açık ayrımcı dil kullanımının artışına dikkat çekmektir. Çalışmada, ayrımcı alt söylem ve konuşma topluluğu kavramları sıkça kullanılarak, bunların İslamofobik nefret söylemi boyutlarının bu iki kavram ile olan ilişkisi ele alınmıştır. Bunun yanında, İslamofobik nefret söyleminin boyutları demografik ve hukuki boyut açısından değerlendirilmiştir. Web of Science (WoS)'da son 10 yıllık zaman diliminde "Islamophobia" ve "Islamophobic" terimleri kullanılarak, 200 çalışmanın araştırma analizi yapılmış ve İslamofobik nefret söyleminin belirli kategorilerde yer aldığı görülmüştür. Araştırma yapılırken betimsel analiz yöntemi kullanılmıştır. Çalışmada "Islamophobia" terimi alanına teorik açıdan katkı olması için, ayrımcı alt söylem kavramı ve söylem topluluğu kavramları ortaya atılmıştır. Bu kavramların İslamofobik ifadenin nefret söylemi boyutları hakkında da detaylı bilgiler vermektedir.

Nefret söylemi açısından Abuşoğlu ve Aydoğan'ın [113] çalışması, sosyal medyada Z-kuşağına yönelik nefret söylemlerini ele almaktadır. Özellikle Ekşi Sözlük'te yer alan Z-kuşağına dair nefret söylemi içeriklerini inceleyerek, bu platformda açılan 227 Z-kuşağı başlık girdisindeki 1.610 kullanıcı yorumu veri kümesine örneklem olması için seçilmiştir. Veri kümesi toplama aracı olarak Selenium test aracı kullanılmış ve Ekşi Sözlük'ten nefret söylemi içerikleri çekilmiştir. Elde edilen bu içeriklerin kategorilendirilmesi için betimsel analiz yöntemi kullanılarak Z-kuşağı nefret söylemi ifadeleri açıklanmıştır. Çalışma genel olarak değerlendirildiğinde, Z-kuşağı için toplumsal değerlere yakın olmayan, politik dışı ve sosyal medyaya bağımlı bireylerden oluştuğu gözlenmiştir. Bu davranışların Z-kuşağına nefret söylemi açısından çeşitli olumsuz davranışlara sebebiyet verdiği ve negatif dile sahip bir sağduyu geliştirdiği ayrımla varılmıştır.

Suçta sevk edilen çocuklara dair haberleri ele alan Doğan'ın [114] bu çalışmada, sosyal medyada kullanıcı yorumlarında nefret söylemleri analiz edilerek incelenmiştir. Bu analizler için araştırma yöntemi olarak, sosyal medyadaki çocuklara dair nefret söylemi haberlerinin belirlenmesinde Betimsel Analiz yöntemi seçilmiştir. Çalışmanın amacına uygun şekilde YouTube platformu kullanılarak, çocuklar için üretilen nefret söylemi verilerinin tespiti yapılmıştır. Nefret söylemlerinin tespiti için YouTube haberlerinden bir örneklem seti oluşturulmuştur. Bu örneklem seti içerisinde haber başlıkları, ara başlıklar, kaynaklar, etiketler/sıfatlar, sorumluluk atıfları, çizelgeler ve istatistikler, haber sonuç paragrafları, video haberler ve haberlerde kullanılan ses ve görüntü unsurları gibi verilerden oluşan toplu bir veri kümesi oluşturulmuştur. Araştırma sonuçlarına göre, suçta karışan çocuklara dair olumsuz bir algının, geleneksel medya kadar sosyal medyada da yaygın olduğu söylenebilir. Bu çocuklara yönelik tehlikeli, tehdit unsuru, dışlayıcı ya da ayrımcı etiketleri verilerek, bu tür nefret söylemi ifadelerinin sosyal medyada dolaşıma sunulduğu sonucuna da varılabilir.

Kobaza'nın [115] 2022 yılında gerçekleştirdiği çalışmada, TikTok platformunda engelli bireyler ile ilgili paylaşılan videoların mizah öğesi olarak kullanıldığına dikkat çekilmektedir. Çalışmada, üretilen içeriklerin toplumsal değerlere uygun olup olmadığı düşünülmeden yapılan paylaşımların engelli bireylere karşı duyarlılığın yitirilmesine neden olduğu belirtilmektedir. Ek olarak çalışmada, engelli bireylerin özne konumuna getirilerek, onların durumları, eylemleri ve söylemleri mizahi öğeler olarak kullanıldığı ve bu içeriklerin sosyal

medya platformlarında sürekli olarak yayıldığı da belirtilmiştir. Özellikle TikTok platformunun bu tür içeriklere daha çok zemin hazırladığı ifade edilmiştir. Çalışmada, engelli bireylere karşı olumsuz davranışların normalleştirilmesine ve bu bireylere yönelik duyarlılığın yitirilme tehlikesine dikkat çekilerek, sorunlu içerik anlayışına vurgu yapan Kobaza, bu tür içeriklerin yaygınlaşmasını engellemek için bu çalışmayı oluşturmuştur. Çalışmada araştırma yöntemi olarak, nitel araştırma metodlarından, betimsel analiz yöntemi kullanılmıştır. TikTok platformunda Necip Ağacan ve Fadime Akkaya isimli engelli bireylerin yer aldığı videolar örneklem olarak seçilmiştir.

Siber zorbalık ve sosyal medya arasındaki bağlantıdan hareket eden Dinç [116] çalışmasında, siber zorbalık konusunu cinsiyet, yaş, psikolojik, duygusal, aile-arkadaş ilişkileri, akademik başarıya etkileri, ebeveynlerin ve okul personellerinin zorbalık konusunda alınması gereken önlemleri ve siber zorbalık hakkındaki çözüm önerilerini ele alarak kapsamlı şekilde incelemiştir. Bu doğrultuda incelenen veriler, 11 lise öğrencisi ile yapılan görüşmeler sonucu nitel araştırma yöntemine göre hazırlanmıştır. Yapılan görüşmeler sonucu, çalışmadaki görüşme formunda yer alan cevaplar incelendiğinde, öğrencilerin en çok kullandığı sosyal medya uygulamaları Twitter, Instagram, YouTube, Snapchat, Pinterest ve Tumblr olduğu belirlenmiştir. Çalışmada, ele alınan konuda toplanan veriler, betimsel analiz ile çözümlenerek yorumlanmıştır. Öğrencilerle yapılan görüşmelerden toplanan verilere göre, siber zorbalık konusunda sosyal medyanın en uygun platformlar olduğuna ve bu platformların genç bireyler üzerindeki etkilerine değinilmiştir. Bunun yanında, birçok yönden siber zorbalığa maruz kalan gençler üzerindeki bireysel etkilerinden de söz edilmiştir.

Altan gözetim, bir toplumda bireylerin veya grupların kendilerini bilinçli bir şekilde denetimleri ve kontrol etmeleri anlamına gelmektedir. Anamur ve Topsakal [117] çalışmalarında, toplumun bu farklı gözetim anlayışı ile ortaya koyduğu davranış ölçmeyi hedeflemişlerdir. Bu sebeple, 28 Kasım 2018 tarihinde trafik polisi ile tartışan bir kadının Instagram platformu üzerinden paylaşılan görüntüleri, sosyal medya kullanıcılarının yaklaşımı ile ele alınmıştır. Araştırmada örneklem toplamak için Eksisozluk.com sitesi incelenmiştir. Site içinde konu ile ilgili 11 başlık ve 1.511 kayıt (entry) açılmıştır. Nitel veri analizi yöntemlerinden olan betimsel analiz yöntemi ile başlıklar ve kayıtlar açıklanarak yorumlanmıştır. Araştırmanın sonuçlarına dayanarak, toplum içinde daha eşitlikçi ve demokratik bir yapı oluşturmak adına, alttan gözetim uygulamalarının geniş çapta henüz benimsenmediği ve bu tür gözetimin siber zorbalık gibi olumsuz neticelere yol açtığı sonucuna varılabilir.

Literatürde incelenen çalışmalara göre, farklı sosyal medya araçlarında yer alan nefret söylemleri için betimsel analiz yöntemini temel alan çalışmalar Tablo 8'de gösterilmiştir.

### 3.2.3. Söylem analizi yapan çalışmalar (Studies using discourse analysis)

Sosyal medya araçlarından olan Twitter üzerinde 65 yaş üstü kişilere yönelik nefret söylemlerini ele alan Öztürk [118] çalışmasında, özellikle Covid-19 sürecinde artan olumsuz söylemlerle 65 yaş üstü bireyleri ayrımcılığa maruz bırakacak şekilde ötekileştirmek istendiğini vurgulamıştır. Bunun yanında, çalışmada, Twitter'ın bu tür nefret söylemlerine dair içerik üretimine aracılık ettiği ve sosyal medyada kullanılan belli başlı hashtaglerle bu tür olumsuz ifadelerin desteklendiği ifade edilmiştir. Belirli tarihler arasında toplanan 1.100 tweet ile veri kümesi oluşturulmuştur. Veri kümesinin oluşturulmasında toplama aracı olarak MAXQDA programı kullanılmış ve nefret söylemleri kategorize edilmiştir. Teun Van Dijk'in (mikro ve makro düzeyde) söylem analizi yöntemi

**Tablo 8.** Nefret söylemi tespitine yönelik betimsel analiz yöntemini kullanan çalışmalar.  
(Studies using descriptive analysis method for the detection of hate speech.)

Ref. No	Yıl	Araştırma konusu	Uygulama platformu	Veri kümesi	Veri toplama aşaması	Yöntem
Anamur ve Topsakal [117]	2019	Altın gazetimin uygulamalarının analizi	Instagram ve eksisozluk	Ekşi sözlükte yer alan konu ile ilgili açılmış 11 başlık ve 1511 entry	-	Betimsel analiz
Dinç [116]	2020	Lise öğrencileri için siber zorbalık	Twitter, Instagram, YouTube, Snapchat, Pinterest, Tumblr	11 lise öğrencisi	-	Betimsel analiz
Çeçen [112]	2022	İslamofobik nefret söyleminin tespiti	Web of Science (WoS)	WoS'daki 200 çalışma	-	Betimsel analiz
Doğan [114]	2022	Çocuklara dair nefret söylemi analizi	YouTube	Haber başlıkları, kaynaklar, ara başlıklar, etiketler/sıfatlar, sorumluluk atıfları, istatistikler ve çizelgeler, haber sonuç paragrafları, video haberler, haberlerde kullanılan ses ve görüntü unsurları	-	Betimsel analiz
Kobaza [115]	2022	Engelli bireylerin mizah konusu olması	TikTok	TikTok platformunda Necip Ağacan ve Fadime Akkaya isimli engelli bireylerin yer aldığı videolar	-	Betimsel analiz
Abuşoğlu ve Aydoğan [113]	2023	Z-kuşağına dair nefret söylemleri	Ekşi Sözlük	227 Z-kuşağı başlık girdisindeki 1.610 kullanıcı yorumu	Selenium	Betimsel analiz

kullanılarak oluşturulan veri kümesinde 218 adet nefret söylemi tespiti yapılmıştır.

Twitter üzerinden Suriyeli mültecileri hedef alan Karamete'nin [119] çalışması, nefret söylemlerinin analizi ve sosyolojik açıdan çözümlenmesi konularını ele almaktadır. Bu doğrultuda Suriyeli, mülteci, göçmen vb. anahtar kelimeleri kullanarak tweet araştırması yapılmıştır. Araştırma sonucunda #Suriyelieriistemiyoruz etiketi seçilerek, 2014-2018 yılları arasındaki 1.383 tweet çalışma kapsamında incelenmiştir. Çalışmada, tweetler içerisinde yer alan nefret söylemlerinin oluşumunu ve vurgularını tespit etmek amacıyla Quirkos programıyla toplanan tweetler, Teun Adrian van Dijk'in eleştirel söylem analizi yöntemine göre değerlendirilmiştir. Çalışma kapsamında kullanılan tweetler "olumlu, tarafsız/belirsiz, olumsuz, nefret söylemi ve diğer" şeklinde kategorize edilmiştir. Bununla birlikte, nefret söylemi kategorisinde incelenen tweetler de kendi arasında "simgeleştirme, çarpıtma/yükleme/abartma, aşağılama/hakaret/küfür, savaş söylemi/düşmanlık ve gizli/dolaylı/diğer" şeklinde kategorilere ayrılmıştır. Çalışmanın geneline bakıldığında, nefret söylemi içeren tweetler ile nefret eylemleri arasında anlamlı bir ilişki olduğu sonucuna varılmıştır. Çalışmanın özü, Twitter'ın, nefret söylemlerinin dile getirildiği uygun bir platform olduğu savını da destekleyici niteliktedir.

