

**TÜRKÇE İÇİN SAHTE HABER TESPİT MODELİNİN
OLUŞTURULMASI**

**A FAKE NEWS DETECTION MODEL FOR TURKISH
LANGUAGE**

UĞUR MERTOĞLU

DR. ÖĞR. ÜYESİ BURKAY GENÇ

Tez Danışmanı

PROF. DR. HAYRİ SEVER

Eş Danışman

Hacettepe Üniversitesi

Lisansüstü Eğitim-Öğretim ve Sınav Yönetmeliğinin
Bilgisayar Mühendisliği Anabilim Dalı İçin Öngördüğü

DOKTORA TEZİ olarak hazırlanmıştır.

2020

Sevgili Eşim, Oğullarım ve Aileme...

ÖZET

TÜRKÇE İÇİN SAHTE HABER TESPİT MODELİNİN OLUŞTURULMASI

Uğur MERTOĞLU

Doktora, Bilgisayar Mühendisliği Bölümü

Tez Danışmanı: Dr. Öğr. Üyesi Burkay GENÇ

Eş Danışman: Prof. Dr. Hayri SEVER

Eylül 2020, 141 sayfa

Basılı medyanın çevrim içi ortamlarda faaliyet gösteren medyaya dönüşmesi, internetin ve sosyal medyanın haber kaynağı olarak giderek artan yoğunlukta kullanımı, medya okuryazarlığı kavramını ve insanların haber edinme alışkanlıklarını çarpıcı şekilde değiştirmiştir. Bu doğrultuda, bireyler ve organizasyonlar haber edinmek için çevrim içi haber sitelerinden ve sosyal medya platformlarından yoğun olarak yararlanmaya başlamıştır. Bu daha hızlı, daha kolay ve nispeten daha ucuz olan fırsat; insanların bilgiye erişimi açısından kolaylık sağlarken, haberlerin potansiyel etkileri düşünüldüğünde önemli problemleri de beraberinde getirmektedir. Daha önce görülmemiş bir hızda, devasa boyut ve çeşitlilikte yayılan haberleri, doğrulamak veya bilgiye dönüşebilecek okumaya değer bir veri olduğunu saptamak; içinde bulunduğumuz çağda insan gücü ile üstesinden gelinemeyecek bir problemi doğurmuştur. İnsanların içerik yönünden muğlak veya sahte bir haberi gerçek bir habere oranla daha hızlı yayma eğilimi, problemi daha da zor bir hale sokmaktadır.

Sahte haberlerden kaynaklı gelişmelerin etkisiyle oluşan olumsuz durumlar arttıkça, değişik amaçlarla bilinçli veya bilinçsiz şekilde vuku bulan bu asimetrik tehdide karşı sosyal medya platformları, ticari organizasyonlar, kurumlar ve hatta devletler kendi önlemlerini almaya başlamışlardır. Bu durum araştırmacıları; çoğunlukla metin türündeki bu haberleri analiz etmeye, bilimsel altyapılar oluşturmaya, sahte haber tespitini amaçlayan akıllı sistemler geliştirmeye motive etmiştir. Temel olarak metinsel aldatma/sahteciliğin (*textual deception*) tespiti, Metin Analizi/Madenciliği ve Doğal Dil İşleme (DDİ) disiplinlerinin kesişim alanına girmektedir. Geniş perspektiften değerlendirildiğinde ise problem; büyük veri analizi, yapay zekâ, makine öğrenmesi, derin öğrenme vb. gibi birçok sistematığın birlikte kullanılmasını gerektiren bir araştırma alanı haline dönüşmüştür.

Haber etiketleme ve doğrulama sistemleri; metinsel aldatma, içerik sahteciliği, video/foto kaynak tespiti, gerçeklik kontrolü, ironi tespiti, kitle kaynak kullanımı vb. birçok boyutuyla değerlendirilen karmaşık ve iç içe geçmiş alt sistemleri kullanmak üzere kurgulanmaktadır. Özellikle metinsel aldatmanın tespitinde başarımla, dile özgü kaynak ve kütüphanelerin gelişmişlik düzeyiyle doğrudan orantılıdır. Sahte Haber Tespiti alanında literatürdeki birçok çalışma İngilizce diline özgüdür. Her ne kadar bazı analiz ve modellemeler genelleştirilebilir olsa da Türkçenin de içinde bulunduğu birçok dilin kendilerine özgü karakteristik analizi doğru sonuçlara ulaşmak için elzemdir.

Özellikle son 5 yıldaki gelişmeler alan için akademik anlamda bir bilgi birikimi oluşturmuş olsa da Türkçe Sahte Haber Tespitine yönelik literatürde bilinen bir çalışma olmadığı görülmektedir. Bu durum, Türkçe gibi sondan eklemeli diller için de dile özgü karakteristiklerin üzerinde durularak konunun ele alınmasını gerekli kılmaktadır. Çalışmamızın ana çerçevesi ve temel hipotezi, potansiyel sahte haberlerin yenilikçi dilbilimsel yaklaşımlar kullanılarak hibrit/melez metodolojilerle tespit edilebileceği fikrine dayanmaktadır. Bu doğrultuda, Türkçe Sahte Haber Tespitine yönelik temel alınabilecek, geliştirilebilir bir model oluşturulması amaçlanmıştır. Tez kapsamında üç temel hedefe ulaşılması için çalışmalar yapılmıştır. Bu hedeflerin ilki; tez kapsamında elde edilen veri setinin alanda temel olabilecek, tüm disiplinlerin erişimine açık bir kaynak olarak sunulacak olmasıdır. İkinci temel hedef, bu veri seti kullanılarak elde

edilen ve probleme farklı bir bakış açısı katmak üzere Dil Modellemesi yapılarak Türkçe Sahte Haber Sözlüğü geliştirilmesidir. Tez kapsamında hedeflenen diğer önemli husus ise, Türkçe Sahte Haber Tespitine yönelik makine öğrenmesini de içine alan hibrit bir yaklaşımla sürdürülebilir bir model oluşturulması ve sonuçların deneysel çalışmalarla gösterilmesidir.

Tez, temel olarak üç fazda kurgulanmıştır. Birinci faz; literatür taraması, problemin Türkçe açısından ele alınmasında kullanılacak analiz çalışmaları, Türkçe için bu amaçla kullanılacak ve halihazırda literatürde eksik olan veri setinin elde edilerek literatüre kazandırılması ve tüm ön hazırlık işlemlerini kapsayan giriş fazıdır. Bu fazda; ana akım haber medyasındaki haber metinleri, haber doğrulama platformu kaynaklı haber metinleri ve manuel yöntemler kullanılarak toplanan verilerin, tamamı etiketlenerek ve doğrulama işlemi yapılarak göreceli büyük bir derlem oluşturulmuştur. Bu faz, ikinci fazda geliştirilmesi hedeflenen Türkçe Sahte Haber Sözlüğü oluşturulması için gerekli tüm ön işlemleri kapsar.

İkinci fazda, ilk fazda elde edilen veri seti istatistiksel analizlere tabi tutulmuş ve FanLexTR olarak adlandırdığımız dört kategoriye sahip Türkçe Sahte Haber Dil Modeli/Sözlüğünün geliştirilmesi hedeflenmiştir. Her kategori için terim sıklıkları ve oransal gözlenme sıklıkları hesaplamaları ile terimlere ait ton değerlerinin bir başka deyişle skor değerleri oluşturulmuştur. Terimlere polarite değeri verilmesi de ayrıca denenmiştir. Fakat ton değerlerinin daha kararlı sonuçları verdiği gözlemlenmiştir. Bu doğrultuda ortaya çıkan Türkçe Sahte Haber Sözlüğü kullanılarak sözlük temelli metinsel aldatma tespiti yapılmıştır. Kullanılan yöntem dikkate alındığında, çalışmadan oldukça başarılı sonuçlar elde edilmiş ve literatürdeki ilk sözlük temelli Sahte Haber Tespiti çalışması olması olarak yerini almıştır.

Üçüncü ve son fazda, Türkçenin karakteristikleri de göz önüne alınarak, probleme farklı bakış açılarından yaklaşmış, sahte haberlerin tespitine yönelik özellik/öznitelik seçimi ve çıkarma işlemleri gerçekleştirilmiştir. Bu fazda ikinci fazda kullanılan sözlük temelli yaklaşım, üçüncü fazdaki makine öğrenmesi ve derin öğrenme yaklaşımına öznitelikler

şeklinde dahil edilmiş ve ikinci fazda elde edilen sonuçlar geliştirilmiştir. Geliştirilen modelin performans değerlendirmesi sunulmuştur.

Ayrıca tez kapsamında problemin çözümünü pratiğe yansıtmak açısından yenilikçi bir çözüm önerisinde bulunulmuş ve dijital kütüphanecilik teknolojisini barındıran bir çerçeve ortaya konmuştur.

Anahtar Kelimeler: Sahte Haber Sözlüğü, Makine Öğrenmesi, Doğal Dil İşleme, Sahte Haber Tespiti, Eklemeli Diller.

ABSTRACT

A FAKE NEWS DETECTION MODEL FOR TURKISH LANGUAGE

Uğur MERTOĞLU

Doctor of Philosophy, Department of Computer Engineering

Supervisor: Dr. Öğr. Üyesi Burkay GENÇ

Co-Supervisor: Prof. Dr. Hayri SEVER

September 2020, 141 pages

The transformation of print media into online media along with the increasing use of the internet and social media as news sources have drastically altered the concept of media literacy and people's habit of getting news. Both individuals and organizations, therefore, have started to make extensive use of online news sites and social media platforms to receive news. This easier, faster, and comparatively cheaper opportunity offers comfort in terms of people's access to information, but also creates important problems when the possible effects/impacts of news are considered. In this era, the news which disseminates at an unprecedented pace, huge size and variety has created a challenge that cannot be overcome with human power in terms of verifying or determining if it is worth reading. The tendency of people to spread ambiguous or fake news more than valid news makes the problem even more difficult.

As negative situations arising from the effects of fake news increase, social media platforms, commercial organizations, institutions and even states have started to take their own measures against this asymmetric threat. This situation motivated researchers to analyze this type of news most of which is composed of text. And accordingly, they have been trying to constitute scientific infrastructures and develop smart systems for detecting fake news. Basically, detection of textual deception is in the intersection of the Text Analysis/Mining and Natural Language Processing disciplines. The problem has recently become a field of research that requires the use of many systematics together such as big data analysis, artificial intelligence, machine learning etc.

News verification and labeling systems are designed to utilize or to detect complex and nested components like textual deception, fake content fabrication, news source, truth verification, sarcasm, crowd-sourcing etc. In particular, performance in detecting text deception is directly related with the level of development of language-specific resources and libraries. Many studies in the field of Fake News Detection are largely specific to the English language. Although some analysis and modeling can be generalized, the characteristic analysis of many languages, including Turkish, is essential to achieve correct results.

Although the developments in the last 5 years have constituted an academic knowledge for the field, there is still a lack of a scholar study in the literature on Turkish Fake News Detection. This situation makes it necessary to address the issue by emphasizing language-specific characteristics for agglutinative languages such as Turkish. The main framework and basic hypothesis of our study is based on the notion that potential fake news can be detected by hybrid methodologies using novel linguistic-oriented approaches. To this end, it is aimed to create a developable framework model on Turkish Fake News Detection. Within the scope of this thesis, studies were carried out to reach three main goals. The first of these objectives is that the data set obtained under the thesis will be presented as an open source for all disciplines. The second main goal is to develop the Turkish Fake News Lexicon, which is obtained by using this data set contributing a different perspective to the problem. Another goal targeted in the thesis is to develop a

sustainable model with a hybrid approach to Turkish Fake News Detection and to present this through experimental studies.

The thesis is basically constructed in three phases. The first phase is the introductory phase; including literature review, examining the analyzes to deal with the problem, the dataset collection and all preliminary preparations for further phases. In this phase; a large collection has been created by labeling and verifying all of the data collected using news texts from the mainstream news outlets, news verification platform and manual methods. This phase covers all the preliminary procedures required to create a Turkish Fake News Lexicon aimed at developing in the second phase.

In the second phase, the data set obtained in the first phase was subjected to statistical analysis and it was aimed to develop a Turkish Fake News Language Model/Lexicon with four categories, which we named FanLexTR. For each category, term frequencies and relative occurrence frequencies were calculated and tone values of the terms, in other words, score values were assigned. Giving polarity value to the terms has also been tried. But, given the fact that assigning polarity value to the terms could confuse by omitting the fluctuations between fake and valid, we used tone values which gave us more stable results. A lexicon based textual deception detection was performed according to the FanLexTR. Considering the method used, the methodology is the first lexicon-based Fake News Detection study in the literature and quite successful in terms of its results.

In the third and the last phase, considering the specific characteristics of Turkish, the problem was approached from different perspectives, and the feature selection and feature extraction procedures for detection of fake news were carried out. In this phase, the lexicon-based approach developed in the second phase was added to the machine learning and deep learning approach as feature, and the results obtained in the second phase were developed. In this phase, performance evaluation of the obtained model obtained has been made.

Additionally, within the thesis an innovative solution proposal was made to reflect the solution of the problem to practice and a framework containing digital librarianship technology was introduced.

Keywords: Fake News Lexicon, Machine Learning, Natural Language Processing, Fake News Detection, Agglutinative Languages.

TEŞEKKÜR

Doktora eğitim programım ve tez sürecim boyunca bana olan inancı ve desteğini her zaman hissettiğim, engin bilgi ve tecrübesiyle eşsiz katkılarda bulunan ve aynı zamanda tez eş danışmanlığımı da yapan, saygıdeğer Hocam Sayın Prof. Dr. Hayri SEVER'e,

Tez çalışmalarım boyunca bilgi ve tecrübesinden azami faydalandığım, yol gösteren, hoşgörülü, yüreklendiren tutumuyla doğru adımları atmamı sağlayarak çalışmalarına liderlik eden tez danışmanım Sayın Dr. Öğr. Üyesi Burkay GENÇ'e,

Tez İzleme Kurullarında yorum ve önerileriyle çok değerli katkılar sunan Sayın Prof. Dr. Hasan OĞUL'a, Sayın Dr. Öğr. Üyesi Murat AYDOS'a, Sayın Dr. Öğr. Üyesi Fuat AKAL'a,

Tez savunmam esnasında değerli tespit ve önerileriyle katkıda bulunan juri üyelerim Sayın Prof. Dr. Suat ÖZDEMİR'e, Sayın Dr. Öğr. Üyesi Öner BARUT'a, Sayın Dr. Öğr. Üyesi Adnan ÖZSOY'a,

Hacettepe Üniversitesi Bilgisayar Mühendisliği Bölümünün çok kıymetli öğretim üyelerine, çalışanlarına, kendi çalışma arkadaşlarıma ve bitmek bilmeyen sorularımı hiç üşenmeden cevaplamaya çalışan Sayın Dr. Öğr. Üyesi Fatih SAĞLAM'a,

Doktora eğitimim ve tüm hayatım boyunca, desteklerini benden hiç esirgemeyen, bugünlere gelmemde ve tüm başarılarımda en büyük pay sahibi olan aileme,

Ders ve yeterlilik dönemleri başta olmak üzere her zaman sabır ve sevgi ile destek vererek yanımda olduğunu hissettiren değerli eşim Nida MERTOĞLU'na ve hayatıma anlam katan oğullarım Mert ve Kerem'e tüm içtenliğimle teşekkürlerimi sunarım.

İÇİNDEKİLER

ÖZET.....	i
ABSTRACT	v
TEŞEKKÜR	ix
İÇİNDEKİLER.....	x
ŞEKİLLER DİZİNİ.....	xiii
ÇİZGELER DİZİNİ	xiv
SİMGELER VE KISALTMALAR.....	xv
1. GİRİŞ	1
1.1. Problemin Tanımı.....	1
1.2. Amaç ve Motivasyon	4
1.3. Özgün Değer	5
1.4. Yaygın Etki	6
1.5. Tez Organizasyonu.....	6
2. ALANYAZIN TARAMASI VE ALAN BİLGİSİ.....	8
2.1. Sahte Haber	8
2.1.1. Metinsel Aldatma ve Sahte Haber Kavramı.....	8
2.1.2. Sahte Haberın Karakteristiđi, Etkileri ve Çeşitleri.....	12
2.1.3. Sahte Haberle Mücadele ve Zorlukları.....	19
2.1.4. Sahte Haber Tespit Çalışmaları.....	22
2.2. Türkçe Doğal Dil İşleme ve Sahte Haber Analiz Yaklaşımları	25
2.2.1. Türkçe Doğal Dil İşleme ve Metin Madenciliđi	25
2.2.2. Sahte Haber Analiz Yaklaşımları	30
2.2.3. Türkçe DĐĐ' de Karşılaşılan Zorluklar	34
2.3. Çevrim İçi Haber Kaynakları ve Sosyal Medya.....	38
2.4. Dijital Kütüphanecilik ve Sahte Haber.....	39
2.4.1. Sahte Haber ile Mücadelede Dijital Kütüphaneciliđin Önemi.....	41
3. SÖZLÜK (DİL MODELİ) TEMELLİ SAHTE HABER TESPİTİ	44

3.1. Giriş ve Terminoloji.....	45
3.2. Veri Setinin Oluşturulması	49
3.2.1. Veri Kaynakları.....	51
3.2.2. Doğrulama ve Geçerleme	54
3.2.3. Veri Toplama ve Derlem Oluşturma	57
3.2.4. Temel Bir Veri Seti Olarak TR_FN.....	61
3.3. Metodoloji.....	64
3.3.1. Türkçe Sahte Haber Sözlüğünün Oluşturulması.....	65
3.3.2. Sözlük ile Sahte Haber Tahmini	70
3.3.3. Çıktılar ve Performans Değerlendirme	75
3.4. Sonuç (Faz-2).....	82
4. MAKİNE ÖĞRENMESİ TEMELLİ SAHTE HABER TESPİTİ	85
4.1. Giriş (Faz-3).....	85
4.2. Makine Öğrenmesi ve Derin Öğrenme	86
4.3. Kelime Yerleştirme Teknikleri ve Metinlerin Vektörlere Çevrilmesi.....	90
4.3.1. Word2Vec Modeli	91
4.3.2. N-Gram Modeli.....	93
4.3.3. Kelime Çantası Modeli (Bag of Words) Modeli	94
4.4. Metodoloji.....	94
4.4.1. Öznitelik Seçimi ve Çıkarımı	95
4.4.2. SMOTE Tekniği	101
4.4.3. Çapraz Doğrulama/Geçerleme (<i>Cross Validation</i>).....	101
4.5. Sınıflandırma Modelleri.....	102
4.5.1. En Yakın k-Komşuluk (KNN).....	102
4.5.2. Multinomial (Çok Terimli) Naive Bayes (MNB).....	104
4.5.3. Destek Vektör Makineleri (DVM).....	105
4.5.4. Lojistik Regresyon (LR)	106
4.5.5. Karar Ağaçları.....	107
4.6. Uzun-Kısa Süreli Bellek Ağları (<i>LSTM</i>)	110
4.7. Tez Blok Diyagramı.....	111
4.8. Çıktılar ve Performans Değerlendirme	113
4.8.1. Keşifsel Veri Analizi Tespitleri	119
4.9. Sonuç (Faz-3).....	121

5. SONUÇLAR ve ÖNERİLER.....	123
5.1. Sonuçlar.....	123
5.2. Öneriler.....	126
6. KAYNAKLAR.....	127
EKLER	136
EK 1 – Veri seti oluşturulurken yararlanılan yaygın kullanıma sahip haber siteleri	136
EK 2 – Görsellerle desteklenen teyitsiz sahte haber örneği (Örnek-1)	137
EK 3 – Sahte bilgiler içeren bilim haberleri ve paylaşımları (Örnek-2)	138
EK 4 - Tezden Türetilmiş Yayınlar	139
EK 5 - Tezden Türetilmiş Bildiriler	140
ÖZGEÇMİŞ	141

ŞEKİLLER DİZİNİ

Şekil 2.1 Sahte haber çeşitlerinin risk seviyesi ve yayılım hızına göre dağılımı.....	16
Şekil 2.2 Sahte haber çalışmalarının esas aldığı temel kıstaslar ve modellemeler.	23
Şekil 2.3 Dilbilimin temel analiz seviyeleri.....	26
Şekil 2.4 Sahte Haber Tespiti Çatı Modeli	42
Şekil 3.1 Sahte haber tespit problemine yönelik hata matrisi tanımları	46
Şekil 3.2 ROC eğrisi altındaki alan gösterimi	49
Şekil 3.4 FanLexTR geliştirme süreci ana sayfaları.	57
Şekil 3.5 Haber kaynakları içerik etkileşimi.....	59
Şekil 3.6 TR_FN veri setindeki haber alanlarının dağılımı	63
Şekil 3.7 Birleştirilmiş (Ensemble) Model Çoğunluk Kararı (Kural-2)	81
Şekil 4.1 Yapay zekâ, makine öğrenmesi ve derin öğrenme kavramları.....	88
Şekil 4.2 Makine öğrenmesinde kullanılan yaklaşımlara genel bakış.	89
Şekil 4.3 Örnek bir cümle üzerinden Skip-gram model mantıksal gösterimi.....	92
Şekil 4.4 Kelime çantası modeli	94
Şekil 4.5 Ham veriye ait kolonların korelasyon matrisi	97
Şekil 4.6 TR_FN veri setindeki ana kaynak dağılımı.	100
Şekil 4.7 Çapraz doğrulama uygulama adımları.....	102
Şekil 4.8 KNN algoritması çalışma modeli	103
Şekil 4.9 KNN algoritması adımları	103
Şekil 4.10 DVM için temsili problemler (Doğrusal ve doğrusal olmayan).....	105
Şekil 4.11 Karar Ağacı Modeli.	107
Şekil 4.12 Uzun-Kısa Süreli Bellek Ağları 4 katmanlı yapısı.	111
Şekil 4.13 Çatı Sistem Modeli – Sahte Haber Tespit Modeli gösterimi.....	112
Şekil 4.14 Tez Blok Diyagramı	112
Şekil 4.15 LSTM modeline ait ROC eğrisi.....	116
Şekil 4.16 LSTM parametrelerinin seçilmesi	118
Şekil 4.17 LSTM ağlarının performans (Spor, Sağlık ve Mizah) değerlendirmesi.....	118

ÇİZGELER DİZİNİ

Çizelge 2.1 Metinsel aldatmanın literatürde ele alınan çeşitleri.	9
Çizelge 2.2 Çeşitli çalışmalardaki “Sahte Haber” kavramı tanımları.	11
Çizelge 2.3 Sahte haberlerin oluşturulmasındaki niyet ve olası etkileri.	13
Çizelge 2.4 Türkçede anlam belirsizliği örnekleri.	35
Çizelge 2.5 Türkçede morfolojik belirsizlik.	37
Çizelge 3.1 Google BigQuery ile veri toplanması	52
Çizelge 3.2 Veri kümelerine ait cümle başı argo ve yazım hatası.	56
Çizelge 3.3 Elde edilen haber derlemine ait istatistiki bilgiler.	58
Çizelge 3.4 TR_FN veri setinin bilgi alanlarının tabular biçimde sunulması.	61
Çizelge 3.5 Haberlerde en çok kullanılan (kök form) sözcükler.	63
Çizelge 3.6 Elde edilen sözlük modellerine ait istatistiki bilgiler.	66
Çizelge 3.7 Raw modele ait bazı örnekler.	67
Çizelge 3.8 Root modele ait bazı örnekler.	68
Çizelge 3.9 Raw+Pos Modele ait bazı örnekler.	69
Çizelge 3.10 Suffix Modele ait bazı örnekler.	70
Çizelge 3.11 Test veri setine ait bir sahte haber örneği.	70
Çizelge 3.12 RAW Form modele göre puanlanmış bazı örnekler.	72
Çizelge 3.13 Eğitim veri setlerine ait hold-out doğrulama sonuçları.	76
Çizelge 3.14 Eğitim veri setlerine ait hold-out doğrulama sonuç ortalamaları.	77
Çizelge 3.15 Test veri setlerine test veri setlerine ait değerlendirme sonuçları.	77
Çizelge 3.16 Raw Model için Hata Matrisi (Confusion Matrix).	78
Çizelge 3.17 Root Model için Hata Matrisi (Confusion Matrix)	78
Çizelge 3.18 Raw+Pos Model için Hata Matrisi (Confusion Matrix).	79
Çizelge 3.19 Suffix Model Hata Matrisi (Confusion Matrix).	79
Çizelge 3.20 Birleştirilmiş Model için Hata Matrisi (Confusion Matrix).	82
Çizelge 4.1 Önerilen/kullanılan öznitelikler, öznitelik grupları ve açıklamaları (Tr _{All}).	98
Çizelge 4.2 Makine Öğrenmesi modellerine ait değerlendirme sonuçları	115
Çizelge 4.3 LSTM modeline ait değerlendirme sonuçları	116
Çizelge 4.4 Sahte haberlerde daha sık karşılaşılan bazı ifadeler.	120

SİMGELER VE KISALTMALAR

Simgeler

μ	Ortalama
Exp(β)	Odds Oranı

Kısaltmalar

DDİ	Doğal Dil İşleme
SHS	Sahte Haber Sözlüğü
GDELT	Global Database of Events, Language (Location) and Tone
RADAR	Relevance, Authority, Date, Appearance, Reason for writing
CRAAP	Currency, Relevance, Authority, Accuracy, and Purpose
BT	Bilgi Teknolojileri
POS	Sözcük Türü (<i>Part of Speech</i>)
CFG	İçerikten Bağımsız Gramer (<i>Context Free Grammar</i>)
RST	Retorik Yapı Teorisi (<i>Rhetorical Structure Theory</i>)
CNN	Evrişimsel Sinir Ağları (<i>CNN, Convolutional Neural Network</i>)
FSA	Sonlu Durum Özdevinirleri (<i>Finite-State Automata</i>)
LSTM	Uzun-Kısa Süreli Bellek (<i>Long-Short Term Memory</i>)
ROC	Alıcı İşletim Karakteristiği (<i>Receiver Operator Characteristics</i>)
EKK	En Küçük Kareler
TF-IDF	Term Frequency-Inverse Document Frequency
PCA	Temel Bileşen Analizi (<i>Principal Component Analysis</i>)
SVM	Destek Vektör Makinesi (<i>Support Vector Machine</i>)
IFCN	International Fact-Checking Network
SMOTE	Synthetic Minority Over-sampling Technique
URL	Uniform Resource Locator
HTML	Hyper Text Markup Language
KNN	En Yakın k-Komşuluk (<i>k-Nearest Neighbour</i>)
KVA	Keşifsel Veri Analizi (<i>Exploratory Data Analysis</i>)
RNN	Tekrarlayan Sinir Ağları (<i>Recurrent Neural Network</i>)
LIS	Library and Information Science

1. GİRİŞ

Bu tez çalışmasının giriş bölümü; problemin tanımı, amaç ve motivasyon, tez çalışmasının özgün değeri, yaygın değeri ve tez organizasyonu bölümlerini içeren alt başlıklarda sunulmuştur. Bu alt başlıklarda sırasıyla problemin tanımı yapılmış, ardından amaç ve motivasyon ele alınmıştır. Tez kapsamında önerilen metodolojiler ve uygulanan yaklaşımlar doğrultusunda tezin özgün ve yaygın değeri hakkında değerlendirmelere yer verilmiş ve çalışmanın detaylı anlatımını kılavuzlayan tez organizasyonu sunulmuştur.

1.1. Problemin Tanımı

Günümüzde yazılı haber medyasının çevrim içi (*online*) yayın modeline evrilmesi ve sosyal ağların haber üreten-tüketen platformlar haline gelmesi hacim ve çeşitlilik anlamında muazzam büyüklükte bir haber ve veri akışına neden olmaktadır. Bu üretim ve dağılım kolaylığı gerçek içeriklerin yanı sıra sahte veya uydurma içeriklerin de dolaşımını hızlandırmıştır. İletişim teknolojilerinde yaşanan bu hızlı değişim ve dönüşüm, insanları maruz kaldıkları haber ve veri akışı karşısında savunmasız bırakmaktadır. Her gün neredeyse her saniye devasa boyuttaki bilgi ve haber, çoğu doğrulanmamış bir şekilde milyonlarca kullanıcıya farklı platformlardan sunulmaktadır. Öyle ki; insanlar çevrim içi olarak karşılaştıkları haberleri çoğu zaman hiçbir doğrulama arayışına girmeden yayarak, sahte/yanlış bilgi ve içeriklerin (*fake/fraudulent information and content*) hızla yayılmasına neden olmaktadır. Bu yayılımın bireyler ve toplumlar üzerinde korku, umutsuzluk, öfke ve önyargı vb. birçok olumsuz etkisi bulunmaktadır. Bütüncül bir bakış açısıyla, bireysel ve toplumsal etkilerinin yanı sıra sosyal medya platformlarının ve şirketlerin itibar ve ticari kayıplar yaşamasından, toplumsal kutuplaşmaya, bölgesel ve uluslararası krizlerin tetiklenmesine, politik ve ticari faaliyetlerin manipüle edilmesine, toplumda güvensizlik yaratılmasına kadar birçok olumsuz etkisi gözlemlenmektedir. Son dönemlerde sahte haberler, sosyal mühendislik unsurlarından yararlanılarak adeta asimetrik bir saldırı aracı haline dönüştürülmeye başlamıştır. Bu probleme çözüm bulma arayışı neticesinde, son 5 yıl içerisinde başta bilgisayar bilimleri olmak üzere birçok disiplinde merak uyandıran sahte haber tespiti, disiplinler arası bir araştırma alanı olarak ortaya çıkmıştır.

Haber içerikleri genellikle başta metin olmak üzere ses, görsel ve video vb. bileşenlerden oluşmaktadır. Çevrim içi ortamdaki verinin yayılma hızı, boyutu, çeşitliliği ve heterojenliği göz önüne alındığında, haberlerin gerçekliğini tespit edebilecek kapsayıcı bir sistematığın geliştirilmesi oldukça zor bir görevdir. Bu bağlamda haberi oluşturan bileşenler içinde en baskın tür olan metin içeriklerinin analizi büyük bir öneme sahiptir. Bu metin içeriklerinin analizi için öncelikli başvuru noktası Doğal Dil İşleme (DDİ) disiplini'dir. Genel anlamda, gelişen ve sürekli değişen teknolojilerin birçok açıdan dünyayı küresel bir köy haline getirmekte olduğu öne sürülse de dil unsurunun hala toplumsal farklılıkları barındıran kültürel bir öğe olarak varlığını sürdürdüğü gözlemlenmektedir. DDİ alanındaki araştırmacıların kendi dillerinde çalışmalar yapması beklenirken özellikle Sahte Haber Tespiti konusunda incelemeler yapan araştırmacıların büyük bir kısmının İngilizce dilini çalıştığı gözlemlenmektedir. Bu durumun nedenlerinden ilki, İngilizce dil kütüphaneleri ve araçlarının gelişmişlik düzeyidir. Bir diğer neden ise nispeten yeni bir çalışma alanı olan “Sahte Haber Tespiti” üzerinde çalışılacak veri setlerinin ve dile özgü kaynakların eksikliğidir. Ülkemize yakın coğrafyadaki bölge ülkelerinde yaşanan bazı olumsuz gelişmeler derinlemesine incelendiğinde, bu gelişmelerde sahte haberlerin etkisinin görülmesi mümkündür. Ülkemizin de son yıllarda dünya ile paralel şekilde sahte haberlerden olumsuz etkilendiği ve gelecekte etkilenebileceği değerlendirilmektedir. Bu kapsamda Türkçe Sahte Haber Tespiti çalışmalarını ve bu doğrultuda elde edilecek bilgi birikimini aktif savunmada kullanılabilecek milli bir güç çarpanı olarak değerlendiriyoruz.

Literatürde Türkçe sahte haber tespitine yönelik çalışmaların eksikliği tez kapsamında belirtilen önemli hususların başında gelmektedir. Bu eksiklik hem neden hem sonuç bağlamında “Türkçe Haber Veri Seti” eksikliği ile doğrudan ilişkilidir. DDİ çalışmalarında başarıyı arttırabilmek için dillere özgü kapsamlı dil kütüphanelerinin geliştirilmesi gerekmektedir. Her ne kadar DDİ bir alt dalı olan duygu analizi ve yazar tanıma problemleri gibi metin sınıflandırma problemlerine yönelik Türkçe çalışmalar artmış olsa da Sahte Haber Tespiti özelinde temel teşkil edebilecek kapsayıcı çalışmaların eksikliği görülmektedir. Tez çalışmaları kapsamında elde edilen Türkçe Sahte Haber Sözlüğü bu eksikliğin giderilmesine yönelik yapılan çalışmaların sonucu olarak ortaya çıkmış bir üründür. Bu sözlük temel alınarak ilerleyen çalışmalarda daha kapsamlı ve gelişmiş hale getirilebileceği değerlendirilmektedir. Ayrıca çalışmamızda geliştirdiğimiz

bu sözlük makine öğrenmesi temelli çözümümüzde faydalandığımız bir kaynak olarak kullanılmıştır. Tezin geri kalan bölümlerinde tekrarlardan kaçınmak ve tezin okunabilirliğini artırmak için SHS (Sahte Haber Sözlüğü), FanLexTr şeklinde isimlendirdiğimiz kısa adı ile kullanılacaktır.

Literatürde sahte haber tespiti üzerine yapılan çalışmalarda, Yapay Zekânın alt dalları olan Makine Öğrenmesi ve Derin Öğrenme teknolojilerinden faydalanılmaktadır. Özellikle probleme uygunluğu açısından makine öğrenmesinde kullanılacak algoritmalar ve dile özgü ortaya konacak öznitelik seçimi ve çıkarımlarının doğru seçimi başarıyı etkilemektedir. Bu çalışmada bu alanda geliştirdiğimiz sözlük temelli yaklaşım ve yararlandığımız makine öğrenmesi temelli yaklaşımlarla ayrı ayrı ve birlikte kullanılarak problemin çözümüne uygun bir metodoloji ortaya konulmuştur. Ayrıca tez kapsamında derin öğrenme tekniklerinden faydalanılmış ve elde edilen sonuçlar karşılaştırmalı olarak sunulmuştur.

Alandaki bir diğer eksiklik, sahte haber tespit çalışmalarında gerçek ve doğru, sahte ve yanlış haber arasında yapılan tanımlama hataları ve eksikliklerinin çalışmaların ilerleyiş yönünü olumsuz etkilemesidir. Bazı çalışmalar başarı yakalamak için sadece alana özgü kalıp genelleştirilemezken, içerik doğruluğu ispatı içerisindeki bazı çalışmalar çok farklı yaklaşımları birlikte kullanmayı deneyerek hedeflerinden uzaklaşmaktadırlar. Bu yaklaşımlar, doğru veya yanlış haberi ayırt etmek için farklı bileşenleri tek potada eritmenin oldukça zor olduğunu göstermektedir. Benzer şekilde, içerik doğrulama çalışmalarında da haberin kaynağı ve içeriği başta olmak üzere foto ve video gibi metin dışı unsurların doğrulanması, ağdaki yayılımı, üst verilerin (*metadata*) zaman içerisinde değişkenliklerinin ele alınması gibi yeteneklere ihtiyaç duyulmaktadır. Konunun zorlayıcı bir diğer özelliği de son zamanlarda sıklıkla kullanılan nesnel doğruların, hakikatlerin, olguların önemini yitirmesi anlamına gelen “post-truth” (gerçek-ötesi) kavramı açısından haberin kime ve neye doğru olduğu gibi eleştirel yaklaşımların ortaya çıkmasıdır. Bu tez çalışmasında, problem hassasiyetle tanımlanarak sahte ve gerçek haber dilinin tespitinin yapılmasına ve sağlanan otomasyonla bir haber değerlendirmesi yapılırken öncelikli olarak harcanan manuel çabanın önüne geçilmesi hedeflenmiştir.

1.2. Amaç ve Motivasyon

Doğal Dil İşleme, İstatistik ve Makine öğrenmesi yöntemlerinin kullanıldığı çalışmada, Türkçe haber metinleri üzerinde sahtelik tespiti konusunda temel alınabilecek, sürdürülebilir ve geliştirilebilir bir model oluşturulması amaçlanmaktadır. Alandaki çalışmaların büyük çoğunluğunun İngilizce diline özgü çalışmalar olması nedeniyle, çalışmanın alanda Türkçe üzerinde çalışmak isteyen araştırmacıların temel başvuru kaynaklarından biri olması hedeflenmektedir. Sayısı gittikçe artan haber doğrulama platformları için bir karar destek sistemi servisi olarak düşünüldüğünde, çalışmanın akademik katkısı yanında pratik hayatta da kullanım alanı bulabileceği öngörülmektedir.

Tez kapsamında sahte haberin tanımı ve çeşitleri titizlikle ele alınmıştır. Elde edilen derlem ilk olarak Keşifsel Veri Analiz (KVA) yöntemleriyle ele alınmıştır. Ortaya çıkarılan çarpıcı bulgular, Türkçe Sahte Haber Dil Modeli/Sözlüğünün, oluşturulmasına da temel teşkil etmiştir. Diğer dillere ve özellikle sondan eklemeli dillere uyarlanabilir bir kılavuz olabilme potansiyeli çalışmamızın motive edici unsurlarından biri olmuştur.

Hem yapılan Türkçe çalışma hem de Türkçe veri seti açısından ilk olması nedeniyle referans alınabilecek bir başlangıç noktası olabileceği düşünülmektedir. Elde ettiğimiz veri setinin, sadece sahte haber tespiti konusunda değil farklı disiplinlerden bu haber verisini kullanmak isteyen araştırmacıların yararlanabileceği şekilde açık kaynak olarak sunulması hedeflenmiştir. Tüm bu özellikleri ile Türkçe özelinde alanda öncü olabilecek bir çalışma ortaya çıkarılması hedeflenmiştir.

Yukarıda belirtilen nedenlerden dolayı, Türkçe Sahte Haber tespitinde kullanılacak dile özgü kaynaklar, metodolojiler, yaklaşımların çalışılmasını elzem gördüğümüzden tez kapsamında Türkçe üzerine yoğunlaşarak problem ele alınmıştır. Bunun yanı sıra dillerden bağımsız olarak kullanılacak literatürdeki makine öğrenme yöntemleri, DDİ ve istatistiksel analizlerle desteklenerek tezin genelleştirilebilir bir kurguda sunulmasına dikkat edilmiştir.

Yapay zekâ ve DDİ disiplinlerinin çalışma alanları çok geniş kapsamlıdır. Bu bağlamda Türkçe için de çalışılması gereken birçok alt alan bulunmaktadır. Tez kapsamında sunulan çözüm yaklaşımlarının farklı metin sınıflandırma problemlerine yönelik makine öğrenmesi ve derin öğrenme çalışmalarında kullanılabileceği, Türkçe Sahte Haber Dil Modeli/Sözlüğünün spor, politika vb. alt alanlara özgü çalışmalarının ortaya konulabileceği ve bu tarz çalışmalar için sağlam bir temel inşa edeceği öngörülmektedir.

1.3. Özgün Değer

Sahte haber tespiti üzerine yapılan bilgisayarlı çözümlere odaklanan ilk Türkçe çalışmalardan olması, Türkçe Sahte Haber Dil Modeli/Sözlüğünün geliştirilmesi ve sahte haber tespitine yönelik kapsamlı bir modelin sunulmasının yanı sıra tezin aşağıda belirtilen özgün değerlere sahip olduğu değerlendirilmektedir.

- Büyük bir derlemeden elde ettiğimiz dört kategoriye sahip FanLexTR Türkçe Sahte Haber Dil Modeli/Sözlüğünün ve veri setinin DDİ çalışması yapacak araştırmacılara sunulması.
- Önerilen modelin ve kullanılan metodolojilerin sadece Türkçe için değil başta sondan eklemeli diller (Fince, Macarca vb.) olmak üzere farklı diller için de genelleştirilebilir ve geliştirilebilir olması.
- Literatürde geleneksel metin sınıflandırma çalışmaları esas alındığında, çalışmanın haberin niteliğine odaklanması neticesinde kullanılan dile bir değer atfetmesi açısından bu tarz çalışmalara referans olabilecek yenilikçi bir bakış açısına sahip olması.
- Belli bir alana (spor, politika, mizah vb.) özgü sahte haber terimler sözlüklerinin geliştirilmesinde FanLexTR sözlük çalışmasından (dil kütüphanesinden) faydalanılabileceği.
- Haberlerin bireyleri ve toplumları etkileme potansiyeli düşünüldüğünde geliştirilen çerçevenin bilimsel seviyeye yapacağı katalizör etkisi.
- Ayrıca problemin çözümünde dijital kütüphaneciliğe yer veren bütünleşik bir çözüm stratejisi önermesi.

1.4. Yaygın Etki

Önceki bölümlerde belirttiğimiz üzere, sahte haberin toplumsal etkilerinin tehlikeli boyutlara ulaşması nedeniyle, çalışmanın bir platform/servis olarak sunulduğunda ve/veya haber doğrulama siteleri tarafından yararlanıldığında aşağıda belirtilen yaygın etkileri olacağı öngörülmektedir.

- Sosyal medya kullanıcılarının haber niteliğinde kendilerine gelen paylaşımların viral yayılımından önce, ön kontrol yapabilmesi.
- Haber doğrulama ve/veya haber sitelerinin bir haberi yayınlamadan önce test edebilmesi.
- Cep telefonu ve e-posta yoluyla gelen özellikle dolandırıcılık, hırsızlık vs. amaçlı kötü niyetli mesajların ve notların ayırt edilebilmesi.
- Yalan haberlerin yayılmasının önlenmesi ve potansiyel toplumsal krizlerin önüne geçilmesi konusunda yardımcı bir karar destek sistemi olarak güvenlik güçlerine hizmet verebilme kapasitesi taşıması.