Covid-19 süresince sağlık çalışanlarına karşı oluşturulan önce destek ve sonra nefret söylemi konularını ele alan Bahçecioğlu'nun [120] bu çalışmada, sağlık iletişiminin önemi vurgulanmıştır. Çalışma, sosyal medya çerçevesinde bir araştırma projesidir ve iki aşamadan oluşmaktadır. İlk aşamada, sağlık iletişiminde pozitif iletişim kurmanın önemi ele alınmıştır. Burada Sağlık İnanç Modeli, bir araç olarak seçilmiş ve sağlık çalışanlarına karşı davranış şekillerinin belirlenmesinde kullanılmıştır. İkinci aşamada ise, Covid-19 sürecinde sağlık çalışanlarına yönelik ortaya çıkan egoist davranışlar ve önyargılı nefret söylemlerine dair düşünceler ifade edilmiştir. Bu araştırma projesinde, Twitter kullanılarak iki hashtag (#sağlıkçılarınıalkışla , #evdekal) üzerinden elde edilen verilerle inceleme yapılmıştır. Yapılan incelemede, seçilen hashtagler ile paylaşılan görsellerde, insanlar sağlık çalışanlarına destek verme

eğilimi göstermiştir. Göstergibilimsel analiz ve eleştirel söylem analizi teknikleriyle toplanan veriler, çözümleme sürecinde kullanılmıştır. Yapılan analizler sonucunda, sağlık çalışanlarına yönelik "benmerkezci" anlayışından sıyrılarak, sosyal medyada bireysel bilinç konusunda bir farkındalık oluşturması beklenmektedir.

Kurt [121] çalışmasında, Türkiye'deki Suriyeli kadın mültecilere ilişkin cinsiyet temelli nefret söylemi konusunu ele almaktadır. Çalışmada, mülteciliğin sosyal deneyimler ve ilişkilerle iç içe geçen boyutu incelenmektedir. Özellikle Suriyeli mülteci kadınların maruz kaldığı nefret söyleminin haberlerdeki boyutları (gazete haberlerindeki örnekler, göç, kadın, ötekileştirme ve nefret söylemi kavramları) Van Dijk'in eleştirel söylem analizi yöntemi kullanılarak ayrıntılandırılmıştır. Bu yöntemin seçilmesindeki amaç, haberlerin toplumsal yöntemlerle bir bütün olduğu ve toplum içindeki sorunlara en iyi şekilde odaklanarak çözümleyici olmasıdır. Geleneksel medyadan elde edilen veriler ışığında, Suriyeli kadınlara yönelik haberlerden elde edilen veriler ayrılarak 10 gazete ve her gazeteden üçer haber örneği seçilip nefret söylemi veri kümesi oluşturulmuştur. Bu veri kümesinde başlıklar, fotoğraflar, kelime seçimleri (anahtar kelimeler) ve haberin retorikliği yer almaktadır. Bu seçimler değerlendirildiğinde Suriyeli mülteci kadınların nefret söylemiyle karşı karşıya kaldığı tespit edilmiştir. Çalışmanın sonuç kısmında, medyanın ataerkil egemenliği altında Suriyeli kadın mültecilerin sorunlarına yeterince ışık tutulmadığı ve nefret söylemine zemin hazırlayan haberlerin kadınlar için çoğunlukta olduğu belirtilmiştir.

Sosyal medyada yaygınlaşan linç kültürü fenomenine odaklanan Karataş [122] çalışmasında, özellikle Twitter platformu içinde bu fenomenin özelliklerini, dinamiklerini ve sosyal medya kullanıcıları üzerindeki etkilerini incelemeyi amaçlamaktadır. Bu amaçla yola çıkan Karataş, Twitter'da çevrimiçi linçlenme örnekleriyle ilgili tweet veri kümesini analiz etmiştir. Bu veri kümesi, 6 ünlü isme yapılan linç tweetlerini içermektedir. Çalışmada veri kümesinin içeriği, linç kültürünün beslediği konulardan oluşturulan örneklem kelimeler ile 6 ünlü isme yapılan linç tweetlerinden oluşmaktadır. Veri kümesinin oluşturulmasında söylem analizi yöntemi kullanılmıştır. Çalışma ayrıca, çevrimiçi linçlenme örneklerini takip ederek kullanıcılara bu

fenomenin doğasını ve kapsamını anlamalarına yardımcı olmak için, geliştirilen yeni bir araç olan "linç atlası analizi"ni içermektedir. Genel olarak çalışma, çevrimiçi linç kültürü fenomeninin karmaşık ve endişe verici doğasına ışık tutmakta ve Twitter gibi sosyal medya platformlarında bu fenomenin nasıl ortaya çıktığına dair içgörüler sağlamaktadır.

Mardin'in [123] çalışmasında, çevrimiçi video paylaşım sitesi üzerinden dijital oyun reklamlarında üretilen nefret söylemleri, nitel bir çalışma şeklinde ele alınmıştır. YouTube kullanıcılarının dijital oyun endüstrisi reklamlarında üretilen nefret ifadeleri ve reklamlara yapılan kullanıcı yorumları, Teun A. van Dijk'in eleştirel söylem analizi tekniğiyle ele alınmıştır. Çalışmada veriler, YouTube'un beğeni sayıları ve reklam yorumları, MAXQDA programı ile çekilmiştir. Veri kümesi, dijital oyunlardaki üç farklı türde beğenme sayısı ve olumsuz yorum alan oyun reklamlarından oluşan bir örneklemi kapsamaktadır. Çalışmanın sonucunda, nefret söylemlerinin online oyun platformlarındaki reklamlarda yaygın şekilde kullanıldığı, reklam dağıtımını yapan firmaya ve firmanın dijital oyunlarına karşı nefret içeriklerinin yoğun şekilde üretilerek tüketici kitlesini olumsuz etkilediği kanısına varılmaktadır.

Literatürde incelenen çalışmalara göre, farklı sosyal medya araçlarında yer alan nefret söylemleri için söylem analizi yöntemini kullanan çalışmalar Tablo 9 ve Tablo 10'da gösterilmiştir.

### 3.2.4. İçerik analizi yapan çalışmalar (Studies using content analysis)

Sosyal medyanın kullanımı üzerinde kriz ve afet yönetimi için bir araştırma çalışması yapan Çanakçı vd. [124] çalışmasında, Türkiye ve dünya genelindeki meydana gelen afetlerde sosyal medyanın kullanım rolü ele alınmıştır. Çalışmada, sosyal medya platformlarından doğru bilgi paylaşımı ile halkın afet bilincinin artırılması ve müdahalede yardım faaliyetlerinin kolaylaştırılması önerilmektedir. Çalışma kapsamında sosyal medyanın kullanımına dair özellikle 1999 yılı

Marmara Depremi ve 2022 yılı Koronavirüs vakalarına kadarki geçen süreçte birkaç afet ele alınarak, Twitter'ın kriz ve afet yönetimi gibi önemli durumlarda verimli şekilde nasıl kullanılabileceği ifade edilmeye çalışılmıştır. Yaklaşık 20 Twitter kullanıcısının paylaştığı içeriklerde, konulara göre birçok hashtag için içerik analizi yapılmış, TT listesinde yer alan hashtagların doğruluğu ve yanlışlığı incelenerek, Twitter'ın daha iyi ve doğru kullanımı için öneriler sunulmuştur.

Pandemi döneminde Çin'de üretilen aşılara yönelik oluşturulan Kuş ve Öztürk'ün [125] çalışmasında, YouTube videolarının altındaki yorumlar aracılığıyla nefret söylemi, kullanıcı etkileşim miktarı ve yanlış bilgi üçgenindeki bağlantı gün yüzüne çıkarılmaya çalışılmıştır. Veri kümesi oluşturma aşamasında, Youtube üzerinden "Sinovac" ve "CoronaVac" anahtar kelimeleri kullanılarak, en çok izlenen beş YouTube videosunun altında yer alan 2.919 kullanıcı yorumu alınmıştır. Yorumların toplanmasında Mozdeh Big Data Text Analysis aracı kullanılmıştır. Çalışma kapsamı, iki aşamalı metod yaklaşımıyla ele alınmıştır. İlk aşamada, tündengelimci nitel içerik analizi yaklaşımı ile YouTube kullanıcı yorumlarının analizi yapılmıştır. İkinci aşamada ise, elde edilen veriler arasındaki bağlantı tespiti yapılmıştır. Elde edilen bu verilere Ki-kare testi uygulanarak, test sonucunda yanlış bilgi ve nefret söylemi arasında pozitif korelasyon olduğu saptanmıştır. Seçilen her iki metod yönteminden elde edilen bulgular değerlendirildiğinde, nefret söylemi ile etkileşim miktarı arasında pozitif bir ilişki olduğu saptanmıştır. Sonuç olarak, geleneksel medya ve yeni medyada yanlış bilgi ve nefret söyleminin birbirini dögüsel olarak beslediği ve yaşanmakta olan COVID-19 sürecindeki sağlık mücadelesini negatif etkileme durumu taşıdığı bu çalışma ile ortaya konulmuştur.

Sosyal bilimler alanında disiplinlerarası çalışmalarda içerik ve söylem analizi yöntemlerini detaylı şekilde ele alan Gül ve Nizam'ın [126] bu çalışması, iki aşamalı yürütülmüştür. Birinci aşamada, içerik analizi türlerinden olan nitel analiz, nicel analiz ve karma analiz teknikleri ele

**Tablo 9.** Nefret söylemi tespitine yönelik söylem analizi yöntemini kullanan çalışmalar (1).  
(Studies using the discourse analysis method for the detection of hate speech (1).)

Ref. No	Yıl	Araştırma konusu	Yöntem
Karamete [119]	2019	Suriyeli mültecilere dair nefret söylemleri	Eleştirel söylem analizi
Öztürk [118]	2021	65 yaş üstü bireylere yönelik nefret söylemi	Eleştirel söylem analizi (makro ve mikro düzeyde)
Bahçecioglu [120]	2022	Sağlık çalışanlarına yönelik nefret söylemi tespiti	Göstergebilimsel analiz, Eleştirel söylem analizi
Karataş [122]	2022	Türkiye'de Twitter alanındaki linç kültürü analizi	Söylem analizi, Linç atlası analizi
Kurt [121]	2023	Suriyeli kadın mülteciler için nefret söylemi analizi	Eleştirel söylem analizi
Mardin [123]	2023	Dijital oyunlarda üretilen reklamlardaki nefret söylemleri	Eleştirel söylem analizi

**Tablo 10.** Nefret söylemi tespitine yönelik söylem analizi yöntemini kullanan çalışmalar (2).  
(Studies using the discourse analysis method for the detection of hate speech (2).)

Ref. No	Uygulama platformu	Veri kümesi	Veri toplama aşaması
Karamete [119]	Twitter	1.383 tweet	Quirkos programı
Öztürk [118]	Twitter	1.100 tane tweet	MAXQDA programı
Bahçecioglu [120]	Twitter	#sağlıkçılarınıalkışla ve #evdekal hashtagleri üzerinden elde edilen veriler	Sağlık İnanç Modeli
Karataş [122]	Twitter	6 ünlü isme yapılan linç tweetleri	Linç kültürünün beslediği konulardan oluşturulan örneklem kelimeler ile yazar tarafından tweetlerin araması
Kurt [121]	Geleneksel medya	10 gazete ve her gazeteden üçer haber örneği seçilerek oluşturulan veriler	Gazeteler
Mardin [123]	YouTube	Dijital oyunlardaki 3 farklı türde beğenme sayısı ve olumsuz yorum alan oyun reklamları	MAXQDA programı

alınmıştır. İkinci aşamada ise, söylem analizi türlerinden olan Foucaultcu ve Van Dijkci analiz teknikleri açıklanmıştır. Araştırma alanlarında tekniksel kısımların doğru belirlenmesi ve uygulama adamlarında ise, içerik ve söylem analizi yöntemlerinin uygulamalı şekilde açıklanması ile konular örneklendirilmiştir. Uygulamalar için gazeteler, filmler, ders kitapları ve yapılan görüşmeler, katılımcıların anlatıları, Atatürk'ün yapmış olduğu meclis konuşmaları, haberler ve sosyal medyadaki "Hashtag"ler kullanılarak veri kümesi oluşturulmuştur. Dolayısıyla geçmişten günümüze bilimsel çalışmalar incelendiğinde, araştırma verilerini analiz ederken araştırma tekniklerinin oldukça önemli olduğu ve analiz yöntemlerinin de bu çalışmayla nicel yaklaşımlardan nitel yaklaşımlara doğru kaydığı söylenebilir.

Bilgiler'in [127] hazırladığı bu çalışmanın ilk amacı, Youtube platformunun kullanıcı yorumları ile tekrardan oluşturulan ve yayılan cinsel kimlikle ilgili (cinsel yönelim) nefret suçlarını [128] tetikleyen ve nefret söylem ifadelerini ortaya çıkaran söylemlere dikkat çekmektir. İkinci amacı ise, cinsel kimlikle ilgili destekleyen iletilerin incelenmesidir. Çalışmada yöntem olarak, araştırma probleminde ait seçilen ilgili Youtube videolarındaki yorumlar, nicel içerik analizi tekniğiyle incelenmiştir. İçerik analizinde, başka bir nefret söylemi tespiti çalışmasında geliştirilen kod cetveli araştırmaya uyarlanarak kullanılmıştır. Çalışmanın veri kümesinde, üç Youtube kanalı seçilmiştir ve bu kanalların içeriklerine yapılan 3.000 tane yorum ele alınmıştır. Bu kanallar seçilirken dikkat edilen konuların en başında, cinsel yönelimi hakkında bilgi veren kanal sahibinin eşcinsel bir birey olmasıdır. Araştırmada seçilen video içeriklerinde kanalın sahipleri, cinsel yönelimlerini ailesi ve sosyal çevresiyle paylaştıklarında aldıkları tepkileri anlatmaktadır. Çalışmanın algoritmasında, üç eşcinsel bireye ait YouTube videolarına yapılan 3.000 tane destekleyici ve nefret içerikli yorumlar incelenmiştir. Üç farklı videoya yapılmış olan nefret içerikli ve destekleyici yorumlar oranlanarak bu konuda sonuca ulaşılmıştır. Sonuç olarak, yeni

medyada nefret söylemlerinin ötekileştirilen kişi ve grupların sesi olacağı ve nefret suçuna [129] dönüşebileceği durumları da ifade edilmiştir.

Facebook'ta nefret söyleminin etnik temelli biçimde nasıl inşa edildiğini araştıran Sağlam [130] bu çalışmasında, örnek olarak BBC News'in Dağlık Karabağ Çatışması'na ilişkin yayınladığı bir haberi seçerek çalışmasını oluşturmuştur. Çalışmada, haberin altındaki kullanıcı yorumları eleştirel söylem analizi, içerik çözümlemesi ve nicel araştırma yöntemlerine tabii tutulmuştur. BBC News'in Dağlık Karabağ Çatışması haberlerindeki Facebook sayfasının 4.000 kullanıcı yorumu veri kümesi olarak seçilmiştir. Veri kümesinin toplanmasında yazar, haber içeriğine göre Facebook kullanıcılarının yorumlarını ele almaktadır. Araştırma çerçevesinde çalışma, Ermeni ve Azeri Facebook kullanıcılarının nefret söylemlerini oluştururken birbirleriyle benzer ve farklı yanlarındaki hususlarına da değinerek, nefret söylemini önlemede yasal düzenlemelerin önemine ve medya kurumlarına düşen görevler de öneri olarak sunmaktadır. Sosyal medyada düşünce ve algı analizi yapmayı hedefleyen Kaygısız'ın [131] bu çalışmasında, mobbing kavramı üzerinde durulmuştur. Bu çalışma için 15.11.2022 - 20.12.2022 tarih aralığı seçilerek iki farklı veri kümesi oluşturulmuştur. Bunlardan ilk veri kümesi, YouTube'daki Türkçe dilinde yapılan mobbing içerikli 2.285 kullanıcı yorumları ve ikinci veri kümesi de Twitter'daki mobbing içeren 3.926 Türkçe içerikli yorumlar şeklindedir. Çalışmada yöntem olarak içerik analizi yöntemi seçilmiş ve MAXQDA 2020 programı kullanılarak iki farklı veri kümesi oluşturulmuştur. Çalışmada elde edilen sonuçlara göre, mobbingin her meslek grubunda doğrudan veya dolaylı şekilde maruz kalılabileceği, kişilerin yaşam kalitesini düşürdüğü ve belli bir kalıbı olmayan bir işyeri virtüsü olarak çalışanlar arasında (özellikle günümüzde) artarak yayıldığı ifade edilmiştir.

Literatürde incelenen çalışmalara göre, farklı sosyal medya araçlarında yer alan nefret söylemleri için içerik analizi yöntemini ele alan çalışmalar Tablo 11'de gösterilmiştir.

**Tablo 11.** Nefret söylemi tespitine yönelik içerik analizi yöntemini kullanan çalışmalar.  
(Studies using content analysis method for the detection of hate speech.)

Ref. No	Yıl	Araştırma konusu	Uygulama platformu	Veri kümesi	Veri toplama aşaması	Yöntem
Kuş ve Öztürk [125]	2021	Çin'de üretilen aşılara dair nefret söylemi ve yanlış bilgi arasındaki bağlantı	YouTube	En çok izlenen beş YouTube videosuna yapılan 2.919 kullanıcı yorumu	Mozdeh Big Data Text Analysis programı	Tümdengelimci nitel içerik analizi, Ki-kare testi
Gül ve Nizam [126]	2021	Sosyal bilimlerde nefret söyleminin içerik ve söylem analizi teknikleri	-	Gazeteler, filmler, ders kitapları ve yapılan görüşmeler, katılımcıların anlatıları, Atatürk'ün yapmış olduğu meclis konuşmaları, haberler, sosyal medyadaki "Hashtag"ler	-	İçerik analizi (nitel, nicel, karma analiz teknikleri) ; Söylem analizi (Foucaultcu ve Van Dijkci analiz teknikleri)
Bilgiler ve İnceli [127]	2021	Nefret söylemi tespiti	YouTube	Üç YouTube videosuna yapılan 3.000 kullanıcı yorumu	-	Nicel içerik analizi
Çanakçı vd. [124]	2022	Kriz ve afet anında sosyal medyanın kullanımı	Twitter	20 Twitter kullanıcısının paylaştığı içerikler	-	İçerik analizi
Sağlam [130]	2022	BBC News - Dağlık Karabağ Çatışması üzerinde nefret söyleminin inşası	BBC News (Facebook sayfası)	BBC News'in Facebook Sayfasındaki Dağlık Karabağ Çatışması Haberlerindeki 4.000 kullanıcı yorumu	Haber içeriğine göre Facebook kullanıcılarının yorumlarının yazar tarafından bulunması	Nitel araştırma yöntemi, İçerik çözümlemesi (içerik analizi), Eleştirel söylem analizi
Kaygısız [131]	2023	Sosyal medyada mobbing kavramı	YouTube, Twitter	YouTube'daki mobbing içerikli 2.285 kullanıcı yorumları ; Twitter'daki mobbing içerikli 3.926 kullanıcı yorumları	MAXQDA programı	İçerik analizi

#### 4. İncelenen Uygulama Tabanlı Türkçe Nefret Söylemi Tespitine Yönelik Çalışmalarda Kullanılan Yaklaşımların Performans Kıyaslaması ve Paylaşılan Açık Erişimli Veri Kümeleri (Performance Comparison of Approaches Used in Reviewed Application-Based Turkish Hate Speech Detection Studies and Shared Open Access Datasets)

Nefret söylemi, toplumlar üzerinde derin etkilere neden olan ciddi bir sosyal sorundur. Bu sorun, günümüzde teknolojinin bir gelişimi olan internet aracılığıyla insanlar arasında da yayılmaktadır. Her dilden, dinden, renkten ve kültürden insanı bir araya getiren internet, küresel dünyanın ana arterini oluştururken, internetin en çok kullanılan araçlarından birisi olan sosyal medya ise, nefret söylemlerinin yayılmasına uygun zemin hazırlamaktadır. Buradan yola çıkarak bu çalışma kapsamında, nefret söylemi üzerine geniş bir tarama yapılmıştır. Çalışma kapsamında, Türkçe nefret söylemi üzerine yapılan akademik çalışmalar incelenmiştir. İncelenen çalışmalar; uygulama tabanlı çalışmalar ve sosyal analiz tabanlı çalışmalar olmak üzere iki başlıkta toplanmıştır. Uygulama tabanlı çalışmalarda, nefret söylemi tespiti için otomatik nefret söylemi tespit sistemleri geliştirilmiştir. İncelenen bu uygulama tabanlı çalışmaların sonuçlarından elde edilen bulgular ve paylaşılan açık erişimli veri kümeleri bu başlık altında değerlendirilmiştir.

##### 4.1. Türkçe Nefret Söylemi Tespiti İçin Kullanılan Yöntemlerin Performans Kıyaslaması ve Genel Değerlendirmesi (Performance Comparison and General Evaluation of Methods Used for Turkish Hate Speech Detection)

Makale kapsamında nefret söylemleri alanında incelenen çalışmalar, uygulama tabanlı çalışmalar ve sosyal analiz çalışmaları şeklinde iki ana başlık altında toplanmıştır. Uygulama tabanlı çalışmalarda, nefret söylemi tespiti için, makine öğrenimi ve derin öğrenme yöntemleri kullanılarak yapılan uygulamalar yer almaktadır. Bu adım için kullanılan veri kümeleri iki farklı şekilde temin edilmiştir. İlki, yazılım geliştirme ve veri bilimi alanında popüler platformlar arasında olan GitHub [132] ya da Kaggle [133] ortamında hazır paylaşılan açık erişimli veri kümelerinin kullanımınıdır. İkincisi ise, veri kümesinin çalışma kapsamında hazırlanmasıdır. Bu yöntem, veri kümesinin hazırlama işlemlerinde kullanılacak veri toplama ortamlarının ve sosyal medyaların belirlenmesiyle başlamaktadır. Ardından çeşitli veri toplama araçları ile nefret söylemi içerikli paylaşımlar, uygun anahtar kelimelerle aratarak veri kümesi oluşturulmaktadır. Çalışma kapsamında değerlendirilen makalelerde nefret söylemi içerikli veri kümelerinin oluşturulmasında en fazla Twitter [134] platformu kullanılmıştır. Bu durum göstermektedir ki, Twitter ortamında kötü niyetli kullanıcılar gerçek kimliklerini gizleyerek ya da gizlemeden, düşüncelerini rahat bir şekilde ifade edebilmekte ve geniş erişim imkanlarına sahip oldukları için, nefret söylemi içerikli mesajları kolayca paylaşabilmektedir. Ayrıca, sosyal medya algoritmaları, kullanıcıların daha fazla etkileşim almasını sağlamak amacıyla nefret içerikli paylaşımları da teşvik edebilmektedir.

Oluşturulan veri kümeleri üzerinde, veri ön işleme adımları uygulanarak veri işlenebilir hale getirilmiştir. Makale kapsamında incelenen çalışmalarda özellikle Twitter API yada SNScrape kütüphanesi kullanılarak Twitter platformundan veri kümesi hazırlandığı için, veri kümesi nefret söylemi içeren tweetlerden oluşmaktadır. Veri ön işleme aşamalarında tweet içerisinde yer alan noktalama işaretleri, URL adresleri, kullanıcı isimleri kaldırılarak temizlenmiştir. Yazım hataları, yazım denetleme servisleri kullanılarak yapılmıştır. Otomatik olarak düzeltilmeyen hatalarda elle düzeltilme işlemleri yapılmıştır.