1.5. Tez Organizasyonu

Tez dokümanı beş ana bölümden ve kaynaklardan oluşmaktadır. Tezin ilk bölümü önceki başlıklarda sunulmuştur. Tezin bundan sonraki kurgusu bu bölümde özetlenmiştir.

Tezin ikinci bölümü, Alanyazın Taraması ve Alan Bilgisi şeklinde isimlendirilmiştir. Bu bölümde “Sahte Haber” kavramı, ortaya çıkışı, çeşitleri, mücadele yöntemleri, tespitine yönelik literatürdeki çalışmalar, metin madenciliği, Türkçe DDİ ve karşılaşılan zorluklar, sahte haber analiz yaklaşımları, haber kaynakları ve sosyal medya, dijital kütüphanecilik konuları incelenmiş, literatürdeki ilgili çalışmalara yer verilmiş ve sonraki bölümlerde kullanılacak terminoloji ve veri seti ile ilgili özet bilgi verilmiştir.

Üçüncü bölümde, çeşitli fazlardan oluşan FanLexTR'nin geliştirilmesi, kullanılan metodoloji, örneklerden yararlanılarak sunulan uygulama esasları, performans ölçüm süreçleri ve çıktılara yer verilmiştir. Bu bölümde veri kaynakları, veri toplama aşamaları,

derlemin oluřturulmasında gerekleřtirilen iřlemler, KVA alıřmaları, veri istatistik ve grselleřtirmeleri sunulmuřtur.

Drdnc blmde makine ğrenmesi yntemleri ile sahte haber tespiti, deęerlendirme metrikleri, kullanılan modeller, znelik seimi ve znelik ıkarımı, szlk ve makine ğrenmesi yaklařımlarının birleřtirilmesinden oluřan hibrit (melez) model, model blok diyagramı, sonuların deęerlendirilmesi, KVA tespitleri ve sorun sahaları ortaya konmuřtur.

Son olarak teze ait genel deęerlendirme ve gelecek alıřmalara ynelik neriler Sonu ve neriler blmnde sunulmuřtur.

2. ALANYAZIN TARAMASI VE ALAN BİLGİSİ

2.1. Sahte Haber

Sahte haberler içinde bulunduğumuz teknoloji odaklı çağda internetin de yaygın kullanımı ile uluslararası toplum açısından ciddi bir sorun alanı haline gelmiştir. Bunun temel nedeni teknolojinin ve sahte haberin politik, ekonomik ve hatta stratejik bir çıkar aracı haline gelmiş olmasıdır. Oldukça hassas yaklaşılması gereken bir hem sosyal hem de sayısal disiplinlerin ilgi ve etki alanına girmektedir. Bu bölümde, sahte haber konusu farklı boyutları ile detaylı bir şekilde ele alınacaktır.

2.1.1. Metinsel Aldatma ve Sahte Haber Kavramı

Aldatma veya kandırma; Buller ve Burgoon [1] tarafından Kişilerarası Aldatma Teorisi (*Interpersonal Deception Theory*) temel alınarak, gönderici tarafından yanlış bir inanış, bilgi veya sonucu alıcıya bilinçli olarak iletmek şeklinde tanımlanmıştır. Tarihsel perspektiften değerlendirildiğinde ise; sahtecilik ve aldatmanın geçmişi çok eski zamanlara dayandırılmaktadır. Somut olarak kanıtlanması zor da olsa, yalan ve aldatma insanların birbirleri ile iletişime geçtikleri ilk zamanlara kadar dayandırılabilir. Konunun tarihi ve sosyal incelemesi tezin kapsamı dışında tutulmakla beraber, tarihin ilk propaganda ve aldatma örneklerini, M.Ö. 13. yüzyıldaki Kadeş Antlaşması¹ ile ilişkilendiren bir çalışma bulunması [2], çarpıcı bir örnek olarak gösterilebilir.

Aldatma ve yalanın, zaman içerisinde metin ortamına uyarlanabilmesi, metinlerde aldatma kavramını ortaya çıkarmıştır. Bu tez kapsamında, “Metinsel Aldatma”, fiziksel veya çevrim içi olarak sunulan sahte bir belgeden ziyade, yazılı ve çevrim içi sahtelik olgusunu ifade eden bir kavram olarak ele alınmıştır. Neden olduğu olumsuz durumlar göz önüne alındığında, insanların yüz yüze iletişimi esnasındaki yalan ve aldatmadan daha yıkıcı etkileri olduğu söylenebilir. Bu doğrultuda, insanların yazılı ve görsel iletişim

¹ Kadeş Antlaşması; MÖ 13. yüzyılda, aynı adlı savaşı bitiren, Mısır Firavunu II. Ramses ile Hitit Kralı III. Hattuşili arasında bir savaş açmazı neticesinde imzalanan ve toprak paylaşımı (Suriye) ile neticelenen barış antlaşmasıdır. Fakat bazı kaynaklar her iki liderin de kendi halklarına savaşın ezici üstünlükle kazanıldığı anlatan propagandalar yaptığı, tapınaklara savaşın kazanıldığı belirten sahneleri tasvir ettirdiği belirtilmektedir (https://en.wikipedia.org/wiki/Fake_news).

açısından, yüz yüze kurulan iletişime göre çok daha savunmasız olduğunu söylemek mümkündür.

Bu kapsamda, metinsel aldatmanın başta dilbilimsel ipuçları olmak üzere birçok farklı yöntem kullanılarak tespit edilebileceği [3, 4], bazıları ilerleyen bölümlerde referans edilecek çeşitli akademik çalışmaların konusu olmuştur. Metinsel aldatma, yazılı iletişimin gelişmesine paralel olarak çok farklı etki alanlarında kendini göstermiştir. Literatürdeki DDİ ve Metin Madenciliği çalışmaları, metinsel aldatmayı çeşitli metin sınıflandırma problemleriyle ilişkilendirir. Çizelge2.1’de metinsel aldatma çalışmalarına ait etki alanları gösterilmiştir.

Çizelge 2.1 Metinsel aldatmanın literatürde ele alınan çeşitleri.

Sözlü İletişimin Yazıya Geçirilmesi	Psikolojik ve Klinik Vakaların incelenmesi Anket ve Röportajlar Durum Senaryoları Müşteri Hizmetleri Konuşmaları Politik söylevler (Parlamento, miting konuşmaları vb.) Deneysel Veriler (Yalan Testi vb.)
Bilgisayar Aracılı İletişim	Elektronik Postalar (istenmeyen e-postalar, dolandırıcılık, robot hesaplar) Sosyal Medya Platformları, Mesajlaşma Uygulamaları Reklamlar, Ürün Yorumları ve Bloglar Çevrim İçi iletişim
Hukuk, Güvenlik ve Bilimsel Metinler	Davalar (Sanık, Tanık, Şahit ifadeleri) Adli Bilim (Polis sorguları) Belge Sahteciliği İntihal Tespit Programları
Habercilik	Çeşitli Sahte Haberler ve Alanları (Politika, Eleştiri ve Hiciv yazıları vb.) Uluslararası ve bölgesel rekabetler, çatışmalar Sosyal olaylar (İsyan, Protesto, Kargaşa, Mülteci Sorunları vb.) Sosyal Medya ve İnternetin Kötüye Kullanılması (Manipülasyon, propaganda vb.)

Haber, Türk Dil Kurumu sözlüğünde² “bir olay, bir olgu üzerine edinilen bilgi, salık”, “iletişim veya yayın organlarıyla verilen bilgi” ve “bilgi” ifadeleriyle tanımlanmaktadır. Haber her ne kadar toplumun bilgi edinme ve merak içgüdülerine karşılık veren doğal bir olgu olarak var olsa da etkileri bakımından değerlendirildiğinde basit bir bilgi yayma faaliyetinden çok daha öte olduğu aşikârdır.

Özellikle son 5 yıldır popülerlik kazanan hakikat-sonrası (*post-truth*), doğruluk kontrolü (*fact-checking*), sahte haber, yalan haber vb. yeni kavramlar haberin tanımını ve yapısını karmaşıklştırmaktadır. Haberin birlikte kullanıldığı tamamlayıcılar, sunum şekli, yöntemleri, haberde kullanılan dil ve hatta seçilen kelimeler gibi birçok faktör haberin kime, neye göre doğru veya yanlış olduğunu belirlemeyi zorlaştırmaktadır. Bunun yanı sıra, teknolojiyi arkasına alarak hızlanan ve yoğunlaşan haber yayılımı, insanlara haberin niteliğini sorgulama fırsatı bile vermemektedir. İşte tam bu noktada, “Sahte Haber” kavramı ortaya çıkışı, nitelikleri, etkileri, tespiti ve teknolojik evrimi bakımından birçok farklı disiplinden araştırmacıların ilgisini çekmeye başlamıştır. Bu kapsamda, tez doğrultusunda tüm bu konular ele alınmış ve sahte haber tespitine yönelik bilgisayar merkezli sayısal çözümler üzerine yoğunlaşmıştır. Sosyal ve fen bilimlerini ilgilendiren geniş kapsamlı bir konu olması nedeniyle, birçok alt araştırma alanına kısaca değinilmiş ve disiplinler arası kesişim ölçülü bir şekilde verilmeye çalışılmıştır. Tezin bu bağlamda incelenmesi, tüm disiplinlerdeki araştırmacılar için hem olası gelecek çalışmaların tespiti hem de tezin okunabilirliği anlamında yerinde olacaktır.

Sahte haber; bir diğer deyişle sözde/aldatmaca haber, insanları yanıltmak, aldatmak için geleneksel haber medyası (yazılı veya basılı) ve sosyal medya aracılığıyla bilinçli veya bilinçsiz şekilde üretilen bir haber türüdür. Birçok akademik makalede sahte haberin tanımı çarpıcı özellikleriyle verilmiştir. Çizelge 2.2’de, Egelhofer’in çalışmasından [5] yararlanarak oluşturulan çeşitli çalışmalarda yapılmış sahte haber tanımları detaylı bir şekilde sunulmuştur.

² <http://www.tdk.gov.tr>

Çizelge 2.2 Çeşitli çalışmalardaki “Sahte Haber” kavramı tanımları.

<i>Kasıtlı oluşturulmuş</i> ve teyit edilebilir şekilde yanlış olan ve okuyucuları yanlış yönlendirebilecek haber makaleleridir. (Alcott ve Gentzkow) [6]	1, 2, 3
Tamamen yanlış veya <i> bilinçli bir şekilde</i> haber içeriğine yanıltıcı öğeler içeren haberlerdir. (Bakir ve McStay) [7]	1, 3
<u>Gerçekmiş gibi hazırlanan ve yayınlanan sahte haber öyküleridir.</u> (DiFranzo ve Gloria-Garcia) [8]	1, 2
İlk olarak 2016 ABD Başkanlık Seçimlerinde ortaya çıkan öne çıkan yeni bir siyasi yanlış bilgilendirme biçimi. (Guess ve arkadaşları) [9]	1
Sahte bir haber tartışmasının altında yatan varsayım, <u>gerçek habermiş gibi yazılmış olmasıdır</u> (...). Oysa sahte haberler <i>aldatma niyeti</i> taşır, bu da okuyucuyu doğru olduğuna inandırır. (Horne ve Adalı) [10]	1, 2, 3
Bu terim yaygın olarak <u>gerçeğe dayalı bir haber hikâyesi gibi görünmek için yapılan yanlış veya yanıltıcı bilgilere</u> atıfta bulunur. (Nelson ve Taneja) [11]	1, 2
<u>Legal kaynaklardan gelmiş gibi sunulan uydurma hikâyelerdir.</u> (Pennycook ve Rand) [12]	1, 2

Tanımlamalarda vurgulanan özellikler: (1) **Gerçeklikten uzak olma (koyu olarak yazılı)**; (2) Gazetecilik/Haber/Basın formatı (altı çizili); (3) *Aldatma niyeti (italik yazı)*

Bu çalışmaların dışında da farklı makalelerde [13, 14] farklı tanımlamalar bulmak mümkündür. Biz çalışmamızda sahte haberin kavramsal tanımından daha ziyade problemin çözümü için gerekli olan matematiksel modellemeye yoğunlaştık. Bu yaklaşımdaki ana neden her ne kadar kavram tanımında fikir birliği sağlanamamış olsa da problemin tanımı için ortak bir noktada buluşulabileceğine olan inancımızdır. Shu'nun çalışmasında [15] yaptığı problem tanımına benzer şekilde kurguladığımız matematiksel modelin bu bağlamda genel-geçer olduğunu değerlendirmekteyiz. Buna göre;

- Bir haberi “h” olarak tanımlayalım. Haberin bir kaynağı (K) ve haber ana metni (M) şeklinde iki temel bileşeni olduğu ve bu bileşenlerin bazı özellik ve alt bileşenleri üzerinde taşıdığını, ayrıca bazı haberlerin de sosyal bağlamda (*social context*) belirli etkileşimler (*tuple*) içinde olduğunu bilmekteyiz.

- Şöyle ki; K_h kaynağa ait yazar, yayın kaynağı, zaman, isim vb. bilgileri içerir. Aynı şekilde, M_h haber metnine ve tamamlayıcı bileşenlerine ait metin bilgisi, başlık, ses, görsel ve video vb. alt parçalardan oluşmaktadır.
- Ayrıca, S_h haberin sosyal bağlamdaki okuyucu, gönderi, paylaşım ve ağ davranışı bilgilerini içerir.
- Buna göre temel bir sahte haber tespit modelinde fonksiyon/model F , kaynaktan yayılan haberi değerlendirir. Örneğin bizim modelimizde $F_{K_h}(M_h)$ modeli M_h 'da (M_h =Metin+Başlık+Kategorik Veri) sahtelik/aldatma tespit ederse “Sahte” değilse “Gerçek” olarak etiketlemektedir. Literatürde sahte haber sınıflandırması yapan hemen hemen bütün modeller sayılan bu bileşenlerin farklı kombinasyonlarını farklı yöntemlerle ele almaktadır.

2.1.2. Sahte Haberin Karakteristiği, Etkileri ve Çeşitleri

Yapısal karakteristiğine bakıldığında; haberi oluşturan ana unsurun bazı çalışmalarda da [16] ifade edildiği şekilde “haber metni” olduğu söylenebilir. Haber metninin tamamlayıcı bileşenleri ses, görsel ve video eklentileridir. Bazı durumlarda bu bileşenler metnin bile önüne geçebilmektedir. Yapısal olarak haberin oluşumunu ve okuyuculara sunumunu sağlayan “haber kaynağı” bir diğer önemli unsurdur. Son olarak, bütünlük bir bakış açısıyla “haber okuyucudan aldığı tepki ve yayılımındaki rolü” yukarıdaki unsurlardan ayrı tutulamaz. Bu bağlamda, sahte haberin tespiti için bu unsurların dikkatle ele alınması gerekmektedir.

Sahte haberler genel olarak hem üretilme hem de yayılması bakımından bilinçli veya bilinçsiz motivasyonlar ile ortaya çıkmaktadır. Farklı çalışmalarda etkisi, çeşitleri ve karakteristiği açısından farklı şekillerde değerlendirilse de tüm çalışmalarda ortak olan görüş sahte haberlerin topluma yararlı bir etkisinin rapor edilmemiş olmasıdır. Çizelge 2.3'te sahte haberin üretilme niyeti ve potansiyel etkisine göre numaralandırılan tehlike boyutu gösterilmiştir. Bu çizelgede bilinçli veya bilinçsiz olarak oluşturulan sahte haberlerin farklı boyutlarda etki potansiyeline sahip olabileceği görülmektedir.

Çizelge 2.3 Sahte haberlerin oluşturulmasındaki niyet ve olası etkileri.

Oluşturulma Şekli	Etki potansiyeli/kapasitesi			
	Çok Zararlı	Zararlı	Etkisiz	Belirsiz
Bilinçli	1	2	6	5
Bilinçsiz	3	4	8	7

Sahte haberlerin özellikleri çeşitlerine göre değişmekle birlikte bazı çarpıcı özellikleri şu şekilde sıralanabilir:

- Sahte haberlerde, kullanılan dil daha fazla okuyucuya ulaşabilmek adına çarpıcı, çoğu zaman resmi olmayan (*informal*) ve uydurmadır.
- Paylaşılması kolay şekilde sunulur.
- Etkileyici bir üslup kullanılır.
- Haberi doğrulayacak zaman, yazar, tarih vb. damga bilgisi yoktur veya yanlıştır.
- Daha sonraki bölümlerde ayrıca ele alınacak olan belli kelimelerin birlikte görülmesi (*co-occurrence*) durumu gözlenir. Diğer bir deyişle; kullanılan dilin basit, dar bir çerçevede belli kalıp ve kelimeler üzerinde kendini tekrar ettiği görülebilir.
- Sahte haber kaynaklarının ve sahte haber yayan hesapların belirli aralıklarla kapanarak farklı adres ve kimliklerle tekrar ortaya çıkması söz konusudur.
- Özellikle içerik sahteliği içeren yanlış haberler genellikle bilinçli bir şekilde oluşturulmaktadır.
- Özellikle toplumun genelini ilgilendiren sahte haberlerde insanlar doğru olmadığını bildiği halde haberden olumsuz etkilenebilmektedir.

Sahte haberlerin, bireylere, toplumlara, kurumlara, organizasyonlara ve devletlere çeşitli şekillerde zarar verdiği ve politika, ekonomi, turizm, toplumsal kutuplaşma vb. birçok alana doğrudan ve dolaylı etkileri gözlemlenmiştir. Bu bakımdan, insanların kişilere, olaylara, kuruluşlara, organizasyonlara ve devletlere bakış açılarını değiştiren bir silaha dönüştüğünü söylemek yerinde olacaktır. Sahte haberi konu alan birçok farklı çalışmada

sahte haberin etkilerinin direk veya dolaylı görüldüğü uluslararası çapta olaylara değinilmektedir. Aşağıda, farklı boyut ve derecelerde sahte haberin neden olduğu, katalizör etki yaptığı veya ilişkilendirildiği dünya yakın tarihindeki bazı önemli olaylar sırasıyla listelenmiştir.

- 2000’li yılların başlarında Irak’ın kitle imha silahları ürettiğine dair raporlar ve bu silahlarla birçok ülkeye saldıracağı yönünde manipülatif haberler üzerine bölgedeki durumun daha karmaşık bir hal alması [17],
- Sri Lanka ve Myanmar gibi ülkelerde etnik azınlıklara karşı şiddeti tetikleyen kışkırtıcı nitelikte haberler [18],
- Arap Baharı ayaklanmasına neden olan elit güçler destekli sahte haberler [19],
- Ukrayna-Rusya Krizi [20],
- Boston Maratonu bombalanması sonrasında sosyal medyada dolaşan yanlış bilgiler [21],
- 2016 Amerikan Başkanlık Seçimleri ile ilişkilendirilen sahte haber hikâyeleri [22],
- Pizzagate skandalı ve komplo teorileri [23],
- Brexit tartışmaları ve propagandalar [24],
- Fransa Cumhurbaşkanlığı seçimleriyle ilgili bilgi çarpıtma (dezenformasyon) ve sahte sosyal medya hesapları (*bot*) operasyonları [25],
- Cambridge Analytica - Facebook olayı sonucunda haberlerin şirketlere verdiği zararlar [26],
- Zeytin Dalı, Afrin harekâtları ile ilgili özellikle dış basında çıkan yanlış bilgiler, karalamalar [27],
- Mülteciler hakkında çevrim içi dolaşım halindeki, toplumu kutuplaştıran söylenti tarzında haberler,
- COVID-19 (*Coronavirus*) yayılımında toplumu huzursuz eden sahte haberler (5G baz istasyonlarının virüs yaydığı iddiaları gibi).

Yukarıda belirtilen olaylar dışında da sahte haberler birçok toplumsal ve bireysel olaya direk veya dolaylı şekilde etki etmektedir. Tüm olumsuz etkilerinin yanı sıra insanların gerçeklik algısını ve dünyaya bakış açılarını sarsıyor olması da gelinen noktada durumun farklı bir sosyolojik sonucu olduğunu göstermektedir.

Sahte haber özellikle bilgisayar bilimleri alanında nispeten yeni çalışılmaya başlayan bir konu olmasından dolayı, sahte haberin çeşitleri için birçok farklı kaynaktan birçok farklı görüş bulunmaktadır. Bu bakımdan üzerinde fikir birliğine varılmış belirli çeşitleri olduğu söylenemez. Literatürde bir kısmı 2.1.4'te referans edilen sahte haber tespiti çalışmalarında çeşitli sahte haber alt başlıklarına odaklanılmıştır.

Sahte haberin birçok alt başlık olarak çeşitlenmesindeki ana etken çok çeşitli amaçlarla oluşturulmasıdır. Rashkin ve arkadaşları [28] tarafından yapılan çalışmada, sahte haberler bilgi kalitesi ve oluşturulma niyetine göre kümelenmiştir. Bir diğer çalışmada [29], Meel ve Vishwakarma sahte haber kategorilerini Venn Şeması şeklinde göstermiş, sahte haber bazı diğer türlerle kesişen ayrı bir başlıkta konumlandırmışlardır. Literatürde sahte haberi tanımlamadaki zorluk, sahte haberi türlerine ayırmak veya hangi tip haberlerin sahte haber kapsamına gireceğini belirlemek hususunda da göze çarpmaktadır.

Tez kapsamında literatür taranarak yapılan gözlemlerimize ve elimizdeki veri setine ait (beğeni, retweet edilme, haberi yayan kaynak, sahte hesap vb.) bazı ayırt edici bilgilerine dayanarak sahte haberlerin tüm olası türlerinin, Şekil 2.1'de gösterilen iki boyut ve dört bölge arasında geniş bir varyasyonda değişken hızda bir yayılım gösterdiği söylenebilir. Sahte haber türlerini noktasal veya bölgesel bir dağılımda konumlandırmanın oldukça zorlu farklı bir çalışma olabileceğini değerlendirmekteyiz. Bu nedenle çalışmamızda sahte haber türleri noktasal olarak konumlandırmak yerine bir kümeleme mantığı ile 4 bölge içerisinde değerlendirilmiştir. Şekilde dört bölge risk ve yayılım hızına göre büyük ve küçük harf olacak şekilde ifade edilmiştir. Buna göre, büyük "Y" harfi "Hızlı yayılım", küçük "y" harfi "Yavaş yayılım" anlamına gelmektedir. Aynı şekilde risk seviyesi de "R" ve "r" gösterimleri sırasıyla "Yüksek Risk" ve "Düşük Risk" anlamına gelir. Bu bölümde, sahte haber türlerinin İngilizce karşılıkları ile kısa tanımları da yapılmıştır. Bunlar:



Şekil 2.1 Sahte haber çeşitlerinin risk seviyesi ve yayılım hızına göre dağılımı.

Yönlendirici Başlıklar (Misleading headings) : Başlık içerik bilgisi arasında ilgi bulunmayan sahte haberlerdir. Özellikle bakış açısı ve görüş bulma çalışmalarında [30, 31] haber başlıklarının (*news header/headline*) kullanıldığını ve haber için önemli bir bileşen olduğunu görmekteyiz.

Tık Tuzağı (Click-baits) : Genellikle reklam yönlendirmeleri taşıyan, ticari veya diğer kazançlar için sadece kullanıcıyı sitede biraz daha fazla tutabilmek için haber niteliği taşımayan ve genelde birbiri ile ilişkisi olmayan farklı içerikler sunan sahte haber çeşididir. Okuyucunun haber sitesine olan güvenini sarsan önemli bir unsurdur.

Asparagas Habercilik (Yellow journalism) : Duyulduğunda veya okunduğunda heyecan yaratan, insanları etkileyen, genellikle herhangi bir kanıt içermeyen sansasyonel ve düzmece haberlere denir. Magazin haberlerinde sıklıkla rastlanan bir türdür.

Yanlı Kurgu (Biased fiction) : Taraflı, çoğu zaman politik kurgu haberlere denir. Genellikle var olan gerçek bir olayın saptırılmasıyla oluşturulur. Bu tarz haberlerin sosyal medyada paylaşımlarında toplumsal kutuplaşma izleri görülmektedir.

Sahte Hesaplar (*Sockpuppets*) : Sahte/Robot hesap (*bot account*) veya argo tabiri ile “troll” şeklinde tanımlanan hesapların yaydığı sahte haber türüne veya yaptığı faaliyete verilen isimdir. Çoğunlukla haberin yayılmasını hızlandırmak için bilinçli şekilde yönetilmektedir.

Şaka, Latife tarzı mizah haberleri (*Hoax, tidbits, humour*) : Bir gerçeğin abartılması veya olmayan bir durum üzerinden eğlendirme ve mizah amaçlı sahte haberlerdir. Çoğu zaman yayımlandığı platform ve yayın bildirisinde gerçek dışı olduğu bildirilir. Tıpkı hiciv tespiti gibi, literatürde ayrıca ele alındığı çalışmalara rastlamak mümkündür [32].

Hiciv, parodi (*Satire, parody*) : Eleştirel bakışla, kinayelerle süslenen veya doğrudan şaka yollu anlatımları kullanan haber biçimidir. Dolaylı anlatım ve benzetmeler sıklıkla kullanılır. Hiciv içeren bir metinde kelime ve ifadeler farklı saklı/mecazi anlamlar yüklendiğinden, hiciv tespiti oldukça zor bir görevdir. Literatürde sahte haber tespitinden ayrı olarak sadece hiciv tespiti üzerine yoğunlaşmış çalışmalar bulunmaktadır.

Partizan haber (*Extreme bias*) : Aşırı şekilde tarafını belli eden, objektif olmayan sahte haberlerdir. Yanlı kurgu haberlerle birçok yönden ortak özellikler göstermektedir.

Komplo Teorileri (*Conspiracy theory*) : Toplumun genelini ilgilendiren önemli bir konu hakkında, toplumda şüphe uyandırmayı hedefleyen gerçekte olandan farklı/uydurma sahte haberlerdir. İnanılmasa dahi toplumu tedirgin etme potansiyeli açısından olumsuz etkileri bulunmaktadır.

Nefret Söylev ve Haberleri (*Hate news and speeches*) : Genel olarak politik motivasyonlarla beslenen seçim mitingleri, parlamento oturumlarında geçen konuşmaları (genellikle belirli bir kısmını) konu edinen haberlerdir. “Deep fake” şeklinde tabir edilen ve teknoloji kullanımı ile yapılan içerik manipülasyonlarında kullanıldığı da gözlemlenebilir.

Uydurma haber (*Newspeak, Make-up*) : Çoğunlukla hiçbir referansı olmayan ve çok kısa sürede hazırlanmış sahte haberlere denir. Bilinçli olarak hazırlandığı ve kötü niyetli olduğu söylenebilir.

Dedikodu haberler (*Rumours*) : Söylenti şeklinde yayılan genellikle tutarsız uydurma haberlerdir. Tevatür, söylenti ve rivayet şeklinde de ifade edilmektedir. Bu tür sahte haberler yayıldıkça içeriği de farklılaşabilir. Dedikodu/söylenti yayılımı (*Rumour diffusion*) tıpkı metinsel aldatma çalışmaları gibi uzunca bir süredir çalışılan konulardan biridir. Bu bakımdan literatürde sahte haber tespiti ve dedikodu tespiti gibi farklı şekilde ele alındıkları görülmüştür.

Yanlış bilgilendirme (*Misinformation*) : Gerçek bir olay ve kişi kurgusu olmasına rağmen içeriğinde hatalı ve yanlış bilgiler barındıran sahte haberlere denir. Yanlış ve yanıltıcı bilgi içeren bu tarz haberlerde genellikle kasıt yoktur.

Provokatif haber (*Provocative news*) : Özellikle kriz dönemlerinde doğru bilgiye ulaşmak zorlaştığında ortaya çıkan kışkırtıcı haberlerdir. Bu tip haberlere, en çok sosyal medya ortamında karşılaşılmaktadır.

Şehir Efsaneleri (*Urban legends*) : Belli dönemlerde tekrar tekrar ortaya çıkan genellikle toplum tarafından şüpheyle yaklaşıldığı için risk potansiyeli az ama oldukça hızlı yayılan sahte haberlerdir.

Tanımsız sosyal medya hesapları (*Social media bots*) : Sahte bir haberi üreten ve/veya yayan sahte hesaplardır. Sahte profil ve grupları yöneten kötü niyetli girişimlerdir. Bir haber çeşidinden ziyade tüm sahte haber türlerine etkisi olan bir aktör olarak ele alınabilir.

Manipülasyon (*Manipulation, deepfake*) : İnsanları etkilemek ve yönlendirmek için üretilen sahte haber türüdür. Bir kavram olarak tek bir haberden ziyade sürekli bir

yönlendirme sürecini ifade eder. Basit birkaç ifadeden video içerikleriyle oynanmasına kadar geniş bir çerçevede kendini gösterebilir.

Propaganda : Hedef kitlelerin bir konu hakkındaki inancı ve motivasyonu etkilemek ve yönlendirmek için yapılan, manipülasyon ile iç içe geçmiş bir süreci ifade eder. Propaganda kavramının çok geniş bir kullanım alanı bulunmaktadır. Bu bakımdan sahte haberden farklı ele alındığı çalışmalar bulunmaktadır.

Sahte Bilim (*Pseudoscience, junk science*) : Şehir efsanelerine benzer şekilde ortaya çıkan akılcı rasyonel yaklaşımlardan uzak, temelinde çoğu zaman ticari çıkar barındıran sahte bilimsel haberleri ifade eder.

Yukarıda literatürde kullanılan isimleri dikkate alınarak tanımlanan sahte haber çeşitlerini çoğaltmak ya da daha genel olarak belirtip azaltmak mümkündür. Tez boyunca verilen bazı çeşitler üzerinden örneklemeler yapılacaktır.

2.1.3. Sahte Haberle Mücadele ve Zorlukları

İnsanoğlunun ilgi çekici ve merak uyandırıcı bilgi yayma eğilimi, sahte haber miktarının giderek artmasına neden olmaktadır. Şüphesiz günümüzde internet kullanımının yaygınlığı nedeniyle bu eğilimin en kolay şekilde tatbik edilebildiği ortamlar çevrim içi metinleri barındıran sosyal medya platformları ve çevrim içi haber siteleridir. Bu doğal eğilimin sebep olduğu ve mücadeleyi zorlaştıran ana nedenlerden biri sahte haberin yayılımının gerçek habere göre çok daha hızlı [33] ve etkileşiminin daha fazla olmasıdır.

Son dönemlerde, kritik etkileri giderek artan sahte haberler hakkında genel bir toplumsal farkındalık oluştuğu söylenebilir. Fakat önceki bölümlerde belirtildiği şekilde bu farkındalığın somut çözümlere katkısı oldukça düşüktür. Son yıllarda akademik çalışmaların yanı sıra sahte haber ile mücadeleyi konu alan uluslararası yarışmalar³,

³ <http://www.fakenewschallenge.org/>

konferanslar⁴, paneller ve seminerler düzenlenmeye başlamıştır. Sahte haberlerle mücadelenin genel olarak üç farklı şekilde yürüdüğünü söyleyebiliriz. Bunlar:

- Okuyucu farkındalığı,
- Haber doğrulama (*fact-checking*) kuruluşları ve örün (*web*) siteleri,
- Otomatik tespit/algılama sistemleri/çalışmaları (*automated detection systems*) ve araçlarıdır.

Bunlardan ilki olan okuyucu farkındalığı, bireylerin ve toplumun sahte haberlere karşı farkındalığını ifade eden koruyucu bir kalkan olarak tanımlanabilir. Bu farkındalığın toplumsal anlamda bazı somut çabalara dönüştüğü de gözlemlenmektedir. Toplumsal farkındalığı arttırabilmek adına bazı resmi organizasyonlar pratik olarak başvurulabilecek yazılı ve görsel rehberler, kontrol listeleri vb. üretmektedirler. Örneğin; IFLA⁵ (Uluslararası Kütüphane Dernekleri Federasyonu), sahte haberleri tespit edebilmek için sekiz adımdan oluşan bir infografik yayınlamıştır. Yine RADAR yaklaşımı [34] ve CRAAP testi [35] okuyucunun farkındalığını arttırmak için ortaya çıkarılmış gayretlerdendir. Fakat, sahte haberin doğası ve insanların sansasyonel bilgileri yaymaya yönelik güdülerine göre akıllıca kurgulanması nedeniyle bu tarz stratejiler sahte haber ile mücadelede kontrolü sağlamak için çok yetersiz kalmaktadır. Bazı çalışmalar, insanların yalanı tespit etmede [36] veya haberin sahte veya gerçek olup olmadığına karar vermede [16, 37] başarılı olmadıklarını belirtmektedir. Bunun daha ötesinde insanlar bu tespitleri yaparken tarafsız olamamaktadır. Şöyle ki; insanlar çoğu zaman özellikle politik ve siyasi görüş içeren tartışmalı sosyal konularda kendi görüşlerine ters düşen gerçekleri görmezden gelmektedir [38].

İkinci strateji olarak tanımlayabileceğimiz haber doğrulama kuruluşları ve örün siteleri ise yoğun insan çabası ve yarı-otomatik araçlar kullanarak yürütülen mücadele yöntemleridir. Son zamanlarda, çevrim dışı ve içi olarak bu yöntemleri kullanmaya çalışan ve sayıları her geçen gün artan haber doğrulama kuruluşları [39], genç yenilikçi

⁴ Bobcatss 2020

⁵ IFLA (International Federation of Library Associations and Institutions), "How To Spot Fake News," 2017.

şirketler⁶ (*start-up*) ve bu amaçla oluşturulan projeler ortaya çıkmıştır. Ne yazık ki, çoğunlukla manuel olarak yürütülen bu sistemler sürekli bir şekilde üretilen çok miktarda dijital içerikle baş edememektedir. Bu nedenle dijital haber içeriklerinin spor, politika, hiciv vb. alt kümelerine odaklanmaktadır. Sahte bilginin yayılım hızı düşünüldüğünde bu alt kümelerde de benzer problemler ortaya çıkmaktadır. Bu nedenle otomatik ve doğrulanmış sistemler kaçınılmaz bir son seçenek olarak ortaya çıkmaktadır.

Üçüncü strateji ise yapay zekâ teknolojileri sayesinde çok daha etkin hale getirilebilecek olan otomatik olarak gerçeklik doğrulaması veya sahte haber tespiti yapmak için tasarlanan sistemlerdir. Sahte haberlerin daha hızlı yayıldığı mecralar olan Facebook, Twitter ve Google gibi sosyal medya platformları ve teknoloji şirketleri saygınlıklarını korumak adına bu tarz çözüm arayışlarına girmişler ve akademik çalışmaları teşvik etmişlerdir. Yapay zekâ ve insan incelemesini birlikte kullanmaya başlamışlardır. Hatta bazı önemli durumlar ve olaylarda hesap ve içerik engellemesi, etiketlemesi yaptıkları görülmektedir. Ayrıca yoğun manuel süreçler yürüten, haber doğrulama (*fact-checking*) organizasyonları ile ortak çalışmalar yürütmektedirler.

Son yıllarda, birçok araştırmacı otomatik çözümleri ve farklı metodolojileri uygulanabilir hale getirmek için önemli çalışmalar da [16, 40-42] bulunmaktadır. Geline nokta da problemin çözümüne yönelik önemli adımlar atılmış olsa da halen tatmin edici seviyelere ulaşıldığı söylenemez. Dünya genelinde haber doğrulama organizasyonlarının ve buralarda çalışan insan sayısının artışı [39, 43] uygun (*optimal*) çözümlerin henüz bulunmadığının göstergesi olarak kabul edilebilir.

Multidisipliner bir konu olması nedeniyle, Lazer ve arkadaşlarının da çalışmalarında [14] belirttiği gibi sunulacak çözümlerinde bu doğrultuda olması önemlidir. Nitekim tez kapsamında geliştirilen otomatik tespit modelimizin Bölüm 2.4 'te anlatılacak olan Dijital Kütüphaneciliği merkezine alan bir sistematikte kullanımı ile daha yararlı hale gelebileceğini değerlendirmekteyiz.

⁶ Storzy, Factmata

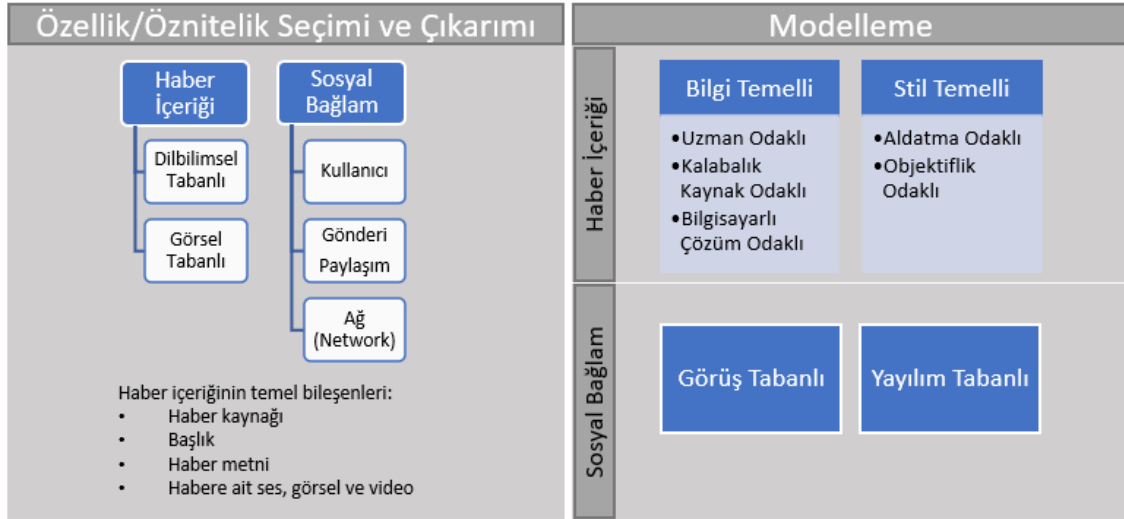
2.1.4. Sahte Haber Tespit Çalışmaları

Sosyal medya ve geleneksel medyanın çevrim içi ortamdaki sunumu, sahte haberin iki ana yayılma ortamıdır. Dolayısıyla tüm çalışmalar bu yayılma ortamlarını temel alarak çözüm bulmaya ve stratejiler geliştirmeye çalışmaktadır. Sahte haber tespit çalışmalarının ortaya çıkışı ve gelişimini; sosyal bilimler açısından sahte haberleri ortaya çıkartan toplumsal/sosyolojik nedenler ve bunlara yönelik çözüm önerileri, sayısal bilimler açısından ise problemin çözümünde kullanılacak matematiksel modellemeler, algoritmalar ve sistemler olarak iki ana çerçevede incelemek mümkündür. Bu bölümde özellikle Bilgi Teknolojileri (BT) bölümlerinin konuya artan ilgisi çerçevesinde konu ile ilgili yapılan çalışmalara değinilecektir.

Toplumsal etkilerinin fark edilebilir düzeye geldiği son 4-5 yılda sahte haber kavramı gittikçe popüler hale gelmiştir. Öyle ki bu terim 2017 yılında Collins sözlüğü tarafından yılın en çok kullanılan terimlerinden biri olarak gösterilmiştir. “Sahte haber” kavramının, birçok çalışmada da ifade edildiği gibi Amerikan Başkanlık seçimlerinden sonra daha da popüler bir kavram haline geldiği ve bu alandaki akademik çalışmaların da aynı doğrultuda hız kazandığı görülmektedir [6]. Bahse konu seçimlerde, sahte haber dolaşımında sosyal medyanın özellikle Facebook’un büyük etkisi olduğu belirtilmiştir [44]. Gröndahl ve Asokan [45], Bilgi Teknolojileri (BT) bölümlerinden araştırmacıların konuya artan ilgisini sahte haber tespitinde uzman olmayan insan performansının oldukça düşük olması ile ilişkilendirmiştir. Bu neden araştırmacıları bilgisayar temelli çözümler bulmaya motive etmiştir.

Probleme bilgisayar temelli çözümleri esas alarak yaklaşan çalışmaların, metinsel aldatma tespiti çalışmalarında izlenen nicel yaklaşımları temel aldığı görülür. Bunun yanında problem klasik metin sınıflandırma problemlerinden bazı noktalarda farklılaşmaktadır. Örneğin, klasik metin sınıflandırma problemlerinde kullanılan kitle kaynak (crowdsourcing) yöntemleriyle yapılabilecek etiketlemelerin bu probleme aynı ölçüde katkı sağlayacağı söylenemez. Problem bir yönüyle sahte ve gerçek haber yazan yazar tanıma problemine benzerken, öte yandan ağ bilgisinden yararlanması açısından istenmeyen e-posta (*spam*) tespiti problemine de benzemektedir. Bu bölümde ve sonrasında kullanılacak terminolojiyi daha iyi anlamak açısından sahte haber tespit

çalışmalarının dayandığı temel kıstas ve modellemeler Şekil 2.2’de şematize edilmiştir. Birçok farklı çalışma da sahte haber tespitindeki modellemeler ve bazı özellikleri çeşitli şekillerde ele almıştır. Bu çalışmalardan birinde, Shu ve arkadaşları tarafından [15], kapsamlı bir literatür incelemesi yapılmıştır.



Şekil 2.2 Sahte haber çalışmalarının esas aldığı temel kıstaslar ve modellemeler.

Alandaki öncü çalışmalardan birinde, Conroy ve arkadaşları tarafından sahte haber tespitine yönelik yaklaşımlar kapsamlı bir şekilde ele alınmıştır [46]. Bu bağlamda ana yaklaşımlar; dilbilimsel ve ağ yaklaşımları şeklinde belirtilebilir. Bunlara ilave olarak kaynak güvenilirliği, anlamsal çözümleme ve melez yaklaşımlar başta olmak üzere çeşitli yaklaşımların kullanıldığı görülmektedir.

Çalışmalarda genel olarak makine öğrenmesi ve derin öğrenme teknikleri kullanılmıştır. Kullanılan yöntemler ise, temel bir dilbilimsel yöntem olan sözcük çantası/kesesi (bag of words) yaklaşımından karmaşık derin öğrenme yaklaşımlarına kadar geniş bir yelpazede değişiklik göstermektedir. Çalışmaların daha etkili performanslar elde edebilmek adına belirli haber alanlarına yöneldiği de gözlemlenmektedir. Aynı şekilde, çalışmalar birçok farklı konu üzerinde özgülleştirilmiştir. Bu konulardan bazıları; sosyal medya gönderilerindeki söylentilerin tespiti [47], istenmeyen gönderilerin (*spam*) tespiti [48], yorumlardaki aldatıcı fikirlerin bulunması [49] vb. şeklinde listeleyebiliriz.

Araştırmacılar tarafından en çok kullanılan yaklaşımlardan biri dilbilimsel yaklaşımlardır. Örneğin; Markowitz ve Hancock çalışmalarında N-gram, sözcük türü (*POS, part of speech*) vb. metodları esas alan dilbilimsel yaklaşımlar kullanmıştır [50]. Dilbilimsel yaklaşımları kullanan bir diğer çalışmada problem, içerikten bağımsız gramer (*CFG, Context Free Grammar*) kullanılarak kural tabanlı sözdizimsel analizlerle ele alınmıştır [51]. Bazı çalışmalarda ise; dilbilimsel yaklaşımlar diğer yaklaşımların tamamlayıcısı olarak kullanılmıştır [52, 53].

Anlamsal (semantik) yaklaşım ve dilbilimsel yaklaşımın kesişim noktasındaki söylev (retorik) temelli tespit için bazı çalışmalarda kullanılmıştır. Rubin, çalışmasında [54] tutarlılık ve yapı bakımından aldatıcı ve doğru hikayeler arasındaki sistematik farklılıkları tanımlamak için Retorik (Söylev) Yapı Teorisini (*RST, Rhetorical Structure Theory*) analitik bir çerçeve olarak kullanmıştır. Benzer şekilde söylem düzeyine odaklanan bir diğer çalışmada [55], İngilizce haberler için retorik yapılar, vektörel alan modelleyiciler olarak kullanılmıştır.

Bazı araştırmacılar, görüş tespitine (*stance detection*) yönelik çalışmalarında derin öğrenme tekniklerinin karmaşık mimarisinin genelleştirilebilir olmadığına işaret ederek geleneksel makine öğrenme metodolojisini kullanmış [31], bazı araştırmacılar ise geleneksel makine öğrenme tekniklerinin bazı yetersizliğini belirterek derin öğrenme teknikleri uygulamıştır. Derin öğrenmeyi kullanan çalışmalardan birinde [56], yazarlar sahte haberlerin alt alanı olan kinayeli haber tespiti için duygu ve kişilik özelliklerine yoğunlaşan önceden eğitilmiş bir Evrişimsel (Konvolüsyonel) Sinir Ağına (*CNN, Convolutional Neural Network*) dayalı modeller geliştirmişlerdir.

Sahte haber tespiti; haber içeriğinin temel bileşenleri göz önüne alındığında, yoğunlukla metin tabanlı olarak çalışılsa da zamanla değişen sunumu ile haber artık metnin yanı sıra ses, görsel ve video öğeleri ile zenginleştirilmektedir. Bu bakımdan haberin tüm bileşenleri ile değerlendirildiği çalışmalara olan ihtiyaç ortaya çıkmaktadır. Ruchansky, bu ihtiyacı ele alan çalışmasında [16], hibrit(melez) bir model önerisinde bulunmuştur. Bu tarz çalışmalarda, sosyal ağlarda kullanıcılar ile haber arasındaki etkileşim gibi zaman içerisinde değişkenlik gösterebilecek üst verilerin çok dikkatli analiz edilmesi,

çalışmaların tutarlılığı ve pratikte uygulanabilirliği açısından çözümü zorlaştırıcı fakat gerekli bir husus olarak ortaya çıkmaktadır.

Bu tür verilerin toplanması ve etiketlenmesinin zaman alıcı ve yorucu bir işlem olduğu düşünüldüğünde, hemen hemen bütün çalışmalarda araştırmacıların halka açık İngilizce veri setlerini kullanmaları şaşırtıcı değildir. Bu durum özellikle DDİ alanında çalışan araştırmacıların ana dilleri üzerinde çalışmamaları ve araştırmaların var olan veri setleri ile kısıtlı kalması sonucunu ortaya çıkarmaktadır. Çalışmalarla ilgili bir diğer husus ise, genel olarak makine öğrenmesi problemlerinde üzerinde çalışılan alan daraltıldığında (spor, politika, mizah, eleştiri vb.) göreceli daha başarılı sonuçlar alınmasıdır. Bu durum sahte haber tespiti içinde geçerlidir. Bu da her ne kadar çalışmalar belli bir alanda başarılı olsa bile tüm alt alanları kapsayan bir sistematığın ortaya konulamamasına neden olmaktadır.

2.2. Türkçe Doğal Dil İşleme ve Sahte Haber Analiz Yaklaşımları

Bu bölümde DDİ kavramı, Sahte Haber Analiz yaklaşımları ve Türkçe DDİ'de karşılaşılan zorluklar anlatılacaktır.

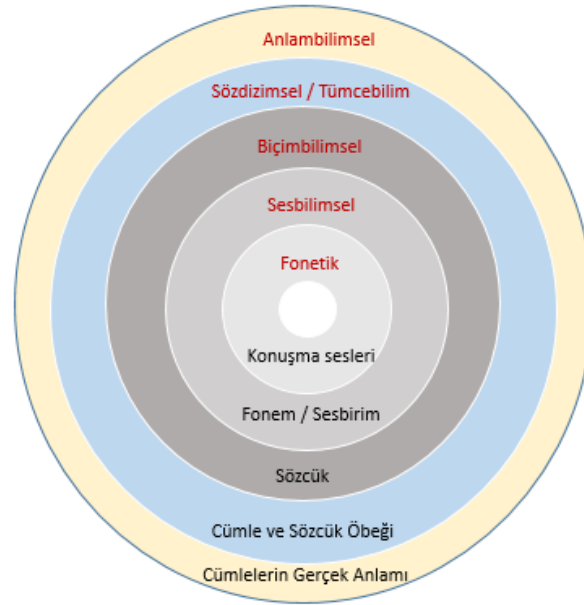
2.2.1. Türkçe Doğal Dil İşleme ve Metin Madenciliği

Doğal Dil İşleme (DDİ), literatürde sık kullanılan İngilizce ifadesiyle Natural Language Processing (NLP), doğal dillerin kurallı yapısının ele alınarak çözümlenmesi veya yeniden üretilmesi için matematiksel modelinin oluşturulmasını amaçlayan bir disiplindir. Bilgisayar bilimlerinin ana disiplinlerinden biri olmasının yanı sıra, dilbilim ve yapay zekâ alanlarının alt kategorisi olarak da tanımlanan bu disiplin matematik, robotik, psikoloji vb. birçok farklı disiplin ile de etkileşim içindedir [57, 58].

Dil uzun bir zaman diliminde oluşmuş ve gelişimine devam eden karmaşık yapıya sahip canlı bir olgudur. Dil zamanla özellikle eski kelimelerin kaybolması ve yeni kelimelerin eklenmesi şeklinde değişime uğrarken, dilin yapısal karakteristiğinin çok fazla değişkenlik göstermeyen kararlı bir yapısı bulunmaktadır. Doğal dil, insanlar arasındaki iletişimin temelini oluşturduğu için kültür ve toplumsal birçok kavramın ana unsuru

olarak görülebilir. Bu kavramların başında ulus kavramı gelmektedir. Ülke isimlerinin çoğunlukla o ülkede kullanılan dilden türemesi buna güzel bir örnektir. Bu doğrultuda dillerin kütüphaneleri, yapısı ve terimleri toplumdan topluma değişen birçok farklılık barındırır. Bu kapsamda, insanların konuştuğu dil anlamına gelen “Doğal Dilin” makineler tarafından işlenmesinde de bu farklılıkların gözetilerek o dile özgü çalışmaların yapılması bir zorunluluk olarak ortaya çıkmaktadır.

Dilbilim; DDİ çalışmalarında da sıklıkla kullanılan, alt fazlarda diğer bir deyişle farklı analiz seviyelerinde ele alınmaktadır. Bu alt fazlar Şekil 2.3'te şematize edilmiş ve aşağıda kısaca anlatılmıştır.



Şekil 2.3 Dilbilimin temel analiz seviyeleri

Fonetik (*Phonetics*), insanların konuşurken kullandığı seslerin incelenmesi ile ilgilidir. Özellikle ses türü girdiler üzerindeki çalışmalarda fonetik analiz kullanılır. Sesler, kelime telaffuzları, vurgular ve seslendirme dalgalanmaları sıklıkla analiz edilen konular arasındadır. Burada, DDİ’de çalışmaların metin ve ses olmak üzere iki ana girdiye yoğunlaştığını belirtmek yerinde olacaktır.

Ses bilimi diye de bilinen Fonoloji (*Phonology*) de benzer şekilde dil içindeki seslerin işlevlerini ve belirli dil seslerini (fonemler) inceler. Dili meydana getiren ses birlikteliklerini ele alır.

Biçimbilimsel (*morphological*) analizin temel inceleme alanı, dilin anlam taşıyan en küçük parçaları olan biçim birimlerdir (*morpheme*). Sözcükleri, diğer sözcüklerle olan ilişkilerini, nasıl oluştuklarını ve sözcüklerin kök, gövde ve ek gibi bileşenlerinin yapısını çözümler. Dildeki kelimelerin yapılarını (kök-ek-gövde) inceleyen katmandır. Türkçenin sondan eklemeli dil yapısı düşünüldüğünde, biçimsel olarak karmaşık yapıda olan dillerde sıklıkla karşılaşılan biçimbirimsel belirsizlik (*morphological ambiguity*) problemi açısından da bu analiz çok önemli bir analizdir.

Tüm diller, morfolojik kurallarla birlikte her dilin gramerini oluşturan temel sözdizimsel kurallara da sahiptir. Sözdizimsel (*syntactic*) analiz, cümlelerin ve kelime öbekleri ile insanların anlamlı bir şekilde iletişim kurabilmeleri açısından nasıl kullandıklarını üzerine yapılan incelemelerdir. Cümle yapılarını inceleyen katmandır. Gramer yapısal analiz olarak da adlandırılabilen bu analiz de cümle seviyesinde dil kuralları işletilir. Doğru anlama ulaşabilmek için sözcüklerin dizilişi, konumları, bağlaçların kullanımı vb. analizlerle anlam çözümlemesi yapılır.

Anlambilimsel/Anlamsal analiz, genel olarak cümlelerin tam olarak neyi ifade ettiğini inceler. Birçok dilde kelimelerin birden fazla anlamı olabildiğinden bu kelimelerin hangi gerçek dünya nesnesini veya kavramını ifade ettiği bu analizin temel araştırma alanına girer. Anlamsal analize, Edimbilim (pragmatics)'te dahil edilebilir. Edimbilim, cümlelerin bağlama nasıl anlam kattığının belirleyen hiyerarşik olarak daha üst seviyede bir inceleme alanıdır.

Doğal dil problemleri bu analiz seviyelerinin yanı sıra sözcüksel analiz (*lexical analysis*), duygu analizi (*sentiment analysis*), söylev analizi (*discourse analysis*), sözcük türü (*POS, part of speech*) bulma vb. ara analizleri kullanmaktadır [59-61]. Bu noktada sözcük türü (*POS*) analizi veya literatürde sıklıkla kullanılan ismi ile Sözcük Türü

Belirleme/İşaretleme (POS Tagging), bir metni oluşturan öğelerin görevini/türünü belirten bir etiketlemenin yapılmasıdır (sıfat, isim, fiil, zarf vb.). Söylev/Söylem analizinde ise, metnin içeriğinden daha öte de gerçek dünyadaki karşılığını bulmak hedeflenir. Metni kimin ifade ettiğini, neye dayandığını, amacını inceleyerek doğru anlamı yakalamaya çalışır [62]. Yazar tanıma problemlerinde de bu analizin sıklıkla kullanıldığı görülmektedir. Duygu analizi de bu iki analiz gibi Türkçe açısından oldukça popüler bir çalışma alanıdır. Bu tez kapsamında özellikle ikinci fazdaki sözlük (*lexicon*) oluşturma metodolojisi, duygu analizi çalışmalarında oldukça sık kullanılan bir yaklaşımdır. Bu bakımdan çalışmamızda izlenen süreç açısından duygu analizinde sözlük kullanan çalışmalarla metodolojik ve terminolojik benzerlikler gözlemlenebilir.

DDİ, doğrudan ve dolaylı olarak çok farklı araştırma alanlarındaki problemlerin çözümünde çalışılmaktadır [61, 63, 64]. Bunlardan öne çıkan bazıları şu şekildedir:

- Metin ve içerik sınıflandırma,
- Metin etiketleme, seslendirme,
- Duygu analizi,
- Varlık İsmi Tanımlama (*NER, Name Entity Recognition*),
- Makine Dili Çeviri (*Machine Translation*),
- Metinlerden bilgi çıkarımı,
- Anlamsal Ayrıştırma,
- Soru-Cevap makineleri,
- Metin özetleme,
- Ses ve karakter tanıma,
- Yazım ve imla hatası denetimi,
- Diller arası çeviri,
- Konuşma üretme ve algılama.

DDİ disiplininde ortaya çıkan bazı zorluklar aşağıdaki başlıklarla ifade edilebilir [63, 65-68]:

- Dilin formal yapısına uymayan içerikleri de ihtiva edebilmesi.
- Sözcük ve morfolojik çözümlenmeler gibi çeşitli seviyelerde muğlaklık (belirsizlik) problemleri.
- Aynı ifadenin farklı bağlamlarda farklı terimlerle edilebilmesinden kaynaklanan anlam tespiti ve eş/zıt anlamlılık (*synonym/antonym*) tespiti problemi.
- Aynı varlığı referans eden tüm ifadeleri bulma zorluğu olarak ifade edebileceğimiz özdeşlik çözümlenme (*coreference resolution*) problemi.
- Mecazi ve dolaylı anlamlara sahip atasözü, deyim vb. kalıpların yarattığı anlam farklılıkları.
- Kinaye, eleştiri, argo, alay ve hiciv vb. kullanımlardaki örtük anlam ve terim karmaşası.
- Farklı dillerin farklı kurallara sahip yapısı nedeniyle ortak kurallar bütünü oluşturulamaması ve Sonlu Durum Özdevinirleri/Makineleri (*FSA, Finite State Automata*) ile dilleri modellemede yaşanan zorluklar.
- Birçok dilde gelişmiş dil kaynaklarının, kütüphanelerinin ve araçlarının İngilizcede olduğu kadar gelişmiş olmaması/yetersizliği.
- Arka plan bilgisine (*background knowledge*) ihtiyaç duyan metinlerin çözümlenmesi.
- Lehçe farklılıklarından kaynaklanan ses verisi çözümlenme sorunsalı.

DDİ ile iç içe geçmiş ve metin tabanlı çalışmalarda sıklıkla başvuru alan diğer bir disiplin ise Metin Madenciliğidir. Metin Madenciliği bir bakıma Veri Madenciliğinde (*Data Mining*) kaynak verinin metin olduğu inceleme alanıdır. Genel olarak Metin Madenciliği (*Text Mining*), düzensiz metinlerden örtük bilgiyi çıkarmayı hedefleyen ve metni işleyerek düzenli, yapısal bir biçimde ele alınması sağlayan süreç olarak tanımlanmaktadır. Diğer bir ifadeyle yapılandırılmamış metin verisinin yapısal bir biçime çevrilmesidir. İşlenerek düzenli hale getirilen bu veri üzerinden yapılan analizlerle metinlerin içinde barındırdığı ilişkiler, örüntüler ve hatta hipotezler elde edilebilir.

Metinlerin sınıflandırılması, özetlenmesi, kümelenmesi, metinlerden konu çıkarımı, duygu analizi, iş zekâsı, varlık ilişki modellerinin oluşturulması, kayıt yönetimi, bilgi erişimi vb. çalışmalarda metin madenciliği yöntemlerine sıklıkla başvurulmaktadır. Esasında; DDİ, Metin Madenciliği, Veri Madenciliği, Yapay Zekâ ve Makine Öğrenmesi gibi disiplinlerin günümüzde birbiri ile kesişen teknik ve yöntemleri kullanarak birbirlerini tamamladığı ve bu kapsamda bu tezde de olduğu gibi birçok çalışmada birlikte kullanıldığı söylemek mümkündür. Birçok çalışmada Metin Madenciliği, DDİ ile birlikte kullanılmaktadır. Buradaki temel farkın; Metin Madenciliğinin biraz daha istatistiksel sonuçlarla ve odağına metni alarak ilerlemesi, DDİ çalışmalarının ise yapay zekâ yöntemlerinden yararlanarak bu bölümün başında tanımlanan dilbilim analizlerine odaklanarak ilerlemesi olduğu söylenebilir. Bu doğrultuda çalışma kapsamında hem Metin Madenciliği hem DDİ teknikleri ve analizlerinin ayrı ayrı ve birlikte kullanıldığı örneklere rastlamak mümkündür.

2.2.2. Sahte Haber Analiz Yaklaşımları

Metinsel aldatma ve sahte haber tespiti üzerine yaptıkları araştırmada [46], Conroy ve arkadaşları tarafından sahte haber analiz yaklaşımları aşağıdaki iki ana başlık altında özetlenmiştir:

- Dilbilimsel Yaklaşımlar
- Ağ (*Network*) Yaklaşımları

Bu yaklaşımlara Kaynak Analizini ana bir yaklaşım olarak eklemek mümkündür. Ayrıca bu ana yaklaşımlarla ilişkili olan Veri Temsili (*Data Representation*), Bağlı Veri (*Linked Data*), Sosyal Ağ Davranışı (*Social Network Behaviour*) ve Bölüm 2.2.1’de anlatılan diğer DDİ analiz seviyeleri de alt yaklaşımlar olarak eklenebilir. Çalışmalardaki genel eğilim birden fazla teknik ve yaklaşımın birlikte kullanılması şeklindedir. Sahte haber konusu, bu analizlerin yanı sıra özellikle sosyal bağlamdaki görüş ve yayılım tabanlı modellemelerde Çizge Tabanlı Çözümlemelere (*Graph-based Reasoning*) oldukça açıktır.

2.2.2.1. Dilbilimsel Yaklaşımlar

Tıpkı yalan söylerken kullanılan beden dili gibi metinsel aldatma ve sahte haberde kullanılan metin dili de sahtelik ve aldatmaya yönelik bazı ipuçları barındırır. Doğal olarak, sahte haber metinlerini bilinçli olarak oluşturanlar haberi mümkün olduğunca gerçeğe yakın bir şekilde aktarmak isteyecektir. Feng'in profil uyumluluğu ile sahte görüşlerin tespitini ele alan çalışmasında [69] belirttiği gibi gözlemlenmesi ilk bakışta zor da olsa sahte haber; zamir, bağlaç vb. cümle öğelerinin kullanımı ve bazı olumsuz anlam taşıyan kelimelerin kullanım sıklığı gibi gerçek haberden farklılaşan örüntüler içerebilir. Bu yaklaşım, birçok çalışmada ana veya Bölüm 2.1.4'te belirtildiği gibi tamamlayıcı olarak kullanılmıştır.

Metinsel sınıflandırma problemlerinde yoğunlukla kullanılmış geleneksel dilbilimsel analiz teknikleri şu şekilde özetlenebilir:

- Köklerine ayırma (*stemming*) ve kök çözümlenme (*lemmatization*),
- Kelime çantası/kesei (*Bag of words*) – kelime frekansları (*countvectorizer*), tf-idf, kelime ve harf bazında n-gram frekansları, Word2Vec gibi metin verilerini vektörlere çeviren çeşitli metin temsil yöntemleri (*text representation*) ve frekans hesaplamada kullanılan çeşitli metotlar,
- Sözcük türlerinin (*POS*) kullanımı, içerikten bağımsız dilbilgisi kurallarının kullanılması,
- İşlev (*Function*) kelime kullanımı,
- Eklerin kullanımı,
- Kelimelerin kümelenmesi, birlikte geçme matrisi (*co-occurrence matrix*), kelime filtreleme,
- Saklı anlam indeksleme, anlamsal uzay kullanımı,
- Noktalama işaretleri, kelime, cümle, devrik cümle, harf, ek, cümlelerdeki ortalama kelime ve harf, kelimelerdeki ortalama harf ve ek sayıları vb. stil işaretçileri (*stylometric*) ve istatistiksel bulguların kullanımı. Örneğin bu çalışmada sayı bazlı özniteliklerin kullanımında haberlerin göreceli olarak uzun veya kısa

olması göz önünde bulundurularak ortalama hesaplarından (μ) faydalanılmıştır.

Sahte haber tespiti çalışmalarında da yukarıda belirtilen başta veri/metin temsil yöntemleri olmak üzere bu teknikler ve çeşitli türevlerini, özellik odaklı (*feature-oriented*) makine ve derin öğrenme yöntemlerini ve diğer güvenilirlik analiz yöntemlerini kullanıldığı görülmektedir. Bunun yanı sıra duygu analizi, semantik ve retorik (söylev, üslup) analiz kullanılarak yapılan çalışmalar temelde sözdizimsel çözümlenmeleri kapsadığından dilbilimsel analiz başlığı altında toplanabilir.

Özellikle sözdizimsel ve biçimbilimsel analizlerin yapıldığı çalışmalarda dile özgü DDİ araçlarının ve kütüphanelerinin gelişmişlik seviyesi ön plana çıkmaktadır. Bölüm 2.2.3'te Türkçeye özgü dil zorluklarına karşı nasıl çözümler bulunabileceği tartışılmıştır.

Salt dilbilimsel yaklaşımların kullanıldığı çalışmalarda önerilen bazı modellerin, bağlamsal bilginin ele alınmamasına rağmen, hassasiyet/doğruluk (*accuracy*) anlamında göreceli iyi performanslar gösterdiği görülmektedir. Buna rağmen bazı çalışmalarda bu yaklaşımın tek başına kullanılmasının yaratabileceği eksiklikler ortaya konmuştur. Örneğin, Conroy ve arkadaşları özellikle gerçek zamanlı takip gerektiren ürün incelemeleri gibi ticari konularda dilbilimsel yaklaşımların genelleştirilebilir olmadığını savunmuştur [46].

Bu bakımdan sahte haber tespitinde sıklıkla hibrit modellerin ve dolayısıyla çeşitli yaklaşımların birleşiminin kullanıldığı söylenebilir. Bu tarz çalışmalardan birinde [70], Braşoveanu ve Andonie anlambilimsel, sözdizimsel analiz ve makine öğrenme yöntemleri ile duygu, varlık ve metin gerçekliği gibi ilişkişel öğeleri ele almış, motive edici sonuçlar elde etmişlerdir.

2.2.2.2. Ağ (Network) Yaklaşımları

Twitter, Facebook, Microblog, Instagram, Whatsapp gibi gerçek zamanlı içerik üretilen sosyal medya platformları, mesajlaşma uygulamaları ve çevrim içi ortamların daha yoğun kullanılmaya başlaması ile sahte haber yayılımının daha hızlandığı aşikârdır. Bu

ortamlardaki metin özelliklerinin geleneksel medyada kullanılan haber metinlerinden farklı karakteristik göstermesi ve sosyal ağların karmaşık yapısı nedeniyle, sosyal ağ verisi kullanan çalışmalarda esas olarak ağ yayılımı tabanlı yaklaşımlar kullanılmıştır. Problemin özellikle eğilim ve örüntü (*trend and pattern*) analizi gerektiren özellikleri nedeniyle, Sosyal Ağ Analizi ve Çizge Teorisine üzerine çalışan araştırmacıların konuya ilgi duymalarını sağladığı söylenebilir.

Bilgi ağlarının kullanımı kanıtlanabilir verilerin teyit edilebilmesi açısından bağlı veri (*linked data*) analizi kapsamında önemli bir kaynaktır. İnternet ağı içerisinde her bilgiyi bir anlama sahip olacak biçimde modelleyerek, bunların birbirleriyle ilişkilerinin ortaya çıkarılması, bir nevi ontoloji ve veri tabanlarının oluşumu sağlayan çalışmalar (Dbpedia, Freebase, GDELT vb.) bunun için kullanılabilir. Bilgi ağlarının yanı sıra Varlık İsmi Tanımlama problemlerinde kullanıldığı şekliyle aktör, nesne, yer, olay vb. bilgilere ulaşabilmek sahte haberin erken tespiti için önemli ipuçları ihtiva edebilmektedir. Örneğin, bu çalışmada GDELT projesinden bir veri kaynağı olarak yararlanılmıştır.

Ayrıca sahte haber içerikleri üreten kaynaklar artık bu içerikleri sahte hesap ve profiller üzerinden yapmaktadırlar. Bu durum sosyal medyada kimlik doğrulamasını [71] özellikle gerçek ve sahte içerik tespiti için çok önemli hale getirmiştir. Son dönemlerde, sosyal medya platformlarının hesap veya paylaşımlara yönelik yaptığı etiketlemeler toplumun bir kısmı tarafından desteklense de bir kısım tarafından da iletişim özgürlüğünü kısıtladığı yönünde eleştirilmektedir.

Ağ yaklaşımı, söylenti yayılımı gibi ağ bilgisine ihtiyaç duyulan çalışmalarda sıklıkla kullanılan yaklaşımlardır. Bu çalışmalardan birinde Zhang ve arkadaşları tarafından tweetler; haber, haberi oluşturan (*creators*) ve haber konusu bakımından incelenmiş derin dağınık ağ modeli (*deep diffusive network model*) kullanılarak çalışılmıştır [72]. Benzer şekilde güvenilirlik değerlendirmeye (*credibility assessment*) yoğunlaşan çalışmalarda farklı ağ yapıları kullanılmıştır. Çevrim içi sosyal ağların ağ yapılarını ve kullanıcı ilişkilerini modelleyebilmek için kullanılan ölçeklenebilir grafik/çizelge üreteçleri (*synthetic graph generators*) araştırmacılara büyük kolaylık sağlamaktadır. Fakat, bu

sentetik modellerin gerçek dünyayı ne kadar modelleyebildiği ayrı bir tartışma konusudur.

Genel olarak hem dilbilimsel hem de ağ yaklaşımları kullanan yaklaşımların, belli noktalarda ve sınırlı alanlarda (*domain*) iyi sonuçlara ulaştığı söylenebilir. Fakat kapsamlı ve geniş ilgi alanında başarımın sağlanması için hibrit modellerin uygulanması gerekmektedir. Bu çalışmalarda gözden kaçan ve dikkat edilmesi gereken husus, karmaşık mimariler kullanılırken çözümün genelleştirilebilir ve uyarlanabilir (*adaptive*) olmasından uzaklaşılmasıdır. Özellikle ağ yaklaşımlarında gerçek zamanlı verinin değişken yapısı çok dikkatle ele alınmalıdır. Literatürde kullanılan temel alınabilecek bazı İngilizce veri setlerinde sadece metin bilgisinin bulunmasının bu çekincelerden kaynaklandığı değerlendirilmektedir.

2.2.3. Türkçe DDİ’de Karşılaşılan Zorluklar

Ural-Altay dil grubunun Türk dilleri ailesine giren Türkçe sondan eklemeli bir dildir. Türkçe, Doğal Dil İşleme açısından ilginç zorluklar içermektedir [73, 74]. Ayrıca eklemeli biçimbirim yapısı ve tümce öğelerinin esnek yer değişiklikleri biçimbilimsel olarak Türkçeyi zengin kılmaktadır. Türkçede iki ana ek yapısından yapım ekleri sonuna geldiği sözcüğün anlam ve türünü değiştirirken, çekim ekleri sözcüğün anlamını değiştirmeden türünü değiştirebilir. Bu yapısı ile Türkçe DDİ çalışmalarında öncelikli olarak dilbilimsel analiz yapılmasını gerekli kılar.

Özellikle biçimbilimsel analizin gelişmişlik seviyesine bağlı olarak türetilmiş bir sözcüğün köküne ulaşabilmekte eklemeli dillerde önemli bir zorluk olarak ortaya çıkmaktadır. Bu tez çalışmasında metinleri çözümlmek ve kelimelerin morfolojilerine ulaşmak için Zemberek kullanılmıştır [75]. Türkçe için Java programlama dilinde geliştirilmiş, açık kaynak kodlu DDİ kütüphanesi olan Zemberek’in, yazım denetimi, hatalı sözcük tespiti gibi özelliklerinden de yararlanılmıştır. Burada Zemberek’in orjinal haline müdahale edilmediğini belirtmekte fayda vardır. Özellikle morfolojik analiz kısmında ve Türkçenin kendine has özelliklerinden kaynaklı zorluklardan ileri gelen hatalı çözümler için kodun okunabilirliği ve elle müdahaleyi ortadan kaldırabilmek adına müdahale edilmemiştir.

Türkçe DDİ çalışmalarında karşılaşılan bir problemlerden biri olan Anlam Belirsizliği (*Word-sense disambiguation*) kavramı temelde bir sözcüğün birden fazla anlama sahip olması durumudur. Bu zorlayıcı durum, tez kapsamında genel amaçlı istatistiksel bir sözlük oluşturulduğu için hissedilmemiş olsa da gelecek çalışmalarda alana özgü (*domain specific*) sözlükler oluşturulmak istendiğinde titizlikle ele alınması gereken bir konudur. Burada duygu analizi çalışmalarında kullanılan tohum kelimeler (*seed words*) tarzında sözlüklerde bir seyreltmeye gidilmesi de ayrıca değerlendirilebilir. Örneğin, “kale” kelimesinin spor alanındaki kullanımı ile tarih veya çevre alanındaki kullanımı arasındaki anlam farkı Çizelge 2.4’te verilmiştir. Bu problem, genellikle metin bütünlüğü içerisinde anlamın belirginleştirilmesi ile çözülmeye çalışılmıştır. Bu doğrultuda, Türkçe için çeşitli anlam belirginleştirme çalışmaları mevcuttur [76, 77].

Çizelge 2.4 Türkçede anlam belirsizliği örnekleri.

Kullanımı	Metindeki Anlamı
Karahisar kalesi savunmaya yönelik inşa edilmişti. (Tarih terimi)	Savunma ve güvenlik amacıyla yapılan ve kullanılan, kalın ve yüksek duvarlı, burçlu, mazgallı yapı.
Bu maçta rakip kaleyi kim savunacak? (Spor (futbol) terimi)	Zemine dik iki direkle, bunları birleştiren ve zemine paralel olan bir üst direktten oluşan futbol terimi.

Özellikle Türkçe duygu sözlüğü oluşturma ve duygu analizi çalışmalarında gözlemlenen Türkçede eş anlamlılık, zıt anlamlılık ve olumsuzluk durumları da dil zenginliğinin beraberinde getirdiği bir zorluk olarak ortaya çıkmaktadır. Çünkü, Türkçede tam olarak eş anlamlısı ya da zıt anlamlısı olmamakla birlikte yakın anlam taşıyan birçok istisnai durum söz konusudur. Bu sorunun çözümünde farklı anlamsal sınıflandırmalar önem kazanmaktadır.

Türkçe için bir diğer zorlayıcı özelliği morfolojik belirsizlik problemidir. Bu özetle morfolojik analizi yapılan ham halindeki sözcüğün, anlamsal belirsizlik ile ilişkili bir

şekilde birden fazla morfolojik çözümlemesinin olması demektir. Tıpkı Zemberekte olduğu gibi biçimbilimsel ayrıştırıcı çözümlenen sözcüğün cümle içindeki anlamsal olarak farklı kullanımlarından dolayı morfolojik açıdan da birden fazla sonuç üretebilir. Burada doğru sonucu seçmek için morfolojik belirsizlik giderici (*morphological disambiguator*) olarak adlandırılan yöntemler kullanılmaktadır. Fakat bu yöntemlerin tam bir başarıyı sağlayabildiğini söylemek doğru olmaz. Bu sonuç dilin yapısal zenginliğine de atfedilebilir. Çizelge 2.5'te verilen sözcük analizleri bu belirsizliklerin tipik bir örneğidir.

Tez çalışmamız kapsamında her ne kadar göz ardı edilebilir seviyede olsa da bu durum Bölüm 3'te geliştirilen Sözlük Modellerinden Suffix ve Raw+Pos modellerinde çalışmayı etkilemiştir ve gelecek çalışmalarda göz önünde bulundurulması gerekir. Bu kapsamda gelecek çalışmalarda alana özgü olarak geliştirebilecek sözlükler için Kutlu ve Çiçekli'nin çalışmalarında elde ettikleri hibrit modelin [78], yapılan ön çalışmadan hareketle fayda sağlanabileceği değerlendirilmektedir. Haber metinlerinin puanlaması yapılırken kullanılan sözlüklerin referans alınması ve az sayıdaki sözcüğün tüm metne etkisinin çok az olduğunu bu etkinin ihmal edilebilir seviyede olduğunu gözlemlemek mümkündür. Problemin tespiti ve çözümüne yönelik farklı yaklaşımlar içeren bazı çalışmalar mevcuttur [79-81]. Bir diğer çalışmada Oflazer, problemin ortadan kaldırılmasına yönelik olarak iki seviyeli bir morfoloji ayrıştırma önermiştir [82]. Sak ve arkadaşları da yaptıkları çalışmada [83], ortalama algılayıcı (*averaged perceptron*) yardımı ile morfolojik belirsizlik giderici (*disambiguator*) geliştirmişlerdir.

Çizelge 2.5 Türkçede morfolojik belirsizlik.

SNo	Örnek
1	<p>Cümle: Telaşına geldi herhalde, anahtarı unuttun. (2. Tekil şahıs)</p> <p>Telaşına geldi herhalde, anahtarı unuttu. (3. Tekil şahıs)</p> <p>Olası çözümlenmeler [telaş: İsim] İsim+A3sg+P3sg+Dat [telaş: İsim] İsim+A2sg+P2sg+Dat</p> <p>(Sistem doğru çıktıya ulaşabilmek için diğer sözcüklerin de çözümlemesini dikkate almalıdır ve diğer sözcüklerin de doğru çözümlendiğinden emin olmalıdır.)</p>
2	<p>Cümle: Gelecekte birgün gelecek.</p> <p>Gelecek yıl mezun olacakmış.</p> <p>Olası çözümlenmeler [gelecek: İsim] İsim+Loc [gelecek: Fiil] Fiil+Fut [gelecek: Sıfat] Sıfat</p> <p>(Örnekte 3 farklı morfolojik analiz görülmektedir. Türkçe oldukça belirsiz bir dil olduğundan, bir kelimenin çok daha fazla sayıda morfolojik çözümlemesi olabilmektedir. Bu belirsizliği gidermek üzere birçok farklı çalışma bulunmaktadır)</p>

Morfemlerin art arda birleştirilmesiyle sözcüklerin oluşturulduğu sondan eklemeli yapısı nedeniyle Türkçede özellikle duygu analizi, makine tercümesi gibi bazı problemlerin çözümünde karşılaşılan bir diğer problem de seyreklik (*sparseness*) problemidir. Özetle, kök-ek kombinasyonları düşünüldüğünde ne kadar büyük bir derleme çalışılırsa çalışılırsa Türkçe gibi dillerde olası tüm sözcükleri listeleyebilmek (*full-listing hypothesis*) [84] pek mümkün değildir. Problemin çözümüne odaklanan bazı çalışmalarda, Arısoy ve arkadaşları tarafından [85] farklı bir yöntem (*sub-lexical units*) kullanılması önerilmiş, bir diğer çalışma da Can tarafından Uzun Kısa Süreli Bellek (LSTM, Long-Short Term Memory) Ağları ile kök çözümleme [86] yöntemine başvurulmuştur.

Türkçeye sonradan giren veya formal bir değeri olmayan yabancı kökenli sözcüklerin, kısaltmaların sosyal ağlarda sıklıkla kullanılması ve ASCII kodlarında karşılaşılan durumlarda, nispeten kısa metinler içeren sosyal medya kaynaklı tespitleri zorlaştıran bir diğer unsurdur.

2.3. Çevrim İçi Haber Kaynakları ve Sosyal Medya

Literatürde birçok şekilde tanımlanan haber, insanın doğası gereği bilme ve etrafında olup biteni anlama, haberdar olmak isteme dürtüsü olarak tanımlanabilir. “Dijitalleşme/sayısallaşma” çeşitli medya bileşenlerinin içeriğini ve bu içeriklerin en önemlilerinden biri olan haberin etki alanını önemli ölçüde değiştirmiştir.

Özellikle sosyal medyanın insanları etkileme gücü ve imkanları arttıkça, yazılı basın diye de adlandırılan geleneksel haber medyası ve özellikle ana akım haber medyası internetteki ve çevrim içi ortamlardaki yerini almıştır. Günümüzde, ana haber ajansları ve önemli basılı gazetelerin hemen hepsinin çevrim içi erişilebilir içerik sunumu da mevcuttur. Bu doğrultuda, geleneksel medyayı oluşturan televizyon, radyo, yazılı basın vb. kitle iletişim araçlarının önemi, dijital medya kaynaklarının artan kullanımı ile eskiye göre azalmıştır [87, 88]. İnsanlar haberi geleneksel yöntemlerden ziyade internetten, özellikle de sosyal medyadan takip etmeye başlamışlardır. 2019 yılında, dünyada yaklaşık 2,82 milyar sosyal medya kullanıcısı kayda geçmiştir ve bu rakamın 2021 yılında 3,1 milyara ulaşacağı tahmin edilmektedir⁷. Böylelikle herhangi bir haberin eriştiği kitle de geleneksel yazılı gazetecilik ile kıyaslanamayacak ölçüde artmıştır. Öyleki, 5-10 yıllık yakın bir zaman diliminde yazılı basının yerini çok büyük oranda dijital medyaya bırakacağını öngörmekteyiz.

İnternet gazeteciliği konusunda yapılan öncü yayınlardan birinde [89], internet gazeteciliği kategorilere ayrılmıştır. Ana akım medya (*mainstream news media*) bu kategorilerin en başında gelenidir. Bu tez kapsamında ana akım medya tabiri alternatif medyanın karşıt anlamında kullanılan baskın medyadan daha çok, sayıca daha fazla kişi tarafından erişilen ve haberi sunarken habere ait imza bilgilerini (yazar, zaman, yer, kişi

⁷ Social Media statistics and facts. URL: <https://www.statista.com/statistics/> [Erişilme: 11 Ağustos 2019].

vb.) açık şekilde sunan ve etik ilkelere özen gösteren medya anlamındadır. Diğer bir ifadeyle, popüler geleneksel medyaya ait çevrim içi haber siteleridir.

Çevrim içi haber kuruluşları hedef kitlelerini büyütmek ve varlıklarını devam ettirmek amacıyla, okuyucu/izleyici/kullanıcılarının ilgisini çekmesi ve etkilemesi gerekir. Bu bakımdan habercilik faaliyetleri yürütülürken hedef kitlenin beklentileri göz önüne alınmak zorundadır. Bunun aksine, sosyal medya ise, bazı denetimlerden uzak yapısı ile her kullanıcıyı adeta bir haberciye dönüştürmesi ve güvenilirliği sürekli sorgulanan bir yapıda olmasına rağmen giderek daha baskın bir haber kaynağına dönüşmektedir. Örneğin, Twitter adeta “manşet tabanlı gazeteciliğin” sosyal medyaya yansımış hali olarak değerlendirilebilir [90]. Bu göstergeler kısa bir süre içerisinde çevrim içi haberciliğinde varlığını sürdürebilmesi için sosyal medyayı daha aktif kullanabileceğini işaret etmektedir.

Her ne kadar bazı sosyal medya platformları sahte haberler için tedbirler alsada internet ortamının doğasına aykırı bir şekilde yasaklayıcı tedbirler almak tepki çeken ve ticari kayıplara neden olan açmazlar ortaya çıkarmıştır. Ayrıca bu organizasyon ve platformların tıpkı insanlar gibi ön yargılı ve taraflı olabilme potansiyelleri, maalesef bunları çözümün merkezinden uzakta tutmaktadır.

2.4. Dijital Kütüphanecilik ve Sahte Haber

Bu bölümde, modelimizi ve otomatik sahte haber tespiti yapan diğer tüm sistemleri sadece sayısal değil aynı zamanda sosyolojik bir açıdan da tamamlayacağını düşündüğümüz ve bu bağlamda bütünsel bir bakış açısı ile yararlanabileceğini öngördüğümüz Dijital (Sayısal) Kütüphaneciliğin konu ile olan bağlantısı anlatılmıştır.

Kütüphaneler/Kütüphanecilik; tarih boyunca bilginin üretimi, erişimi ve paylaşımının sağlandığı, insanoğlunun en önemli kültürel varlıklarından biri olarak sürekli bir devinim içerisinde olmuştur. Bunun yanında bilginin saklandığı yer anlamına gelen “kütüphane”, çok eski zamanlardan beri toplumun saygı duyduğu ve gerçek bilgiye erişimi çağrıştıran sosyolojik bir kavramdır. Günümüzde ise geleneksel ortamlar yerine, dijitalleşme

sürecine uyum göstererek daha dinamik ve dijital alanlar üzerinden faaliyetlerini sürdürmektedir. Halen tarihsel süreçte kazandığı saygınlığı sürdürdüğü de söylenebilir.