Metin madenciliğinde, metin verilerinden anlamlı bir bilgi ve sonuç çıkarımı için kullanılan çeşitli özellik kümeleri mevcuttur. İncelenen çalışmalarda veri kümesi içerisinde yer alan kelimelerin morfolojik özelliklerine göre analizi için, açık kaynak kodlu Türkçe doğal dil

işleme kütüphanesi olan Zemberek [135] kullanılmıştır. Zemberek, Türkçe dilindeki kelime köklerini analiz edebilen bir kütüphanedir. Bu kütüphane, metin madenciliği uygulamalarında, morfolojik analiz açısından kelimenin kökleri, ekleri ve dilbilgisel özellikleri üzerinde inceleme yapılmaktadır. Bunun yanında, kelimenin duygu yükünü veya nefret söylemi içerip içermediğini belirlemede de yardımcı olduğu Doğal Dil İşleme uygulamalarındaki kullanımı oldukça önemlidir. Tweet içerisinde ardışık gelen kelime veya karakter gruplarını temsil etmek için kullanılan özellik kümesine N-gram [136] denilmektedir. N- gram tekniğindeki teknikteki amaç, metin içerisindeki dilbilgisel yapıları ve ifade biçimlerini anlamak, ardından bu yapıları sayısal bir temsil ile modele aktarmaktır. İncelenen uygulamaya yönelik çalışmalar için en sık kullanılan özellik kümeleri dokümandaki her bir kelimenin tek başına temsili anlamına gelen Unigram (1-GRAM), ardışık iki kelimeyi içeren Bi-gram (2-GRAM) ve Ardışık n kelimeyi içeren N-gram ( $N \geq 1$ ) olmuştur. En sık kullanılan metin temsil yöntemleri ise Word to Vector (Word2Vec) ve Bag of Words (BoW) [137] olmuştur.

Son olarak, veriler üzerinde uygun yapay zeka yöntemleri kullanılarak model eğitimi gerçekleştirilmektedir. İncelenen çalışmalar arasında gerçekleştirilen modellerin eğitimi için makine öğrenmesinde en çok Destek Vektör Makineleri (Support Vector Machines, SVM) algoritması ve derin öğrenmede en çok Uzun-Kısa Vadeli Hafıza (Long Short Term Memory, LSTM) algoritması kullanılmıştır. Performans metriği olarak F1-skör sonucuna göre en başarılı algoritmaların SVM ve LSTM olduğu gözlemlenmiştir.

##### 4.2. Türkçe Nefret Söylemi Kapsamında İncelenen Çalışmalarda Kullanılan Açık Erişimli Veri Kümeleri (Open-Access Datasets Used in Studies Examined Within the Scope of Turkish Hate Speech)

Türkçe nefret söylemi, Türkçe dilinin çeşitli iletişim platformlarında ayrımcılık, ötekileştirme, hoşgörüsüzlük, önyargı, kalıp yargı gibi nefret içeren ifadelerle kullanılmaktadır. Bu tür söylemler, dijital ortamda toplumsal barışı tehdit etmekte ve bireylerin psikolojik ve duygusal açıdan sağlığını olumsuz yönde etkilemektedir. Bu nedenle, Türkçe nefret söylemi üzerine yapılan akademik çalışmalar mevcuttur. Yapılan çalışmalar ortaya çıkan olumsuz etkileri anlamak, önlemek ve mücadele etmek adına nefret söylemlerinde önemli bir rol oynamaktadır.

Türkçe nefret söylemi alanında gerçekleştirilen akademik çalışmaların sonuçlarını değerlendirebilmenin temel koşulu, açık erişimli veri kümeleridir. Açık erişimli veri kümeleri, verilerin herkes tarafından kolay şekilde erişilebilmesi, paylaşılabilmesi ve kullanılabilmesi için oluşturulan ve çeşitli dijital platformlardan elde edilen verilerdir [138]. Özellikle son yıllarda veri temelli politikaların daha çok yaygınlaştığı düşünüldüğünde, veri kümelerinin açık erişimli hale getirilmesi giderek önem kazanmıştır. Araştırmacıların analiz yapması, model geliştirmesi ve sonuçlarını değerlendirmesi için bu veri kümelerine ücretsiz erişim sağlanmaktadır. Türkçe nefret söylemi üzerine geliştirilen akademik çalışmalarda, genellikle sosyal medya platformlarından elde edilen veri kümeleri tercih edilmektedir. Bu veri kümeleri, nefret söylemi içeren metinler, kullanıcı paylaşımları ve yorumları gibi verileri içermektedir. Veri kümelerinde makine öğrenimi, derin öğrenme ve doğal dil işleme yöntemleri kullanılarak nefret söyleminin tespiti, sınıflandırılması ve önlenmesi konularında çalışmalar gerçekleştirilmektedir.

Türkçe nefret söylemi üzerine yapılan akademik çalışmalarda nefret söylemiyle mücadele etmek, çevrimiçi ortamın daha güvenli bir yer haline gelmesine katkıda bulunmak ve toplumda hoşgörüyü teşvik etmek adına açık erişimli veri kümelerinin kullanımını oldukça önemlidir. Bu çalışma kapsamında, yapılan literatür taraması ile

incelenen makalelerde kullanılan veri kümelerinin boyutu, veri kümelerinin toplanma aracı gibi veri kümelerine dair bilgiler, incelenen başlıklara göre ayrılarak Tablo 4, Tablo 5, Tablo 6, Tablo 7, Tablo 8, Tablo 9, Tablo 10 ve Tablo 11’de gösterilmiştir. Tablo 12 ve Tablo 13’te ise tüm bu çalışmalarda kullanılan veri kümelerinin açık erişimli olanları paylaşılmıştır ve erişim linki paylaşılan veri kümelerinin türü, hangi sosyal medya ortamından toplandığı, oluşturulma yılı, veri kümesi dili ve veri kümesi boyutu bilgileri yer almaktadır.

##### 5. Güncel Zorluklar, Gelecek Yönlendirmeler ve Sonuç (Current Challenges, Future Guidance and Conclusion)

Sosyal medya, bireylerin fikir, düşünce ve eylemlerini rahatça dile getirdiği çeşitli platformlardır. Bu platformlarda, benzer veya aynı

düşüncelere sahip bireyler kolayca bütünleşebileceği gibi, farklı kimlik ve düşüncelerle karşılaşıldığında ise nefret söylemine varan ifadeler ortaya çıkabilmektedir. Sosyal medyanın yapısı, geniş ve sürekli kendini yenileyen bir platform olduğu için, nefret söylemlerini üretmeye müsait ortamlardır. Bu nedenle, farklı sosyal medya platformlarını kullanan farklı dillere sahip bireyler üzerinde nefret söylemlerine dair olumsuz etkiler oluşabilir.

Sosyal medyada nefret söylemlerinin güncel zorlukları arasında, farklı diller örnek verilebilir. Örneğin, Türkçe gibi sondan eklemeli dil ailesi grubunda yer alan diller, bu zorluklar arasında gösterilebilir. Bu durum, hece ve sözcük açısından Türkçe dilinin yapısal zenginliğinden kaynaklanmaktadır. Sosyal medyada Türkçe dili için, kültürel açıdan dilde kullanılan özel ifadeler, yöresel sözcükler, şive, ağız, lehçe gibi farklılıklar, eş anlamlı, eş sesli, mecaz anlamlı

**Tablo 12.** Nefret söylemi tespitine yönelik makale kapsamında değerlendirilen çalışmaların açık erişimli veri kümeleri (1). (Open access datasets of the studies evaluated within the scope of the article on hate speech detection (1).)

Yazar	Yıl	Veri Kümesi Türü	Dil	Boyut
Bohra vd. [28]	2018	Tweet (Metin)	İngilizce, Hintçe	4.575
Santosh ve Aravind [29]	2019	Tweet (Metin)	İngilizce, Hintçe	3.800
Ishmam ve Sharmin [31]	2019	Facebook Yorumları (Metin)	Bengali	5.126
Corazza vd. [30]	2020	Tweet (Metin)	İngilizce, İtalyanca, Almanca	İngilizce 16k, İtalyanca 4.000 (( <i>EVALITA 2018</i> ), Almanca 5.009 ( <i>KONVENS2018</i> ))
Çöltekin [94]	2020	Tweet (Metin)	Türkçe	36.232
Modha vd. [26]	2021	Tweet ve Facebook Yorumları (Metin)	Hintçe, İngilizce	18k Tweet 21k Facebook yorumu
Mandl vd. [27]	2021	Tweet (Metin)	İngilizce, Hintçe, Marathi	Eğitim için veri boyutu: 10.311, Test için veri boyutu: 3.438
Mayda vd. [40]	2021	Tweet (Metin)	Türkçe	1k
Baydoğan ve Alataş [78]	2021	Tweet (Metin)	İngilizce	40.623
Baydogan ve Alatas [97]	2021	Tweet ve Web Forum (Metin)	İngilizce ve İspanyolca	Veri kümesi 1: 25k Veri kümesi 2: 4k, Veri kümesi 3: 13k
Ghosh vd. [34]	2022	Youtube ve Facebook Yorumları (Metin)	Bangla	30k
Şahiner Yılmaz vd. [83]	2022	Tweet (Metin)	Türkçe	14.752
Bayrak [102]	2022	Sanal platformların tümü	Türkçe	TÜİK 2021 Verileri
Nesiboğlu ve Gencer [92]	2023	Tweet (Metin)	Türkçe	15.110

**Tablo 13.** Nefret söylemi tespitine yönelik makale kapsamında değerlendirilen çalışmaların açık erişimli veri kümeleri (2). (Open access datasets of the studies evaluated within the scope of the article on hate speech detection (2).)

Yazar	Kaynak
Bohra vd. [28]	<a href="https://github.com/deepanshu1995/HateSpeech-Hindi-English-Code-Mixed-Social-Media-Text">https://github.com/deepanshu1995/HateSpeech-Hindi-English-Code-Mixed-Social-Media-Text</a>
Santosh ve Aravind [29]	<a href="https://github.com/deepanshu1995/HateSpeech-Hindi-English-Code-Mixed-Social-Media-Text">https://github.com/deepanshu1995/HateSpeech-Hindi-English-Code-Mixed-Social-Media-Text</a>
Ishmam ve Sharmin [31]	<a href="https://github.com/IshmamAlvi/Hate-Speech-for-Bengali-language">https://github.com/IshmamAlvi/Hate-Speech-for-Bengali-language</a>
Corazza vd. [30]	<a href="http://github.com/zeerakw/hatespeech">http://github.com/zeerakw/hatespeech</a>
Çöltekin [94]	<a href="https://huggingface.co/datasets/offenseval2020_tr">https://huggingface.co/datasets/offenseval2020_tr</a>
Modha vd. [26]	<a href="https://github.com/kmi-linguistics/trac-1">https://github.com/kmi-linguistics/trac-1</a>
Mandl vd. [27]	<a href="https://hasocfire.github.io/hasoc/2023/dataset.html">https://hasocfire.github.io/hasoc/2023/dataset.html</a>
Mayda vd. [40]	<a href="https://github.com/imayda/turkish-hate-speech-dataset-1">https://github.com/imayda/turkish-hate-speech-dataset-1</a>
Baydoğan ve Alataş [78]	<a href="https://www.kaggle.com/usharengaraju/dynamically-generated-hate-speech-dataset">https://www.kaggle.com/usharengaraju/dynamically-generated-hate-speech-dataset</a>
Baydogan ve Alatas [97]	Veri kümesi 1: <a href="https://github.com/t-davidson/hate-speech-and-offensive-language">https://github.com/t-davidson/hate-speech-and-offensive-language</a> Veri kümesi 2: <a href="https://www.kaggle.com/c/detecting-insults-in-social-commentary/data">https://www.kaggle.com/c/detecting-insults-in-social-commentary/data</a> Veri kümesi 3: <a href="https://competitions.codalab.org/competitions/19935">https://competitions.codalab.org/competitions/19935</a>
Ghosh vd. [34]	<a href="https://www.kaggle.com/datasets/naurosromim/bengali-hate-speech-dataset">https://www.kaggle.com/datasets/naurosromim/bengali-hate-speech-dataset</a>
Şahiner Yılmaz vd. [83]	<a href="https://www.kaggle.com/datasets/eymaahner/trke-saldragan-dil-derlemi">https://www.kaggle.com/datasets/eymaahner/trke-saldragan-dil-derlemi</a>
Bayrak [102]	<a href="https://data.tuik.gov.tr/Bulten/Index?p=Istatistiklerle-Cocuk-2021-45633">https://data.tuik.gov.tr/Bulten/Index?p=Istatistiklerle-Cocuk-2021-45633</a>
Nesiboğlu ve Gencer [92]	<a href="https://github.com/ezgisubasi/turkish-tweets-sentiment-analysis">https://github.com/ezgisubasi/turkish-tweets-sentiment-analysis</a> <a href="https://github.com/d35k/Turkish-Swear-Words">https://github.com/d35k/Turkish-Swear-Words</a>

kelimeler/ söz öbekleri, atasözleri, deyimler nefret söylemi üzerine güncel zorluklar arasında söylenebilir. Tüm bu farklılıklar Türkçe dilini az bilen ya da dili hiç bilmeyen biri tarafından yanlış algılanarak nefret söylemi üretimine neden olabilir.

Çalışma kapsamında incelenen makalelerde, nefret söylemi üretimine dair uygulama tabanlı çalışmaların değerlendirilmesine detaylı olarak Bölüm 4'te yer verilmiştir. Konuya yönelik yapılan sosyal analiz çalışmaları ise kendi içerisinde konu temelli dikkat uyandıran çalışmalar, betimsel analiz yapan çalışmalar, söylem analizi yapan çalışmalar ve içerik analizi yapan çalışmalar şeklinde dört ayrı alt başlıkta değerlendirilmiştir. Konu temelli dikkat uyandıran çalışmalarda, sosyal medyanın gerçek hayat üzerindeki bireylere etkilerinden bahsedilmiştir. Burada, dijitalleşme ile sosyal medyada nefret söylemlerinden homofobi, şiddet, ayrımcılık, hakaret, troll kültürü, saldırgan mizah olgusu, Metaverse ortamların etkileri gibi kavramların toplumsal bir sorun olarak nasıl görüldüğü ve bireylere yan etkilerinden bahsedilmiştir. Betimsel analiz yapan çalışmalarda, çocuklar, Z kuşağı bireyleri, engelli bireyler ve lise öğrencileri gibi kişiye yönelik veya islamofobik açıdan konu tabanlı nefret söylemi analizleri yapılmıştır. Ekşi sözlük, Youtube ve TikTok gibi sosyal medya uygulamaları aracılığıyla veriler toplanmış, veri dağılımları ve ilişkileri analiz edilerek sonuçlar yorumlanmıştır. Söylem analizi çalışmalarında, sosyal medyada yaşlı bireyleri, sağlık çalışanlarını, mültecileri ve bireyler üzerinden linç kültürü fenomenine odaklanılarak, nefret söylemlerinin analizi ve sosyolojik açıdan çözümlenmesi konuları ele alınmıştır. Son olarak içerik analizi çalışmalarında ise, Türkiye ve dünya genelindeki meydana gelen önemli olaylarda (afet ve kriz yönetimi, depresyon, pandemi, Covid aşısı, bölgesel çatışmalar, mobbing vb.) sosyal medyanın kullanım rolü ele alınmıştır. Sosyal analiz tabanlı çalışmaların genelinde, Twitter ve YouTube platformlarından elde edilen veriler üzerinden analizlerin yapıldığı görülmüştür. Tüm bu alt başlıklar toplanıp değerlendirildiğinde, makaleler genel olarak nefret söylemine maruz kalan sosyal medya kullanıcılarının olumsuz yönde etkilendiğini göstermektedir.

Genel olarak, Türkçe nefret söylemi üzerine yapılan araştırmalar, nefret söylemi konusunun ne kadar ciddi olduğunu ortaya koymaktadır. Toplum üzerinde birçok olumsuz yönde etkileri olan nefret söylemlerinin, sosyal medya gibi dijital platformlarda kullanımıyla hızla arttığı tespit edilmiştir. Bu sonuçlar, nefret söylemi ile mücadele adına, etkili stratejilerin geliştirilmesi ve toplumun bilinçlendirilmesi adına önemli bir temel oluşturmaktadır.

### 5.1. Türkçe Dilinde Nefret Söylemi Tespitinin Zorlukları (Challenges of Detecting Hate Speech in Turkish Language)

Dilbilim açısından Türkçe, sondan eklemeli bir dil ailesi grubunda yer almaktadır. Türkçenin sondan eklemeli bir dil olmasından dolayı, bir sözcüğe farklı ekler eklendiğinde pek çok yeni sözcük üretilebilmektedir [139]. Bu durum, hece ve sözcük açısından Türkçe dilinin yapısal farklılıklarından ötürü zengin bir dile sahip olmasından kaynaklanmaktadır. Dilin sahip olduğu yapısal özelliklerin birleşiminden meydana gelen farklı sözcük gruplarının da farklı anlamlara gelecek şekilde dili zenginleştirdiği söylenebilir. Türkçe'de nefret söylemleri ele alındığında, dildeki yapısal özelliklerden kaynaklı bir kelimenin birden çok anlamı olabileceği için, nefret söylemi tespitinin yapılması diğer dillere göre daha zordur. Bunun yanında, yabancı dillerden Türkçe diline geçerek yerleşmiş birtakım sözcükler de yine nefret söylemi tespitini zorlaştırmaktadır. Ayrıca, kültürel açıdan kullanılan özel ifadeler, yöresel sözcükler, şive, ağız, lehçe gibi farklılıklar da nefret söylemi tespitinin önünde bir engeldir. Türkçe dilini zenginleştiren atasözleri ve deyimler de göz önüne alındığında, günlük kullanımda mecaz anlamlı ifadeler olsa da Türkçe dilini az bilen ya da dili hiç bilmeyen biri tarafından yanlış algılanarak

nefret söylemi olarak nitelendirilebilir. Yine, mecaz anlamlı sözcükler, eş anlamlı sözcükler veya eş sesli sözcükler de Türkçe dilinde farklı anlamlara gelebileceği için farklı kesimler tarafından yanlış değerlendirilerek nefret söylemi olarak algılanabilir. Türkçe dilinin bu denli dilsel ve kavramsal zorluklarının yanında, sosyal medya dili de Türkçe açısından nefret söylemi için olumsuz bir neden oluşturmaktadır. Sosyal medyada, Türkçe dilinde bir kesim tarafından nefret söylemi olmayan ifadeler, diğer kesim için nefret söylemi ifadesi olarak algılanabilmektedir. Tüm bu zorluklar düşünüldüğünde, özellikle sosyal medya alanında, Türkçe dili için nefret söylemlerinin tespitini zorlaştırdığı görülmektedir.

### 5.2. Türkçe Dilinde Nefret Söylemi Örneklerinin Kısıtlılığı (Limitation of Samples of Hate Speech in Turkish Language)

Türkçe dilinde nefret söylemi üzerine yapılan akademik araştırmalar için en büyük kısıtlılık, bu alanda oluşturulmuş veri kümelerinin sınırlı olmasıdır. Bu durum, araştırmacıların konuyu derinlemesine anlamalarını ve etkili çözümler üretmelerini zorlaştırmaktadır. Türkçe dilinde nefret söylemini temsil eden çeşitli ve büyük boyutlu veri kümelerinin eksikliği, model geliştirme, analiz yapma ve doğrulama süreçlerini de olumsuz yönde etkilemektedir.

Türkçe nefret söylemi üzerine yapılan akademik çalışmalarda, veri kümelerinin etik kullanımı ve oluşturulması büyük önem taşımaktadır. Veri kümesi oluştururken nefret söylemi içerikli metinlerin toplanması ve sınıflandırılması süreci zordur. Bu tür verilerin elde edilmesi genellikle zaman alıcı ve maliyetli olabilmektedir. Özellikle nefret söylemi konusuna yönelik hazırlanmış veri kümeleri, veri mahremiyetini ve güvenliğini tehlikeye atmamak adına veri paylaşımı sınırlı yapılmakta ya da yapılmamaktadır. Bu durum, Türkçe dilinde nefret söylemi örneklerinin kısıtlılığına sebep olmaktadır. Bunun yanında, doğru bir şekilde etiketlenmiş veri kümelerinin az bulunması ya da oluşturulan veri kümelerinin farklı insani duygulardan ötürü erişime açık hale getirilmemesi, Türkçe dilindeki nefret söylemi örneklerinin kısıtlılığına sebep olan diğer faktörlerdendir.

Çalışma kapsamında "Türkçe Dilinde Nefret Söylemi Tespitinin Zorlukları" başlığında anlatılan Türkçe dilinin yapısal farklılıkları, yabancı dillerden Türkçe diline geçerek yerleşmiş birtakım sözcükler, kültürel açıdan kullanılan özel ifadeler, yöresel sözcükler, şive, ağız, lehçe gibi farklılıklar, atasözleri ve deyimler ile dildeki nefret söylemi tespitini zorlaştıran durumlar da bu alanda örneklem oluşturulmasına sebep olabilmektedir. Türkçe dilindeki nefret söylemi üzerine çalışan araştırmacılar, karşılaşılan bu kısıtlılıklar için çözüm arayışlarını sürdürmelidir. Özellikle araştırmacılar, kişisel verilerde gizlilik ve veri güvenliği konularına dikkat etmeli, veri kümelerini kullanırken ve paylaşırken, veri mahremiyetini ve güvenliğini sağlamak için gerekli önlemleri almalıdır. Bu alanda yeni veri kümeleri oluşturmak, mevcut veri kümelerini genişletmek, etik kurallara uygun şekilde veri paylaşımını teşvik etmek ve veri mahremiyetini korumak adına önemli adımlar olacaktır.

### 5.3. Türkçe Dilinde Nefret Söyleminin Geleceği (The Future of Hate Speech in Turkish Language)

Geçmişten günümüze bakıldığında, iletişim teknolojilerinde meydana gelen değişimler, interneti hayatımızın önemli bir parçası haline getirmiştir. Teknolojinin hızla gelişmesi, geleneksel medyanın yerini günümüzde sosyal medyaya bırakmıştır. Geleneksel medya, genellikle egemen bir ideolojiyi yansıtırken, sosyal medya ise toplumun tüm farklı kesimlerini temsil edebilme özelliğini yansıtmaktadır [140]. Bu iki açıdan bakıldığında, nefret söyleminin ana kaynağı toplumun kendisidir denilebilir. Toplumdaki belli kalıplarla/ideolojilerle üretilen nefret söylemi ifadeleri, sosyal

medyada olumlu/olumsuz şekilde kanıksanarak nefret söylemlerinin yayılmasına destek vermektedir. Tam da bu noktada, sosyal medyanın nefret söylemlerinin tümü, bireylerin ve toplumun temel ve özel değerlerini aşarak, bireylerin olumlu gelişimi ve toplumun ilerleyişi önünde büyük bir engeldir. Geleneksel ve sosyal medyanın temel amacı, insanlar arasındaki iletişimi sağlamaktır. İletişimin ana gayesi unutulurken, görsel/işitsel/yazınsal olumsuz öğelerle desteklenen sosyal medya ortamları, birçok insanı olumsuz şekilde etkileyerek bireylerin günlük yaşantısını ve sosyal medya üzerinden iletişim hakkını da engellemektedir. Genelde bireylerin farklı veya eksik yönleri üzerinden oluşum yayan nefret söylemlerindeki tüm kavramlar, sosyal medyalar aracılığıyla ifade özgürlüğü çerçevesinde diğer kullanıcılara ulaştırılmaktadır. Bu durumda, sosyal medya kullanıcılarının haysiyetlerine saygı duyulacak şekilde, sosyal medya araçları kullanılmalı ve bu tür nefret söylemi ifadelerinin de engellenmesi gerekmektedir.