Elbette çağlar boyunca güvenilir bilgi kaynakları olarak bilinen kütüphaneler de sahte haberlerden etkilenmiştir. Öyle ki, tüm dünyada kütüphanelerin dijitalleşme sürecinden geçtiği ve kullanıcılarına dijital içerik sunduğu düşünüldüğünde, dijital kütüphanelerin kullanıcılarına sundukları devasa boyuttaki içeriklerin anlık kontrolü oldukça zordur. Dolayısıyla sahte ve/veya yanlış bilgi ve haberler içermesi doğaldır. Medya/bilgi okuryazarlığı (*media/information literacy*) konusunda dijital kütüphaneciliğine oldukça önem veren birçok çalışma mevcuttur. Kattimani ve arkadaşları [91] kütüphanelerin tarih boyunca yeniden konumlanma ve rol kazanma anlamında başarılı olduğunu belirtmektedir. Benzer şekilde Connaway ve arkadaşları [92] kütüphaneciler, arşivciler, gazeteciler gibi bilgi profesyonellerinin toplumun haberlerle ilgili ürün ve hizmetleri uygun bir şekilde kullanmasına yardımcı olmadaki kilit rolüne işaret etmektedirler. Fakat, bilgi kirliliğinin en üst düzeyde olduğu günümüzde Kütüphane ve Enformasyon Bilimi araştırmacıları (*LIS, Library and Information Science Community*) arasında sahte haber ile mücadelede dijital kütüphaneciliğin somut rolü konusunda bir fikir birliği olmadığı görülmektedir.

Çağdaş bilimsel çalışmalar, ilk zamanlarda sadece pazarlama ve çıkar aracı olarak görülen sahte haberin artık BT başta olmak üzere birçok farklı disiplinden araştırmacıların ilgisini çeken toplumsal bir tehlike olarak göstermektedir. Bu çözüm arayışı içerisinde kütüphanecilikte, araştırmacılar tarafından irdelenen konular arasında yerini almıştır. Kavramsal boyutta çerçeveler çizen bazı çalışmalarda, konu ile ilgili çeşitli tespitler yapılmıştır. Bu çalışmalardan birinde; Sullivan, Kütüphane ve Enformasyon Bilimi topluluğunun 2016 ABD Başkanlık Seçimleri sonrasında sahte bilginin dolaşımı anlamında derinden etkilendiğini belirtmektedir [93]. Aynı çalışmada sahte haberin ve kütüphaneciliğin ilişkisinin doğru kurgulanmadığı ve problemin iyi anlaşılmadığı belirtilmiştir.

Bunun yanı sıra sahte haber ile mücadele ile kütüphaneler arasında sıkı bir bağ kuran çalışmalar mevcuttur. Örneğin, Finley ve arkadaşları [94] kütüphanelerin sahte haberlerin

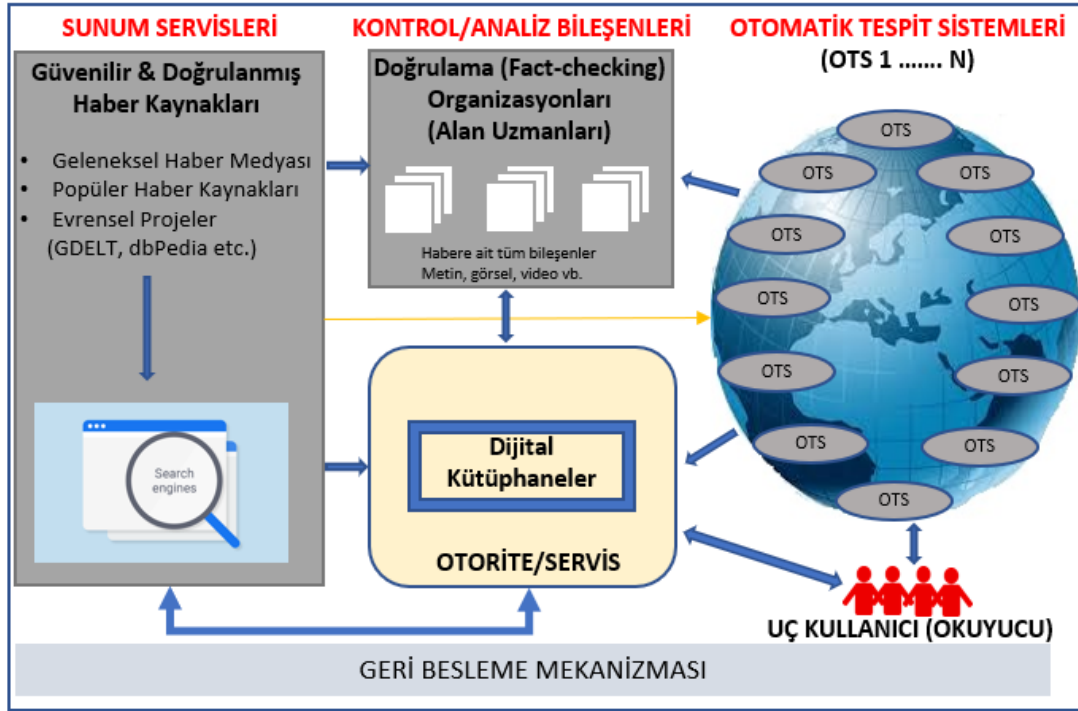
doğrudan antitezi olduğunu söylemektedir. Amerikan Kütüphane Derneği (*ALA, American Library Association*) sahte haberlerle kütüphanecilik etiğinin “afroz edildiğine” işaret etmektedir [95]. Rochlin kütüphanecilerin bu mücadeledeki rolünü vurgulamış ve sahte haberi kütüphanecilik için temel bir kaygı olarak [96] belirtmiştir. Birçok başka araştırmacı da kütüphanecilerin sahte haberlere karşı mücadelenin ön saflarında [97-99] olması gerektiğini ifade etmişlerdir.

Dünyanın dört bir yanındaki kütüphaneler dijital içerik hizmetlerine geçiş yaparken, sahte ve yanıltıcı içeriğin yayılmasını önlemek için bilgi profesyonellerinin rehberliğinde mekanizmalar uygulamalıdır. Uygun çözüm stratejileri belirlemek için Kütüphane ve Bilgi Bilimi topluluğunun “dijital kütüphaneler” ve “sahte haberler” arasındaki ilişkiyi daha somut bir şekilde ele alması gerekmektedir. Daha önce de belirttiğimiz gibi bu birçok yaklaşımın bir potada eritilerek oluşturulacak kolektif bir çaba gerektirmektedir. Bu doğrultuda dijital kütüphaneciliğin nitelikli haberler için hizmet sunma zamanının geldiğini öngörmekteyiz. Bu açıdan gerçek bilgi/haber denilince ilk aklımıza gelen kavram olan kütüphanelerin, bilgiyi değerlendirme konusunda uzman olarak değerlendirebileceğimiz kütüphanecilerin ve dijital kütüphaneciliğin hem bir servis hem de otorite olarak çatı modelimizde yer alması gerektiğini değerlendirdik.

2.4.1. Sahte Haber ile Mücadelede Dijital Kütüphaneciliğin Önemi

Bu bölümde çatı model (çerçeveyi) açıkladıktan sonra, diğer bölümlerde bu çerçeve için çok kritik olan otomatik algılama sistemi çözümümüzü Türkçe açısından açıklayacağız.

Dijital Kütüphanecilikten servis ve otorite olarak faydalanılmasını hedeflediğimiz çerçevede/çatı model Şekil 2.4'te görselleştirilmiştir. Bu yenilikçi çözüm önerisi, esasen tez çalışması sonucunda elde edilen Tez Blok Diyagramını da içine alan ve birçok farklı otomatik sahte haber/bilgi tespitini yapan sistemleri bütünleştirebilecek üst seviye bir bakış açısı ile sunulmuştur. Bu yüzden, birbirine bağlı, esnek katmanlardan oluşur. Birden fazla otomatik tespiti sistemini barındıran bu yapı, sanallaştırma teknolojilerinde kullanılan bir konteyner mimarisine benzetilebilir.



Şekil 2.4 Sahte Haber Tespiti Çatı Modeli

Bu modelde, halihazırda çevrim içi haber imkânı sunan erişilebilir haber siteleri, projeleri haberin sunulduğu platformlardır. Sunulan haberlerin arama motorlarında yeterince indekslendiği varsayılmaktadır. Haber doğrulama organizasyonları, ana kaynaklarını kendilerinin belirlediği haberlerin, uzman kişiler tarafından çoğunlukla manuel yürütülen sınıflandırmasının yapıldığı mecradır. Otomatik tespit sistemleri ise bu sınıflandırmayı otomatik yapan çekirdek modellerdir. Burada dijital kütüphanelerin rolü iki kanaldan kendilerine gelen sınıflandırılmış verinin nitelikli haber olarak arşivlemesinin yapılmasıdır. Bu noktada en mühim hususlardan biri arama motorları ile dijital kütüphaneler arasındaki geri besleme mekanizmasının doğru işletilmesidir.

Haber doğrulama organizasyonlarının da otomatik tespit sistemleri kullanabileceği düşünüldüğünde, içinde buldukları yoğun manuel iş gücü gerektiren süreçte kolaylaşacaktır. Bu arada sahte haberin geldiği kaynaklar her seferinde arşivlendiğinden, gerçekten kullanıcılara hizmet sunmak isteyen haber platformlarını da daha titiz davranmaya sevk edecektir.

Ana bileşenleri Şekil 2.4'te sunulan çatı modelin çok genel bir çerçeve çizdiği düşünüldüğünde, eklenebilecek farklı bileşenler, mekanizmalar ve katmanlarla daha

kararlı bir yapıya kavuşabilir. Burada sistemin bir otoriteye ihtiyaç duyması bir eleştiri noktası olabilir. Fakat, sadece teknoloji odaklı, mod moda bir çözümün böylesine karmaşık bir problemi çözmesi beklenemez. Sahte haber ile mücadele bu noktada doğru yönetilmesi gereken bir süreç olarak ele alınmıştır. Otorite olarak dijital kütüphanelerin seçilmesi ise, halihazırda teknoloji devleri ve sosyal medya platformlarının kişisel veriler de dahil olmak üzere bilgi yönetimi ve sahte haberler ile mücadele konusunda yanlı olmak ve ticari kaygılar taşımak noktasında sürekli olarak eleştirilmekte olmasından kaynaklanmaktadır.

Küresel seviyede bir çabaya her zamankinden daha çok ihtiyaç olduğu açıktır. Dijital kütüphanelerin teknolojik bir gelişmeden daha ötede olduğunu değerlendiriyoruz. Yazılı medyanın gündend güne eridiği günümüzde yazılı olmasa da kuralları olan yeni medyanın yerini dijital kütüphaneler alabilecektir. Sahte haberlerin neden olduğu tehditler gerçek ve tehlikelidir, bu nedenle bu çatı modelden daha da önemlisi sosyoloji, eğitim, bilişim, hukuk veya politik gibi farklı disiplinlerden araştırmacıların tarafsız şekilde çözüme katkıda bulunması gerekmektedir.

3. SÖZLÜK (DİL MODELİ) TEMELLİ SAHTE HABER TESPİTİ

Bu bölümde anlatılacak olan sözlük geliştirme çalışmasında, literatürde sözlük temelli sahte haber tespiti yapan Türkçe çalışma olmaması nedeniyle, olası diğer çalışmalara yol göstermesi açısından sıralı kurgu takip edilmiştir. Bu yüzden, öncelikli olarak keşifsel veri analizi uygulanmış ardından sözlük üretme tabanlı (*lexicon generation based*) bir yaklaşımla, genel amaçlı Sahte Haber Sözlüğü (SHS) oluşturulmuştur. Oluşturulan derlemin Türkçe dilini ve çevrim içi haberciliği gerçeğe en yakın şekilde temsil edebilecek bir derlem olması göz önünde bulundurulmuştur. Müteakip aşamada sözlük makine öğrenmesi ve derin öğrenme yöntemleri ile daha genel-geçer bir çözüm modeli sunulmuştur. Sosyal bilimlerden çözüm modelinde faydalanmak ve pratik olarak uygulanabilir bir çerçeve oluşturmak için dijital kütüphaneler sürece dahil edilerek bütünleşik bir çözüm Bölüm 2.4.1’de önerilmiştir.

Tez çalışmamızda, Bölüm 1.5’te belirtilen şekilde öncelikli olarak SHS’nün ortaya çıkarılması ardından makine öğrenmesi ve derin öğrenme yöntemleriyle problemin ele alınması hedeflenmiştir. Son olarak bütünleşik hibrit bir yöntem ile nihai model oluşturularak yenilikçi bir çözüm modeli şeklinde sunulması amaçlanmıştır. Bu kapsamda, ilk hedef SHS oluşturulmasıdır. Bu bölümde SHS oluşturulması hedefine yönelik yapılan çalışmalar sunulmuştur.

SHS oluşturulması ve özellikleri bakımından sentiment analizi çalışmalarında sıklıkla kullanılan sözlük (*lexicon*) temelli çalışmalara benzerlik göstermektedir. Tezin bundan sonraki bölümlerinde Sahte Haber Sözlüğü (SHS) veya FanLexTR şeklinde kullanılarak ifade tekrarlarının önüne geçilmesi hedeflenmiştir. FanLexTR sözlüğünün geliştirilme süreci üç ana fazda yürütülmüştür. Bunlar sırasıyla; veri setinin oluşturulması, FanLexTR sözlüğünün oluşturulması ve FanLexTR sözlüğü ile sahte haber tahmini aşamalarıdır.

3.1. Giriş ve Terminoloji

Tezin özellikle bu bölüm ve bundan sonraki bölümlerinde, terminolojinin daha iyi anlaşılması için sıklıkla kullanılan bazı terimlerin açıklamaları aşağıda kısaca sunulmuştur:

Fonem : Sözcükte anlam farkı meydana getirebilen en küçük ses unsuru, dildeki en basit ses unsuru: “İKİ ve İYİ kelimelerinde K- Y”.

Biçimbirim (*Morpheme*) : Anlam ayırıcı olarak özelliğe sahip en küçük dilsel birim şeklinde ifade edilebilir. Diğer bir deyişle, bir sözcük ufak parçalarına ayrıldığında, kendisine anlam atfedilebilecek (yapım, çekim ekleri de dahil) tüm parçalar bir morfem yani biçimbirimdir.

Lemma : Eklenerak türetilmiş bir sözcüğün temel biçimine (*canonical form*) verilen addır. Diğer bir deyişle, sözcüğün morfolojik anlamda köküne inilmesidir.

Lexeme : Fonolojik ve anlamsal olarak farklı, bir dildeki en küçük anlamlı birimdir. Örneğin; “koş” sözcüğü bir lexeme olarak ele alındığında bu kelimeden türetilen formlar: “Koş, koşu, koşucu” sözcükleri “koş” lexeme’nin çeşitleridir.

Token : Tokenlar, anlamlı küçük birimlerdir, semboller, kelimeler, deyimler token’lara örnek verilebilir. Kullanılan tokenizer’a göre ayrıştırma değişir. Bu doğrultuda “tokenizing” bir metnin/cümlenin istenilen özelliklere göre parçalara (daha küçük anlamlı birimlere) ayırmak olarak tanımlanabilir.

Dilbilimsel terimleri detaylı olarak incelemek için Cruse’un [100] çalışmasına bakılması uygun olacaktır.

Sahte ve Doğru Skoru : Tez kapsamında oluşturulan sözlüklerdeki her bir terimin sahtelik ve doğruluk puanlamasını ifade eder (0 ile +1.0 arası). Bu puanlama belirtilen değerler arasında değişen bir ton değeri hesaplamasıdır. Duygu analizi çalışmalarında kullanılan polarite hesaplamasında terimler daha keskin çizgilerle -1 ve +1 değerler alabilmektedir. Bir terime salt sahte veya gerçeklik atamak şeklinde keskin bir ayırım bu çalışmanın kapsamı dışında tutulmuştur.

Makine öğrenmesi temelli problemlerde, her bir modeli değerlendirmek için sabit bir bakış açısı ile bakmamak açısından birden fazla metrik kullanılmalıdır. Bu bölümde de sistem çıktıları birden fazla metrik ile değerlendirilmiştir. Bilgi Getirimi ve Çıkarımı (*Information Retrieval*) çalışmalarında sıklıkla kullanılan Hata Matrisi kullanılarak, aşağıda problemin özet gösterimler sunulmuştur. Biz bu problemde pozitif değeri sahte haberin doğru şekilde tespit edilmesi olarak belirlediğimiz Hata Matrisini kullandık.

Hata (Karmaşıklık) Matrisi (*Confusion Matrics*) : Hata Matrisi, veri setindeki var olan gerçek durum ile sınıflandırma modelinin yaptığı doğru ve yanlış tahminlerin tablo olarak gösterilmesidir. Şekil 3.1’de Sahte Haber Tespit problemine ait Hata Matrisi verilmiştir.

<i>SAHTE HABERİN BULUNMASI (P)</i>		GERÇEK DURUM		<i>GERÇEK HABERİN BULUNMASI (P)</i>		GERÇEK DURUM	
		SAHTE (P)	GERÇEK (N)			GERÇEK (P)	SAHTE (N)
SİSTEM TAHMİNİ	SAHTE	TP	FP	SİSTEM TAHMİNİ	GERÇEK	TP	FP
	GERÇEK	FN	TN		SAHTE	FN	TN

TP=True Positive, FP=False Positive, FN=False Negative, TN=True Negative

Şekil 3.1 Sahte haber tespit problemine yönelik hata matrisi tanımları

Bir sahte haber tespit probleminde genel olarak pozitif olarak değerlendirilen durum, Şekil 3.1’de sol tarafta belirtildiği üzere sahte haberin bulunmasıdır. Problemin ele alınış şekline bağlı olarak pozitif (olumlu) durum gerçek haberin tespiti olarak ifade edilebilir. İkili sınıflandırma yapan ideal bir tahminleme modelinde her iki sınıfın da doğru tespit edilebilmesi önemlidir. Bu bakımdan sadece sahte haberi doğru bulan değil aynı zamanda gerçek haberi de ayırt edebilen bir modelleme yapılmasına dikkat edilmelidir. Bu bakımdan aşağıda tanımları verilen kavram ve metriklerin çok iyi anlaşılması gerekmektedir.

Dođru Pozitif (*TP, True Positive*) : Tahmin pozitif ve dođru bir tahmindir. Őekil 3.1’de soldaki sisteme gore, haber ‘‘sahte’’ tespit edilmiŐ (tahmin) ve esasında (gerek durum) bu haberin ‘‘sahte’’ olmasını ifade eder.

Dođru Negatif (*TN, True Negative*) : Tahmin negatif ve dođru bir tahmindir. Őekil 3.1’de soldaki sisteme gore, haber ‘‘gerek’’ olarak tespit/tahmin edilmiŐ ve esasında bu haberin ‘‘gerek’’ olmasını ifade eder.

YanlıŐ Pozitif (*FP, False Positive*) : Tahmin pozitif ve yanlıŐ bir tahmindir. Őekil 3.1’de soldaki sisteme gore, haber ‘‘sahte’’ olarak tespit edilmesine rađmen esasında bu haberin ‘‘gerek’’ olmasını ifade eder.

YanlıŐ Negatif (*FN, False Negative*) : Tahmin negatif ve yanlıŐ bir tahmindir. Őekil 3.1’de soldaki sisteme gore, haber ‘‘gerek’’ olarak tespit edilmesine rađmen esasında bu haberin ‘‘sahte’’ olmasını ifade eder.

Kesinlik (*Precision*) : Sistem tarafından pozitif olarak tahmin edilen deđerlerden gerekte ka adedinin pozitif olduđunu gosterir. Bu durumda, hangi sınıfın seimini pozitif olarak deđerlendirdiđimize gore sistemin iki adet kesinlik deđerine sahip olduđunun unutulmaması gerekir. EŐitlik (1)’de Kesinlik metriđinin hesaplaması gosterilmiŐtir. Bu metrik literaturde Hassasiyet olarakta kullanılmaktadır.

$$Kesinlik = \frac{TP}{TP + FP} \quad (1)$$

Duyarlılık (*Recall*) : Gerekte pozitif olarak bulunan deđerlerden, ne kadarının sistem tarafından pozitif olarak tahmin edildiđini gosteren metriktir. Aynı Őekilde pozitif olarak seilen sınıfa gore iki adet duyarlılık deđerleri bulunmaktadır. EŐitlik (2)’de Duyarlılık metriđinin hesaplaması gosterilmiŐtir. Literaturde, Geri ađırma olarakta kullanımını gormek mumkundur.

$$Duyarlılık = \frac{TP}{TP + FN} \quad (2)$$

Duyarlılık, Doğru Pozitif Oran (DPO)⁸ ismi ile de geçmektedir. Bu bakış açısı ile benzer bir hesaplama oranı olan Yanlış Pozitif Oranını da (YPO)⁹ tanımlamak uygun olacaktır. Olumsuz tahminlerin hesaplanmasında kullanılır. Bu iki metrik birlikte Alıcı İşletim Karakteristiği (*ROC, Receiver Operator Characteristics*) eğrisinin çizilmesinde kullanılmaktadır.

F1 Skor : Duyarlılık ve hassasiyet değerlerinin harmonik ortalaması hesaplanarak elde edilen metriktir. Uç durumlar da hesaba katıldığı için, özellikle eşit dağılmayan veri kümelerinde hatalı bir model seçimi yapmamak adına doğruluk metriğinin yerine tercih edilir. Eşitlik (3)'de F1 Skor metriğinin hesaplaması gösterilmiştir.

$$F1 = 2 * \frac{Kesinlik * Duyarlılık}{Kesinlik + Duyarlılık} \quad (3)$$

Doğruluk (Accuracy) : Model tarafından doğru tahmin edilen sınıflandırmaların, toplam sınıflandırılma sayısına bölümü ile ifade edilmesidir. Eşitlik (4)'de Doğruluk metriğinin hesaplaması gösterilmiştir.

$$Doğruluk = \frac{TP + TN}{TP + FP + TN + FN} \quad (4)$$

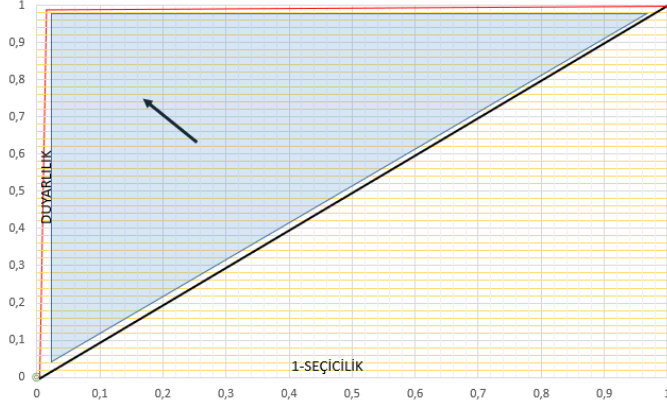
ROC Eğrisi (ROC Curve) : Alıcı İşletim Karakteristiği Eğrisi, ikili sınıflandırmada makine öğrenmesi algoritmalarının performansını değerlendirmek için yaygın olarak kullanılan ölçütlerden biridir. Dengesiz veri setlerinin kullanıldığı çalışmalarda sıklıkla tercih edilir. Ayrım eşik değeri farklılık arz ettiğinde, hassasiyetin kesinliğe oranı olarak değerlendirilebilir. Bir diğer tanımla, ROC eğrisinde X ekseninde YPO ve Y ekseninde DPO vardır.

ROC Eğrisi Altındaki Alan (AUC, Area Under Curve) : ROC Eğrisi altındaki alana "AUC" denir. Bu alanın büyüklüğü, makine öğrenme modellerinin sınıflandırmadaki başarısını gösterir. Örneğin, tahmin sistemi sahte ve gerçek haberi birbirinden tamamen ayırt edebiliyorsa bu tahmin sistemi mükemmeldir. Bu durumda eğri altında kalan alan maksimum değeri yani "1" değerini almıştır. İdealde alabileceği minimum değer ise

⁸ DPO=TPR (*True Positive Rate*)

⁹ YPO=FPR (*False Positive Rate*)

“0.50” ‘dir ki bu nerdeyse şans eseri bulmaya eşdeğerdir. İdeal bir modelde, Şekil 3.2’de temsili ROC grafiği üzerinde eğrinin altında kalan alanın bahsedilen iki değer arasında (0.5-1) ok yönünde artması beklenir.



Şekil 3.2 ROC eğrisi altındaki alan gösterimi

3.2. Veri Setinin Oluşturulması

Türkçe SHS'nün oluşturulması için gerekli olan veri seti işlemleri bu bölümde ele alınacaktır. Öncelikle derlemin oluşması için haber verisinin toplandığı haber kaynakları anlatılmış, veri doğrulama ve geçişleme işlemleri ele alınmıştır. Sonrasında veri toplama süreci ve ortaya çıkarılan derlemin özelliklerine yer verilmiştir.

Veri setinin oluşturulması, birçok sınıflandırma probleminde olduğu gibi, sahte ve gerçek haber içeren veri setleri oluştururken de kesin referans doğrulaması yapılmasının zorluğu, doğruluk kontrolü gibi hassas birçok faktörden dolayı oldukça güç, maliyetli ve yoğun emek isteyen bir süreçtir [101, 102]. Bu nedenden dolayıdır ki, açık kaynak olarak erişilebilir sahte haber veri setleri İngilizce hariç birçok dilde ya çok sınırlı ya da hiç yoktur. Bu doğrultuda Bölüm 1.2'de belirtildiği üzere tez kapsamındaki motivasyonlarımızdan biri de bu Türkçe veri setinin açık kaynak olarak literatüre kazandırılması olmuştur.

Bu doğrultuda çalışmamızda, Peres ve arkadaşlarının da çalışmalarında [37] kullandığı gibi, Rubin ve arkadaşlarının bir sahte haber derlemi oluşturmak için gerekli ana

maddeleri bildiren kılavuz çalışmasından [103] yararlandık. Bu kılavuz çalışmada sahte haber veri seti oluşturmak için belirtilen 9 (dokuz) gereksinim şu şekildedir:

- Sahte ve gerçek haber maddelerini içermeli,
- Sadece metinlerden oluşan haberleri içermeli,
- Doğrulanabilir temel bir gerçekliğe sahip olmalı (*verifiable ground-truth*)
- Uzunluk bakımından homojen olmalı,
- Yazı stili bakımından homojen olmalı,
- Önceden belirlenmiş zaman dilimine ait olmalı,
- Sahte ve gerçek olaylar için aynı şekilde ve aynı amaçla sunulmalı,
- Herkese açık kaynak olarak sunulmalı,
- Dil ve kültür farklılıklarını dikkate almalıdır.

Çalışmamızda, mümkün olduğunca yukarıdaki yönergelerin tümünü ele almayı amaçladık. Bu gereksinimlerden “Doğrulanabilir temel gerçeklik” maddesi içerik doğrulamasını kapsadığından oldukça zorlayıcı ve başlı başına bir araştırma alanıdır. Temel hedefimiz içerik doğrulamasından ziyade kullanılan dilin sahteliğinin tespiti olmasına rağmen gerektiğinde çapraz doğrulama ve diğer doğrulama yöntemleri ile bu gereksinim de adreslenmeye çalışılmıştır.

Genel amaçlı ve etkili bir sahte haber dili sözlüğü oluşturmak için veri setinde olabildiğince fazla sayıda sahte haber türüne yer verilmeye çalışılmıştır. Türkçe DDİ motoru olan Zemberek (<https://code.google.com/p/zemberek>) ile kelimelerin köklerini, POS etiketlerini ve eklerini bulmak için kullanılmıştır.

Türkçe için kapsamlı ve etiketli bir derlem oluşturarak, 4 farklı terim kategorisinde ele aldığımız FaNLexTR adını verdiğimiz sözlüğü geliştirdik. Bu bölümde, Türkçe için genel amaçlı sahte haber dili modelini geliştirirken kullandığımız yenilikçi metodoloji sunulmuştur. Öncelikli olarak, sahte haber derlemine oluştururken kullandığımız 3 kaynak anlatılacaktır. Bu kaynaklar şu şekildedir:

- Büyük bir küresel medya arşivi olan ve dünya çapında araştırmacılara açık kaynak erişimi sağlayan GDELT Projesi¹⁰,
- Bir gerçeklik doğrulama organizasyonu olan teyit.org referans alınarak elde edilen veri kaynakları,
- Ve MVN (*Manually Verified News*) olarak adlandırdığımız ekibimiz tarafından yukarıdaki iki kaynak havuzunun içinde olmak şartıyla toplanan ve etiketlenen popüler haber sitelerinden elde edilen verilerdir.

3.2.1. Veri Kaynakları

Tez kapsamında geliştirilen ve kullanılan veri seti 3 ana kaynaktan toplanmıştır. Bu üç kaynak seçilirken, sahte haberlerle mücadelenin genel olarak üç farklı şekilde yürüdüğünü anlattığımız Bölüm 2.1.3 ile uyumlu olmasına ve böylelikle oluşturulan veri setinin gerçek dünyayı daha iyi temsil etmesine özen gösterilmiştir.

GDELT :

Araştırmacı Kalev H. Leetaru tarafından 2013 yılında ortaya çıkarılan GDELT (*Global Database of Events, Language (Location) and Tone*) Projesi [104, 105], halen geliştirilmeye devam etmekte olan dinamik bir projedir. Projenin amacı, dünyadaki çevrim içi bilgi kaynaklarının taranması suretiyle belli formatta saklanan devasa bir veri arşivi oluşturmaktır. Bu açıdan bakıldığında, geniş kapsamlı bir dijital haber kütüphanesi olarak düşünülebilir. Sayısı giderek artan akademik çalışmalarda GDELT verilerinin kullanılmasına başlanması projenin başarısına atfedilebilir.

GDELT projesinde haberler direkt olarak haber metni biçiminde sunulmamıştır. Haberin URL'si ve işlenmiş meta verilerini içeren bir formatta veriler tutulur. Yaklaşık 70 dil için gerçek zamanlı olarak İngilizceye çevrilen haberler, DDİ mekanizmaları vasıtası ile işlenerek GDELT veri tabanlarında arşivlenmektedir. Anlık bir çeviri sistematğine sahip olması, farklı dillerde DDİ alanında çalışan araştırmacılar için önemli fırsat olarak görülebilir. Araştırmacıların bu veriyi kullanabilmeleri için çeşitli şekillerde sunan proje,

¹⁰ <https://www.gdeltproject.org/>

yaklaşık 5 yıldır da veri setlerini Google'un bulut bilişim servisi Google BigQuery (<https://bigquery.cloud.google.com/>) üzerinden de sunmaktadır. GDEL T projesinde farklı veri setlerinin yanı sıra, 3 temel veri seti (*events*, *eventmentions*, *gkg*) araştırmacılar için dikkate değerdir. Çizelge 3.1'de Google'un söz konusu servisinden yararlanırken kullanılan bir sorgu örneği verilmiştir. Elde edilen bu ham veriler üzerinden haberlerin sunulduğu adreslere gidilerek metin bilgileri elde edilmiştir.

Çizelge 3.1 Google BigQuery ile veri toplanması

```
SELECT * FROM
(SELECT DATE, DocumentIdentifier, TranslationInfo, V2Tone AS Tone, NTH (7,
SPLIT(V2Tone,',')) AS WordCount
FROM [gdelt-bq: gdel tv2.gkg])
WHERE TranslationInfo CONTAINS 'src: tur'
AND (DocumentIdentifier CONTAINS ('aa.com.tr')
OR DocumentIdentifier CONTAINS ('dha.com.tr')
OR DocumentIdentifier CONTAINS ('iha.com.tr'))
AND (CAST (WordCount AS INTEGER)>= 30
AND (DATE>20170101000000 AND DATE<=20191231234500))
```

GDEL T gelişmiş dil kaynakları ile kapsamlı içerik analizi, yapmaktadır. Bu gelişmiş İngilizce dil kaynakları Sağ lam'ın tez çalışmasında sunulmuştur [106]. Fakat, sistem herhangi bir haber metninin geçerliliğini kontrol etmez. Bu yüzden, GDEL T aracılığıyla otomatik olarak elde ettiğimiz haberlerin geçerliliğini en üst düzeye çıkarmak için, yalnızca 3 büyük haber ajansının (AA, DHA, İHA)¹¹ ettiği haberleri kullandık. Doğ rulanması daha zor olan, yerel ve daha az bilinen kaynaklardan gelen haberler göz ardı edilmiştir. Buna ilave olarak, GDEL T yoluyla elde edilen olası sahte haberleri temizlemek için teyit.org verileriyle çapraz doğrulayarak elde edilen metinlerin bu kontrolden sonra metin ön işlemeye sokulması sağ lanmıştır.

Teyit.org :

İkinci olarak kullandığımız kaynak teyit.org; çevrim içi kanallar aracılığı ile yayılan sahte/yanlış bilgileri, sosyal medya kaynaklı doğrulanmamış bilgileri, şehir efsaneleri

¹¹ AA: Anadolu Haber Ajansı
DHA: Demirören Haber Ajansı
İHA: İhlas Haber Ajansı

gibi yanlış bilgilerin yayılmasını önlemek amacıyla IFCN (Uluslararası Gerçek Kontrol Ağı) ilkelerine uygun, Türkiye'de yerleşik bir haber doğrulama organizasyonudur.

Büyük sosyal medya platformlarının sahte haber ile mücadele kapsamında, haber doğrulama organizasyonları ile yaptıkları iş birliklerinin bir örneği de teyit.org'un¹² bu kapsamda Facebook ile yaptığı iş birliğidir.

Her bir haber metnini uzman personel vasıtası ile analiz eden bu organizasyon, haberleri yanlış, doğru, karma veya belirsiz olarak etiketlemektedirler. Haber analizi açısından insan merkezli bir süreç işletilmektedir. Bu kaynak sahte haber ile mücadelede 2.1.3'te anlatılan haber doğrulama (*fact-checking*) kuruluşları ve örün (*web*) sitelerine karşılık gelmektedir. İlgili site takip edildiğinde her geçen gün yanlış/sahte olarak etiketlenen haber sayısının arttığı gözlemlenebilir. Şekil 3.3'te bu kaynaktan toplanan bir sahte haber örneği ve sahte haberin yayınlandığı belirtilen kaynaklar gösterilmektedir. Sistemimiz tarafından haber kaynaklarının orantılı dağıtımı göze alınarak otomatik ve rastsal bir şekilde toplanan ham veri üzerinden haber metin bilgisi elde edilmiştir.

Dünya genelinde haber doğrulama araçları, platform ve organizasyonlarının sayıları artmaktadır. Bunlardan akademik çalışmalarda kaynak olarak kullanılan bazıları; COMPA [107], Snopes.com, FakeNewsTracker[108], PolitiFact.com, FactCheck.org, TruthOrFiction.com, CredFinder [109], ClaimBuster [110] şeklinde belirtilebilir. Son yıllarda ülkemizde de buna yönelik çabaların arttığı görülmektedir.

Elle Derlenen Veri :

Son kaynağımız ise, çeşitli çevrim içi kaynaklardan elde edilen uzun bir sürece yayılarak toplanan ve çalışma ekibimizce doğrulanmış sahte (*fake*) ve gerçek/geçerli (*valid*) haberlerden oluşan, MVN (*Manually Verified News*) olarak adlandırdığımız veri setidir.

¹² Tez çalışması kapsamında teyit.org organizasyonuna, bir haber doğrulama organizasyonunun nasıl çalıştığını gözlemlememize (Mart 2019) ve açık kaynak olarak sundukları ve titizlikle yürüttükleri analiz sonuçlarını/kaynaklarını veri setimizi oluşturmak için otomatik olarak çekmemize izin verdikleri için ayrıca teşekkürlerimizi sunarız.

Bu set, GDEL ve teyit.org kaynaklarından alınan haberlerle örtüşmeyen sahte ve gerçek/geçerli haberleri içermektedir.



Şekil 3.3 Teyit.org kaynağından elde edilen haber örneği

3.2.2. Doğrulama ve Geçerleme

Bu tarz sınıflandırma işlemlerinde veri toplama ve hazırlama safhaları genellikle zor ve zaman alıcıdır. Bunun en temel sebebi, geliştirilen sistemin geçerli ve doğrulanmış bir başarıml elde etmesi için çalışılan verinin gerçek dünyayı temsil etmesi gerekliliğidir. Bu problem özelinde herhangi bir haberin ne zaman gerçek ne zaman sahte olarak kabul edildiğinin ve etiketlenmenin nasıl yapıldığı hususunda titiz bir ön çalışma gerektirir. Aynı şekilde veri doğrulama ve veri geçerleme işlemleri aynı hassasiyetle doğru şekilde yapılmalıdır.

İçerik doğrulaması yapan çalışmaların bu anlamda çok titiz davranması gerekir. Bu çalışmada içerik doğrulaması yapılmamasına karşın objektif olabilmek adına, haber kaynaklarında bir filtrelemeye gidilmemiştir. Üç ana haber ajansı dışında, haber

kaynakları popüler haber siteleri, haber yayını yapan kaynaklar, mizah haber sitesinin yanı sıra Facebook ve Whatsapp kaynaklı anonim haberlere de yer verilmiştir. Bu kaynaklar teyit.org'un veri tabanında haberlerini etiketlediği kaynaklardır. Kaynak tutarlılığı ve kaynak seçim etiği açısından çalışma ekibimizce etiketlenen haberlerde de bu kaynakların dışına çıkılmamış ve kaynaklar arası dengeli dağılım gözlemlenmiştir.

Veri setimizde hâlihazırda GDELT ve teyit.org tarafından geçirilmiş iki küme vardır. Üçüncü küme ise, çalışma ekibimiz tarafından geçirilmesi yapılan ve verinin yaklaşık %1'ini oluşturan MVN diye adlandırdığımız veri setidir.

GDELT aracılığıyla toplanan haberler sadece ilgili üç büyük ana haber ajansı ile sınırlı tutulmuştur. Her ne kadar bu üç kaynak yazarsız, tarihsiz ve kaynağı belli olmayan haberler yayınlamasa da bu durum sahte haberin bu kaynaklardan hiç yayınlanmayacağı anlamına gelmez. Bu yüzden öncelikle veri setimizin tutarlılığı açısından veri kümeleri arasında çapraz-geçerleme yaptık. Bu kontrol veri kümeleri içinde aynı haberlerin geçip geçmediğini ve hatalı etiketlemeleri içerir. Örneğin, GDELT'ten elde edilen haberlerin sahte haber veri setlerinde bulunan haberleri içerip içermediğini kontrol ettik. GDELT'teki kaynağın ana akım haber ajansları olduğu düşünüldüğünde, onlar tarafından iletilen herhangi bir sahte haber içeriğinin teyit.org gibi gerçek kontrol organizasyonları için bir ilgi odağı olacağı ve inceleneceğini varsaymak mümkündür.

Bu işlemleri kolaylaştırmak adına metin kontrollerinin yanı sıra haber etiketlerinden de yararlandık. Etiketli verisi bulunmayan haberlerin için etiket havuzu oluştururken otomatik haber etiketleme çalışmamızdan yararlanarak [111] mümkün olabildiğince otomatik eleme ve filtreleme işlemleri yapılmıştır.

Doğrulama aşamasında, veri setinin yeterli veri kalitesine sahip olup olmadığını kontrol etmek için bazı istatistiksel değerlendirmeler yapılmıştır. Bu işlemlerde veri temizlemenin yanı sıra veri toplama sürecinde gözlem-temelli yapılan tespit ve varsayımların teyit edilmesine odaklanılmıştır. Böylelikle başlangıç varsayımları doğrulanarak daha sağlam bir temel ile probleme yaklaşmıştır.

Bu kapsamda, ilk olarak “argo sözcüklerin sahte haberlerde kullanılmasının daha muhtemel olduğu” varsayımımızı bazı istatistiki hesaplamalarla teyit edilmesi yoluna gidilmiştir. Bunun için Türkçe Argo Sözlüğü’nden¹³, yararlanılmıştır. Bu hesaplamada, 782 Türkçe argo sözcük ve deyim listesi için veri kümelerinde cümle başına argo sözcüklerin ortalaması hesaplanmıştır.

Ayrıca, benzer şekilde “yazım hatalarının özensiz gazeteciliğin bir işareti olabileceği ve sahte haberlerin daha özensiz hazırlandıkları” doğrulamak üzere yazım hataları için benzer hesaplama yapılmıştır. Bu hesaplamalar doğrultusunda yapılan etiketlemenin doğruluğuna atfedilebilecek yüksek korelasyon içeren sonuçlar elde edilmiştir. Çizelge 3.2’de bu sonuçlar gösterilmiştir.

Çizelge 3.2 Veri kümelerine ait cümle başı argo ve yazım hatası.

Veri Seti	Cümle başına argo ifade kullanımı	Cümle başına yazım hatası kullanımı
GDELT (Gerçek)	0.032	0,057
MVN (Gerçek)	0,071	0,102
Teyit.org (Sahte)	0,115	0,396
MVN (Sahte)	0,098	0,275

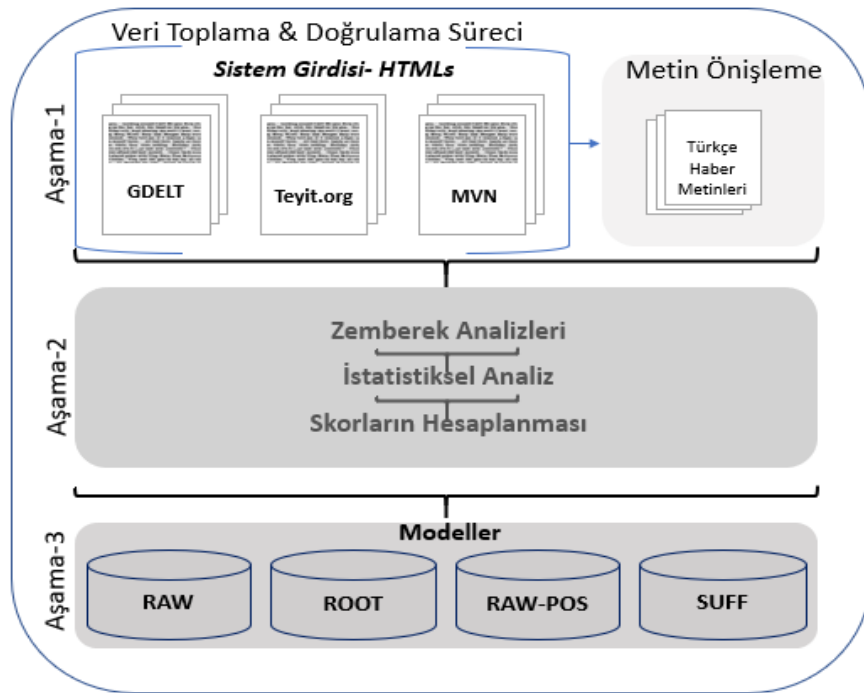
Bu çizelgedeki istatistikler bize titiz ve özensiz gazetecilik eğilimlerini, resmi ve gayri resmi dil kullanımının açık işaretlerini vermektedir. Bu tespit esasında hazırlanış itibari ile sahte olan haberin içerik olarak sahte olabileceğinin işaretidir. Bu kapsamda örneklemeler üzerinden yapılan kontrollerimizde bu varsayımın da çok büyük bir oranda (100 örnekte 93) doğru olduğu tespit edilmiştir. Fakat tüm veriyi içerik sahteliği açısından etiketlemek için bunun tek başına yeterli olmadığını değerlendirdiğimizden bu hesaplamayı tez kapsamında genelleştirerek içerik doğrulaması yaptığımızı iddia etmediğimizi tekrar belirtmekte fayda görmekteyiz. Problemin tanımında belirttiğimiz

¹³ Türkçe Argo Sözlüğü <https://www.turkedebiyati.org/argo-sozlugu.html>

üzere tez kapsamında içerik doğrulamasından ziyade haber dilinin gerçek (*valid*) ve sahte/uydurma (*fake*) oluşu üzerine çalışılmıştır.

3.2.3. Veri Toplama ve Derlem Oluşturma

Bu bölümde genel amaçlı bir SHS olan FanLexTR sözlüğünün geliştirilme aşamaları sunulmuştur. Veri setinin toplama adımları ve sürecin ana aşamaları Şekil 3.4'te özetlenmiştir.



Şekil 3.4 FanLexTR geliştirme süreci ana safhaları.

İlk aşamada; tipik bir metin madenciliği çalışmasına benzer şekilde haber metni, başlığı ve sınıflandırmada kullanabileceğimiz kategori bilgileri dışında kalan habere ait HTML etiketlemeleri, diğer işaretçiler vb. unsurlardan haberler arındırılmış ve metin dosyaları halinde kayıt altına alınmıştır. 10 farklı alana ait, 2017-2019 yılları arasında dengeli şekilde dağılmış 84734 tekil haber kayıt altına alınmıştır. Oluşturduğumuz haber derlemine ait bazı istatistiki bilgiler Çizelge 3.3'te gösterilmiştir.

Çizelge 3.3 Elde edilen haber derlemine ait istatistiki bilgiler.

Veri Seti	Sınıf	Kaynak	Sayı	Ortalama Kelime	Ortalama Cümle
Eğitim (Train)	GEÇERLİ (VALID)	GDELT	82708	243	17
		MVN	838	169	11
	SAHTE (FAKE)	Teyit.org	904	119	9
		MVN	284	102	8
Test	GEÇERLİ (VALID)	GDELT	22	188	16
		MVN	188	170	10
	SAHTE (FAKE)	Teyit.org	89	114	10
		MVN	121	106	9

İkinci aşamada, elde edilen haber metinleri Zemberek kullanarak, her bir metne ait kelimeler için kök, cümle içi pozisyonu, sözcük türü etiketleri (POS Tags) ve ek bilgisi elde edilmiştir. Ardından derlemin istatistiki analizi yapılmıştır. Müteakiben elde edilen sahte ve gerçek haber veri kümeleri kullanılarak terimlere sahte ve geçerli skor ataması yapılmıştır.

Son aşamada, ortaya çıkan modeller için, SHS vasıtası ile haber metinleri puanlanmış ve sahte haber tahmini yapılmıştır.

3.2.3.1 Derlemin (Sahte-Gerçek Haber) Oluşturulması

Bu ilk aşamada; GDELT, teyit.org ve MVN veri kümelerinden yaklaşık 110000 habere ait adres ve linklere ulaşılmıştır. Haberlere ait metinler elde edilmiş, çeşitli filtreleme ve eleme işlemlerinden sonra 84734 adet habere indirgenen, genel maksatlı bir Türkçe haber derlemi oluşturulmuştur. Bazı haberler derleme henüz dahil edilmeden ön eleme aşamasında çıkarılmıştır. Bu haberler:

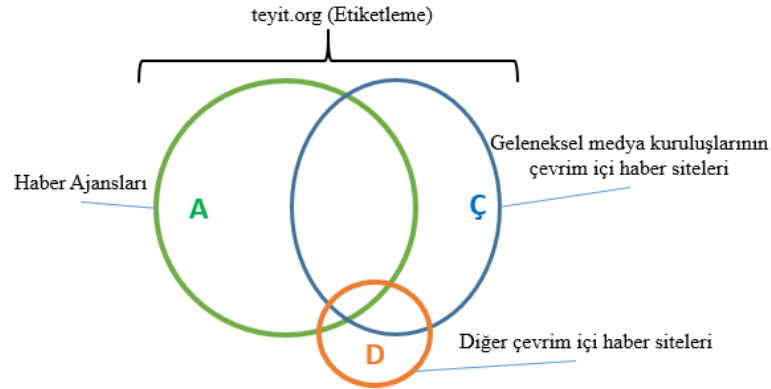
- Farklı kaynaklarda birebir tekrarlanan haberler (tekil haber kalacak şekilde),
- Doğrulanması zor olan yerel haberler,
- Çok fazla rakam içeren hava durumu ve loto çekilişi haberleri,
- Ve genel ortalamanın çok altında kısa, çok üstünde uzun metinler içeren haberler derlemde elenmiştir.

- Ayrıca bazı haberlere ait link ve içerik bilgileri bu süreçte internet ortamından tekrar erişilemediğinden çıkarılmıştır.

Oluşturulan bu derlem, tez kapsamında FaNLexTR'nin geliştirilmesi ve makine öğrenmesi çalışmamızda kullanılmıştır. Bu derlemin gelecekteki diğer olası araştırma çalışmaları için bir eğitim seti olarak kullanabileceğini değerlendirilmektedir.

3.2.3.2 Metin Ön İşleme

Veri kaynaklarımız arasında, ana akım medyayı oluşturan farklı haber ajansları ve popüler çevrim içi haber sitelerinin yanı sıra teyit.org'tan referans edilmiş haber siteleri, sosyal medya paylaşımlarından oluşan haber kaynakları ve diğer çevrim içi haber siteleri bulunmaktadır. Veri kümelerini oluşturan haber kaynakları Şekil 3.5'te sunulmuştur.



Şekil 3.5 Haber kaynakları içerik etkileşimi

Şekil 3.5'te küme gösterimleri; A (Haber Ajansı), Ç (Çevrim içi haber siteleri) ve D (Diğer kaynaklar) olarak tanımlanmıştır. Kesişim kümesi okuyucuya bu kaynaklardan sunulan haberlerin direk veya dolaylı (haber içerik benzerliği) olarak kesiştiğini temsil etmektedir. Şüphesiz haber ajanslarının bir taraftan çevrim içi haber sitelerini besleyen bir kaynak olması kesişimdeki ana etkidir. Fakat sunulan haber içeriği bakımından kümeler birbirinden farklılık göstermektedir. Örneğin; D kümesinde bulunan Zaytung.com bir mizah haber sitesi, dogrula.org ise teyit.org benzeri bir haber doğrulama sitesidir.

Çeşitli kaynaklardan verilerin toplanması durumu beraberinde metin ön işleme sürecini zorlaştırmıştır. Bazı kaynaklar belli formatlarda standart içerikler sunarken bazıları ise istisnai olarak ele alınan biçimlerde haber sunumlarını yönetmektedirler. Veri toplama sürecinde format güncelleyen kaynaklarda veri kazıma/alma, link toplama (*web scraping-web crawling*) açısından durumu zorlaştıran diğer bir husus olmuştur. Özellikle sosyal medyadaki haber içerikleri ele alınırken bu durum daha da karmaşıklaşmaktadır.

Ayrıca, metin ön işleme süreci de Sözlük Temelli ve Makine Öğrenmesi Temelli çözüm önerilerinde farklı olarak ele alınmıştır. Özellikle Makine Öğrenmesi Temelli yaklaşımda özellik çıkarımı adına stil işaretçileri açısından önem arz eden noktalama işaretleri, yazım yanlışları ve kelimelerin ham hallerini ön işlem sürecinde hem elenerek hem elenmeden ele alınması buna örnek olarak gösterilebilir.

Metin Ön İşleme sürecinde ele alınan konular şu şekilde özetlenebilir:

- Öncelikle haber metinleri, HTML ve kod (javascript) formatlarından, noktalama işaretlerinden, bazı sembollerden ve sayısal (numeric) ifadelerden filtrelenerek arındırılmıştır.
- Daha sonra Türkçe için karakter problemi yaratacak karakter setleri problemi ele alınmıştır. Ön işlem süreci, Makine öğrenmesi yaklaşımdan farklı olarak bu aşamada noktalama işaretleri de süzülerek, tamamlanmıştır.

Son olarak, tez kapsamında biçimbilimsel (morfolojik) analiz için Türkçe DDİ kütüphanesi Zemberek [75] kullanılarak, kelimelerin kök formları, ekler ve sözcük türü etiketleri (*POS Tags*) elde edilmiştir. Bu tarz çalışmalarda, özellikle sondan eklemeli diller için sözcüklerin morfemlerine ayrılması işlemi analiz yapılan sistem/aracın başarımından etkileneceğinden, modelimizin bundan sonraki safhalarda sunulacak başarım ölçümleri, Zemberek'in morfolojik çözümleme gücü ile doğru orantılıdır. Bu bağlamda Zemberek herhangi bir kelime için muhtemel kök adayları bulduğunda, sistemin kök aday seçici mekanizmasına göre yapılan sıralamanın takip edildiğini söylemek yerinde olacaktır. Bu tez kapsamında ilk sıradaki önerilen aday kök kelime olarak kullanılmıştır.

Öncelikli olarak terim ve doküman frekans hesaplamaları yapılmıştır. Daha sonra Bölüm 3.3’te anlatıldığı şekliyle terimlere ait sahte-gerçek ton değerleri ortaya çıkarılmıştır. Ayrıca bir terimi sahte veya gerçek şeklinde nitelemek yani polarite değerleri (+1, -1, 0) atama hususu da çalışılmış fakat arzu edilen sonuçları vermediği görülmüştür. Bu bakımdan çalışmanın duygu analizi çalışmalarından farklılaştığı söylenebilir.

3.2.4. Temel Bir Veri Seti Olarak TR_FN

Tez kapsamında elde edilen veri seti Türkçe Sahte haberleri içermesinden dolayı TR_FN (*Turkish Fake News Data Set*) veri seti olarak adlandırılmış ve tezin bundan sonraki bölümlerinde bu şekilde kullanılacaktır. Bu veri setine ait bazı tanımlayıcı bilgiler önceki bölümlerde verilmiştir. Bu bölümde, TR_FN veri setine ait temel bilgilerin yanı sıra, genel-geçerliği, temsil gücü ve dengesiz dağılımı gibi özgün özellikleri anlatılmıştır.

Makine Öğrenmesi temelli çözümümüzde ve gelecek çalışmalarda da kullanılabilmesi açısından TR_FN’de başlık ve metin bilgisi (işlenmiş ve işlenmemiş olarak) dışında bazı meta bilgileri de kayıt altına alınmıştır. Bunlar alanlar içerikleriyle beraber Çizelge 3.4’te sunulmuştur. TR_FN veri seti işlenmiş verileri ile direk olarak veya metin madenciliği yöntemleri ile daha da detaylandırılarak diğer sınıflandırma problemlerinde de kullanılabilir.

TR_FN veri setine ait önemli bir diğer husus sadece bir alana özgü değil, çevrim içi Türkçe haberlerin temel karakteristiğini veren genelleştirilebilir bir veri seti olmasıdır. Bu bakımdan veri seti sadece sahte haber açısından değil diğer sınıflandırma problemi açısından da çalışılabilir.

Çizelge 3.4 TR_FN veri setinin bilgi alanlarının tabular biçimde sunulması.

Alan Bilgisi	Açıklama
Haber Kaynağı	Ana akım haber medyası, popüler haber siteleri, sosyal medya kaynaklı paylaşımlar.
Tarih Bilgisi	Habere ait gün, ay, yıl bilgisi (Yoksa tarih bilgisi yok diye geçilmiştir).

Sınıf Bilgisi	Gerçek/Geçerli (Valid), Sahte (Fake).
Haber Sitesi	<ul style="list-style-type: none"> • Haber ajansları (AA, İHA, DHA), • Popüler haber siteleri (EK-1’de sunulmuştur), • Facebook, whatsapp ve sosyal medya viralleri, • Belirsiz (anonymous) kaynak, • Diğer çevrim içi haber kaynakları.
Kategori	<ul style="list-style-type: none"> • Dünya, Ekonomi, Sağlık, Politika, Popüler Haberler, Bilim-Teknoloji-Çevre-Doğa-Yaşam, Spor, Türkiye, Eğitim-Kültür-Sanat-Turizm, Diğer konular <p>(Ayrıca bu kategorilere ait alt kategori bilgileri de bulunmaktadır).</p>
Metin tanımlayıcı	Haberin internetteki esas kaynağından ulaşılabilecek URL (Uniform Resource Locator, Birörnek Kaynak Konumlayıcı) bilgisini içerir.
Stilistik çıkarımlar	Yukarıdaki bilgilerin dışında analizler sonucunda erişilen hatalı yazım, argo kelime kullanım, durak kelime (stop words) frekansları, cümle, işlev kelime, noktalama işaretleri, uzun kelime sayısı, cümle karmaşıklık skoru gibi hesaplanması gereken bilgiler sürekli tekrar hesap etmemek adına hesaplanarak kayıt halinde tutulmuştur.

Sınıflandırma problemlerinde özellikle metin sınıflandırmasına odaklanan problemlerde, veriyi belli alanlar ile kısıtlamak başarıyı arttırır. Fakat verinin genelleştirilebilir olmasını engeller. Veri setimizi oluştururken bu hususa dikkat ederek haber konularını (sadece politika, ekonomi, spor) sınırlı tutmayarak daha genel-geçer bir veri seti elde etmeye çalıştık. Veri setimizdeki bu alanların dağılımı Şekil 3.6’da gösterilmiştir. Haber alanları incelendiğinde okuyucuların günlük hayatta internette karşılaşılabileceği bir dağılımın yakalandığı söylenebilir. Benzer şekilde sahte haber türleri içinde haber alanlarında olduğu gibi birkaç tür ile sınırlandırılmamış, Bölüm 2.1.2’de belirtilen türlerin hepsine yer vermeye çalışılmıştır.



Şekil 3.6 TR_FN veri setindeki haber alanlarının dağılımı

Çalışılan veri setlerinin gerçek hayattaki verileri temsil edip etmediği çalışma sonuçları ve elde edilen istatistiki bilginin daha büyük derlemlerle yapılan çalışmalarla kıyaslanması ile yorumlanabilir. Bu kapsamda TR_FN veri setinden elde edilen Çizelge 3.5’te sunulan en çok kullanılan kelimelerin, Zemberek geliştirici ekibinin yaklaşık 5 milyon kelime ile gerçekleştirdiği çalışma ile uyumlu olduğu, bu bakımdan iyi bir temsil olduğu söylenebilir¹⁴. Sözcük ham formu biçiminde toplamda yaklaşık 20,7 milyon kelime (token) havuzundan oluşan veri setimizde geçiş frekansına göre en yüksek 8 kelime (kök form) Çizelge 3.5’te sunulmuştur. Bu istatistiki gösterimde durak/etkisiz kelimeler (*stop words*) filtresi uygulanmamıştır.

Çizelge 3.5 Haberlerde en çok kullanılan (kök form) sözcükler.

En çok kullanılan kelimeler	
Kelime	Geçiş sıklığı
ol	146787
et	124784
ve	84712
yap	82681
ver	66497
gel	65037
bu	64819
bir	62840

¹⁴ <http://zembeknlp.blogspot.com/2006/11/kelime-istatistikleri.html> (Erişim tarihi: 21 Haziran 2020)

Sözlük Temelli Sahte Haber Tespitini anlatan bu bölümde problem kelime-çantası/kesesi (*bag of words*) yaklaşımı ile ele alınmıştır. Veri setindeki dengesiz dağılım her ne kadar bu bölümdeki yaklaşıma büyük etkisi olmasa da özellikle eğitim veri setlerindeki dengesizliğe karşı hassas olan makine öğrenmesi modellerinde ayrıca ele alınması gereken bir husus olarak belirlenmiştir. Bu bakımdan TR_FN veri seti açısından Çizelge 3.3'te istatistiki bilgileri verilen çizelgede de gözlenebilecek olan dengesiz bir görünüm sergileyen sahte-gerçek haber oranlarından bahsetmek yerinde olacaktır. Veri kümesinin dengesiz doğasının elverişsiz bir durum (*handicap*) yarattığı söylenebilir, ancak haber verisinin gerçek dünyadaki dağılımı düşünüldüğünde bu beklenmedik bir durum değildir. Çünkü gerçek dünyadaki veri kümelerinde verinin çoğunluğu normal gözlemlerden oluşurken sadece küçük bir oranda anormal ve ilginç örneklerden oluşmaktadır. Gerçek hayatta verilere ait sınıf dağılımındaki dengesizliğe sıklıkla rastlanılmaktadır [112]. Sahte haber özelinde, Vosoughi ve arkadaşları çalışmalarında [33] sahte içeriklerin tüm kategorilerde gerçek içeriklerden önemli ölçüde daha geniş, daha hızlı, daha derin ve daha geniş yayılım gösterdiği belirtilmiştir. Makine öğrenmesi temelli çözümde daha ayrıntılı işlenen dengesiz veri seti problemi ve bu probleme yönelik çözüm yöntemi Bölüm 4'te ayrıca sunulmuştur.

Çizelge 3.3'te, sahte haberlerde ortalama kelime ve cümle sayısı istatistiklerinin gerçek haberlere göre düşük olduğu görülmektedir. Bu bilgiden yola çıkarak; sahte haberlerde daha basit bir dil kullanıldığı, geçerli/gerçek haberlerde ise karmaşık (*sophisticated*) ve sözcüksel zenginliğe (*lexical richness*) sahip bir dilin kullanıldığı söylenebilir.

Bundan sonraki bölümlerde gerçek haberlere ait eğitim seti D_{tr}^V , test seti D_{ts}^V ; sahte haberlere ait eğitim seti D_{tr}^F , test seti D_{ts}^F olarak ifade edilecektir.

3.3. Metodoloji

Türkçe, bir cümlenin anlamını ve kullanımını değiştirmek için bir kelimenin sonuna son eklerin eklendiği sondan eklemeli bir dildir. Metodolojiyi kavramak açısından temel bileşenleri kısaca şu şekilde açıklamak önemlidir:

- Lemma, eklenerek türetilmiş bir sözcüğün temel biçimine (canonical form) verilen addır (koş (*run*)).
- Bir lexeme ise; türetilen sözcüğü temsil etmektedir. Genel olarak Türkçe bir sözcük aşağıda görülen yapıya sahiptir (koşucu (*runner*)).

$$Lexeme = Lemma + S_1 + S_2 + \dots + S_k$$

Bu ifadede, Lexeme sözcüğün metinde kullanıldığı hali başka bir deyişle henüz işlenmemiş ham formudur. Lemma kelimenin kökü/gövdesidir. S_i ile ifade edilen eklentiler son ekleri ifade etmektedir. Genellikle kök kelime Lemma, sözcüğün anlamı üzerinde etkiliyken, belli sonekler sözlük anlamı üzerinde etkili olabilmektedir. Hatta cümlenin öğelerinin belirlenmesi eklerle doğrudan ilgilidir. Çalışmamızda farklı karakterler gösterdiğini tespit ettiğimiz Kelimenin Ham Hali (*Raw*), Kök/Gövde (*Root/Stem*), Kelimenin Ham Hali+Öge bilgisi (*Raw+POS*) ve Ekler (*Suffixes*) olacak şekilde dört farklı kategoride inceleme yapılmıştır. Bu sınıfların her biri için ayrı bir model oluşturmak ve sonunda hangi modelin Türkçe sahte haber tespiti için daha etkin bir sınıflandırıcı olduğunu bulmak hedeflenmiştir. Bu analizler, aynı zamanda Makine Öğrenmesi kullanılmadan sahte haberin ne oranda tespit edilebileceğini göstermesi açısından da önemlidir. Ortaya çıkarılan modellerin bundan sonraki bölümlerde İngilizce isimleri kullanılacaktır.

3.3.1. Türkçe Sahte Haber Sözlüğünün Oluşturulması

SHS geliştirilirken kullanılan metodoloji bu bölümde anlatılmıştır. Literatürde sahte haberler veya terimlere sahtelik/gerçeklik değeri atayarak sözlük oluşturan bir çalışma bulunmamaktadır. Bu bakımdan, sözlük oluştururken izlenen metodolojinin, duygu sözlüğü oluştururken izlenen derlem temelli yöntemlere yakınsadığı söylenebilir.

İlk olarak D_{tr}^V setinde her bir terim (RAW, ROOT, RAW+POS, SUFFIX) için frekans hesaplaması yapılmıştır. Benzer şekilde D_{tr}^F veri seti içinde terim frekansı hesaplamaları yapılmıştır. Daha sonra Sahte/Gerçek Değer (*Fake/Valid Score*) olarak adlandırdığımız yeni bir metrik sayesinde haber dokümanında geçen terimler için sahte/gerçek değerlerin toplanması ile habere sahte ve gerçek bir doküman skor değeri atanmıştır. Kullanılan terimler geliştirilen modellere ismini vermiştir.

Bu doğrultuda, dört farklı sözlüğü oluşturmak için o sözlüğe karşılık gelen terimlerin veri kümelerinden tespit edilmesi gerekmektedir. Bu amaçla D_{tr}^V ve D_{tr}^F veri kümeleri RAW, ROOT, RAW+POS, SUFFIX (SUFF) sözlük modellerine ait terimlerin çıkarılması için işlenmiştir. Çıkarılan her bir terim için biri D_{tr}^V diğeri D_{tr}^F kullanılarak 2 skor elde edilmiştir.

$S_{t,C}^V$ ve $S_{t,C}^F$ değerleri, model sınıfını temsil eden C (*Class*) ait her bir t terimi elde edilen skorları temsil eder ve Eşitlik (5) ve (6)'de gösterildiği şekilde hesaplanmaktadır.

$$S_{t,C}^V = \frac{\sum_{d \in D_{tr}^V} f_d(t)}{\sum_{d \in D_{tr}^V} \sum_{x \in T_C} f_d(x)} \quad (5)$$

$$S_{t,C}^F = \frac{\sum_{d \in D_{tr}^F} f_d(t)}{\sum_{d \in D_{tr}^F} \sum_{x \in T_C} f_d(x)} \quad (6)$$

Bu denklemlerde T_C ; RAW, ROOT, RAW+POS, SUFFIX sınıflarını temsil eden C sınıfına ait modeldeki tüm terimlerin kümesini ifade etmektedir. Ayrıca $f_d(x)$ 'de x teriminin, d haber dokümanı içerisindeki frekansını temsil etmektedir.

Tüm modeller için terimlere ait gerçek/sahte skor hesaplamalarını yaptıktan sonra elde edilen bazı istatistiki sonuçlar Çizelge 3.6'da gösterilmiştir.

Çizelge 3.6 Elde edilen sözlük modellerine ait istatistiki bilgiler.

Model	Tekil Terim	Ortak Terim	Sadece Sahte	Sadece Gerçek
RAW	443174	10166	456	432552
ROOT/STEM	63237	3830	77	59330
RAW+POS	457187	10301	466	446420
SUFFIX	9930	898	12	9020

RAW formu, sözcüğün haber dokümanında geçen ham halinin ele alındığı, modellerimiz içinde en sezgisel olanıdır. Bu nedenle modelde çok sayıda tekil (benzersiz) sözcük bulunmaktadır. Bu modeldeki sahte ve gerçek skor değerleri bulunan ortak sözcükler ele

alındığında bu sayı 10166'ya düşmektedir. Bu durum iki nedenden kaynaklanmaktadır: Birincisi, D_{tr}^F veri kümesi D_{tr}^V 'ye göre oldukça küçük olması nedeniyle daha az sözcük içermesi ve kullanılan dilin basit olmasıdır. Öte yandan bu modelde 456 adet sözcüğün sadece sahte derlemde bulunması incelenmesi gereken bir husus olarak ortaya çıkmaktadır. Bu durum incelendiğinde, bu sözcüklerin çoğunlukla resmi olmayan (*informal*) kullanım veya şaşkınlık, hayret, sevgi, korku, sesleniş, coşku, heyecanlanma ve sitem ifade eden cümlelere ait ünlem sözcükleri olduğu görülmektedir. Bu tarz kullanımların ciddi habercilikte pek yeri olmadığı düşünüldüğünde, bu bulgular sahte habercilik açısından önemli işaretlerdir. Çizelge 3.7'de modele ait bazı ilginç örnekler skorlarıyla birlikte sunulmuştur. Çizelgede sağda gerçek değeri yüksek solda sahte değeri yüksek olan terimler gösterilmiştir. Tüm değerler daha okunabilir olması açısından 10.000 ile çarpılmıştır¹⁵.

Çizelge 3.7 Raw modele ait bazı örnekler.

Ham Sözcük (RAW)	Sahte Ton Değeri	Gerçek Ton Değeri	Ham Sözcük	Sahte Ton Değeri	Gerçek Ton Değeri
Ardından	14,80	21,80	Zaten	60,62	18,60
Bulunan	7,85	17,09	Bulunmuş	0,6	0,1
İlk	10,29	20,12	Son	17,21	15,59
Tutuklanarak	0	1,19	Serbest	3,62	2,95
Yok	5,74	6,22	Değil	11,79	7,07
Doğrultusunda	0	1,59	Gibi	32,32	17,60
Korktu	0	0,01	Tırstı	0,3	0
Güvenilmez	0	0,05	Kaypak	0,6	0
Mantıksız	0	0,01	Saçma	1,20	0,06

Kelimelerin kök/gövdelerini ele aldığımız ROOT modelde ise, benzersiz terimlerin sayısı doğal olarak azalarak 63237'ye düşmüştür. Ortak terimlerde de belirgin bir azalma gözlemlenmektedir. Bu modelin analizinde 77 adet kök/gövde biçimindeki terimin sadece sahte derlemde gözlemlenmiş olduğunu tespit ettik. Bu kelimelerin çok büyük bir çoğunluğunun argo (*slang*) ve uydurma (*make-up*) terimler olduğu görülmektedir. Bu bulgular RAW modelde tespit edilenler gibi sahte haberciliğin ipuçlarını veren işaretlerdir. Çizelge 3.8'de modele ait bazı ilginç örnekler skorlarıyla birlikte

¹⁵ Bundan sonraki Bölüm 3 içerisindeki çizelgelerde ve hesaplamalarda daha okunaklı olabilmesi açısından aynı şekilde 10.000 ile genişletilme yapılmıştır.

sunulmuştur. Çizelgede solda gerçek değeri yüksek, sağda sahte değeri yüksek olan terimler gösterilmiştir.

Çizelge 3.8 Root modele ait bazı örnekler.

Kök/Gövde (ROOT)	Sahte Ton Değeri	Gerçek Ton Değeri	Ham Sözcük	Sahte Ton Değeri	Gerçek Ton Değeri
Süs	0	0,29	Janjan	1,21	0
Ara	28,70	38,90	Bul	8,15	7,26
Değer	6,64	13,37	Haber	17,82	7,49
Sahtekâr	0	0,03	Sahte	1,81	1,56
Ehil	0	0,12	Tıs	0,6	0
Hilekâr	0	0,004	Fırıldak	2,41	0,001
Net	0,91	1,55	Gibi	32,31	17,86
Düzgün	0	0,29	Paçoz	0,3	0
Yüzde	0	4,07	Tahmini	1,21	0,03

RAW+POS Form Modelinde, kelimelerin ham formları ve hangi cümle ögesine karşılık geldiğini gösteren POS etiketleri ile eşleştirilmiş hali bir terim olarak ele alınmıştır. Bu modelde bir sözcük farklı cümle ögesi eşleşmeleriyle nedeniyle birden fazla kez geçebilmektedir.

Çizelge 3.9’da görüldüğü üzere bu durum Raw+Pos modelde, Raw modeldeki ham sözcük sayısından daha fazla tekil terim yani “sözcük+sözcük türü” içermesine neden olmuştur. Modelde; 457187 tekil terime karşılık her iki derlemde de geçen 10301, sadece sahte derlemde geçen 466 tekil terim bulunmaktadır. Çizelge 3.9’da modele ait bazı ilginç örnekler skorlarıyla birlikte sunulmuştur. Çizelgede sağda gerçek değeri yüksek solda sahte değeri yüksek olan terimler gösterilmiştir.

Çizelge 3.9 Raw+Pos Modele ait bazı örnekler.

Sözcük + Cümle Ögesi (RAW+POS)	Sahte Ton Değeri	Gerçek Ton Değeri	Sözcük + Cümle Ögesi (RAW+POS)	Sahte Ton Değeri	Gerçek Ton Değeri
Teröristlerce (İsim)	0	0,27	Terörist (İsim)	3,32	2,11
Apaçık (Sıfat)	0	0,02	Alenileşen (Sıfat)	1,21	2,33
Bitmemiş (Fii)	0	0,01	Bitmiyormuş (Fill)	0,3	0
Sımsıkı (Zarf)	0	0,001	Yapışırçasına (Zarf)	0,3	0
Uyanık (İsim)	0	0,07	Uyanık (Sıfat)	1,21	0,1
Olay (İsim)	6,04	15,59	Olaylara (İsim)	0,6	0,3
Biz (Zamir)	11,17	12,33	Siz (Zamir)	1,81	0,1
Sınır (İsim)	1,21	1,38	Hadleri (İsim)	0,6	0
Olduğunu (İsim)	18,72	22,70	Olarak (Noun)	44,70	41,04

Son modelimiz ise, SUFFIX (SUFF) Form Modelidir. Bu tarz bir modelin sadece sondan eklemeli dillerde kurgulanabileceği göz önüne alındığında Türkçe için metin tabanlı sınıflandırma yapılırken ham sözcük ve kök biçimlerinin yanı sıra eklerin de incelenmesinin önemli olduğunu göstermektedir.

Literatürde son ek kombinasyonlarının ele alındığı bir çalışma olmaması nedeniyle bu model yenilikçi ve diğer çalışmalarda da yararlanılabilecek oldukça ilginç bir çıkış noktası olabileceği değerlendirilmektedir.

Bu modelde, bir kelimenin son eklerinin tüm olası alt dizileri dikkate alınmıştır. Örneğin, $R + S_1 + S_2 + S_3$ RAW Form Modelinde bulunan bir kelimeyse, SUFFIX Form Modelinde $S_1, S_2, S_3, S_1+S_2, S_2+S_3$ ve $S_1+S_2+S_3$ terimler olarak kabul edilmektedir. Bu modelde 9930 adet ek kombinasyonu elde edilmiştir. 898 adet ortak kombinasyon dizisi varken, 12'si sadece sahte haberlere aittir. Çizelge 3.10'da modele ait bazı ilginç örnekler skorlarıyla birlikte sunulmuştur. Bu bölümde sözlük modelleri oluşturulduktan sonra, bu modellerle yapılan sahte haber tahmini, Bölüm 3.3.2'de anlatılacaktır.

Çizelge 3.10 Suffix Modele ait bazı örnekler.

Ekler (SUFFIXes)	Sahte Ton Değeri	Gerçek Ton Değeri	Sözcük
Caus-Caus-Neg-FutPart-A3pl-Acc	4,11	0	ARA -t-tır-ma-yacak-lar-ı
Inf2-P3pl-Narr	1,54	0	AT -ma-ları-ymış
A3pl-Loc-Rel-P2sg-Abl	3,86	0	ZAMAN -lar-da-ki-n-den
PresPart-P3sg-Narr	1,54	0	GİD -en-i-ymiş
A3pl-Dat	111,91	9,49	İNSAN -lar-a
With-A3pl-P2sg	0	5,8	TALİH – li-ler-in
Able-Aor-A1pl	0	5,00	KAÇ -abil-ir-iz

3.3.2. Sözlük ile Sahte Haber Tahmini

Bu bölümde, bir haber metninin sahte olup olmadığını değerlendirmek için oluşturduğumuz modellerin nasıl kullanıldığı açıklanmıştır. Özetle, geliştirilen FanLexTR sözlüğündeki farklı terimlere dayalı olarak oluşturulan 4 farklı model ile 4 farklı analiz yapılmıştır. Analizleri nasıl yapıldığını göstermek amacıyla incelenen örnek haber metni Çizelge 3.11’de gösterilmiştir.

Çizelge 3.11 Test veri setine ait bir sahte haber örneği.

Haber Başlığı	İNANILMAZ AMA DOĞRU
Haber Metni	Ta Küba! Kim gidecek demeyin! Heralde bu yaz tatil listenizdeki yer Küba olmalı. 47 yıldır cinayet işlenmedi. 58 yıldır tecavüz ve istismar suçu işlenmedi. Hatta 5 yıldır hırsızlık bile olmadı. Herkese eşit maaş. Vergi yok. Hemen herşey ücretsiz ya da sudan ucuz. Gezin görün!

Şimdi, bir haber metnini analiz etmek ve değerlendirmek izlenen adımlar sırasıyla anlatılacaktır. İlk olarak, metnin morfolojik analizi yapılarak sözlük modelleri için gerekli tüm terimler çıkarılmakta, ardından modellerdeki sahte ve gerçek ton değerlerine göre ilgili metin puanlanmaktadır. Haber için elde edilen sahte ton puanlamasına haber dokümanı sahte skoru, $S_{D,C}^F$ ve gerçek ton puanlamasına haber dokümanı gerçek skoru, $S_{D,C}^V$ olarak ifade edilmiştir. Bu değerler, Eşitlik (7) ve (8)’de bu hesaplanmıştır.

$$S_{D,C}^V = \sum_{t \in T_C^D} S_{t,C}^V \quad (7)$$

$$S_{D,C}^F = \sum_{t \in T_C^D} S_{t,C}^F \quad (8)$$

Eşitlik (7)'de T_C^D , D haber dokümanında C model sınıfına ait tüm terimlerin kümesini ifade eder. Bu iki değer tüm modeller için hesaplandıktan sonra karşılaştırması yapılmaktadır. Bu karşılaştırmada haber dokümanı her bir model için eğer $S_{D,C}^V$ değeri $S_{D,C}^F$ değerinden büyükse gerçek, tam tersi durumda sahte olarak değerlendirilmektedir. Eşitlik (9)'te bu karşılaştırma verilmiştir.

$$L_D = \begin{cases} GERÇEK, & \text{eğer } S_{D,C}^V > S_{D,C}^F \\ SAHTE, & \text{değilse} \end{cases} \quad (9)$$

Çizelge 3.11'de verilen örnek haber metninin değerlendirilmesi yukarıda belirtilen yönergeler doğrultusunda gerçekleştirilmiştir. İlerleyen bölümlerdeki anlatımı detaylara boğmamak ve takip kolaylığı sağlaması adına nispeten kısa bir haber seçilmiştir. Bu haber, tık tuzağı (*click-bait*) diye bilinen ve okuyucuyu haber sitesinde tutmayı hedefleyen ve bilginin kısa kısa verilerek bir anda sunulmadığı bir haber türüdür. Bu tarz haberlerde genellikle çok az da olsa doğru bilgi kırıntıları bulunmakta ve okuyucuda bir sonraki bilgiye ulaşma isteği yaratması hedeflenmektedir. Örneğin bu haberde, Küba'nın bir tatil noktası olarak ilgi çekebileceğini tespit eden sahte haber yazarı, devamında doğrulayıcı referanstan uzak ama yine kullanıcının dikkatini çekebilecek bilgiler vermektedir. Ayrıca, aynı site içerisinde Latin ülkelerine tatil kampanyaları sunan tatil tur reklamlarının (*pop-up*) açılması, haberin ticari bir kaygı taşıdığını göstermektedir.

3.3.2.1. RAW Model ile Yapılan Değerlendirme

Bu modele ilişkin olarak haber metninde 39 adet terim bulunmuştur. Bu terimlerden $S_{t,C}^F$ ve $S_{t,C}^V$ değerleri bağlamında çarpıcı birkaç örnek üzerinden açıklamalarda bulunmak uygun olacaktır. İlk olarak yazım hatası içeren bir sözcükten örnekleme yapalım. İnceleyeceğimiz ilk terim metinde “heralde”¹⁶ şeklinde geçen doğru yazımı “herhalde”

¹⁶ Bu tarz yazım yanlışları ve hatalı kullanımların büyük bir kısmını tespit edebildiğimiz halde düzeltmeden orjinal halleri ile bırakılmıştır. Çünkü, bu tarz kullanımların özensiz haberciliğin bir özelliği olduğunu düşünmekteyiz.

olan cümleye olasılık anlamı katmaya yarayan bir ifadedir. Burada dikkat edilmesi gereken bir hususta sözcüğün Türkçede günlük konuşmada sıklıkla kullanılan haliyle yani sözcükteki “h” harfi yutulmuş bir şekilde haber metninde kullanılmış olmasıdır. Yanlış bir kullanım olmasına rağmen bu terimi gerçek haber derleminde de gözlemlemek mümkündür. Bu gerçek haberlerde de hatalı kullanımlar olabileceğini göstermesi açısından önemlidir. Fakat hesaplanan skor değerlerindeki fark bu terimin sahte haberi belirlemede seçici bir özelliğinin olduğunu gösterir. Şöyle ki “heralde” teriminin sahte derlemede 600 kat daha sıklıkla geçtiği söylenebilir. Çizelge 3.12’de gözlemlenebileceği üzere $S_{t,C}^F$ ve $S_{t,C}^V$ değerleri arasındaki fark 0.603’tür.

Bu metinde göze çarpan diğer bir terimde “Hatta 5 yıldır hırsızlık bile olmadı.” cümlesinde geçen “bile” terimidir. Aynı zamanda, dahi anlamına gelen terim bu cümleye şaşkınlık katan bir işlev kelimesidir (*function word*). Modelden elde edilen skorlara bakıldığında terimin her iki derlemede de sıklıkla kullanıldığı gözlemlenmektedir. Fakat, $S_{t,C}^F$ ve $S_{t,C}^V$ değerleri arasındaki farkın göze çarpan bir düzeyde olduğu da görülmektedir. Çizelge 3.11’deki haber örneği incelenerek elde edilen diğer ilginç örnekler Çizelge 3.12’de sunulmuştur.

Çizelge 3.12 RAW Form modele göre puanlanmış bazı örnekler.

Terim	$S_{t,C}^F$	$S_{t,C}^V$
heralde	0,604	0,001
listenizdeki	0,302	0,0003
herşey	0,604	0,0286
görün	0,906	0,0418
tecavüz	2,7181	0,281
bile	8,1544	2,5641

Bu analizlerden yola çıkarak sahte haberlerde, Zipf kanunlarında (Zips Law) [113] belirtilen bir ilke olan konuşmacıların (bu örnek özelinde sahte haber yazarının) mümkün olabildiğince az gayret ile farklı kavramları ifade etmeye çalışması göze çarpmaktadır.

Zipf Kanunları (Yasası): Zipf yasası, K. Zipf (1902–1950) tarafından ortaya konulan istatistiki hesaplamalar kullanılarak metinlerde kelimelerin geçiş sıklıkları ile ilgili belirlenmiş bir genellemedir [114]. Temel olarak sözcükler kullanılma sıklığına göre en çoktan en aza doğru sıralandığında, ortaya çıkan sıralama listesindeki tek bir sözcüğün sıra numarası ile o sözcüğün geçiş sıklık sayısının her zaman sabit bir sayı olduğu temel varsayımına dayanır. Daha basit bir ifade ile geçiş sıklığı değerlendirmesinde N. Sırada olan bir sözcüğün sıklığının $1/N$ şeklinde ifade edilmesidir. N'nin Türkçede bulunan sözcük sayısı olduğu varsayılırsa aşağıdaki Eşitlik (10)'da verilen fonksiyon, k sıralama numarası gösterilen sözcüklerin dağılımını göstermektedir.

$$f(k; s, N) = \frac{1 / k^s}{\sum_{i=1}^N (1 / i^s)} \quad (10)$$

N : Sözcük sayısı

k : Sözcük sıralaması

s : Dağılımı ifade eden üssel değer

Sonuç bölümünde de belirtileceği üzere, aynı kanunlarda belirtilen dinleyicilerin (bu örnek üzerinde okuyucu) ise farklı kavramlar için farklı kelime beklemek ve böylece anlama gayretlerini en aza indirmek istemesi durumu ise, internet okuryazarlığı nezdinde tam tersine çalışıyor görünmektedir. Şöyle ki, genel olarak ilginç gelen sansasyonel içeriği yayma güdüsü taşıyan çevrim içi okuyucu kitlenin aynı zamanda hızlı ve kolayca okuyabileceği basit bir dil ile beslenmek istediği söylenebilir. Bu tespit modelin, semantik yaklaşımlarla birlikte kullanıldığında daha tutarlı sonuçlar verebileceğinin göstergesidir.

Haberdeki tüm terimler RAW model sözlüğe göre puanlandığında, $S_{D,C}^V$ 50.20, $S_{D,C}^F$ ise 70.37 toplam skor değerleri elde edilmiştir. Aradaki bu fark dikkat çekicidir. Bu modele göre bu haber net bir fark ile “sahte” olarak etiketlenmiştir.