Sosyal medyada nefret söylemi tartışmalarında çözümü en zor problemlerden birisi de ifade özgürlüğü kavramıdır. Sosyal medya kullanıcıları tarafından “ifade özgürlüğü” [141] olarak adlandırılan nefret söylemlerinin tümü, bireylerin ve toplumun temel ve özel değerlerini aşarak, bireylerin olumlu gelişimi ve toplumun ilerleyişi önünde büyük bir engeldir. Geleneksel ve sosyal medyanın temel amacı, insanlar arasındaki iletişimi sağlamaktır. İletişimin ana gayesi unutulurken, görsel/işitsel/yazınsal olumsuz öğelerle desteklenen sosyal medya ortamları, birçok insanı olumsuz şekilde etkileyerek bireylerin günlük yaşantısını ve sosyal medya üzerinden iletişim hakkını da engellemektedir. Genelde bireylerin farklı veya eksik yönleri üzerinden oluşum yayan nefret söylemlerindeki tüm kavramlar, sosyal medyalar aracılığıyla ifade özgürlüğü çerçevesinde diğer kullanıcılara ulaştırılmaktadır. Bu durumda, sosyal medya kullanıcılarının haysiyetlerine saygı duyulacak şekilde, sosyal medya araçları kullanılmalı ve bu tür nefret söylemi ifadelerinin de engellenmesi gerekmektedir.

Sonuç olarak, sosyal medya ile nefret söylemi arasında önemli bir ilişki olduğu görülmektedir. Bu ilişki, teknolojinin gelişimiyle daha da belirgin hale gelmiştir. Geçmişte geleneksel medya aracılığıyla gerçekleştirilen iletişim şekli, günümüzde yerini sosyal medya araçlarına bırakmıştır. Sosyal medyanın iletişimin temel yapısını oluşturduğu söylenebilir. Bu durum, beraberinde nefret söylemlerini de aklı getirmektedir. Pek çok insan, artık günümüzde düşüncelerini olumlu veya olumsuz ayırt etmeksizin sosyal medya araçları vasıtasıyla paylaşmaktadır. Bu paylaşımların genelde ifade özgürlüğü olarak nefret söylemi ifadelerini içerdiği görülmüştür. Halbuki sosyal medya ortamları ifade özgürlüğü adı altında nefret söylemlerinin yayıldığı bir yer değil, bir iletişim şekli ve kullanım aracı olarak amacına uygun kullanılması gerekir. Geçmişte olduğu gibi gelecekte de toplumun olduğu her yerde iletişim kavramının olacağı söylenebilir. Teknolojinin gelişimine bağlı olarak sosyal medya araçları iletişimi hızlandırırsa da nefret söylemlerini azaltmamaktadır; aksine sosyal medyanın farklı bireyleri/grupları/toplulukları bir araya getirmesi, nefret söylemi üretimini sürekli artıracaktır. Bu sebeple, sosyal medya araçlarının tümü kutuplaştırıcı gönderilerle viral içerikler oluşturmaya devam ettiği müddetçe, nefret söylemi kavramlarında üretimin devam edeceği ve gelecekte de hızla yayılma riski taşıyacağı söylenebilir.

#### Teşekkür (Acknowledgement)

Bu çalışma, Türkiye Bilimsel ve Teknolojik Araştırma Kurumu (TÜBİTAK) tarafından 120E187 numaralı hibe ile kısmen desteklenmiştir. Yazarlar destek olan TÜBİTAK'a teşekkürlerini sunar.

#### Kaynaklar (References)

1. Türk Dil Kurumu Sözlükleri. Güncel Türkçe Sözlükte Ara: Nefret. Erişim adresi: <https://sozluk.gov.tr>. Erişim tarihi: Mayıs 15, 2023.
2. Nockleby, J.T., Encyclopedia of the American Constitution. Hate Speech. volume 1. 2nd ed., edited by Leonard W. Levy, Kenneth L. Karst et al. New York: Macmillan, New York: Macmillan. 2000.
3. Kandır, M.O., Sosyal Medyanın Bilişim Suçlarına Etkisi, Eskişehir Barosu Dergisi, 8 (1), 150-169, 2023.
4. Öztekin, H., Yeni Medyada Nefret Söylemi: Ekşi Sözlük Örneği, Journal of International Social Research, 8 (38), 925-936, 2015.
5. Dilbirliği, M., Din ve İnanç Temelinde Nefret Söylemi: “İslamofobi”, Nefret Söylemi ve Nefret Suçları Sempozyumu, Ankara - Türkiye, 40-65, 27 Ekim, 2021.
6. Weber, A., Manual on hate speech. Council of Europe Publishing, F-67075 Strazburg Cedex, Çeviri: Metin Çulhaoğlu, Düzelti: Naci Temeltaş, Bişeng Özdiç, 2019.
7. Adalet Bakanlığı. Nefret Söylemi, Avrupa Konseyi Bakanlar Komitesi'nin Üye Devletlere R 97(20) sayılı Tavsiye Karar. Erişim adresi: <https://inhak.adalet.gov.tr/Resimler/Dokuman/10122019130702nefret.pdf> Yayın tarihi: Ekim 30, 1997. Erişim tarihi: Mayıs 15, 2023.
8. Yıldırım, S., Ayrımcılık Temelli Dışlanma: Türkiye’de Çalışma Hayatında Kadınlar, SGD-Sosyal Güvenlik Dergisi, 2, 321-346, 2021.
9. Erdem, A.R., Ayrımcılığın Kavramsal Analizi, International Journal of Humanities and Art Research, 4 (4), 12-20, 2020.
10. Dirini, İ., Okur Yorumlarıyla Yeniden Üretilen Nefret Söylemi, Tuğrul Çomu (der.), Yeni Medyada Nefret Söylemi, İstanbul: Kalkedon Yayınlar, 2010.
11. Vardal, Z.B., Nefret Söylemi ve Yeni Medya, Maltepe Üniversitesi İletişim Fakültesi Dergisi, 2 (1), 132-156, 2016.
12. Bilge, R., Sosyal Medyada Nefret Söyleminin İnşası ve Nefret Suçlarına İlişkin Yasal Düzenlemeler, Hakemli, Akademik, E-Dergi Gazi Üniversitesi İletişim Fakültesi, 1, 2016, Güz, 2016.
13. İlhan, R.S. ve Çevik, A., Önyargıların Psikolojisi: Psikodinamik Bir Gözden Geçirme, 1 (1), 52-67, 2013.
14. Suveren Y., Bilinçsiz Önyargı: Tanımı ve Önemi. Current Approaches in Psychiatry, 14 (3), 414-426, 2022.
15. Eniç, D., COVID-19 Pandemisinde Önyargı ve Ayrımcılık, İnsan ve Toplum Bilimleri Araştırmaları Dergisi, 10 (3), 2694-2716, 2021.
16. Cansız, U. ve Bolat, H., Yabancı Dil Olarak Türkçe ve Almanca Öğretimi Ders Kitaplarında Kalıp Yargılar Bağlamında Türk İmgesi, 21. Yüzyılda Eğitim Ve Toplum Eğitim Bilimleri Ve Sosyal Araştırmalar Dergisi, 11 (33), 527 - 558, 2022.
17. Yaşar, H. ve Öksüz, O., Nefret Söyleminin İnşasında Sosyal Medyanın Rolü: Ekşi Sözlük Örneği, Erciyes İletişim Dergisi, 7 (2), 1383-1408, 2020.
18. Ünlü, D.G., Kişilerarası İletişim Sürecinde Bireyin Fiziksel Görünüşüne Dayanan Toplumsal Cinsiyet Kalıp Yargılarını Belirlemek, Global Media Journal TR Edition, 11 (22), 2021.
19. Karaca, E., Çokkültürlülük Çerçevesinde Avrupa Birliği'nin Değişen Yüzü, Yüksek Lisans Tezi, İstanbul Gelişim Üniversitesi, Sosyal Bilimler Enstitüsü, İstanbul, Türkiye, 2018.
20. Gören, K. B. B., Çokkültürlülük Bağlamında Türkiye'nin Göç Politikası, Güvenlik Çalışmaları Dergisi, 21 (1), 3-19, 2019.
21. Aslan, A. ve Öney Doğan, B., Çevrimiçi Şiddet: Bir Siber Zorbalık Alanı Olarak “Potinss” Örneği, Marmara İletişim Dergisi, 27, 95-119, 2017.
22. Yüctürk, C. ve Ağın, B., Sporcuların Perspektifinden Sosyal Medyada Siber Zorbalık, TRT Akademi, 7 (14), 154-181, 2022.
23. Karaboğa, D. ve Işık, M., Dijital Şiddet ve Nefret Söylemi. Spor, Eğitim ve Çocuk, 2 (1), 1-26, 2022.
24. Atatamır, H., İş Yaşamında Siber Zorbalık, Anasay, 22, 459 - 486, 2022.
25. Akgün, İ., Teknolojik Yıldırma Yahut Siber Mobbing, Erişim adresi: <http://mobbingderneği.org.tr/teknolojik-yildirma-yahut-siber-mobbing/>. Erişim tarihi: Kasım 20, 2023.
26. Modha, S., Majumder, P., Madl, T. and Mandalia, C., Detecting and visualizing hate speech in social media: A cyber Watchdog for surveillance, Expert System with Applications, 161, 2020.
27. Mandl, T., Modha, S., Shahi, G.K., Madhu, H., Satapara, S. and et al., Overview of the HASOC Subtrack at FIRE 2021: Hate Speech and Offensive Content Identification in English and Indo-Aryan Languages, Computation and Language, 2021.
28. Bohra, A., Vijay, D., Singh, V., Akhtar, S.S. and Shrivastava, M., A Dataset of Hindi-English Code-Mixed Social Media Text for Hate Speech Detection, In Proceedings of the Second Workshop on Computational Modeling of People's Opinions, Personality, and Emotions in Social Media, pp.36-41, New Orleans, Louisiana, USA, 2018.
29. Santosh, T.Y.S.S. and Aravind, K.V.S., Hate Speech Detection in Hindi-English Code-Mixed Social Media Text, CODS-COMAD '19: Proceedings of the ACM India Joint International Conference on Data Science and Management of Data, 3-5 January 2019, 310-313, Kolkata, India, 2018.