3.3.2.2. ROOT Model ile Yapılan Değerlendirme

Bu analizde RAW Modeline benzer şekilde haber metnindeki kök/gövde biçimindeki terimler bulunarak hesaplamalar yapılmıştır. Burada değinilmesi gereken önemli bir husus, Türkçe sözcüklerde kök/gövde çıkarımı yapmanın kolay bir görev olmadığıdır. Zemberek kütüphanesi her ne kadar gelişmiş bir morfolojik çözümleme gücüne sahip olsa da yüzde yüz doğrulukla köklerin çözümlendiği söylenemez. Analiz aşamasında, bazı yanlış çözümlenmeleri ve “heralde” örneğindeki gibi hatalı yazımlardan kaynaklı durumları tespit edilmiştir. Fakat, sistemin olabildiğince otomatik işlemesi için müdahalede bulunulmamıştır. Hatalı yazımların sahte derlemdeki yoğunluğu düşünüldüğünde bu durumun tespiti kolaylaştırıcı bir etkisinin olduğu da söylenebilir.

Puanlama sonucunda, $S_{D,C}^V$ 493.46, $S_{D,C}^F$ ise 504.06 olarak hesaplanmıştır. RAW modelle kıyaslandığında değerler birbirine daha yakın olarak çıkmıştır. Bu modelde de haber “sahte” olarak etiketlenmiştir. Bu modelde $S_{D,C}^V$ ve $S_{D,C}^F$ değerleri bağlamında farklar diğer modellere göre daha tutarlı bir şekilde seyretmiştir.

3.3.2.3. RAW+POS Model ile Yapılan Değerlendirme

Bu üçüncü model RAW modele, sözcüklere, sözcük türlerini diğer bir deyişle cümle içindeki görevlerini atayarak çeşitlilik katmıştır. Sonuçları açısından RAW modele benzerlik göstermektedir. Puanlama sonucunda, $S_{D,C}^V$ 44.634, $S_{D,C}^F$ ise 62.517 olarak hesaplanmıştır. Farkın ilk analizdekine benzer şekilde açık olduğu bu modelde haberi “sahte” olarak etiketlemiştir. Hem RAW+POS model hem de SUFFIX modelin çözümleme belirsizlikleri ortadan kaldırıldığında çok daha başarılı tahminler yapacağı değerlendirilmektedir.

3.3.2.4. SUFFIX Model ile Yapılan Değerlendirme

Son modelimiz ise son ek kombinasyonlarından oluşan sözlüğü içeren ve SUFFIX Model¹⁷ olarak adlandırdığımız modeldir. Daha önce belirttiğimiz üzere bu model son ek morphemelerinin sıralı kombinasyonlarını içerdiğinden (biçimbirim), sözcüğün kökünü dikkate alınmamıştır.

¹⁷ Bazı şekil ve grafiklerde “SUFFIX” Model “SUFF” şeklinde kısaltılmıştır.

Analiz edilen bu haber özelinde 25 adet son ek dizisi elde edilmiştir. Bunlardan ikinci çoğul şahıs (second plural) A2Pl ton değeri açısından ilginçtir. Bu son ekin, $S_{t,C}^F$ değeri 16.980 ve $S_{t,C}^V$ değeri 4.077 ‘dir. Oluşturulan derlemlere göre bu ekin sahte haberlerde daha sıklıkla kullanıldığı gözlemlenmektedir. Gerçekten de ciddi habercilikte haberler direk okuyucunun şahsına anlatılır gibi sunulmadığı bilinmektedir. Bu haber örneğinde, “demeyin” ifadesi direk olarak okuyucuyla ikili konuşma içindeymiş gibi sunulmuştur. Özellikle sosyal medyadaki haber niteliğindeki paylaşımlarda benzer şekilde duyuru veya ikili konuşma biçimine sıklıkla rastlanmaktadır.

Diğer modeller ile kıyaslandığında bu model genel olarak daha düşük bir başarı oranına sahiptir. Bu haber özelinde bu modelde de göreceli büyük bir fark ile $S_{D,C}^V$ 954.4907, $S_{D,C}^F$ ise 981.9913 hesaplanmış ve “sahte” olarak etiketlenmiştir.

3.3.3. Çıktılar ve Performans Değerlendirme

Bölüm 3’te, FanLexTR olarak adlandırdığımız 4 farklı modeldeki genel amaçlı SHS geliştirilmiştir. Sözlüklerde terimlere ait sahte ve gerçek olmak üzere iki farklı ton değeri mevcuttur. Önceki bölümlerde, örnek bir haber üzerinden bir haber dokümanının değerlendirilmesinin nasıl gerçekleştirildiği gösterilmişti. Bu bölümde amacımız, FanLexTR’nin tüm kategorilerde sahte haber tespit performansının kapsamlı değerlendirilmesidir. Bu değerlendirmede, gelecek çalışmalara ışık tutması açısından, hiçbir makine öğrenmesi metodunun kullanılmadan sadece dilin temel bileşenleri kullanılarak, Türkçe için Sözlük Temelli Sahte Haber tespiti çalışması ölçümlenmiştir. Tüm modellerin performansını ölçümlemek ve karşılaştırmalı bir analiz sunmak için deneylerimizin sonuçları özetlenmiştir.

Geliştirilen modellerde sahte haberin tespiti daha ön planda olduğu için yanlışlıkla gerçek olarak etiketlenen sahte haber sayısını en aza indirmek kurgulanan metodolojideki temel hedeftir. Bölümün bundan sonraki kısımlarında, test veri seti üzerinde yapılan değerlendirmelerin istatistikleri sunulacaktır. Test veri seti eğitim sürecine dahil edilmediğinden (*hold-out cross validation*), bu veri setinde sözlüklerde bulunmayan

terimlerin ortaya çıkması muhtemeldir. Böyle bir durum ile karşılaşıldığında terimin sahte ve gerçek her iki değeri de 0 olarak hesaplanmıştır.

Bu değerlendirme sürecinde eğitim setine hiç katılmadan ayrılan test seti kullanılmıştır. Ayrıca sonuçların genel-geçer olup olmadığını olduğunu da ölçümlenerek sonuçlar değerlendirilmiştir. Bunu sağlayabilmek için, testlerimizi değerlendirebileceğimiz istatistiksel olarak sağlam bir karşılaştırma ölçütü (*benchmark*) oluşturulmuştur. Bu doğrultuda, D_{tr}^F ve D_{tr}^V eğitim veri setlerini test ve eğitim şeklinde rastgele iki parçaya bölerek her seferinde birini test kalanları eğitim veriseti olarak hold-out doğrulamayla kullanılmıştır. Hold-out yönteminde sıklıkla, verinin eğitim için %80'ini ve test için %20'si kullanılır. Burada sınıf dağılımındaki dengesizlikten dolayı bu oranı %90'a, %10 olacak şekilde ele alınmıştır. Bu tarz bir yöntem seçilmesinin iki nedeni vardır: Birincisi, olabildiğince fazla haberi kullanarak daha istikrarlı ton değerleri üretmek, bir diğer nedende makine öğrenmesi temelli bir yaklaşım yürütülmediğinden çapraz-doğrulama işlemlerinin kullanılan yöntemeye uygun olmama durumudur. Bu deneylerden elde edilen karşılaştırmalı sonuçlar (Eğitim veri setlerine ait (D_{tr}^F ve D_{tr}^V)), Çizelge 3.13 ve 3.14'te sunulmuştur.

Çizelge 3.13 Eğitim veri setlerine ait hold-out doğrulama sonuçları.

Veri setleri	Sözlük Modelleri	Kesinlik (Precision)	Duyarlılık (Recall)	Doğruluk (Accuracy)	F1 Değeri
Set-1	Raw	78,846	85,416	81,25	82
	Root	75,564	83,75	78,333	79,44
	Raw+Pos	75,806	78,333	76,667	77,049
	Suff	73,19	71,667	72,708	72,421
Set-2	Raw	79,347	91,25	83,75	84,883
	Root	75,94	84,167	78,75	79,841
	Raw+Pos	75,572	82,5	77,917	78,884
	Suff	74,059	73,75	73,958	73,903
Set-3	Raw	78,445	92,5	83,541	84,894
	Root	75,357	87,917	79,583	81,153
	Raw+Pos	74,532	82,917	77,292	78,501
	Suff	74,089	76,25	74,792	75,154
Set-4	Raw	79,71	91,667	84,167	85,271
	Root	75,547	86,25	79,167	80,544
	Raw+Pos	75	81,25	78,083	78
	Suff	74,477	74,167	74,375	74,322
Set-5	Raw	78,677	89,167	82,5	83,593
	Root	77,737	88,75	81,667	82,88
	Raw+Pos	76,245	82,917	78,541	79,441
	Suff	75,294	80	76,875	77,575

Burada makine öğrenmesinde tercih edilen çapraz doğrulamayı (*cross-validation*) yerine bu metodu (*hold-out validation*) kullanmamızın nedeni modelleri değerlendirme parametrelerinin her seferinde yeniden hesaplanması yani modelin yeniden oluşturulmasıdır. Makine öğrenmesi kütüphanelerinde oldukça hızlı bir şekilde yapılabilecek hesaplamalar bu yöntem özelinde sözlüklerin baştan hesaplamasını da içeren oldukça maliyetli bir çaba gerektirmiştir. Modelin ilk geçerlemesi için bir kez yapıldıktan sonra tez çalışmasının sonunda bu işlem 4 kez daha tekrarlanmıştır. Böylelikle geçeleme işlemi farklı kombinasyonlarla 5 kez tekrarlanarak bir nevi çapraz-doğrulamaya benzer şekilde kullanılmıştır.

Çizelge 3.14 Eğitim veri setlerine ait hold-out doğrulama sonuç ortalamaları.

Veri Setleri	Sözlük Modelleri	Kesinlik (Precision)	Duyarlılık (Recall)	Doğruluk (Accuracy)	F1 Değeri
Doğrulama	Raw	79	90	83,04	84,12
	Root	76,02	86,8	79,5	80,77
Setleri	Raw+Pos	75,43	81,58	77,5	78,37
Ortalama	Suffix	74,15	75,16	74,5	74,64

Çizelge 3.14 incelendiğinde sonuçların Çizelge 3.15'teki test veri setindeki (D_{ts}^F ve D_{ts}^V) sonuçlarla benzeştiği görülmektedir. RAW Form modelde %83,04 doğruluk değeri elde edilmiştir. Doğruluk (*Accuracy*) açısından ikinci sırayı ROOT Form modeli ve RAW+POS'un üçüncü sırada yer aldığı görülür. SUFFIX modelinde %74,5 accuracy ile son sırada yer almaktadır.

Çizelge 3.15 Test veri setlerine test veri setlerine ait değerlendirme sonuçları.

Veri Seti	Sözlük Modelleri	Kesinlik (Precision)	Duyarlılık (Recall)	Doğruluk (Accuracy)	F1 Değeri
Test Seti	Raw	80	92,9	85	86,1
	Root	75,1	91,9	80,07	82,7
	Raw+Pos	76,1	89,5	80,07	82,3
	Suff	74	81,4	76,4	77,6

Çizelge 3.15'te başlangıçta belirlenen test setine ait sonuç değerlendirmesi verilmiştir. Her iki veri setindeki sonuçların birbiri ile doğru korelasyonda olduğu görülmektedir. Test setindeki sonuçların daha iyi olması, geliştirilen sözlüklerin daha geniş bir derlem elde edilmesinden kaynaklanmaktadır. Çizelge 3.14'te her seferinde sözlükler derlemin yaklaşık 4/5'i kadar bir veri kaynak alınarak geliştirilmiştir. Bundan sonraki bölümde test setine ait sonuçlar sözlük bazında değerlendirilecektir. Çizelge 3.16, 3.17, 3.18 ve 3.19'da

tüm modellere ait, hata matrisi sonuçları ve kesinlik, duyarlılık, doğruluk ve F1 Skor metrikleri sunulmuştur. Bu matrislerde 210 sahte, 210 gerçek haberden oluşan test veri setine yönelik tahminler ele alınmıştır.

Çizelge 3.16 Raw Model için Hata Matrisi (Confusion Matrix).

Sistem Tahmini	Gerçek Durum		Kesinlik	Duyarlılık	Doğruluk	F1-Skor
	Sahte	Gerçek				
Sahte (Fake)	195	48	0,802	0,929	0,85	0,861
Gerçek (Valid)	15	162				

Çizelge 3.16’da görüldüğü üzere; RAW Form sözlüğüyle, 210 sahte haberin 195’i sahte olarak, 15 tanesi ise hatalı bir şekilde gerçek olarak tahminlemiştir. Kesinlik (*Precision*) değerinin Duyarlılık (*Recall*) değerine göre düşük olmasına rağmen 0,802 gibi nispeten yüksek bir orandır. Doğruluk değeri (*Accuracy*) 0,85 olarak ölçülmüştür. Bu değer sadece sözlük kullanılarak elde edildiği düşünüldüğünde oldukça yüksek bir başarımın yakalandığı söylenebilir.

Bu problem özelinde Duyarlılık değerinin diğer değerlere oranla fark edilebilir düzeyde yüksek olması, başlangıçtaki motivasyonla örtüşmektedir. Bir başka deyişle sistemin haberi gerçek ve sahte diye ayırt etmesinden ziyade, sahte haberi daha büyük oranda doğru tespit edebilmesi ön plana çıkmaktadır. Bu duruma, Bölüm 3.4’te (Sonuç Bölümü) değinilen eleştirel bir bakış açısıyla yaklaşmıştır. Ayrıca modelin bu eleştirel kısımları Bölüm 4’teki hibrit çözümde, bütünsel yaklaşım içeren makine öğrenmesi yaklaşımı ile giderilmiş çok daha kararlı bir yapıya ve yüksek başarım oranlarına ulaşılmıştır. Sonuç olarak yaklaşık %93 duyarlılık performansı ile Raw Form sözlüğün en tüm sözlüklerin içerisinde en başarılı sözlük modeli olduğu görülmektedir.

Çizelge 3.17 Root Model için Hata Matrisi (Confusion Matrix)

Sistem Tahmini	Gerçek Durum		Kesinlik	Duyarlılık	Doğruluk	F1-Skor
	Sahte	Gerçek				
Sahte (Fake)	193	64	0,751	0,919	0,807	0,827
Gerçek (Valid)	17	146				

Root modeline ait Hata Matrisi Çizelge 3.17’de gösterilmiştir. Sahte haberin tespit etme hususunda Raw Form Model’e yakın bir performans (-%1), fakat gerçek haberin doğru oranda tespit edilmesinde %5 düşük bir performans sergilemiştir. F1-Skor’da %82 ile Raw Form Modele göre biraz düşük bir başarıyı elde edilmiştir. Bu modelde diğer modellerin arasında başarı oranları açısından ikinci olarak yerini almıştır. Buradan yola çıkarak sözcüklerin kök/gövde hallerinin haberin sahteliğine yönelik anlamlı ipuçları barındırdığı söylenebilir.

Çizelge 3.18 Raw+Pos Model için Hata Matrisi (Confusion Matrix).

Sistem Tahmini	Gerçek Durum		Kesinlik	Duyarlılık	Doğruluk	F1-Skor
	Sahte	Gerçek				
Sahte (Fake)	188	59	0,761	0,895	0,807	0,823
Gerçek (Valid)	22	151				

Çizelge 3.18’de RAW+POS modele ait Hata Matrisi gösterilmiştir. Başarı oranı ele alındığında üçüncü sıradaki bu model, ROOT Form model ile kıyaslandığında daha düşük duyarlılık oranına sahipken, daha yüksek bir kesinlik oranına sahiptir. Ve çok yakın bir doğruluk oranına sahiptir. Sözcüklerin ham hallerini ve öge bilgisi olmak üzere iki önemli bilgiyi taşımasına rağmen, üçüncü sırada yer almıştır. Taşıdığı bilgi göz önüne alındığında cümle öğelerinden bir öznelik (*feature*) olarak faydalanılmıştır. Bölüm 4.3.3 ve 4.4.1.2’de bu husus anlatılmıştır.

Çizelge 3.19 Suffix Model Hata Matrisi (Confusion Matrix).

Sistem Tahmini	Gerçek Durum		Kesinlik	Duyarlılık	Doğruluk	F1-Skor
	Sahte	Gerçek				
Sahte (Fake)	171	60	0,74	0,814	0,764	0,776
Gerçek (Valid)	39	150				

Son olarak, Çizelge 3.19’da Suffix modele ait Hata Matrisi gösterilmiştir. Diğer modellere göre başarı oranı düşüktür. Fakat oluşturulan sözlük modelinin kök ve sözcükten bağımsız olarak ek kullanımı ve eklerin sıralı kombinasyonlarını sergilemesi açısından diğer çalışmalara da ışık tutabilecek farklı bir model olduğu aşikârdır.

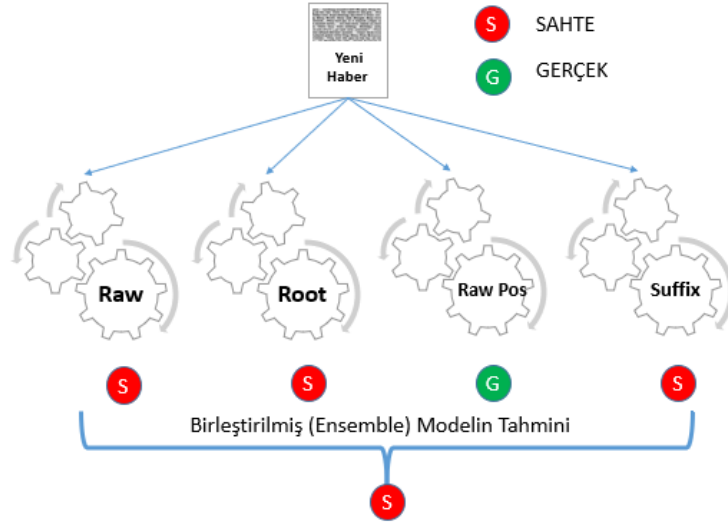
Tüm modellerin test seti ve doğrulama yapılarak elde edilen sonuçları değerlendirildiğinde, istikrarlı ve motive edici sonuçlar elde edilmiştir. Doğrulama elde edilen sonuçlardaki nispeten ufak sapmaların, daha az veri ile çalışılmasından kaynaklandığı görülmektedir. Daha büyük bir test seti çalışıldığında daha kararlı sonuçlar elde edilebileceği değerlendirilmektedir. Bölüm 4'te, Makine öğrenmesi yöntemleri kullanılarak Bölüm 3'te geliştirilen modellerin sınıflandırmadaki tutarlılığı da ölçülmüştür.

3.3.3.1. Topluluk/Birleştirilmiş Öğrenme Modeli (Ensemles Model) Çıktıları

Özellikle Raw ve Root Form Modellerden elde edilen çıktılar, Türkçe açısından metin tabanlı sahte haber tespiti hususunda henüz makine öğrenmesi gibi daha karmaşık metodolojiler uygulanmamasına rağmen motive edici sonuçlar alınabileceğinin bir göstergesidir. Bu sonuçlar Türkçe haberler için özellikle gerçek olmamasına rağmen sonradan üretilen uydurma (*fabricated*) tarzı haberlerde kullanılan dildeki sahteliğin yakalanabileceğini göstermektedir. Ayrıca N-gram modellerinin makine öğrenmesi metotları ile kullanıldığında çok daha başarılı sonuçlar alınabileceğini öngörmemizi sağlamıştır.

Topluluk (*Ensemble*) modeli, aynı hedef değişkeni tahminleyen sistemlerin birlikte daha iyi çalışacağı mantığından ortaya çıkmıştır. Genellikle benzer sınıflandırıcı modeller birlikte kullanıldığında, modellerin içinde en iyi sonucun alındığı modelden daha sonuçlar üretir. Modellerin birbirlerinin eksiklerini tamamladığı bu grup tahmini algoritmalarına Modellerin Birleştirilmesi (*Ensemble*) Yöntemi denir. Bu öğrenme tekniği "Ensemble Learning" olarak geçer. Oldukça esnek olan bu yöntemde ön şart modellerin benzer metodolojiler kullanılarak geliştirilmiş ve tutarlı olmasıdır. Ensemble teknikleri temelde Toplama (*Bagging*) ve Arttırma (*Boosting*) olarak ayrılabilir. Bu çalışmada Toplama (*Bagging*) tekniği kullanılmıştır. Bu teknik, model ortalama teknikleri ile sınıflandırıcıların birleştirildiği basit bir toplamaya benzetilebilir (Ağırlıklı ortalama, çoğunluk oyu, normal ortalama).

İlerleyen konularda sınıflandırıcı modellerinde görülecek olan Rastsgale Orman modeli bir toplama (*bagging*), Gradyan Artırma modeli ise bir arttırma (*boosting*) örneği olarak değerlendirilebilir. Bu çalışmada daha iyi bir sınıflandırıcı geliştirebilmek için sınıflandırıcılara tahminlerini sorarak oylama yöntemi (*voting classifier*) ile daha iyi sonuç veren bir birleştirilmiş modelin ortaya çıkarılması hedeflenmiştir.



Şekil 3.7 Birleştirilmiş (Ensemble) Model Çoğunluk Kararı (Kural-2)

Dört model ayrı ayrı ele alındığında elde edilen sonuçlar, modeller birlikte ele alındığında daha başarılı bir performans sergileyip sergilemediğini görmek için D_{ts}^F ve D_{ts}^V test veri setlerindeki her bir haber için modellerin tespitlerini birleştiren bir değerlendirme ortaya çıkarılmıştır. Her bir modelimizi bir karar verici olarak etiketleme yapan sistemlere benzetirsek, Birleştirilmiş/Topluluk (*Ensemble*) modeli bu sistemlerden en yüksek faydayı sağlayıcı şekilde tüm sistemleri birlikte kullanmayı ifade eder. Topluluk/Birleştirilmiş (*Ensemble*) modelinde sezgisel yaklaşımla 2 kural ortaya konmuştur:

- Bu kurallardan ilki KURAL-1 (Esnek Karar-Ortalama) olarak adlandırılan dört modelin en az ikisinin (2) aynı etiketlemeyi yapmasına,
- KURAL-2 (Çoğunluk Kararı) olarak adlandırılan diğer kural ise, modellerden en az üçünün (3) aynı etiketlemeyi yapmasına göre karar veren birleştirilmiş (*ensemble*) model ortaya çıkarılmıştır.

Şekil 3.7’de KURAL-2’ye ait çalışma mantığı görselleştirilmiştir. Bu modelde her bir sınıflandırma modeli eşit olarak ağırlıklandırılmıştır. Bu modellerin ortaya çıkardığı sonuçlar Çizelge 3.20’de sunulmuştur.

Çizelge 3.20 Birleştirilmiş Model için Hata Matrisi (Confusion Matrix)

Sistem Tahmini		Gerçek Durum		Kesinlik	Duyarlılık	Doğruluk	F1-Skor
		Sahte	Gerçek				
KURAL-1	Sahte	197	44	0,818	0,938	0,865	0,874
	Gerçek	13	166				
KURAL-2	Sahte	193	52	0,787	0,919	0,836	0,848
	Gerçek	17	158				

3.4. Sonuç (Faz-2)

Tez çalışması kapsamında, sözlük tabanlı Türkçe STS için bilinen ilk bilimsel çalışmalardan biri sunulmuştur. Türkçenin eklemeli dil yapısından yararlanılarak üretilen sözlük modelleri ile sahte haber tespiti yapılan çalışmada motive edici sonuçlar elde edilmiştir. Özellikle eklerin birbirleri ile kullanımını da ortaya koyan sözlüğün diğer metin sınıflandırma çalışmalarında da kullanılabilmesi değerlendirilmektedir.

Özellikle Raw ve Root Form Modellerden elde edilen sonuçlar, Türkçe açısından sahte haberlerin tespiti hususunda sadece N-gram modellerine ait özellik vektörleri kullanıldığında bile motive edici sonuçlar alınabileceğinin bir göstergesidir. Makine öğrenmesi yöntemleri ile sahte haber tespiti, tez kapsamında Bölüm 4’te değerlendirilmiştir.

Genel başarımlarına bakıldığında, Sahte Haber içerik doğrulaması zor bir görev olmasına karşın, sahte haber dilini yakalamanın Türkçe için mümkün olduğu söylenebilir. Yazar tespiti çalışmalarında yazarın üslup ve söylevinin konuya göre şekillendiğini söylemek mümkünken, sahte haberlerde konudan ve alandan bağımsız ifadelerin kullanılabilmesi gözlemlenmiştir. Öte yandan, sahte haberin bilinçli işlenmesi ve

yayılması için farklı yöntemlerin kullanılması, konuyu standart kalıplara indirgenebilecek bir problemden uzaklaştıran zorlayıcı durumlar olarak ortaya çıkmaktadır.

Sözlük temelli çözümde modellerimizin başarımını ilk aşamada Duyarlılık (*Recall*) sıralanmıştır. Bunun nedeni, sahte haberin tespitinin ön planda tutulmasıdır. Bu modelde, sahte haberleri doğru tespit edebilmek hedeflendiğinden, gerçek haberlerin kabul edilebilir bir oranda sahte olarak işaretlenip yanlış alarm üretilmesi durumu ortaya çıkmıştır. Diğer bir deyişle FN oranı, FP oranından daha kritik tutulmuştur.

Fakat özellikle sınıflandırma problemlerinde sadece hassasiyet ya da doğruluk gibi tek bir metrik üzerinden karar verilmesi sistemin genelini başarısını göstermez. Önceki bölümlerde de ifade edildiği gibi sahte haberi iyi oranda tespit edip, gerçek haberi benzer oranlarda tespit edememesi model için bir eksiklik olarak değerlendirilmelidir. Kesinlik ve Duyarlılık değerlerinde bir ödünleşim (*trade-off*) olduğu gözlemlenmektedir. Bu bakımdan uç değerleri cezalandırmak için aritmetik ortalama yerine harmonik ortalamayı kullananan F1 skora bakılması yerinde olacaktır. Ayrıca daha detaylı incelemeler içeren ve genellikle makine öğrenme problemlerinde kullanılan bazı metrikler bir sonraki bölümde daha detaylı verilecektir.

Ayrıca, Türkçenin biçimbirimsel belirsizlik özelliği göz önüne alınarak gerçek anlamlı çok az sıklıkla geçen sözcükler ile yanlış yazım/ kullanımdan kaynaklı çok az sıklıkla geçen sözcüklerin belirlenmesi gelecek çalışmalar için önemlidir. Bu doğrultuda daha büyük bir derlem kullanılarak, geçiş sıklıklarının bir eşik değeri ile ele alınması planlanmıştır. Sözlük çalışmasının derinleştirilmesi kapsamında elde edilen sözlüklerin konu veya alan bazında da çalışılabileceği öngörülmektedir.

Özellikle Birleştirilmiş model kullanıldığında KURAL-1 ile yapılan değerlendirmede başarımın daha da arttığı gözlemlenmiştir. Bununla beraber KURAL-2 çıktılarında görülen hem sahte hem gerçek haberi açısından elde edilen tutarlı sonuçlar modellere ağırlıklandırma (*model weighting*) uygulandığında çok daha kararlı ve tutarlı bir başarımla elde edilebileceğini işaret etmektedir.

Bu bölümde elde edilen sözlük modelleri Bölüm 4’te Makine Öğrenmesi tabanlı çözümde öznitelik olarak ele alınarak hibrit çözüme dahil edilmiştir.

4. MAKİNE ÖĞRENMESİ TEMELLİ SAHTE HABER TESPİTİ

Sahte haber tespiti önceki bölümlerde ifade edildiği şekliyle bir sınıflandırma ve regresyon problemi olarak değerlendirilmektedir. Bu alanda bir kısmı gelecek çalışmalar için sonuç bölümünde önerilen bazı alt konular farklı makine öğrenmesi problemleri olarak ele alınabilir. Örneğin, sahte haber türlerinin belirlenmesi makine öğrenmesi ile ele alınabilecek bir kümeleme çalışması olabilir. Sahte haber tespiti hususunda literatürdeki birçok çalışma olduğu gibi en temel gereksinim, haber verisinin büyük bir kısmını oluşturan metinlerin dilbilimsel analizidir. Bu tez çalışması kapsamında bu analiz Bölüm 3'te istatistiki temellere dayalı olarak geniş kapsamlı bir şekilde ele alınmıştır. Yapılan değerlendirmede makine öğrenmesi yaklaşımları ile daha iyi sonuçlar alınabileceği öngörülmüştür. Bu maksatla, bu bölümde konu makine öğrenmesi ve derin öğrenme yöntemleriyle ele alınmış, geliştirilen çözüm modelleri sunulmuştur.

Bu doğrultuda, makine öğrenmesi ve derin öğrenme yöntemleriyle konunun ele alınması için öncelikli olarak dilbilimsel analizin derinleştirilmesi gerekmektedir. Elde edilen modelin, gerçek hayatta etkin bir şekilde kullanılabilmesi için Bölüm 2.4.1'de çok temel ve esnek bir yapıda önerilen bütünleşik çerçevenin gelişime açık tasarlanması şarttır. Bu çerçevede uydurma/sahte haberin tespitinin yanı sıra içerik doğrulaması (*fact-checking*) modelleri ve geri bildirim mekanizmaları gibi sistemi daha kararlı ve tutarlı hale getirecek tamamlayıcı bileşenler gereklidir. Sonuç olarak geliştirilen hibrit model, Bölüm 2.4.1'de kurgulanan kavramsal çerçevede, otomatik tespit sistemlerine karşılık gelmektedir.

4.1. Giriş (Faz-3)

Bölüm 3'te, Türkçe dili için sahte haber dilbilimsel analizi ve ardından sözlük temelli yöntemle sahte haberin tespiti anlatılmıştır. Akılda kalıcı bir tabirle, Türkçe çevrim içi haberciliğin sahte haber perspektifinden metinsel fotoğrafı çekilmiştir. Bu bölümde; daha ileri ve karmaşık yöntemlerle, Bölüm 3'te tespit edilen bulgulardan da faydalanılarak, otomatik sahte haber tespiti için esas çerçeveyi oluşturmak ve Türkçe için gelecek çalışmalara yön vermek hedeflenmiştir. Bu doğrultuda, problem makine öğrenmesi

temelli bir çözüm ile ele alınmış ve Bölüm 3.'te elde edilen sistem başarımı/performansı geliştirilmiştir. Bu bölümde genel anlamda bir makine öğrenmesi problemi yaklaşımına uygun biçimde örneklemeler yapılarak üzerine çalıştığımız problem bir bütünlük içerisinde sunulmaya çalışılmıştır. Bu kapsamda; ilk olarak kavramsal bütünlük içinde konular sunulmuş, kullanılan yöntem ve teknikler açıklanmış ardından özellik seçimi ve çıkarımı, model seçimi, tez blok diyagramı, sistem performans çıktıları anlatılmış ve son olarak sonuç değerlendirmesi yapılmıştır.

4.2. Makine Öğrenmesi ve Derin Öğrenme

Makine öğrenmesi ve derin öğrenme konuları bilgisayar mühendisliklerinde lisans ve lisansüstü eğitimde ders olarak okutulmaya başlanan çok kapsamlı konulardır. Bu bölümde, olabildiğince öğretici ve genel hatları ile bu konular işlenecektir. Bu iki konunun anlatıldığı bir ortamda yapay zekâ kavramından bahsedilmemesi anlatımı eksik bırakacaktır. Bu nedenle bu bölümde kısaca yapay zekâ kavramı, makine öğrenmesi ve temel yaklaşımları, derin öğrenme konuları işlenecek, sahte haber tespit problemi özelinde kullanımı anlatılacaktır. Bazı ifade ve hususların daha anlaşılabilir kılınması açısından teknik ve kavramsal olarak tespitler ve tanımlamalar yapılacaktır.

Yapay zekâ, en genel tanımı ile “bir bilgisayarın veya bilgisayar kontrolündeki bir robotun çeşitli faaliyetleri zeki canlılara benzer şekilde yerine getirme kabiliyeti” [115] şeklinde özetlenebilir. Literatürde sıklıkla “*AI, Artificial Intelligence*” olarak geçen yapay zekânın günümüzde sayısız çalışma, proje ve uygulama alanı bulunmaktadır. Çağımızın belki de en önemli kavramlarından biri olan yapay zekâ, başlı başına incelenebilecek çok detaylı bir alandır. Başta insan olmak üzere canlı varlıklarda gözlemlenen ve akıllı davranış olarak nitelenen davranışları sergileyebilen sistemler, diğer birçok disiplinde olduğu gibi Doğal Dil İşlemede de çok çeşitli uygulama alanları bulmuştur.

Kronolojik gelişimi açısından bakıldığında yapay zekâ kavramı, 1950’lerde Alan Turing’in “Turing Testi” ve sonrasında John McCarthy tarafından akademik bir konferansla kavram olarak ortaya çıkarılmış, sonrasında çok daha kapsayıcı bir çerçeve

olarak gelişim göstermiş ve bir araştırma alanına evrilmiştir. Yapay Zekâ, 1980’li yıllara kadar kavramsal bir çerçeve olarak varlığını hep güncel tutsa da sürdürülebilir teknolojilerle teknik boyuta ve pratik hayata yansımalarının 1970’li yılların sonuna doğru olduğu söylenebilir. Yapay zekâ kavramı; sonrasında makine öğrenmesi, derin öğrenme vb. gibi alt başlıklarda büyük gelişmeler gösterdiğinden 2000’li yılların başlarından itibaren daha kapsayıcı genel bir kullanımı ile teknik bir kavramdan daha çok üstün teknolojik gelişmeleri ifade eden sosyal bir olguya dönüştüğü söylenebilir. Günümüzde birçok önemli kişi ve kuruluşun konuya yaptığı güncel atıf ve söylemler kavramın önemini artması ve sürekli gündemde olmasını da beraberinde getirmektedir. Bu söylemlere örnek olarak, Rusya Lideri Putin’in “Yapay Zekâ konusunda güçlenenlerin dünyayı yöneteceği” veya ünlü girişimci Elon Musk’ın “3. Dünya Savaşı’nın yapay zekâ nedeniyle çıkacağı” [116] gibi demeçleri vermek mümkündür.

Makine öğrenmesi ise, uzun süre veri madenciliği, istatistiksel vb. çalışmaların gölgesinde kalsa da kullanılan algoritmaların gelişmesi ve ticari rekabet ortamının bu tarz çalışmaların önünü açması neticesinde 1980’lerde ortaya çıkmış, 1990’lı yıllar ve özellikle 2000’li yılların başında oldukça yoğun olarak kullanılmıştır ve halen kullanılmaktadır. Bu bakımdan makine öğrenimi yapay zekânın bir alt kolu ve teknik yansıması olarak görülebilir.

2010’lu yıllarda gelişim gösteren derin öğrenme sistemleri ve algoritmaları ise, oldukça büyük ve çeşitli veri üzerinde çalışan, birden fazla katmansal yapı kullanan ve yüksek hesaplama gücü gerektiren problemlere yoğunlaşan sistemlerdir. Bu özellikleri ile derin öğrenmeyi makine öğrenmesinden keskin çizgilerle ayırmak zordur. Fakat, derin öğrenmenin daha karmaşık problemleri daha güçlü bilgisayarlarla ele alması nedeniyle popülerliği giderek artmaktadır. Şekil 4.1’de sunulan ve yapay zekânın kapsayıcı bir çerçeve olarak tanımlandığı görsel; yapay zekâ, makine öğrenmesi ve derin öğrenme kavramları ve gelişim tarihçesine ait özet bir sunum niteliğindedir.



Şekil 4.1 Yapay zekâ, makine öğrenmesi ve derin öğrenme kavramları

Bu tez kapsamında metin odaklı çalışmamızda makine öğrenmesi yöntemleri ve ayrıca derin öğrenme tekniklerinden biri olan LSTM Ağları kullanılmıştır. Makine öğrenmesi, matematiksel ve istatistiksel yöntemler ile geçmiş bilgi ve veriler üzerinden [117] çıkarımlar yaparak tahminleme yapan sistemlerin bilgisayarlar ile modellenmesidir. Literatürde, makine öğrenmesi için geliştirilmiş birçok algoritma, yöntem ve kütüphane kullanılmaktadır. Makine öğrenme problemleri öğrenme ve geri bildirimlere göre şekillenen üç temel öğrenme yaklaşımı ile ele alınmaktadır. Bu yaklaşımlar aşağıda kısaca tanımlanmıştır:

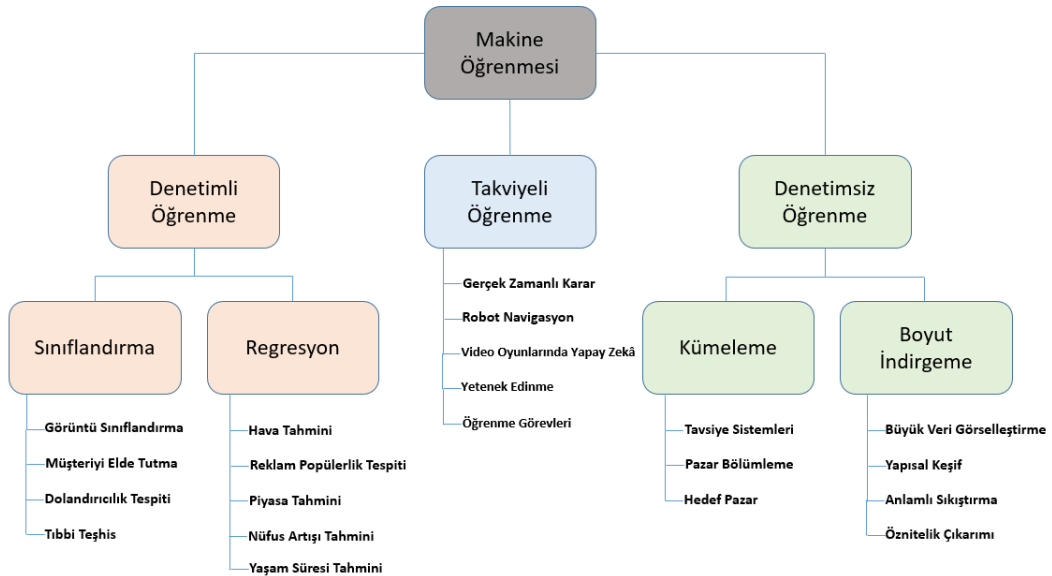
Denetimli Öğrenme (*Supervised Learning*): Bu öğrenme yaklaşımında algoritmalar, öğrendiklerinden yani etiketli (*labelled*) verileri kullanarak tahminleme yapar. Eğitimde kullanılacak veri ve bu veriye ait kategori, etiket vb. sınıf bilgisi önceden bilinmektedir. Denetimli öğrenmede, var olan bilgiyi kullanarak öğrenme yapan sistem, yeni gelen veri/durum için tahminleme yapar. Gözetimli öğrenme diye de adlandırılmakla beraber tez boyunca “denetimli öğrenme” şeklinde kullanılacaktır. Genel olarak; Sınıflandırma (*Classification*) ve Regresyon (*Regression*) olarak 2 temel yöntemde incelenebilir.

Denetimsiz Öğrenme (*Unsupervised Learning*): Denetimli öğrenmedeki gibi önceden etiketlenmiş bir eğitim verisi bulunmaz. Öğrenme mantığı, önceden eğitilmemiş verileri kullanarak, veriler arasında çeşitli ilişkiler/bağıntılar tespit edilmesi ve verilerin yakın

içerik/değer/anlam vb. gruplarda toplanması kümelenmesi şeklinde işlenmesine dayanır. Bu kapsamda yeni gelen veri/durum algoritma tarafından oluşturulan en uygun gruba atanır. Gözetimsiz öğrenme diye de adlandırılmakla beraber tez boyunca “denetimsiz öğrenme” şeklinde kullanılacaktır. Genel olarak; Kümeleme (*Clustering*), Birliktelik Kuralı (*Association Rule Mining*) ve Boyut Azaltma (*Dimensionality Reduction*) olarak üç temel yöntemde incelenebilir.

Takviyeli Öğrenme (*Reinforced Learning*): Bu öğrenme yöntemi diğer iki yönteme göre farklılık arz eder. Bu yöntemde öğrenme mantığı, çevreden alınan geri bildirim (feed-back) mekanizmasına dayanmaktadır. Olası durumlara hedef olarak ulaşıp ulaşılmadığı denetlenir. Sistem kullandığı ödül-ceza yöntemi ile sürekli ve pekiştirmeli bir öğrenme işlevi amaçlar [118].

Bu üç ana yaklaşım dışında, bu yaklaşımlarla veya tek başına kullanılabilen farklı alt yaklaşımlarda bulunmaktadır. Bu tip yaklaşımlara örnek olarak boyut küçültme (*dimension reduction*) ve konu modelleme (*topic modelling*) verilebilir. Bu yaklaşımlar ele aldığı ana problemleri de gösteren açıklayıcı bir şekilde Şekil 4.2’de gösterilmiştir. Elbette ele alınan problemler bu şekilde gösterilenlerle sınırlı değildir.



Şekil 4.2 Makine öğrenmesinde kullanılan yaklaşımlara genel bakış.

Derin öğrenme, özellikle son yıllarda çok popüler hale gelen makine öğrenme yöntemlerindedir. Makine öğrenmesinin bir alt kolu olarak ortaya çıksa da gelişmişlik düzeyi ve hesaplama gücü açısından yapay zekâ teknolojilerinin en etkin lokomotifleri olarak birçok problemin çözümünde kullanılmaktadır. Makine öğrenme yöntemleri, derin öğrenme yöntemleri ile kıyaslandığında genel olarak özneliklerin ortaya çıkarılması hususunda çok daha titiz ve ayrıntılı bir çaba gerektirmektedir. Derin öğrenme teknikleri kullandığı mimari sayesinde makine öğrenmesinde yakalanamayan metinde gizli kalan bazı öznelikleri ortaya çıkarabilir. Bu özellikleriyle derin öğrenme teknikleri günümüzde makine öğrenmesi alanında en sık kullanılan yaklaşımlar olarak belirtilmektedir [119]. Daha çok görsel veri üzerinde çalışılan derin öğrenme son yıllarda bazı çalışmalarda metin verisi üzerinde ve sahte haber tespiti özelinde kullanıldığı görülmektedir. Derin öğrenmede kullanılan bazı önemli teknikler ve bu teknikleri esas alarak sahte haber konusunu ele almış bazı çalışmalar şu şekildedir:

- Evrişimsel Sinir Ağları (*CNN, Convolutional Neural Network*) [120, 121],
- Tekrarlayan Sinir Ağları (*RNN, Recurrent Neural Network*) [122-125],
- Geçitli (Kapılı) Tekrarlayan Birimler (*GRU, Gated Recurrent Unit*) [126],
- Çok Katmanlı Algılayıcı (*MLP, Multi-Layer Perceptron*) [127],
- Uzun Kısa Süreli Bellek Ağları (*LSTM, Long-Short Term Memory*) [128, 129].

Bunların dışında; Bulanık Sinir Ağları (*FNN, Fuzzy Neural Network*), Yoğun Sinir Ağları (*Dense Neural Network*) ve Özyineli Sinir Ağları (*Recursive Neural Network*) gibi tekniklerde sık kullanılan teknikler arasında sayılabilir.

4.3. Kelime Yerleştirme Teknikleri ve Metinlerin Vektörlere Çevrilmesi

Metin sınıflandırma üzerine çalışılan makine öğrenmesi problemlerinde ilk olarak metnin vektörlere çevrilerek temsil edilmesi şarttır. Makine öğrenme algoritmaları doğrudan verilerle çalışamayacağından verilerin vektörel veya sayısal temsilleri önemlidir. Aynı şekilde metinlerin de vektörlerle temsil edilerek, problemlerin matematiksel olarak daha anlamlı hale gelmesi gerekmektedir. Kategorik değişkenlerin temsil edilmesinde kullanılan “one-hot encoding”, “label encoding” gibi teknikler; günümüzde scikit-learn,

keras vb. makine öğrenmesi ve derin öğrenme kütüphaneleri sayesinde kolaylıkla yapılabilmektedir.

Sahte haber probleminde hareketle, haber dokümanları (H) ve haber dokümanlarında geçen farklı kelimeleri (K) içeren $H \times K$ lık bir M matrisinde, $M_{xy}=0$ ifadesi ile x haber dokümanında y kelimesinin bulunmadığı ifade edelim. Bu temel gösterimimizde; kelimelerin sıklıkları, birlikte geçme sıklıkları anlamsal uzaydaki yerleri vb. özelliklere sahip olmadığından sadece bu kodlama ile metinleri sınıflandırmak anlamsızdır. Bu yüzden metin verisi kullanan bu problemi daha anlamlı kılabilmek için yapılacak ilk işlem M_{ij} ifadesinde terimin hangi sıklıkla geçtiğine karşılık gelen terim frekansını ifade eden tf ve terim frekansı-ters doküman frekansını ifade eden $tf-idf$ hesaplamalarının yapılmasıdır. Bu hesaplamalar yapılarak problemin ilk aşama çözülmesine başlanmıştır. Terim frekansı, bir doküman içerisinde geçen terim sıklığını hesaplamak için kullanılır. Eşitlik (11) 'te terim frekansının hesaplaması gösterilmiştir.

$$tf_{ij} = \#(K_j | K_j \in H_i) \quad idf_j = \log\left(\frac{\# H}{\#(H_i | K_j \in H_i)}\right) \quad (11)$$

Burada tf_{ij} , j kelimesinin i haber dokümanında kaç kez geçtiğini göstermektedir. Diğer hesaplamada idf_j , j kelimesini içeren dokümanlar ile tüm dokümanlarının oranının log hesaplamasını ifade eder. Bu iki hesaplamadan Eşitlik (12)'de gösterildiği şekliyle, metin sınıflandırmada sıklıkla kullanılan $tf-idf$ parametresi hesaplanmaktadır.

$$tf - idf_{ij} = tf_{ij} * idf_j \quad (12)$$

Bundan sonraki bölümlerde, öncelikli olarak kullanılan kelime yerleştirme teknikleri, ne şekilde kullanıldıkları anlatılmış ve daha sonra bu tekniklerle birlikte kullanılacak olan özniteliklerin anlatımı yapılmıştır. Çeşitli makine öğrenmesi modelleri ve derin öğrenme teknikleri (LSTM) kullanılarak veriler eğitilmiş ve elde edilen sonuçlar karşılaştırmalı bir şekilde sunulmuştur.

4.3.1. Word2Vec Modeli

Word2Vec modeli, sözcüklerin vektör uzayında ifade edildiği tahmin temelli bir modeldir. Thomas Mikolov ve arkadaşları tarafından geliştirilen modelde [130, 131] birbirlerinden giriş-çıkış verileri açısından farklılaşan, Skip-gram ve CBOW (*Continuous*

Bag of Words) diye adlandırılan iki ana teknik bulunmaktadır. Bu çalışma kapsamında Skip-gram modeli kullanılmıştır.

Bu yöntemde haber metinleri için gereken ön işlemlerden sonra, her haberdeki tüm cümleler için merkez olarak belirlenen sözcük giriş olarak alınıp merkezde olmayan sözcükler çıkış olarak tahmin edilmeye çalışılmaktadır. Bu modelde önemli bir parametre olan pencere-aralığı (*windows-size*) değeri 2 olarak alınmıştır. Bu değer merkez olarak seçilen kelimenin sağ ve solundaki sözcük uzaklığını ifade eder. Yapay sinir ağlarına benzer bir çalışma mantığı olan bu modelde yapay sinir ağlarında kullanılan aktivasyon fonksiyonları kullanılmaz. Bitcoin hakkındaki bir sahte haber örneğinden alınan bir cümlenin Skip-gram tekniği ile analizinin mantıksal modeli Şekil 4.3'te gösterilmiştir.

Haber Metni						Eğitilecek Kelime Çiftleri
Önemli Bilgi: Yarından itibaren tüm dünyada yeni iletişim kuralları yürürlüğe girecek.						
ÖNEMLİ	BİLGİ	YARINDAN	İtibaren tüm dünyada yeni iletişim kuralları yürürlüğe girecek.			[ÖNEMLİ, BİLGİ] [ÖNEMLİ, YARINDAN]
ÖNEMLİ	BİLGİ	YARINDAN	İTİBAREN	tüm dünyada yeni iletişim kuralları yürürlüğe girecek.		[BİLGİ, ÖNEMLİ] [BİLGİ, YARINDAN] [BİLGİ, İTİBAREN]
ÖNEMLİ	BİLGİ	YARINDAN	İTİBAREN	TÜM	dünyada yeni iletişim kuralları yürürlüğe girecek.	[YARINDAN, ÖNEMLİ] [YARINDAN, BİLGİ] [YARINDAN, İTİBAREN] [YARINDAN, TÜM]

Şekil 4.3 Örnek bir cümle üzerinden Skip-gram model mantıksal gösterimi.

Literatürde, metin sınıflandırma problemlerinde Türkçe için TD-IDF ve N-gram temelli çalışmaların, Word2Vec'e göre daha sık tercih edildiği söylenebilir. Bunun bir nedeni bahsedilen yöntemlerle büyük derlemlerde daha iyi sonuçlar verdiğinin gözlemlenmesidir. Bir diğer neden ise, sözcüklerin vektör uzayındaki temsillerinin, sözcüğün derlem açısından önemi istenen ölçüde yansıtması olduğunu değerlendirilmektedir. Bu bakımdan Word2Vec tekniğini daha etkin kullanabilmek adına, sözcük vektörlerini TF-IDF değerlerine göre ağırlıklandırarak güncelledik. Elde edilen sonuçlar yapılan bu güncelleme ile Word2Vec yöntemi kullanılarak Türkçe için de başarılı sonuçlar alınabileceğini göstermiştir.

4.3.2. N-Gram Modeli

N-gram veya bazı kaynaklarda küçük harfle yazımı tercih edilen “n-gram” modeli, belli bir dizide n öğeden oluşan alt diziyi tanımlayan, olasılık ve tahminlemeye dayalı bir dil modelidir. Bu çalışmada verilen bir haber dokümanı için kelime bazlı n-gram modeli kullanılmıştır. Bu modelde “n” kontrol edilen değer sayısına karşılık gelmektedir. En çok kullanılan modeller; unigrams (n=1), bigrams (n=2) and trigrams (n=3) olarak geçmektedir.

Unigrams, n-gram yaklaşımı için en basit modeli sunar. Metinde bulunan tüm bağımsız sözcüklerden oluşur. Bigram modeli, bir çift bitişik kelimeyi tanımlar. Her kelime çifti tek bir bigram oluşturur. Daha büyük gramlar, n bitişik kelime bir araya getirilerek benzer şekilde oluşturulabilir. Daha büyük n-gramlar, kelime konumunun daha iyi anlaşılmasını sağladıkları için bağlamı yakalamada daha etkili olduğu söylenebilir. Aşağıdaki örnekte n-gram model bir örnek ile anlatılmıştır:

Metin: “Gerçek, ayakkabılarını giyene kadar yalan, dünyayı 3 kez dolaşır.”¹⁸

Unigrams: “Gerçek”, “ayakkabılarını”, “giyene”, “kadar”, “yalan”, “dünyayı”, “kez”, “dolaşır”.

Bigrams: “Gerçek Ayakkabılarını”, “ayakkabılarını giyene”, “giyene kadar”, “kadar yalan”, “yalan dünyayı”, “dünyayı kez”, “kez dolaşır”.

Trigrams: “Gerçek ayakkabılarını giyene”, “ayakkabılarını giyene kadar”, “giyene kadar yalan”, “kadar yalan dünyayı”, “yalan dünyayı kez”, “dünyayı kez dolaşır”.

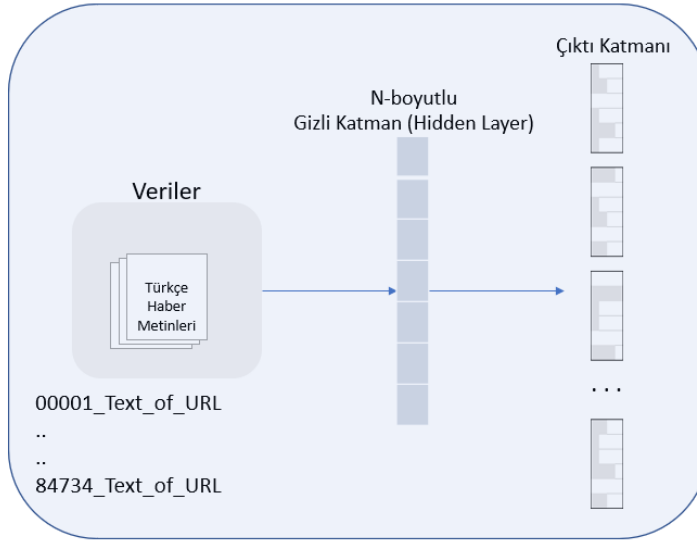
Literatür ve küçük bir veri seti üzerinde prototip bir model kurguladığımız ve motive edici sonuçlar elde ettiğimiz çalışmamızda [27] n-gram modelleri ve kombinasyonları (unigram, bigram, trigram) kullanılmıştır. Benzer mantıkla, büyük derlemlerde iyi sonuçlar vereceği değerlendirilerek bu çalışmamızda uygulanmıştır. Bu doğrultuda haber metinleri bu tekniğin ve seçilen özniteliklerle birlikte kullanılarak eğitilmişlerdir.

¹⁸ Birçok kaynakta ABD’li yazar Mark Twain’e (1835-1910) ait olduğu belirtilen bu söz hakkında birçok kaynakta da yazarın eserlerinde rastlanılmadığı ve kendisine ait olmadığı iddia edilmektedir. Sözün anlamı düşünüldüğünde bu belirsiz durum oldukça ironiktir.

4.3.3. Kelime Çantası Modeli (Bag of Words) Modeli

Kelime çantası modelinde haber metni çoklu sözcük kümeleri olarak tutulurken (çoksallık), gramer ve kelime sırası göz ardı edilir. Şekil 4.4'te örneğimize uyarlanmış temsil mantığı gösterilen BOW modeli sıra içermeyen bir metin temsildir.

Çalışmamızda, farklı bir bakış açısıyla sözcüklerden ziyade bu teknikle birlikte sözcük türlerinin kullanılabileceği gözlemlenmiştir. Sözcük türleri (*POS, Part of Speech*), çeşitli çalışmalarda Türkçe için metin sınıflandırmada yararlanılan ve metin hakkında özellikle retorik açıdan ipuçları verebilecek bir dilbilim kavramıdır. Bu problem özelinde cümle öğelerinin haber metinlerinde çok değişken bir dağılımda olduğu ve çok etkili bir öznitelik olmadığı gözlemlenmiştir. Bu nedenle, Robinson ve arkadaşlarının da tweet metinleri açısından sözcük türlerinin ne kadar bilgi taşıdığını anlamak için yaptıkları çalışmalarında kullandığı şekilde “Bag of POS” modeli şeklinde kullanılmıştır [132]. Bu yöntemin Bölüm 2.2.3'te belirtilen Türkçedeki biçimbilimsel (*morphological*) belirsizlik konusundaki çözümlerle birlikte düşünülmesi gerektiğini belirtmekte fayda vardır.



Şekil 4.4 Kelime çantası modeli

4.4. Metodoloji

Bu bölümde sahte haber tespiti için Bölüm 4.3'te belirtilen teknikler; kategorik veriler, geliştirilen sözlük ve seçilen öznitelikler birlikte kullanılarak çeşitli modellerle eğitilmiş ve elde edilen sonuçlar sunulmuştur. Bu kapsamda geliştirilen model Tez Blok

Diyagramında sunulmuştur. Elde edilen sisteme ait çıktılar sonuç bölümünde değerlendirilmiştir.

4.4.1. Öznitelik Seçimi ve Çıkarımı

Öznitelik seçimi ve çıkarımı, makine öğrenme yöntemlerini kullanabilmek için veriye ait en uygun özniteliklerin belirlenmesi sürecidir. Bu birçok gerçek dünya probleminde önceden hazır olmayan ve hassas analiz gerektiren bir süreçtir. Veri madenciliği çalışmalarında olduğu gibi ön işleme ve veriyi filtreleme/temizleme işlemlerinden sonra uygulanır. Bu problemde belirgin öznitelikleri, verilerin bir tabular forma (tablo/matris olarak) dönüştürüldüğünde her satırı tanımlayan kolonlar olarak ifade edilebiliriz. Kullanılacak öznitelikler ve alt öznitelikler bu kolonlardan en uygunlarının seçilmesi ve öznitelikler arasındaki bağıntılarla ve/veya bu bağıntılardan yola çıkarak yapılacak hesaplamalarla yeni ortaya çıkabilecek özniteliklerin kullanılması şeklinde olabilir.

4.4.1.1. Öznitelik Seçimi

Veri içerisindeki sistem tahminine etki etmeyen ve göreceli olarak daha az etkili olan bazı özniteliklerin ayıklanması işlemine öznitelik seçimi (*feature selection*) denilir. Bu işlem, sisteme daha fazla hesaplama yükü getirilmemesi ve sistem başarımını arttırmak için gereklidir. Öznitelik seçimi, aşırı öğrenmenin olmasını da engeller.

TR_FN veri seti de işlendikten ve tabular forma dönüştürüldükten sonra çok boyutlu bir veriye dönüştüğü gözlemlenmiştir. Diğer kullanılan tekniklerden de elde edilen öznitelikler hesaba katıldığında, nitelik sayısının artması makine öğrenmesi yöntemlerinin maliyetini arttırmıştır. Bu bakımdan boyut indirgemeyi sağlayan tekniklere başvurulmuştur. Buradaki temel hedefimiz, elimizdeki verinin içindeki gereksiz ve sonuca göreceli etkisi en az olan niteliklerin ayıklanması olmuştur. Bunun için Temel Bileşenler Analizi (*PCA, Principal Component Analysis*) tekniği kullanılmıştır. Bu analiz temel olarak, çok boyutlu verilerde en yüksek varyansın bulunmasını ve verideki tüm bilgiyi en yüksek oranda korunarak daha küçük boyutlara indirgenmesini sağlar.

İki değişkenin birlikte ne kadar değiştiği diğer bir ifade ile korelasyon (kovaryans) matrisinin özvektörleri ve özdeğerleri, bu tekniğin temelini oluşturur. Verinin özellik vektörlerini temsil eden matrisin, özvektör¹⁹ ve özdeğer çarpımı olduğu varsayımına dayanır.

4.4.1.2. Öznitelik Çıkarımı

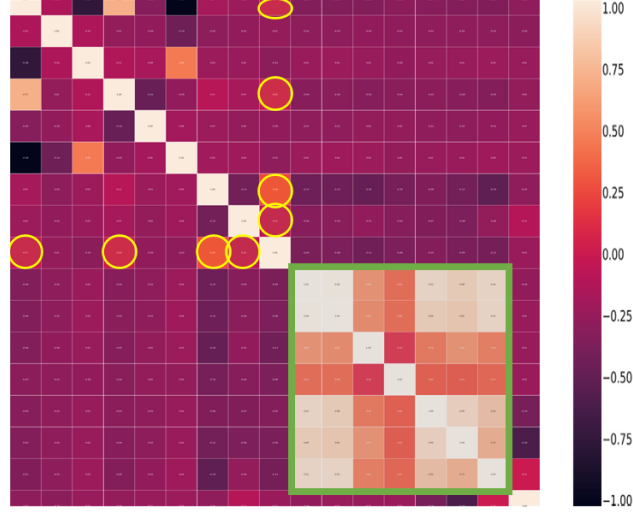
Daha etkin özniteliklerin bulunmasına yönelik işlemler sadece öznitelik seçimi ile sınırlı değildir. Öznitelik çıkarımı (*feature extraction*) işlemi de eski özniteliklerden yeni öznitelikler ortaya çıkararak öznitelik sayısı artırılıp/azaltılabilmektedir. TR_FN veri setimizi ön işlemlerden sonrası tabular bir biçime dönüştürdüğümüzde birçok öznitelik oluşmuştur. Bu doğrultuda, bu özniteliklerin hangi ölçü ve derecede sonuca etki ettiklerini görselleştirmek yapılacak işlemleri görebilmek açısından önemlidir. Bu kapsamda verinin tabular haline dönüştürüldükten sonraki 17 kolondan oluşan ham halinde kolonların ilişkisini göstermek açısından elde edilen korelasyon matrisi²⁰ (*correlation matrix*) matrisi Şekil 4.5'te gösterilmiştir.

Bu matristen işlenmemiş kolon verilerinin sınıfın belirlenmesine büyük etkisi olmadığı görülmektedir. Grafik üzerindeki açık renkli bölgeler güçlü pozitif korelasyonu göstermektedir. Sınıf değişkeni ile arasında çok fazla olmasa da pozitif korelasyon içeren kesişim noktaları sarı çember ile işaretlenmiştir. Şekilde, değişkenlerin (kolonlar) birbirine bağlılık derecesini daha net bir şekilde görüldüğü alan yeşil kare ile işaretlenmiştir. Bu kolonlar incelendiğinde durak kelime (*stop words*), toplam karakter sayısı vb. sayısal toplamların olduğu birbiri ile doğru orantılı olan kolonlar olduğu görülmektedir. Ham veriden çıkarım yapmanın zor olacağını betimleyen bu görselden, problem ve veri özelinde öznitelik çıkarımının itina ile yapılmasının zaruri olduğu anlaşılmaktadır.

¹⁹ Bir vektör/yöney (*vector*) dönüşüme uğradığında boyutundan bağımsız olarak yönünün aynı kaldığı yöneye özvektör (*eigen vector*), bu yön değiştirmeyen uzunluk/genlik değiştiren vektöre ait bu değişim hesabına/miktarına özdeğer (*eigen value*) ismi verilir.

²⁰ Matrisinde elde edilmesinde kullanılan kod parçacığı:

```
plt.figure(figsize=(50,50))
sns.heatmap(data=data.corr(), annot=True, linewidths=.25, fmt="1.2f")
plt.show()
```



Şekil 4.5 Ham veriye ait kolonların korelasyon matrisi

Gerek görselleştirme kütüphaneleri gerekse özyinelemeli (*recursive*) özellik eliminasyonu kullanılarak özniteliklerin göreceli performansı hesaplanarak en uygun öznitelikler seçilmeye çalışılmıştır. Ayrıca, derin öğrenme yöntemlerinde kullanılan özelliklerden hariç olarak, geliştirilen sözlüklerden dört ve diğer özniteliklerin kullanılarak elde edilmesi ile iki olmak üzere 6 adet yeni öznitelik çıkarımı yapılmıştır. Çıkarımı yapılan iki öznitelik aşağıda tanımlanmıştır. Mevcut çalışmada önerilen ve kullanılan öznitelikler, ilgili gruplar ve tanımları Çizelge 4.1’de verilmiştir. Bazı öznitelikler/kolonlar ön deneylerde aşırı öğrenmeye neden olduğu ve/veya göreceli olarak performansa olan katkılarının önemsiz olması nedeniyle çıkarılmıştır.

Ayrıca Kazanım Oranı (*Gain Ratio*), Ki Kare (*Chi-Square*), Bilgi Kazancı (*Information Gain*), Gini Katsayısı (*Gini Index*) vb. öznitelik seçim yöntemlerinin mevcut kütüphanelerle uygulanması artık nispeten basit kodlamalarla problemlere uygulanabilmektedir. Bu yöntemlerin detaylı anlatımı Kaynar ve arkadaşlarının çalışmasında bulunabilir [133].

Çizelge 4.1 Önerilen/kullanılan öznitelikler, öznitelik grupları ve açıklamaları (Tr_{All}).

Id	Öznitelik	Öznitelik Kümesi	Açıklama	TR _{Opt}
L ₁	Haber Raw skor	Sözlük Temelli	Bölüm 3.3.1’de anlatılan yöntemlerle 4 farklı sözlüğe göre haberlerin puanlanmasında elde edilen değer (haber sahte/gerçek) polaritesi.	+
L ₂	Haber Root skor			+
L ₃	Haber Raw+Pos			
L ₄	Haber Suffix			
K ₁	Haber Kategorisi	Kategorik Haber Etiketleri	Haber kaynağındaki etiketlemelere göre yapılan kümeleme (10 ana kategori)	
K ₂	Haber Alt Kategorisi		Haber kaynağındaki elde edilen ana kategorilerin sınıflandırılması ile elde edilen 55 alt kategori	
K ₃	URL veya Kaynak		Haberin tanımlayıcı adresi veya kaynağı	-
K ₄	Ulaşılabilir Link		Haber içerisinde ulaşılabilir link var, yok şeklinde.	
E ₁	Dil karmaşıklığı (zenginliği) skoru	Öznitelik Çıkarımı	Bölüm 4.4.1.2’de anlatılan bu ölçüt, kullanılan dilin karmaşıklığı bir diğer deyişle dil zenginliği-basitliğini değerlendiren hesaplama	+
E ₂	Yazım hatası skoru	Öznitelik Çıkarımı	Bölüm 4.4.1.2’de anlatılan bu ölçüt, haber metninde yapılan yazım hataları için kullanılan doğrusal hesaplama	+
S _{1,11}	Noktalama İşaretleri	Sayı Temelli	Üç nokta*, Ünlem*, Soru İşareti, Tırnak İşareti, Virgül, Nokta, Noktalı Virgül, İki Nokta, Kısa Çizgi, Ayraç ve Köşeli Parantez sayıları	(*) işaretli olanlar
S _{12,16}	Sözcük Türü	Sayı Temelli	İsim, Sıfat, Fiil, Zarf, İşlev Sözcükleri* (Zamir, Edat, Bağlaç) sayıları	(*) işareti
S _{17,22}	Metin Karakteristikleri	Sayı Temelli	Harf, kelime, cümle, kısaltma, paragraf sayıları, haber başlığı kelime sayısı	

Yukarıda belirtilen özelliklerden en uygun öznitelik setine giren ve yeni geliştirdiğimiz iki yeni hesaplamanın literatürde diğer metin sınıflandırma problemlerinde de kullanılabileceği değerlendirilmektedir. Bu iki öznitelik tek başlarına kullandıklarında dahi önemli ipuçları vererek metin karakteristiği hakkında bilgi vermektedir. Bu öznitelikler şu şekildedir:

Dil Karmaşıklığı/Zenginliği (*Lexical richness*) : Dil Karmaşıklığı puanı temelde metnin okunabilirliğini temsil etmektedir. Literatürde Türkçe için iyi bir okunabilirlik ölçütü belirlemeye yönelik çalışmalar mevcuttur [134]. Bu çalışmalar incelendiğinde, metriklerin uzun metinlere göre kurgulandığı görülmektedir. Bu bakımdan nispeten roman ve makalelere göre kısa olan haber metinlerinde sağlıklı bir uygulaması

gerçekleştirilememiştir. Bu noktada farklı yöntemlere başvurulmuştur. Bazı denemeler sonucu, Gunnig-Fog metriğini [135] kelime uzunluğu ve cümle uzunlukları açısından yenileyerek Türkçe için uyarlanmış yeni bir versiyonunu geliştirdik ve performansa olan olumlu katkısını gözlemledik. Bu güncelleme metriğin orijinal halinde kullanılan hece sayısı yerine bir sözcüğü bütün bir parça olarak ele alan ve hece sayısı yerine kelime uzunluğunu kullanmayı içermektedir. Bunu uygularken izlenen mantık, her ne kadar Türkçe’de sonsuz sayıda sözcük türetilbilse de günlük konuşma dili ve formal yazım dilindeki kullanımlar arasındaki uzun kelime seçme/seçmeme farklılığını istatistiksel olarak yakalamamızdan ileri gelmektedir. Bu güncellenmenin yanı sıra değerleri normalize etmek adına formülde bazı değişiklikler yapılmıştır. Ateşman, Türkçe’de okunabilirlik üzerine yaptığı çalışmada ortalama hece sayısı 2,6 olarak vermiştir [136]. Bizde bu formülde bir kelimenin uzun bir kelime sayılabilmesi için sezgisel olarak 9 harften fazla harfe sahip olmasını formülümüze ön şart olarak belirledik. Buna göre bir D haber dokümanı için, karmaşıklık hesaplaması (*CS, Complexity Score*) Eşitlik (13)’te gösterilmiştir.

$$D_{CS} = \left(\frac{\frac{\#W_d}{\#S_d} + \frac{\#LW_d * 100}{\#W_d}}{10} \right) \quad (13)$$

$\#W_d$: Dokümandaki kelime sayısı

$\#S_d$: Dokümandaki cümle sayısı

$\#LW_d$: Dokümandaki uzun kelime sayısı

Böylelikle tek başlarına performansı çok etkilemeyen bazı sayı temelli özelliklerin birlikte kullanılarak yararlı bir metriğe ve özneliğine dönüştüğü söylenebilir.

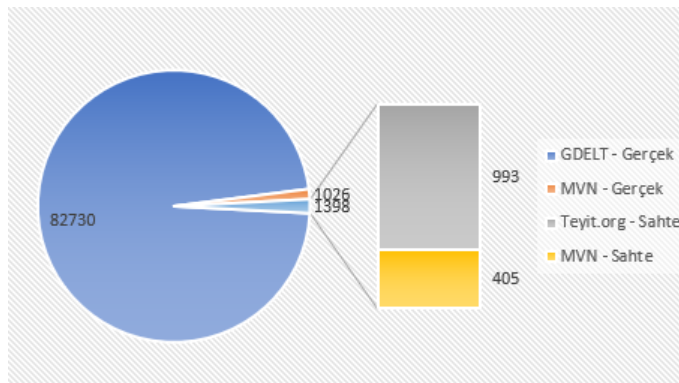
Yazım Hatası Skoru (*Spell Error score*) : Benzer şekilde sayma temelli hesaplamaların bir özneliğe dönüştürülerek sistemin faydalanmasını sağladığımız bir diğer hesaplama da “Yazım Hatası Skorudur”. Önceki bölümlerde sahte ve özensiz hazırlanmış haberlerde yazım hatalarının daha çok göze çarptığı anlatılmıştı. Zemberek kütüphanesinin “Turkish SpellChecker” sınıfı kullanılarak olası yazım hataları hesaplanmıştır. Elbette manuel kontrollü veya kural tabanlı bir yaklaşımla çok daha kesin sonuçlara erişilebilir fakat sistemin olabildiğince otomatize olarak kalmasını sağlamak açısından kural tabanlı veya

farklı şekilde bir müdahalede bulunulmamıştır. Eşitlikte (14)'te gösterilen hesaplama Bölüm 3.3.2'de de kullanılan oldukça basit bir doğrusal normalizasyon hesaplamasıdır.

$$D_{SE} = \left(\frac{\#SE_d}{\#S_d} \right) \quad (14)$$

$\#SE_d$: Dokümandaki hatalı yazım sayısı

Çizelgede gösterilen öznitelikler dışında Word2Vec, kelime çantası Bow_{All} (bag-of-words ve bag of POS olarak iki ayrı biçimde), N-gram teknikleri kullanılarak ortaya çıkarılan öznitelikler birlikte kullanılmıştır. Bölüm 4.8'de ağırlıklı olarak TR_{Opt} olarak adlandırılan optimize edilmiş öznitelikler ve tekniklerin birlikte kullanımından ortaya çıkan en iyi sonuçlar ortaya konmuştur. Ayrıca, TR_{FN} veri setinin alt kümelerinde TR_{Opt} dışındaki özniteliklerle yapılan bazı ilginç deney sonuçları yine aynı bölümde sunulmuştur. Bu özniteliklerden haber verisinin kaynağını belirleyen (URL ve Kaynak) kategorik haber etiketi ön işleme aşamasında ayrıştırılarak (*html parsing*) kategorik olarak elde edilmiştir (<https://www.hurriyet.com.tr/> gibi). Fakat Çizelge 4.1'de gösterilmesine rağmen ana kaynaklardaki dengesiz dağılım göz önüne alındığında bu etiketin kullanılması aşırı öğrenmeye (*overfitting*) neden olacağından kullanılmamıştır. Bu iki kolonun Şekil 4.5'te hedef sınıf değişkeni ile (0.25 – 0.5) aralığında pozitif korelasyona sahip olduğunu belirtmek gerekir. Bu doğrultuda, makine öğrenmesi ve derin öğrenme modellerimiz kaynak bilgisinden haberi olmadan eğitim ve teste (*train-test*) tabi tutulmuştur. Ana veri kaynaklarının dağılımı Şekil 4.6'da gösterilmiştir.



Şekil 4.6 TR_{FN} veri setindeki ana kaynak dağılımı.

4.4.2. SMOTE Tekniđi

TR_FN veri setinde hedef sınıf, sahte-gerçek olarak iki etiketleme yapılan haber metnine deđer atayan sınıftır. TR_FN veri setinin dengesiz hedef sınıf dađılımına sahip bir veri seti olduđuna Bölüm 3.2.4'te deđinilmiřti. Bu bakımdan dengesiz veri setlerinde veriler direkt kullanıldıđında performans deđerlerinde ıkan tutarsız sonuçların bizim veri setimizde de ortaya ıkması olasıdır.

Dengesiz veri problemini özömlmek için veri setindeki örnekleme yapısını düzenleyen (undersampling, oversampling, SMOTE vb.) teknikler kullanılmaktadır [137]. Örneđin TR_FN veri setinde ham veriler üzerinden işlem yapıldıđında, gerçek haberlerin neredeyse tamamını dođru tahmin ederken, sahte haberlerin ok büyük bir kısmını yanlış tahmin etmiřtir. Bu sorun, SMOTE (*Synthetic Minority Over-sampling Technique*) tekniđi [138] kullanılarak ařılmıřtır.

SMOTE tekniđinde azınlık sınıfı örneklerinin “k” komřularına bakılarak sentetik örnekler üretilmektedir.

4.4.3. apraz Dođrulama/Geerleme (*Cross Validation*)

Elde ettiđimiz sistemin verilerin ođunu yanlış tahmin etmesine “eksik öđrenme” (*underfitting*) veya sistemin aşırı eđitilmesi sonucu esnekliđinin kaybedilmesine “aşırı öđrenme” (*overfitting*) denir. Bu gibi sonuçların ortaya ıkmasından kaçınmak ve tutarlı modeller kurgulamak için apraz dođrulama kullanılır. Kurguladıđımız veri setinin aslında bađımsız bir veri üzerinde nasıl alıřacađının denenmesi anlamına gelir. Eđittiđimiz modelin yeni veriler için genelleřtirilebilir olup olmadıđını, aşırı öđrenme, eksik öđrenme ve varyans-yanlılık (*variance-bias*)²¹ eliřkisi gibi problemleri tespit etmeye yarar [139].

Veri kümemiz rastgele 10 (k) gruba bölünerek makine öđrenmesi modelleri için 10-kat apraz dođrulama (10-fold cross validation) yapılmıřtır. řekil 4.7'de gösterildiđi gibi,

²¹ Varyans: Tahmin edilen deđerlerin gerçek deđerden ne kadar dađınık olduđunu ifade eder.
Yanlılık: Tahmin edilen deđerlerin gerçek deđerlerden ne kadar uzak olduđunu ifade eder.

gruplardan biri test, diğeri eğitim seti olarak kullanılmıştır. 10 grup bu şekilde tekrarlanarak model eğitmiştir.

	Bölüm 1	Bölüm 2	Bölüm 3	Bölüm 4	Bölüm 5	Bölüm 6	Bölüm 7	Bölüm 8	Bölüm 9	Bölüm 10
Aşama 1	Test	Eğitim	Eğitim	Eğitim	Eğitim	Eğitim	Eğitim	Eğitim	Eğitim	Eğitim
Aşama 2	Eğitim	Test	Eğitim	Eğitim	Eğitim	Eğitim	Eğitim	Eğitim	Eğitim	Eğitim
Aşama 3	Eğitim	Eğitim	Test	Eğitim	Eğitim	Eğitim	Eğitim	Eğitim	Eğitim	Eğitim
Aşama 4	Eğitim	Eğitim	Eğitim	Test	Eğitim	Eğitim	Eğitim	Eğitim	Eğitim	Eğitim
Aşama 5	Eğitim	Eğitim	Eğitim	Eğitim	Test	Eğitim	Eğitim	Eğitim	Eğitim	Eğitim
Aşama 6	Eğitim	Eğitim	Eğitim	Eğitim	Eğitim	Test	Eğitim	Eğitim	Eğitim	Eğitim
Aşama 7	Eğitim	Eğitim	Eğitim	Eğitim	Eğitim	Eğitim	Test	Eğitim	Eğitim	Eğitim
Aşama 8	Eğitim	Eğitim	Eğitim	Eğitim	Eğitim	Eğitim	Eğitim	Test	Eğitim	Eğitim
Aşama 9	Eğitim	Eğitim	Eğitim	Eğitim	Eğitim	Eğitim	Eğitim	Eğitim	Test	Eğitim
Aşama 10	Eğitim	Eğitim	Eğitim	Eğitim	Eğitim	Eğitim	Eğitim	Eğitim	Eğitim	Test

Şekil 4.7 Çapraz doğrulama uygulama adımları

4.5. Sınıflandırma Modelleri

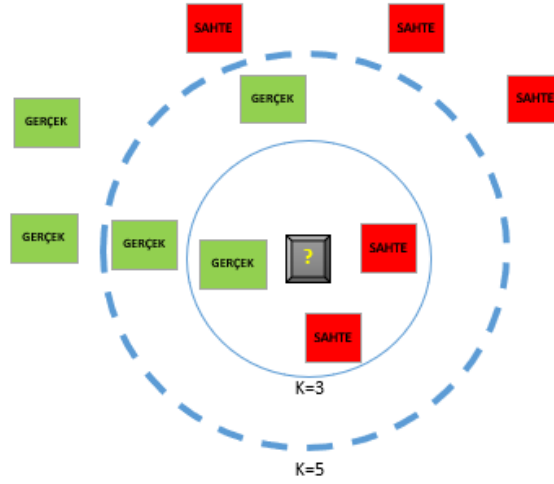
Sınıflandırma, veri kümesi içerisinde veriyi çeşitli sınıflara atama/dağıtma işlemidir. İkili bir sınıflandırma problemi olarak ele aldığımız “Sahte Haber Tespiti” problemi, sınıflandırılmak üzere sisteme verilen yeni bir haberin “Sahte” veya “Gerçek” olarak tahmin edilmesidir.

Sınıflandırma algoritmaları, elimizde mevcut olan veri kümesinden sınıf dağılımını öğrenirler. Regresyon ve sınıflandırma algoritmaları kullandıkları hesaplamalar ve teknik yöntemlerle test verisi için sınıf tahmini/etiketleme yapan modeli oluştururlar.

Modellerin daha iyi anlaşılması için, temel bazı kavramlar bu bölümde verilmiştir. Bu doğrultuda hemen sonraki bölümlerde tez kapsamında kullanılan sınıflandırma modelleri, temel çalışma prensipleri ve kullandıkları yöntemler ilerleyen bölümlerde anlatılmıştır.

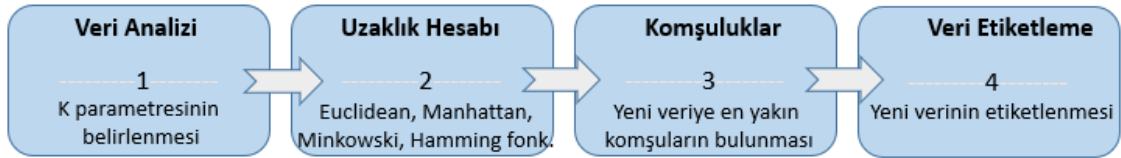
4.5.1. En Yakın k-Komşuluk (KNN)

En Yakın k-Komşuluk (*KNN*, *k-Nearest Neighbour*) algoritması sınıflandırılacak yeni verinin, var olan veri içerisinde en yakın komşu olduğu k komşu/komşuların sınıfına atandığı bir algoritmadır [140]. Genellikle sınıflandırma problemlerinde kullanılan bu denetimli öğrenme algoritması regresyon problemleri için de kullanılmaktadır.



Şekil 4.8 KNN algoritması çalışma modeli

KNN algoritmasında esas “k” değerini belirlemektir. Şekil 4.8’de²² [141], verilen gösterim üzerinden k=3 ve k=5 iken çizilen komşuluk sınırları dikkate alındığında sınıflandırma algoritması tarafından “?” işareti ile gösterilen haberi ilk durumda “Sahte”, ikinci durumda ise “Gerçek” olarak tanımlanacaktır.



Şekil 4.9 KNN algoritması adımları

Şekilde 4.9’da KNN algoritmasının adımları gösterilmiştir. Buna göre öncelikle k parametresi belirlenir. İlgili uzaklık fonksiyonları yardımı ile sınıflandırılacak yeni verinin, mevcut verilere olan uzaklıkları hesaplanır. Bunun için temel olarak “Euclidean”, “Manhattan” ve “Minkowski” olarak adlandırılan üç uzaklık hesaplaması vardır. Bu çalışmada “Minkowski” uzaklığı kullanılmıştır. Bu uzaklıklardan en yakın k komşu ele alınır. Seçilen sınıf, yeni verinin sınıfı olur.

²² Şekil, Hatko’nun tez çalışmasında kullandığı şekilden yola çıkılarak tasarlanmıştır.

4.5.2. Multinomial (Çok Terimli) Naive Bayes (MNB)

Olasılık kuramına ait Bayes teoremine dayanan bu model bir çeşit koşullu olasılık hesaplama algoritmasıdır. Bu algoritmada, hedef sınıfın gerçekleşme ihtimali incelenir ve temelde Eşitlik (15)'te belirtilen Bayes formülünü kullanılır. Tüm sınıfların istatistiksel olarak bağımsız olduğu modelde her sınıfın hedefe katkısını eşit bir şekilde ele alınır.

$$p(C | f) = \frac{p(C) p(f_1, \dots, f_n | C)}{p(f_1, \dots, f_n)} \quad (15)$$

f : Tahmin eden sınıf (öznitelik, özellik)

C : Tahmin edilen sınıf

$p(C | f)$: f olayı gerçekleştiğinde C olayının gerçekleşme olasılığı

$p(f | C)$: C olayı gerçekleştiğinde f olayının gerçekleşme olasılığı

$p(f)$: f olayının gerçekleşme olasılığı

$p(C)$: C olayının gerçekleşme olasılığı

Sınıflandıracağımız veri ikili bir sınıflandırma içeriyorsa (var-yok, evet-hayır) Naive Bayes'in Bernoulli sınıflandırıcısı daha sık kullanılmaktadır. Esasında bu sınıflandırıcı doküman sınıflandırmasında daha çok kullanılan Multinomial Naive Bayes sınıflandırıcıya çok benzemektedir. Tahmin edilecek veri sürekli bir veri ise Gaussian, çoklu kategorilerin içinde bir sınıflandırma yapılacaksa Multinomial Naive Bayes kullanılır. Bu çalışmada Multinomial Naive Bayes sınıflandırıcısı kullanılmıştır.

Değişkenler arası ilişkilerin modellenemediği bu sınıflandırıcı da karşılaşılabilecek bir diğer problem de Sıfır Frekans Problemidir. Eğitim kümesinde gözlemlenemeyen bir değeri diğer bir deyişle olasılık test setinde karşımıza çıkınca olasılık değeri olarak 0 verir yani tahmin yapılamaz. Bu durum genellikle Sıfır Frekans Problemi (*Zero Frequency*) olarak adlandırılır. Bu durumu düzeltmek için genellikle Laplace tahmini kullanılmaktadır. Bu bilgidен hareketle Bölüm 3'teki modelimizde de derlemde bulunmayan terimler için benzer bir durum ile karşılaşıldığını hatırlayalım.

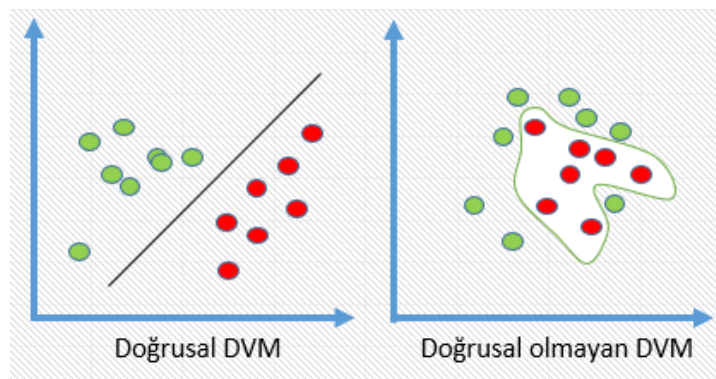
4.5.3. Destek Vektör Makineleri (DVM)

Destek Vektör Makineleri (DVM) (*SVM, Support Vector Machine*), fikri temelleri 1960'lı yıllarda dayanan fakat uygulamalı olarak Vapnik ve arkadaşları tarafından 1990 yıllarda geliştirilen sağlam bir istatistiki teorik temele sahip bir öğrenme algoritmasıdır [142].