30. Corazza, M., Meninin, S., Cabrio, E., Tonelli, S. and Villata S., A Multilingual Evaluation for Online Hate Speech Detection. *ACM Transactions on Internet Technology*, 20 (2), 1–22, 2020.
31. Ishmam, A. M., and Sharmin, S., Hateful speech detection in public facebook pages for the bengali language, In 2019 18th IEEE international conference on machine learning and applications (ICMLA), IEEE, 555-560, 2019.
32. Hettiarachchi, N., Weerasinghe, R. and Pushpanda R., Detecting Hate Speech in Social Media Articles in Romanized Sinhala, 20th International Conference on Advances in ICT for Emerging Regions (ICTer), Colombo, Sri Lanka, 04-07 November 2020.
33. Roy, P.K., Bhawal, S. and Subalalitha, C.N., Hate speech and offensive language detection in Dravidian languages using deep ensemble framework, *Computer Speech & Language*, 75, 2022.
34. Ghosh, T., Chowdhury, A.A.K., Banna, M.H.A., Nahian, M.J.A., Kaiser, M.S. and et al., A Hybrid Deep Learning Approach to Detect Bangla Social Media Hate Speech, *Proceedings of International Conference on Fourth Industrial Revolution and Beyond 2021*, 711-722, 04 October 2022.
35. Al-Hassan, A. and Al-Dossari, H., Detection of hate speech in Arabic tweets using deep learning. *Multimedia Systems*, 28, 1963–1974, 2022.
36. Aldjanabi, W., Dahou, A., Al-qaness, M.A.A., Elaziz, M.A., Helmi, M.A. and et al., Arabic Offensive and Hate Speech Detection Using a Cross-Corpora Multi-Task Learning Model, *Informatics*, 8 (4), 69, 2021.
37. Sultan, D., Mussiraliyeva, S., Toktarova, A., Nurtas, M., Iztayev, Z. and et al., Cyberbullying and Hate Speech Detection on Kazakh-Language Social Networks, 7th IEEE Intl Conference on Big Data Security on Cloud (BigDataSecurity), IEEE Intl Conference on High Performance and Smart Computing, (HPSC) and IEEE Intl Conference on Intelligent Data and Security (IDS), NY, USA, 15-17 May 2021.
38. Bilal, M., Khan, A., Jan, S. and Musa, S., Context-Aware Deep Learning Model for Detection of Roman Urdu Hate Speech on Social Media Platform, *IEEE*, 10, 121133-121151, 21 October 2022.
39. Febriana, T. and Budiarto, A., Twitter Dataset for Hate Speech and Cyberbullying Detection in Indonesian Language, 2019 International Conference on Information Management and Technology (ICIMTech), Jakarta/Bali, Indonesia, 19-20 August 2019.
40. Mayda, İ., Diri, B. ve Yıldız, T., Türkçe Tweetler üzerinde Makine öğrenimi ile Nefret Söylemi Tespiti, *Özel Sayı 24*, 328-334, 2021.
41. Alpaydin, E., Introduction to machine learning, MIT press, 2020.
42. Yazgılı, E. & Baykara, M., Türkçe metinlerde makine öğrenmesi yöntemleri ile siber zorbalık tespiti, *Gümüşhane Üniversitesi Fen Bilimleri Dergisi*, 12 (2), 443-453, 2022.
43. Vishwanathan, S. V. M., and Murty, M. N., SSVN: a simple SVM algorithm, In *Proceedings of the 2002 International Joint Conference on Neural Networks, IJCNN'02 (Cat. No. 02CH37290)*, 3, 2393-2398, May 2002.
44. Bounsaythip, C. and Rinta-Runsala, E., Overview of data mining for customer behavior modeling, *VTT Information Technology Research Report*, Version, 1, 1-53, 2001.
45. Taşkın, Ç. ve Emel, G.G., Veri madenciliğinde karar ağaçları ve bir satış analizi uygulaması, *Eskişehir Osmangazi Üniversitesi Sosyal Bilimler Dergisi*, 6 (2), 221-239, 2005.
46. Beken, M., Eyecioglu, O., and Kurt, N., Prediction of Oil Consumption and Oil Access of Countries in The European Union Region with Machine Learning, *International Journal of Smart Grid-ijSmartGrid*, 6 (3), 79-83, 2022.
47. Bai, B., Tan, Y., Donchyts, G., Haag, A., Xu, B., Chen, G., and et al., Naive Bayes classification-based surface water gap-filling from partially contaminated optical remote sensing image, *Journal of Hydrology*, 616, 128791, 2023.
48. Yegnanarayana, B., *Artificial neural networks*, PHI Learning Pvt. Ltd., 2009.
49. Li, Z., Liu, F., Yang, W., Peng, S., and Zhou, J., A survey of convolutional neural networks: analysis, applications, and prospects, *IEEE transactions on neural networks and learning systems*, 2021.
50. Hochreiter, S., and Schmidhuber, J., Long short-term memory, *Neural computation*, 9 (8), 1735-1780, 1997.
51. Gillioz, A., Casas, J., Mugellini, E., and Abou Khaled, O., Overview of the Transformer-based Models for NLP Tasks, In 2020 15th Conference on Computer Science and Information Systems (FedCSIS), 179-183, 2020.
52. LeCun, Y., Bengio, Y. and Hinton, G., Deep learning. *nature*, 521 (7553), 436-444, 2015.
53. Kat B., Natural language processing for the Turkish Academic texts in the engineering field and development of a decision support system: the case of TUBITAK project proposals, *Journal of the Faculty of Engineering and Architecture of Gazi University*, 38 (3), 1879-1892, 2023.
54. Y. Goldberg, O. Levy., Word2vec Explained: deriving Mikolov et al.'s negative-sampling word-embedding method, 2014.
55. T. Mikolov, I. Sutskever, K. Chen, G. Corrado, J. Dean., Distributed Representations of Words and Phrases and their Compositionality, 26th International Conference on Neural Information Processing Systems, 3111-3119, Nevada, A.B.D., 2013.
56. Aydoğan, M. ve Karıcı, A., Kelime Temsil Yöntemleri ile Kelime Benzerliklerinin İncelenmesi. *Çukurova Üniversitesi Mühendislik Mimarlık Fakültesi Dergisi*, 34 (2), 181-195, 2019.
57. Pennington, J., Socher, R. and Manning, C.D., GloVe: Global Vectors for Word Representation, *Empirical Methods in Natural Language Processing (EMNLP)*, 1532-1543, 2008.
58. Gezici, G. and Yanıkoğlu, B., Sentiment Analysis in Turkish. In: Oflazer, K., Saraçlar, M. (eds) *Turkish Natural Language Processing, Theory and Applications of Natural Language Processing book series (NLP)*, 255–271, Springer, Cham, 2018.
59. Delibaş, A., Doğal dil işleme ile Türkçe yazım hatalarının denetlenmesi, *Yüksek Lisans Tezi*, İstanbul, 2008.
60. Li, J., Sun, A., Han, J. and Li, C., A Survey on Deep Learning for Named Entity Recognition, in *IEEE Transactions on Knowledge and Data Engineering*, 34 (1), 50-70, 1 Jan. 2022.
61. Jia, Y. and Xu, X., Chinese Named Entity Recognition Based on CNN-BiLSTM-CRF, 2018 IEEE 9th International Conference on Software Engineering and Service Science (ICSESS), Beijing, China, 1-4, 2018.
62. Wu, F., Liu, J., Wu, C., Huang, Y. and Xie, X., Neural Chinese Named Entity Recognition via CNN-LSTM-CRF and Joint Training with Word Segmentation, *WWW '19: The World Wide Web Conference*, 3342–3348, 2019.
63. Zhu, Y., Wang, G. and Karlsson, B.F., CAN-NER: Convolutional Attention Network for Chinese Named Entity Recognition, *Computation and Language*, 2020.
64. An, Y., Xia, X., Chen, X., Wu, F.X. and Wang, J., Chinese clinical named entity recognition via multi-head self-attention based BiLSTM-CRF, *Artificial Intelligence in Medicine*, 127, 2022.
65. Yadav, H., Ghosh, S., Yu, Y., and Shah, R.R., End-to-end Named Entity Recognition from English Speech, *Computation and Language*, 2020.
66. Yan, R., Jiang, X. and Dang, D., Named Entity Recognition by Using XLNet-BiLSTM-CRF. *Neural Processing Letters*, 53, 3339–3356, 2021.
67. Affi, M. and Latiri, C., BE-BLC: BERT-ELMO-Based Deep Neural Network Architecture for English Named Entity Recognition Task, *Procedia Computer Science*, 192, 168-181, 2021.
68. Sun, P., Yang, X., Zhao, X. and Wang, Z., An Overview of Named Entity Recognition, 2018 International Conference on Asian Language Processing (IALP), Bandung, Indonesia, 273-278, 2018.
69. S. Kul and A. Sayar., Entity Name Recognition in Job Postings and Resumes, 2021 3rd International Congress on Human-Computer Interaction, Optimization and Robotic Applications (HORA), Ankara, Turkey, 1-6, 2021.
70. Camcı, H. ve Eryiğit, G., Türkçe Zamansal İfadelerin Yakalanması ve Tanımlanması. *Bilişim Teknolojileri Dergisi*, 14 (3), 337-343, 2021.
71. Nasiboğlu, R. ve Gençler, M., Adlandırılmış Varlık Tanıma Modelleri ile Türkçe Sosyal Medya Metinlerinde Küfurlü Sözlerin Sansürlenmesi, *Afyon Kocatepe Üniversitesi Fen ve Mühendislik Bilimleri Dergisi*, 23 (1), 72-88, 2023.
72. Yiğit, G. ve Amasyalı, F., Soru Cevaplama Sistemleri Üzerine Detaylı Bir Çalışma: Veri Kümeleri, Yöntemler ve Açık Araştırma Alanları, *Bilgi Teknolojileri Dergisi*, 14 (3), 239 - 254, 2021.
73. Guven, Z. A., and Unalir, M. O., Improving the BERT Model with proposed named entity recognition method for Question Answering, In 2021 6th International Conference on Computer Science and Engineering (UBMK), 204-208, 2021.
74. Doğan, S. ve Diri, B., Türkçe Dokümanlar İçin N-gram Tabanlı Yeni Bir Sınıflandırma (Ng-ind): Yazar, Tür ve Cinsiyet. *Türkiye Bilişim Vakfı Bilgisayar Bilimleri ve Mühendisliği Dergisi*, 3 (1), 11-19, 2016.

75. Tantuğ, A.C., Metin Sınıflandırma. Türkiye Bilişim Vakfı Bilgisayar Bilimleri ve Mühendisliği Dergisi, 5 (2), 2016.
76. Sokolova, M., Japkowicz, N., Szpakowicz, S., Beyond Accuracy, F-Score and ROC: A Family of Discriminant Measures for Performance Evaluation. In: Sattar, A., Kang, Bh. (eds) AI 2006: Advances in Artificial Intelligence. AI 2006. Lecture Notes in Computer Science(), vol 4304. Springer, Berlin, Heidelberg, 2006.
77. Goutte, C., Gaussier, E., A Probabilistic Interpretation of Precision, Recall and F-Score, with Implication for Evaluation. In: Losada, D.E., Fernández-Luna, J.M. (eds) Advances in Information Retrieval. ECIR 2005. Lecture Notes in Computer Science, vol 3408. Springer, Berlin, Heidelberg, 2005.
78. Baydoğan, V.C. ve Alataş, B., Çevrimiçi Sosyal Ağlarda Nefret Söylemi Tespiti İçin Yapay Zekâ Temelli Algoritmaların Performans Değerlendirmesi, Fırat Üniversitesi Mühendislik Bilimleri Dergisi, 33 (2), 745-754, 2021.
79. Kaggle. Dynamically Generated Hate Speech Dataset. Erişim adresi: <https://www.kaggle.com/datasets/usharengaraju/dynamically-generated-hate-speech-dataset> . Erişim tarihi: Ocak 02, 2024.
80. Github. snowballstem/snowball. Erişim adresi: <https://github.com/snowballstem/snowball>. Yayın tarihi: Kasım 09, 2023. Erişim tarihi: Ocak 02, 2024.
81. Eğin, F. ve Bulut, V., Mültecilere Yönelik Nefret Söyleminin Tespitinde Makine Öğrenmesi Modellerinin Kullanılması, European Journal of Science and Technology, Özel sayı:48, 19-22, 2023.
82. The Python Package Index (PyPI). SNScrape. Erişim adresi: <https://pypi.org/project/snscrape/> . Yayın tarihi: Haziran 22, 2023. Erişim tarihi: Ocak 02, 2024.
83. Yılmaz, Ş. Ş., İlyas, Ö., ve Gökçen, H., Twitter Platformundan Elde Edilen Türkçe Saldırgan Dil Derlemi, Mühendislik Bilimleri ve Araştırmaları Dergisi, 4 (2), 304-316, 2022.
84. Mikolov, T., Chen, K., Corrado, G., and Dean, J., Efficient estimation of word representations in vector space, arXiv preprint arXiv:1301.3781, 2013.
85. Mikolov, T., Sutskever, I., Chen, K., Corrado, G. S., and Dean, J., Distributed representations of words and phrases and their compositionality, In Advances in neural information processing systems, 3111-3119, 2013.
86. Google Code. Word2Vec. Erişim adresi: <https://code.google.com/p/word2vec/>. Erişim tarihi: Ocak 02, 2024.
87. Tantuğ, A.C. and Güler, G., Comparison of Turkish Word Representations Trained on Different Morphological Forms, Computation and Language, arXiv:2002.05417, 2020.
88. Atatanır, H., Çalışma yaşamında bir ayrımcılık türü olarak mobbing ve siber boyutu, İstanbul Ticaret Üniversitesi Sosyal Bilimler Dergisi, 20 (42), 1621-1639, 2021.
89. Yazgılı, E. ve Baykara, M., Siber Zorbalık Tespit Yöntemleri Potansiyel Uygulama Alanları ve Zorluklar, DÜMF Mühendislik Dergisi, 12 (1), 23-35, 2021.
90. Altay, E. V., and Alataş, B., Detection of cyberbullying in social networks using machine learning methods, In 2018 International Congress on Big Data, Deep Learning and Fighting Cyber Terrorism (IBIGDELFT), 87-91, December, 2018.
91. Zois, D. S., Kapodistria, A., Yao, M., and Chelms, C., Optimal online cyberbullying detection, In 2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2017-2021, April 2018.
92. Nesiboğlu, R. ve Gencer, M., Adlandırılmış Varlık Tanıma Modelleri ile Türkçe Sosyal Medya Metinlerinde Küfürlü Sözlerin Sansürlenmesi, Afyon Kocatepe Üniversitesi Fen ve Mühendislik Bilimleri Dergisi, 23 (1), 72 - 88, 2023.
93. Natural Language Toolkit (NLTK). Extracting Information from Text. Erişim adresi: <https://www.nltk.org/book/ch07.html> . Erişim tarihi: Ocak 02, 2024.
94. Çöltekin, Ç., A Corpus of Turkish Offensive Language of Social Media, In Proceedings of the Twelfth Language Resources and Evaluation Conference, 6174-6184, Marseille, France. European Language Resources Association, 2020.
95. Github. A corpus of Turkish offensive language. Erişim adresi: <https://coltekin.github.io/offensive-turkish/guidelines-tr.html> . Erişim tarihi: Ocak 02, 2024.
96. Hugging Face. Datasets: offenseval2020\_tr. Erişim adresi: [https://huggingface.co/datasets/offenseval2020\\_tr](https://huggingface.co/datasets/offenseval2020_tr) . Erişim tarihi: Ocak 02, 2024.
97. C. Baydoğan and B. Alataş., Metaheuristic Ant Lion and Moth Flame Optimization-Based Novel Approach for Automatic Detection of Hate Speech in Online Social Networks, in IEEE Access, 9, 110047-110062, 03 August 2021.
98. Github. t-davidson/hate-speech-and-offensive-language. Erişim adresi: <https://github.com/t-davidson/hate-speech-and-offensive-language> . Erişim tarihi: Ocak 02, 2024.
99. Kaggle. Detecting Insults in Social Commentary. Erişim adresi: <https://www.kaggle.com/c/detecting-insults-in-social-commentary/data> . Erişim tarihi: Ocak 02, 2024.
100. CodaLab. Multilingual detection of hate speech against immigrants and women in Twitter (hatEval). Erişim adresi: <https://competitions.codalab.org/competitions/19935> . Erişim tarihi: Ocak 02, 2024.
101. Çoban, Ö., Özel, S.A. ve İnan, A. Detection and Cross-domain Evaluation of Cyberbullying in Facebook Activity Contents for Turkish, ACM Transactions on Asian and Low-Resource Language Information Processing, 22 (4), 1-32, 2023.
102. Bayrak, N.G., Sanal Platformlarda Çocukları Bekleyen Tehlike: "Çevrimiçi Cinsel İstismar", Doğu Karadeniz Sağlık Bilimleri Dergisi, 1 (3), 48-57, 2022.
103. Bülbül, M. N., Sosyal medyada imaj fetişizmi: Bauman bağlamında imgelerin tüketimi, Yüksek Lisans Tezi, Pamukkale Üniversitesi, Sosyal Bilimler Enstitüsü, 2021.
104. Ayçiçek, M., Sosyal medya ve sosyal medya fenomenlerinin homofobiye etkisi, Yüksek Lisans Tezi, İstanbul Gelişim Üniversitesi, Lisansüstü Eğitim Enstitüsü, 2022.
105. Acay, F., Sosyal Medya Aracılığıyla Hakaret Suçu ve Suçun Tespitine İlişkin Uygulamalar, İstanbul Aydın Üniversitesi Hukuk Fakültesi Dergisi, 7 (1), 71-140, 2021.
106. Yıldırım, Z.M., Ayrımcılık ve Şiddetin Toplumsallaşması: Sosyal Adalet ve Toplumsal Barın, Çalışma ve Toplum, 3 (70), 1813-1838, 2021.
107. Özdemir, G., Twitter'da Gerçekleştirilen Saldırgan Mizah: Troller Üzerine Sosyolojik Bir Araştırma, Doktora Tezi, Anadolu Üniversitesi, 2021.
108. Özgür, H.O. ve Özkul, M., Risk Toplumunda Kuşakların Yeni Şiddet Deneyimi: Dijital Şiddet, Electronic Turkish Studies, 17 (3):551-568, 2022.
109. Kuş, O., Metaverse ve dijital nefret söylemi: toksik içeriğin potansiyel yayılım örüntüleri ve proaktif çözüm önerileri hakkında bir tartışma, Hakemli, Uluslararası, E-Dergi Ankara Hacı Bayram Veli Üniversitesi İletişim Fakültesi, 12, 355 - 368, 2022.
110. Bostancı, M., & Gonca, U., Metaverse: Sanal Mı Gerçek Mi?, 2022.
111. MatthewBall.co. Framework for the Metaverse. Erişim adresi: <https://www.matthewball.co/all/forwardtothemetaverseprimer> Yayın tarihi: Haziran 29, 2021. Erişim tarihi: Ocak 02, 2024.
112. Çeçen, A.F., İslamofobi Bağlamında Kodlanmış ve Açık Ayrımcı Dile Bakmak: İslamofobik Nefret Söylemi Açısından Bir Kategorizasyon Çabası, TRT Akademi, 7 (15), 408-433, 2022.
113. Abuşoğlu, H.Ö. ve Aydoğan, B.B., Sosyal Medyada Z Kuşağına Yönelik Nefret Söylemi: Ekşi Sözlük Örneği. ODÜ Sosyal Bilimler Araştırma Dergisi, 13 (1), 707 - 728, 2023.
114. Doğan, E., Suça Sürüklenen Çocuklara Yönelik Medya Pratikleri ve Nefret Söylemi: Video Haberler ve Kullanıcı Yorumları Örneği, TRT Akademi, 7 (15), 676-701, 2022.
115. Kobaza, O.S., Sosyal Medyada Duyarlılığın Yitimi: TikTok Örneği. Socrates Journal of Interdisciplinary Social Studies, 18, 82-99, 2022.
116. Dinç, E.S., Sosyal Medya Ortamlarında Siber Zorbalık: Lise Öğrencilerinin Siber Zorbalık Deneyimlerinin İncelenmesi. Yeni Medya Elektronik Dergi - eJNM, 4 (1), 24-39, 2020.
117. Anamur, D.D. ve Topsakal, T., Herkesin Herkesi Gözetlediği Toplum: Siberuzamda Altın Gözetimin Farklı İnşası Olarak Siber Zorbalık, İnönü Üniversitesi İletişim Fakültesi Elektronik Dergisi (İnif E-Dergi), 4 (2), 37-58, 2019.
118. Öztürk, F.E., Salgınla Yayılan Nefret: Twitter'da 65 Yaş Üstü Bireylere Yönelik Nefret Söylemi, Ege Üniversitesi İletişim Fakültesi Yeni Düşünceler Hakemli E-Dergisi, 15, 143-159, 2021.
119. Karamete, F., Twitter'da Suriyeli Mültecilere Yönelik Nefret Söylemi. Yüksek lisans Tezi, Anadolu Üniversitesi Sosyal Bilimler Enstitüsü, Eskişehir, Türkiye, 2019.
120. Bahçecioğlu, E.H., Sosyal Medyada Kovid-19 Salgını Sürecinde Sağlık Çalışanlarına Yönelik Önce Destek Sonra Nefret Söylemi, TRT Akademi, 7 (15), 498 - 531, 2022.

121. Kurt, İ., Nefret Söylemi ve Ötekiler: Türkiye’de Suriyeli Kadınlar, Yüksek Lisans Tezi, Pamukkale Üniversitesi, Denizli, Türkiye, 2023.
122. Karataş, Z., Sosyal medyada yaygınlaşan linç kültürünün Twitter mecrası özelinde incelenmesi ve linç atlası uygulaması, Yüksek Lisans Tezi, Sakarya Üniversitesi, Sakarya, Türkiye, 2022.
123. Merdin, M., Dijital Oyun Reklamlarında Oyuncular Tarafından Üretilen Nefret Söylemleri. Selçuk İletişim Dergisi, 16 (1), 34-62, 2023.
124. Çanakçı, M., Şaşmazlar, C. ve Öztürk, S., Afet ve Kriz Yönetiminde Sosyal Medyanın Kullanımı Üzerine Bir Araştırma: Twitter Örneği, Gümüşhane Üniversitesi Sağlık Bilimleri Dergisi, 11 (3), 882 - 897, 2022.
125. Kuş, O. ve Öztürk İ.D., Çin’de Üretilen Aşılara Yönelik YouTube Tartışmaları Çerçevesinde Yanlış Bilgi ve Nefret Söylemi İlişkisi Üzerine Bir İnceleme, Türkiye İletişim Araştırmaları Dergisi, 38, 96-117, 2021.
126. Sallan Gül, S. ve Kâhya Nizam, Ö., Sosyal Bilimlerde İçerik ve Söylem Analizi, Pamukkale Üniversitesi Sosyal Bilimler Enstitüsü Dergisi, 42(özel sayı:1), Ö181-Ö198, 2021.
127. Bilgili, H.A. ve İnceli, P., Yeni Medyada Cinsel Kimlik Temelli Nefret Söyleminin Üretimi ve Yayılımı: YouTube Örneği, Global Media Journal TR Edition, 12 (23), 28-57, 2021.
128. Tekin, F., Türay, A., ve Yıldırım, M., Türkiye’de Din, İnanç veya İnançsızlık Temelli Nefret Suçları 2021 Raporu. 2022.
129. Derneği, İ. H., Türkiye’de Nefret Suçları ve Son Dönemde Yaşanan İrkçi Saldırıları Özel Raporu, 2020.
130. Sağlam, H., Facebook’ta Etnik Temelli Nefret Söylemi: BBC News’ in Dağılık Karabağ Çatışması Haberlerindeki Kullanıcı Yorumları Üzerine Bir İnceleme, Kastamonu İletişim Araştırmaları Dergisi, 9, 154-178, 2022.
131. Kaygısız, E.G., Sosyal medyada mobbing: Youtube ve Twitter kullanıcı yorumları analizi. BMIJ, 11 (1), 226-242, 2023.
132. Cosentino, V., Izquierdo, J.L.C. and Cabot, J., A systematic mapping study of software development with GitHub, Ieee access, 5, 7173-7192, 2017.
133. Quaranta, L., Calefato, F. and Lanubile, F., Kgtorrent: A dataset of python jupyter notebooks from kaggle, In 2021 IEEE/ACM 18th International Conference on Mining Software Repositories (MSR), 550-554, 2021.
134. Kwak, H., Lee, C., Park, H., and Moon, S., What is Twitter, a social network or a news media?, In Proceedings of the 19th international conference on World wide web, 591-600, 2010.
135. Akın, M. D. and Akın, A.A., Türk dilleri için açık kaynaklı doğal dil işleme kütüphanesi: ZEMBEREK. Elektrik mühendisliği, 431, 38-44, 2007.
136. Amasyalı, M.F., Davletov, F., Torayew, A. I. and Çiftçi, Ü., Text2arff: Automatic feature extraction software for Turkish texts, In 2010 IEEE 18th Signal Processing and Communications Applications Conference, 629-632, 2010.
137. Shao, Y., Taylor, S., Marshall, N., Morioka, C., and Zeng-Treitler, Q., Clinical text classification with word embedding features vs. bag-of-words features, In 2018 IEEE International conference on big data (big data), 2874-2878, 2018.
138. Karaca, Y., Özsalmanlı, A.Y., Kamu Yönetiminde Açık Veri Yönetimi ve Şeffaflık: ABD ve İngiltere Uygulamaları, Aksaray Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi, 14 (1), 121-140, 2022.
139. Yakar, Ö., Aşhyan, R., Saklı Markov Modeli Kullanarak Türkçe Konuşma Tanıma, Akademik Bilişim, 16, 1-7, 2016.
140. Vardal, Z.B., Nefret Söylemi ve Yeni Medya, Maltepe Üniversitesi İletişim Fakültesi Dergisi, 2 (1), 132-156, 2015.
141. Beyhan, T., İnternetin ifade özgürlüğünün geleceğine etkileri: Türkiye incelemesi, Salgın Sonrası Dönemde İnsan Hakları Gündemi, TİHV Akademi, 35-44, 20-23, Ankara, Türkiye, Mayıs 2021.