Yapısal risk minimizasyon prensibi/teorisine [143] ve dolayısıyla istatistiksel öğrenme teorisine dayanan bir denetimli öğrenme algoritması olan DVM, model karşılaştırmalı çalışmalarda sıklıkla kullanılmaktadır.

Destek Vektör Makineleri, sınıflandırma problemleri için önerilen ve verileri sınıflarına uygun bir şekilde birbirinden ayırmak için hiper düzlem/karar sınırları kullanan, uygulaması nispeten kolay olan denetimli bir öğrenme algoritmasıdır. Kullanılan veri setine göre Doğrusal ve Doğrusal Olmayan Destek Vektör Makineleri olarak ikiye ayırmak mümkündür.

Doğrusal Destek vektör makinelerinde, sınır düzlemlerine eşit uzaklıkta bir hiper düzlem kullanılır. Teorik olarak verileri birbirinden ayırmak için sonsuz sayıda doğru belirlenebilir DVM'in amacı temelde bu doğrular içerisinde en uygununu (optimal) belirlemektir. Doğrusal olmayanlarda çekirdek numarası (*kernel trick*) diğer bir deyişle en uygun karar fonksiyonunu bulmak için farklı çekirdek fonksiyonlar (*kernel functions*) kullanılır. Sigmoid, Polynomial, Gaussian Radyal (*Radial*) tabanlı kerneller en bilinen çekirdek yöntemlerindedir.



Şekil 4.10 DVM için temsili problemler (Doğrusal ve doğrusal olmayan)

İki sınıflı işlemlerde, Şekil 4.10'da doğrusal olarak ayrılan DVM çiziminde, verileri birbirinden ayırmak için çok sayıda hiper düzlem çizilebilir. İşte burada DVM'in amacı kendisine en yakın noktalar arası uzaklığı en yükseğe çıkarabilecek hiper düzlemi bulmaktır. Bu hesaplama en basit hali ile, k sayıda örnek içeren bir eğitim verisinde (x_i, y_i) , $i=1, \dots, k$ varsayıldığında, en uygun hiper düzlem için gerekli eşitsizlik gösterimi şeklinde Eşitlik (16)'da ifade edilebilir.

$$L_D = \begin{cases} w \cdot x_i + b \geq 1 & y = +1 \text{ için} \\ w \cdot x_i + b \leq -1 & y = -1 \text{ için} \end{cases} \quad (16)$$

$x \in \mathbb{R}^N$ olup N-boyutlu bir uzayı,

$y \in (-1, +1)$ ise sınıf etiketlerini,

w ağırlık vektörünü (hiper düzlemin normali) ve b eğilim değerini göstermektedir.

4.5.4. Lojistik Regresyon (LR)

Denetimli öğrenme algoritmalarından biri olan Lojistik Regresyon (*Logistic Regression*) genellikle regresyon ve sınıflandırma problemlerinin çözümünde kullanılır. İsmi regresyon olarak geçse de bir sınıflandırma söz konusudur ve bağımlı değişkenin (sonuç) kategorik olduğu doğrusal sınıflandırma problemlerinde kullanılır. Esas amaç, kategorik bağımlı değişkeninin sınıflandırılmasıdır. İstatistiksel bir bakış açısı ile bir diğer temel amacının bağımlı ve bağımsız değişkenlerin arasındaki ilişkileri ele almak olduğu söylenebilir [144].

Doğrusal regresyon modelinde, bağımlı değişken sürekli kabul edilirken, lojistik regresyonda kategorik, yani ayrıktır. Doğrusal regresyon EKK (En Küçük Kareler) yöntemini kullanırken, lojistik regresyon parametreler Maksimum Olabilirlik Çıkarımı (*MLE, Maximum Likelihood Estimation*) yöntemiyle hesaplanır [145]. Bu bakımdan, lojistik regresyonun matematik temelini olasılık, odds oranı (*OR, odds ratio*) ve odds'un logaritmasına dayandığını söylemek mümkündür. Lojistik regresyonda odds; "bir olayın olma olasılığının, o olayın olmama olasılığına bölümü olarak" tanımlanmaktadır [146]. Eşitlik (17)'de odds oranı hesaplaması gösterilmektedir.

$$odds = \frac{p(x)}{1 - p(x)} \quad (17)$$

$p(x)$: Bir x olayının gerçekleşme olasılığı / bir karakteristik özelliğin var olma olasılığıdır.

Odds oranı genellikle $-p$ veya $\text{Exp}(\beta)$ (*Exponentiated logistic coefficients*) ile gösterilir [144]. Bir diğer önemli kavram olan logit ise odds oranının doğal logaritmasıdır. Eşitlik (18)'de verilen denklem $p(x)$ 'in lojistik dönüşümünü ifade etmektedir.

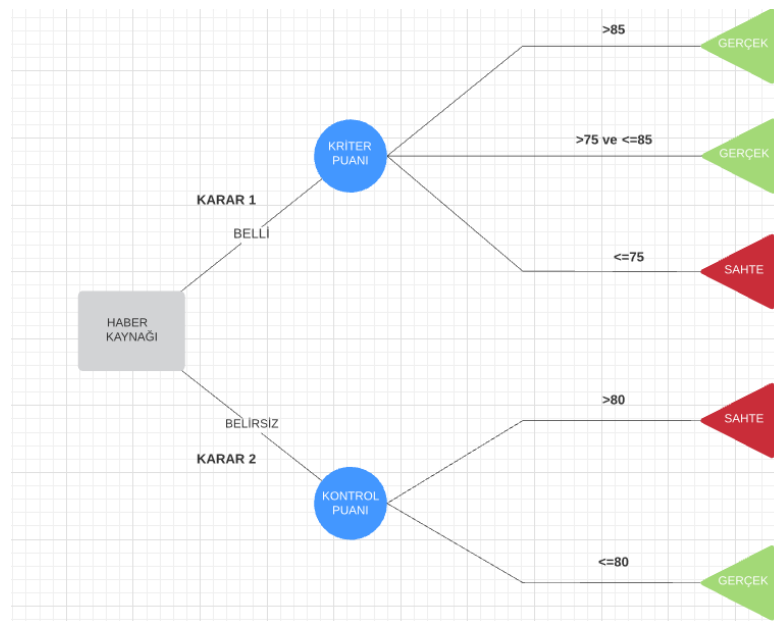
$$\text{logit}(p(x)) = \log\left(\frac{p(x)}{1-p(x)}\right) \quad (18)$$

Logit modelin verildiği Eşitlik (19)'da x 'ler bağımsız değişkenleri göstermektedir.

$$\text{logit}(p) = b_0 + b_1X_1 + b_2X_2 + b_3X_3 + \dots + b_kX_k = \sum_{k=0}^p b_kX_{ik} \quad (19)$$

4.5.5. Karar Ağaçları

Karar ağaçları ağaç tabanlı öğrenmeye dayalı sıklıkla veri madenciliği çalışmalarında tercih edilen, regresyon ve sınıflandırma problemlerinin çözümünde kullanılan denetimli öğrenme algoritmalarındandır. Yapıları itibariyle tepeden aşağıya (*top-down*) bir hiyerarşik yaklaşım içerirler [147]. Genel olarak iç karar düğümleri (*interior*), dallar (*leaf*) ve uç yapraklarla (kategorik sınıflar) temsil edilir. Şekil 4.11'de jenerik bir sahte haber problemi bir karar ağacı ile temsil edilmiştir.



Şekil 4.11 Karar Ağacı Modeli.

Bir karar ağacında temel bazı kavramları şu şekilde tanımlayabiliriz. Her iç düğüm öznitelik üzerinde yapılan bir “testi” temsil etmektedir (Örneğin; bir haberin yazım hatası puanının belli eşik değerinin üzerinde olup olmaması). Her bir dal ise yapılan testin sonucunu ifade eder. Her bir yaprak düğüm ise tüm nitelikler hesaplandıktan sonra alınan kararı yani bir sınıf etiketini temsil eder. Çocukları olmayan bir düğümün bir yaprak olduğu varsayılır.

Veriyi ayrıntılı bir şekilde temsil etmeye çalışırken ağaç dallanmalarının takip edilemediği aşırı karmaşık ağaçlar üretilmesi olasıdır. Ya da aşırı/ezbere öğrenmeye (*overfitting*) neden olan durumlar ortaya çıkabilir. Bahsedilen problemleri çözmek için parametre eksiltilmesi veya budama (*pruning*) yöntemleri kullanılır.

Karar ağaçlarında bölümlenme kriterleri temelde sınıflandırma ve regresyon problemi olmasına bağlı olacak şekilde değişkenlik gösterir. Algoritma seçiminde hedef değişkenin tipi de önemlidir. Hedef değişkenin tipine göre algoritmayı belirlenir. Karar ağaçlarında en sık kullanılan algoritmalar; kategorik değişkenler için Entropi, Gini, Sınıflandırma Hatası; sürekli değişkenler için ise EKK (En Küçük Kareler) yöntemi şeklindedir.

İdeal bir karar ağacında bölümlenmelerin entropiyi en aza indirgemesi beklenir. Bu ideal bölümlenmeyi sağlamak için bölümlenme öncesi veri setinin entropisi ile her bir özniteliğin entropisi arasındaki fark olarak tanımlayabileceğimiz bilgi kazancı kullanılmaktadır. Eşitlik (20) ve Eşitlik (21)’de sırasıyla Entropi (*Entropy*) ve Bilgi Kazancı (*Information Gain*) hesaplamaları gösterilmiştir.

$$H(X) = - \sum_i p(x_i) \log p(x_i) \quad (20)$$

$$\text{Bilgi Kazancı}(S, D) = H(S) - \sum_{V \in D} \frac{|V|}{|S|} H(S) \quad (21)$$

Burada, S veri kümesini ve D ise verinin bölünmüş bir parçasını ifade eder. V’ler ayrıktır ve S’yi oluşturan alt kümelerdir. En sık yararlanılan karar ağacı algoritmaları; C4.5, ID3,

CHAID, CART şeklinde sıralanabilir. Tez kapsamında 3 farklı karar ağacı modeli kullanılmıştır.

4.5.5.1. Rastgele Orman (RO)

Rastgele Orman (*Random Forest*) sınıflandırıcısı, birden fazla karar ağacı üreterek sınıflandırma oranını yükseltmeyi hedefleyen bir algoritmadır. İsmi aldığı karar ormanı rastlantısal şekilde seçilen karar ağaçlarının bir araya gelmesi ile oluşur. Karar ağaçları büyürken, düğümler parçalara ayrılır ve en önemli özneliği arandığı değil, rastgele bir özellik/öznelik alt kümesi arasında en iyi özelliğin arandığı bir kurgu izler.

Genellikle DVM'lerden daha doğru sonuçlar elde ettiği, ayrıca dengesiz bir dağılım sergileyen veri setlerinde iyi sonuçlar elde edildiği söylenebilir [148]. “N_estimators”, “Max_features”, “min_sample_leaf”, “N_jobs” gibi önemli parametrelerin probleme uygun seçilmesi çok önemlidir.

4.5.5.2. Aşırı Rassal Orman (ARO)

ExtraTrees sınıflandırıcısı olarak geçen bu modelde ağacın karar düğümlerini rastgele üretilmesi mantığına dayanmaktadır. Rastgelelik arttıkça öğrencilerin kararları da o kadar farklılık (*diversity*) gösterecektir.

Bir Rastgele Ormanın “*RandomForestClassifier*” Aşırı Rassal Ormandan daha iyi veya daha kötü performans gösterip göstermediğini ancak problem üzerinde deneyerek ve çapraz doğrulama ile karşılaştırarak anlayabiliriz. Scikit-Learn kütüphanesinde kullanımı rastgele orman sınıflandırıcısına çok benzerdir.

4.5.5.3. Gradyan Güçlendirme/Artırma (GG)

Gradyan Güçlendirme/Artırma (*Gradient Boosting, XGBoost*) algoritması, tahmin hataları göz önüne alınarak yeni ağaçlar oluşturulması ve böylelikle önceki ağaçların hatalarını azaltmaya çalışmaktadır. Bu işlem belirlenen ağaç sayısına bağlı olarak veya daha fazla gelişme sağlanamaması halinde son bulur. Özetle Eşitlik (22)'de verilen ortalama karesel hatayı (*MSE, Mean Squared Error*) en aza indirmek hedeflenir.

$$Kayıp = MSE = \sum (y_i - y_i^p)^2 \quad (22)$$

Bu eşitlikte y_i ; i. sıradaki hedef değer, y_i^p ; i. tahmin ve $L(y_i - y_i^p)$ kayıp fonksiyonudur.

4.6. Uzun-Kısa Süreli Bellek Ağları (LSTM)

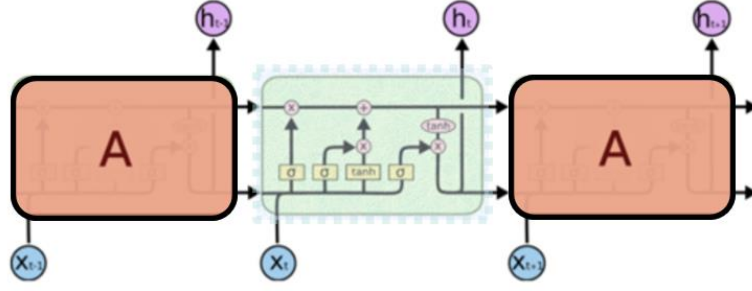
Tekrarlayan Sinir Ağlarının (*RNN, Recurrent Neural Network*) özel bir türü olan Uzun Kısa Süreli Bellek (*LSTM, Long-Short Term Memory*), uzun geçmiş verilerden öğrenebilen ağlardır. Bu özelliği sayesinde girdi yığınının büyüdüğü DDİ çalışmalarında bu Sinir Ağlarını kullanmak oldukça iyi sonuçlar vermektedir. Bunun dışında Evrişimsel (Konvolüsyonel) Sinir Ağları (*CNN, Convolutional Neural Network*) ve Yıgınlı Oto-kodlayıcılar (*DAE, Stacked/Deep Autoencoder*) modellerinin de DDİ çalışmalarında başarılı olduğu söylenebilir [149].

Metin temelli sahte haber tespiti için LSTM kullanan bazı çalışmalarda metin, görüntü, haber başlığı, tweet metni, mesaj metinleri gibi veri içerikleri kullanılmıştır [129, 150-152]. Metin temelli problemlerde LSTM modellerin tercih edilmesindeki ana etken, özellikle geçmiş bilginin artması ile yaşanabilecek gradyan tabanlı geri yayılım (*gradient-based backpropagation*) gibi problemlere kullanılan özel hafıza birimi ile çözüm bularak geçmiş veriden öğrenebilmesidir.

Bu problem özelinde sistemin temel girdisi işlenmiş haber metni (diğer özniteliklerle birlikte) ve çıktı ise iki elemanlı (Sahte-Gerçek) bir vektördür. Eşitlik (23) ve (24)'te LSTM'e ait temel formülasyonlar verilmiştir. Bu eşitliklerde, her bir t adımı için, a^t aktivasyon ve y^t çıktıyı ifade etmektedir. Paylaşılan katsayılar (coefficient) W_{ax} , W_{aa} , W_{ya} , b_a , b_y ve g_1 , g_2 aktivasyon fonksiyonlarıdır. LSTM ağına ait aralarında etkileşim halindeki 4 katmanlı yapı Şekil 4.12'de gösterilmiştir.

$$a^t = g_1 (W_{aa}a^{t-1} + W_{ax}x^t + b_a) \quad (23)$$

$$y^t = g_2 (W_{ya}a^t + b_y) \quad (24)$$



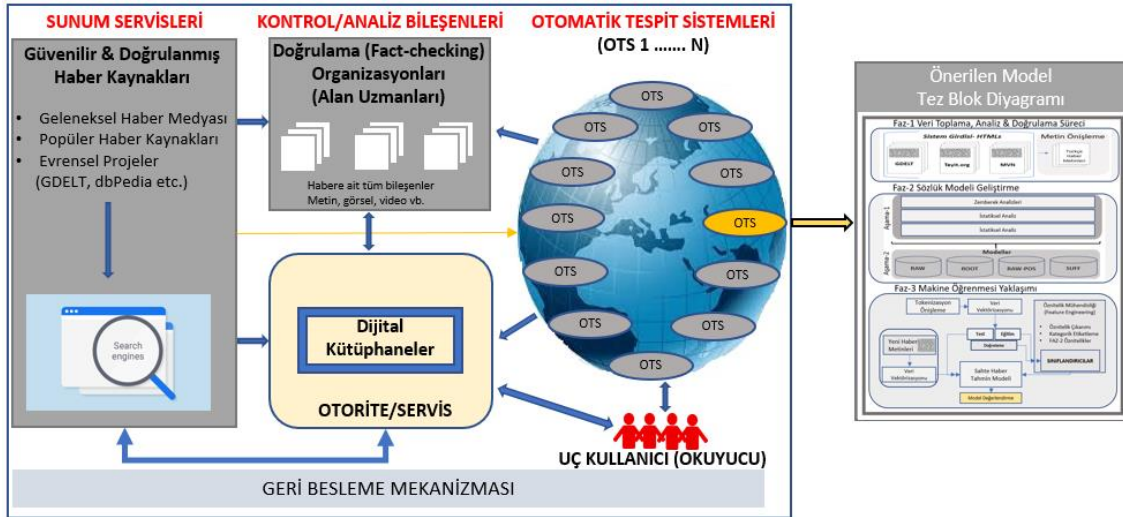
Şekil 4.12 Uzun-Kısa Süreli Bellek Ağları 4 katmanlı yapısı.

Kelime çantası (*bag of words*) modelindeki çok boyutluluk ve bağlamsal ilişkilerin yakalanamaması gibi eksiklikleri görerek, LSTM modelimize girdi olarak 4.3.1’de anlattığımız Word2Vec kelime yerleştirme modelini kullanılarak eğitilmiştir.

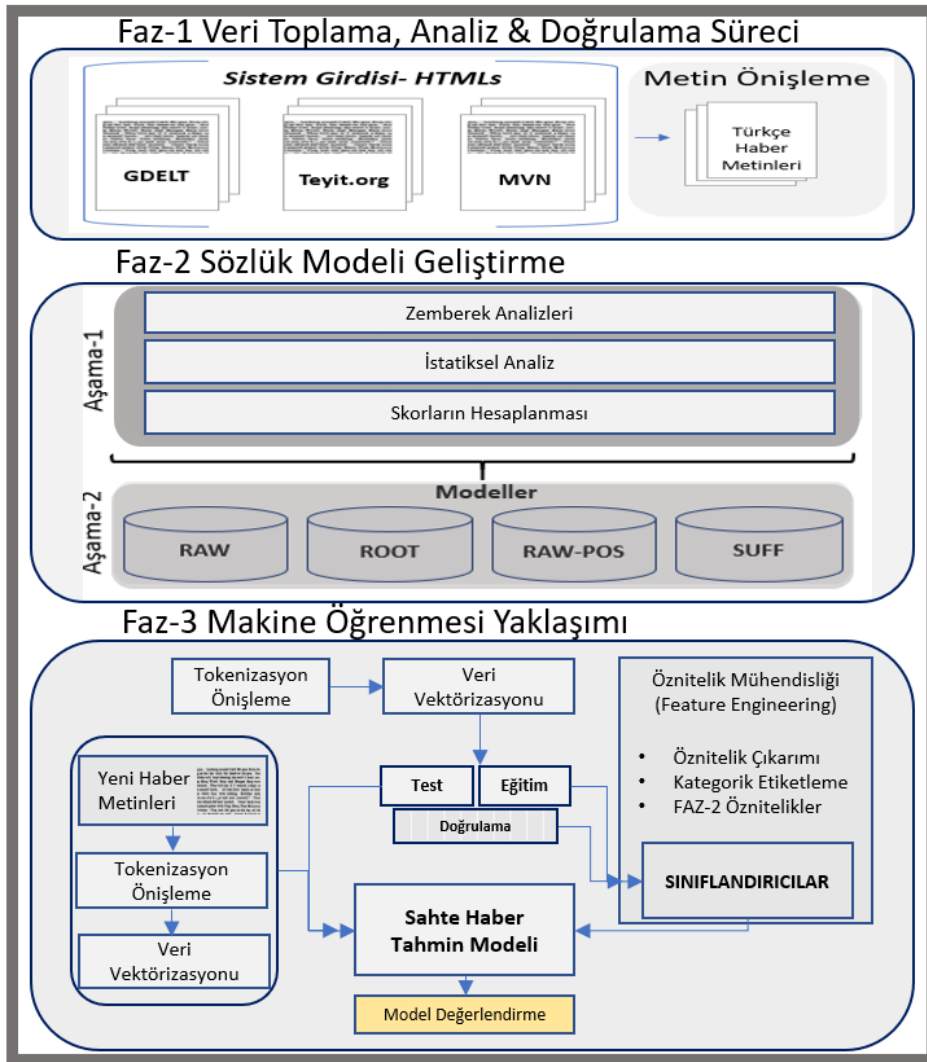
Uygulanan yöntem özetle, her kelimenin Word2Vec modeli tarafından eğitilmiş karşılığı olan vektörlerle temsil edilmesi şeklindedir. Bu kelime dizilişleri (*sequence*), sırasıyla LSTM modelimize Şekil 4.12’de belirtildiği gibi girdi olarak verilmektedir. Yöntemimizde nispeten kısa olan metin bilgisi taşıyan haberler, LSTM’in sıralı giriş düzenine uyum sağlaması açısından üç boyuta indirgenmiş bir matrise dönüştürülmektedir. İlk boyut Word2Vec modelin boyutu, diğer boyutlarda sırasıyla haberdeki kelime sayısı ve toplam haber sayısı olacak şekilde hesaplanmıştır.

4.7. Tez Blok Diyagramı

Bu bölümde, Bölüm 2.4.1’de önerilen Çatı Sistem Modeli Şekil 4.13’te ve Tez Blok Diyagramı Şekil 4.14’te gösterilerek, tez çalışmasını özetleyen iki görsel sunulmuştur. Ayrıca bu bölümde, tez çalışması ana hatları ile özetleyen bir biçimde sunulmuştur. Şekil 4.13’te OTS (Otomatik Tespit Sistemi), Şekil 4.14’te gösterilen tez sonucunda elde edilen model ile gerçekleştirilmiştir.



Şekil 4.13 Çatı Sistem Modeli – Sahte Haber Tespit Modeli gösterimi



Şekil 4.14 Tez Blok Diyagramı

4.8. Çıktılar ve Performans Değerlendirme

Bu bölümde kurulan hibrit modelin performansının ölçüldüğü deneysel süreç ve sonuçlar sunulmuştur. Tüm deneyler Scikit-learn kütüphanesi kullanılarak gerçekleştirilmiştir. Ayrıca, LSTM kullanabilmek için Keras kütüphanesinden faydalanılmıştır. Farklı sınıflandırma modelleriyle bir dizi deneysel çalışma yürütülmüştür. Veri setinin %70'i eğitim, %30'u test verisi olarak ayrılmıştır. Modellerin ve önerilen özniteliklerin performansını değerlendirmek için Bölüm 3.1'de anlatılan kesinlik (hassasiyet), duyarlılık (geri çağırma), F1 skoru ve doğruluk ölçütlerini kullanılmıştır. Geniş öznitelik havuzundan nihai olarak belirlediğimiz ve kullanım kolaylığı açısından grupladığımız öznitelikleri ayrı ayrı ve birlikte olacak şekilde farklı kombinasyonlarını farklı sınıflandırma modelleri ile kullanarak, modellerle olan uyumlarını ve etkililikleri karşılaştırılmıştır.

Sınıflandırma modellerinden en etkin sonuçları alabilmek için modellere uygun çeşitli parametreler denenerek en uygun (optimal) parametrelerin kullanılması performansı arttırmanın önemli bir yoludur [153]. Parametre veri içerisinde tahmin edilebilirken, hiper parametreler modellere özgüdür ve eğitim süreci başlamadan belirlenmesi gerekir. Örneğin k-Yakın Komşu için k, DVM için c parametrelerin belirlenmesi hiper parametre optimizasyon (*hyperparameter tuning*) sürecine örnek olarak gösterilebilir. Burada iki yöntem ile bu optimizasyon deneyimlenmiştir. Öncelikli olarak Grid Arama (*Grid Search*) ve Rastgele Arama (*Randomized Search*) metotları [154] TF_RN veri setine ait küçük bir alt kümede (4000 haber) ele alınmıştır. Her ne kadar bu küçük veri setinde Grid Arama ile çok az farkla daha iyi sonuçlar elde edilse, tüm veri üzerinde düşünüldüğünde zaman ve hesaplama açısından, Scikit-Learn kütüphanesinin de önerdiği Rastgele Arama ile uygun parametrelerin bulunması sağlanmıştır. Bu iki yöntemin yanı sıra sezgisel, Bayes yöntemi ile, genetik algoritmalarla vb. hiper parametre arama yöntemleri mevcuttur. Model TF_RN veri seti üzerinde 10 kat-çapraz doğrulama ile geçerlenmiştir.

Ayrıca probleme derin öğrenme yöntemlerinden metin karakteristiğine en uygun olarak değerlendirdiğimiz bir Yinelenen/Tekrarlayan Sinir Ağı (RNN, Recurrent Neural Network) çeşidi olan Uzun-Kısa Süreli Bellek (*LSTM*) Ağları mimarisi ile de yaklaşılmıştır. Bu noktada derin öğrenmede hiper parametre bulma yöntemlerinin halen

gelişim ve değişim sürecinde olduğunu belirtmek faydalı olacaktır. Sonuçların problemlere göre değişiklik gösterdiği söylenebilir. Ayrıca sadece tek bir mimari kullandığımız için literatürdeki en iyi pratiklerden (*best practice*) yararlanarak optimizasyon algoritması, aktivasyon fonksiyonu, katman sayısı, nöron sayısı, kernel boyutu vb. birçok hiper parametre türü için sırasıyla hiper parametreler sezgisel olarak denenmiş, modelin başarımı gözlemlenmiştir. Elde edilen başarılı sonuçlar değerlendirildiğinde, Grid Search veya genetik algoritmalar gibi parametreleri otomatize şekilde bulmamızı sağlayan tekniklere bu yöntemde ihtiyaç duyulmadığı söylenebilir.

7 farklı makine öğrenmesi modeli ve bir Yinelene Sinir Ağlarının (RNN, Recurrent Neural Network) özel bir çeşidi olan LSTM ile eğitilmiştir. Bu modeller Bölüm 4.5 ve 4.6'da detaylı şekilde anlatılmıştır. Sonuçlar özellikle TR_{Opt} olarak adlandırdığımız optimize edilmiş özniteliklerin farklı modellerdeki birbirine yakın tutarlı sonuçları bu özelliklerin genel-geçer ve etkili özellikler olduğunu göstermektedir.

Birçok farklı öznitelik kombinasyonu ve verinin alt kümeleriyle çok sayıda deney gerçekleştirilmiştir. Bunlardan tüm veri esas alınarak yaptığımız değerlendirme, Çizelge 4.2 ve Çizelge 4.3'te sırasıyla makine öğrenmesi modellerinden ve LSTM'den elde edilen sonuçlar şeklinde sunulmuştur. Bu çizelgelerde her bir model özniteliklerin kombinasyonlarından en iyi sonuç veren 3 tanesinin yanı sıra, her bir modelde teknikler ayrı kullanıldığında elde edilen bazı ilginç bulgular gösterilmiştir.

Elde edilen verilere göre en iyi sonuçların, Uzun Kısa Süreli Bellek Ağlarından alındığı gözlemlenmiştir. Bu sonuca ait ROC eğrisi grafiği Şekil 4.15'te gösterilmiştir. Bunun ardından Karar Ağaçlarının sırasıyla Aşırı Rassal (ExtraTrees) Orman, Rassal Orman ve Gradyen Güçlendirme sınıflandırıcılarından iyi sonuçlar alınmıştır. Ardından KNN, MNB, Lojistik Regresyon ve DVM algoritmalarından alınan sonuçların sıralandığı görülmektedir.

Yapılan deneysel sonuçlara göre eğitilen verilerde ses, görsel ve video öğelerinin yoğun olduğu hibrit çözüm gerektiren problemlerde tercih edilen derin öğrenme yöntemlerinin metin tabanlı çözümlerde de oldukça başarılı sonuçlar verdiğini söylemek mümkündür.

Genel çerçeveden bakıldığında yapılabilecek bir çıkarım da tüm modellerin en iyi sonuçları ele alındığında F1 skor ve Doğruluk metriklerindeki tutarlılıktır. %90 ile %98 arasında ölçümlenen başarılı performanslar kurulan modellemenin ve özellikle kullanılan optimize edilmiş özniteliklerin oldukça yerinde olduğuna atfedilebilir.

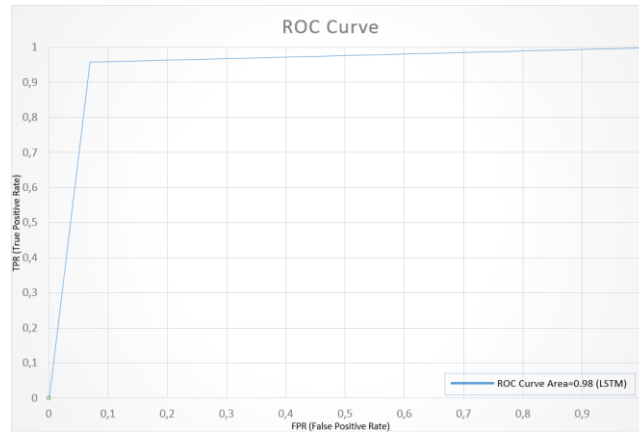
Çizelge 4.2 Makine Öğrenmesi modellerine ait değerlendirme sonuçları

Model	Öznitelik Seti	Kesinlik % (0,1)		Duyarlılık % (0,1)		Doğruluk	F1 Skor
<i>MNB</i>	BOW _{All} + TROpt	92,92	93,41	93,37	92,95	93,16	93,14
	Word2Vec + TROpt	93,91	94,03	94,02	93,91	93,97	93,96
	(Unigram+bigram (TF)) + TROpt	93,85	93,80	93,81	93,85	93,83	93,83
	Word2Vec	91,12	92,77	91,72	91,18	91,45	91,42
<i>KNN</i>	BOW _{All} + TROpt	93,66	94,05	94,03	93,68	93,85	93,84
	(Unigram-(TF)) + TROpt	93,65	93,50	93,52	93,63	93,58	93,58
	Word2Vec + TROpt	94,03	94,08	94,07	94,03	94,05	94,05
	bigram (TF)	85,30	84,73	84,89	85,15	85,02	85,09
<i>DVM</i>	(Unigram-(TF)) + TROpt	87,83	88,92	88,99	87,75	88,37	88,41
	Word2Vec + TROpt	89,70	88,96	89,04	89,62	89,33	89,37
	(Unigram-bigram (TF)) + TROpt	90,85	91,26	91,22	90,89	91,05	91,03
	TR _{All}	88,99	86,05	86,45	88,66	87,52	87,70
<i>LR</i>	(Unigram-(TF)) + TROpt	91,56	92,28	92,22	91,62	91,92	91,89
	Word2Vec + TROpt	90,66	92,28	92,22	90,73	91,46	91,43
	(Unigram-bigram (TF)) + TROpt	92,25	92,90	92,86	92,30	92,57	92,55
	E _{1,2} + L _{1,2}	89,05	87,04	87,29	88,82	88,04	88,16
<i>RO</i>	(Unigram-(TF)) + TROpt	93,71	94,44	94,40	93,75	94,07	94,05
	Word2Vec + TROpt	93,87	95,00	94,94	93,94	94,44	94,41
	(Unigram-bigram (TF)) + TROpt	93,90	95,14	95,12	93,92	94,52	94,51
	E _{1,2} + L _{1,2,3,4}	91,05	91,84	91,78	91,12	91,44	91,41
<i>GG</i>	(Unigram-(TF)) + TROpt	93,95	94,59	94,56	93,99	94,27	94,25
	Word2Vec + TROpt	93,85	95,08	95,03	93,91	94,47	94,44
	(Unigram-bigram (TF)) + TROpt	94,94	95,24	95,23	94,95	95,09	95,08
	(bigram (TF-IDF)) + E _{1,2} + L _{1,2}	93,27	93,80	93,77	93,30	93,53	93,52
<i>ARO</i>	(Unigram-(TF)) + TROpt	94,05	94,92	94,88	94,10	94,49	94,46
	Word2Vec + TROpt	94,09	94,94	94,90	94,14	94,51	94,49
	(Unigram-bigram (TF)) + TROpt	97,90	95,72	95,81	97,86	96,81	96,85
	(Unigram-bigram-trigram (TF)) + TROpt	92,62	95,26	95,13	92,81	93,94	93,86

Çizelge 4.3 LSTM modeline ait değerlendirme sonuçları

Model	Öznitelik Seti	Kesinlik % (0,1)		Duyarlılık % (0,1)		Doğruluk	F1 Skor
<i>LSTM</i>	BOW _{All} + TROpt	94,26	94,99	94,96	94,30	94,63	94,61
	(Unigram+bigram (TF)) + TROpt	97,85	95,40	95,51	97,80	96,63	96,67
	Word2Vec + TROpt	98,20	95,83	95,93	98,16	97,02	97,05
	E _{1,2} + L _{1,2,3}	91,91	87,63	88,13	91,54	89,77	89,98

Ağırlıklandırarak kullandığımız Word2Vec modelin sistemin başarımına katkısı göze çarpmaktadır. Bu maksatla, sadece Word2Vec modelin öznitelikleri sistem eğitildiğinde, Çizelge 4.2’de gözlemlenebildiği üzere MNB sınıflandırıcısı ile %91’e yakın bir doğruluk değeri elde edilmiştir. MNB sınıflandırıcısı ile elde edilen en yüksek metrik değerlerine bakıldığında Word2Vec modelin baskın bir şekilde etkisini hissettirdiğini söylemek mümkündür. Gözlemlenen bu durum diğer sınıflandırıcılarda da aynı şekildedir. Her modelden elde edilen en yüksek başarımla sadece Word2Vec kullanılarak elde edilen başarımlar arasında %2 ile %8 arası bir fark gözlemlenmiştir.



Şekil 4.15 LSTM modeline ait ROC eğrisi

Nispeten daha düşük başarımların elde edildiği Lojistik Regresyon sınıflandırıcısında, sözlük temelli öznitelikler (L_{1,2}), dil karmaşıklığı ve yazım hata skoru (E_{1,2}) ile değerlendirildiğinde hem sahte hem gerçek haberin istikrarlı bir biçimde yakalanması, Bölüm 3’te küçük bir test seti ile test ettiğimiz dil modellerinin/sözlüklerin beklenildiği üzere makine öğrenmesi yöntemleri ile daha tutarlı davrandığı söylenebilir.

Sınıflandırıcılar birbirleriyle kıyaslandığında Destek Vektör Makinelerinin (DVM) daha fazla yanlış alarm (FP , FN) ürettiği görülmektedir. Yine DVM’de tüm öznitelikler birlikte kullanıldığında başarıyı azalttığı görülmüştür.

Karar ağaçlarından diğer makine öğrenmesi modellerine göre daha iyi sonuçlar elde edilmiştir. Çizelge 4.2’de; Rassal orman, gradyan güçlendirme ve aşırı rassal orman için ilk üç satırda aynı öznitelik grupları kullanılmıştır. 4. Satırda farklılıkları gözlemleyebilmek için sırasıyla; “ $E_{1,2} + L_{1,2,3,4}$ ”, “(bigram (TF-IDF)) + $E_{1,2} + L_{1,2}$ ”, “(Unigram-bigram-trigram (TF)) + TR_{Opt} ” öznitelik grupları ile eğitilmişlerdir. Burada gradyan güçlendirme ile kullanılan TF-IDF hesaplamasına ait sonuç dikkat çekicidir. Durak kelimeler (*stop words*) çıkarıldıktan sonra yapılan hesaplama ikili terimlerin doküman içerisindeki önemini gösterdiğinden tamamlayıcı bir görev üstlendiği ve terim frekanslarına ağırlık veren sözlükleri tamamladığını söylemek uygun olacaktır. Diğer iki öznitelik grubunda TF daha ağırlıklı kaldığından sonuca benzer etki gösteren öznitelikler gibi davranmışlardır.

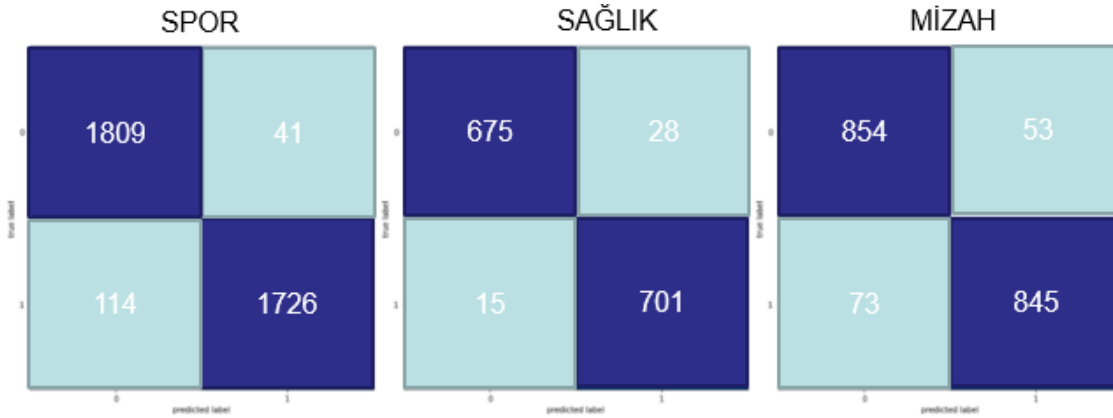
Algoritmaların çalışma sürelerinde olumsuz olarak kaydedilebilecek bir yavaşlık gözlemlenmemiştir. Sadece LSTM beklenildiği üzere daha uzun sürmüştür. Makine öğrenmesinden farklı olarak, LSTM modeli Keras kütüphanesinden faydalanarak kullanılmıştır. Ayrıca, Çizelge 4.3’ten elde edilen sonuç, Şekil 4.16’da gösterilen parametrelerle çalıştırılmıştır. Epoch sayısı, model eğitilirken verilerin kaç tekrar ile ele alınacağı belirtilir. Bu değeri olabildiğince küçük tutarak eğitim süresi kısa tutulmaya çalışılmıştır. İlk denemelerde daha büyük sayılar verildiğinde modelin her seferinde epoch=5’te gelişimini tamamladığını gözlemlenmiştir.

Ayrıca veri seti içerisinde belli alanlardaki haberlerde (spor, sağlık, sosyal medya paylaşım haberleri) LSTM ile daha da başarılı sonuçlar alındığı görülmektedir. Sonuç olarak makine öğrenmesi yöntemlerinde Aşırı Rassal Orman (*ExtreTrees*) ve derin öğrenmede LSTM modellerinin bu problemde ortaya çıkarılan özniteliklerle yüksek bir uyumda çalıştığı gözlemlenmiştir. Kurulan hibrit model oldukça başarılı doğruluk oranları elde edilmiştir. LSTM kullanılarak yapılan alan bazlı tahminleme sonuçları Şekil 4.17’de verilmiştir.

```
[109] def RNN():
    inputs = Input(name='inputs', shape=[features_len])
    layer = Embedding(max_number, 50, input_length=features_len)(inputs)
    layer = LSTM(64)(layer)
    layer = Dense(256, name='FC1')(layer)
    layer = Activation('relu')(layer)
    layer = Dropout(0.5)(layer)
    layer = Dense(1, name='out_layer')(layer)
    layer = Activation('sigmoid')(layer)
    model = Model(inputs=inputs, outputs=layer)
    return model

[111] history=model.fit(X_train,y_train,batch_size=80,epochs=5,
    validation_split=0.1,callbacks=[EarlyStopping(monitor='val_loss',min_delta=0.0001)])
```

Şekil 4.16 LSTM parametrelerinin seçilmesi



Şekil 4.17 LSTM ağlarının performans (Spor, Sağlık ve Mizah) değerlendirmesi

Şekil 4.17’de verilen sonuçlara göre sırasıyla Spor haberlerinde %95.8 doğruluk oranı ve %95.89 F1 skor yaklanırken, Sağlık alanında hem gerçek hem sahte haber tahminleme daha dengeli şekilde %96.97 doğruluk ve %96.91 F1 skor değerleri elde edilmiştir. Modelimiz mizah haberlerinde ise biraz daha düşük oranlar yakalamıştır. Mizah haberleri için her iki oranda %93’te kalmıştır. Bu nispeten düşük başarımların oranları mizah haberlerinin tespitinde yoğun anlamsal (semantik) analiz kullanılmasını doğrular şeklindedir. Bir diğer neden ise mizah haberlerinin direk bir haber türü değil bir üst tür olmasıdır. Spor, kültür, politika vb. her alandaki haberler mizah boyutunda ele alınabilir. Tezin önceki bölümlerinde de belirtildiği üzere mizah, alay, hiciv tarzındaki dil kullanımlarında sözcüklerin dolaylı anlamlar barındırması, bu tür haberlerde sahteliği salt biçimbilimsel analizle tespit etmenin yetersiz olduğunu göstermesi açısından önemlidir.

4.8.1. Keşifsel Veri Analizi Tespitleri

Bu çalışmada deneysel sonuçların yanı sıra bazıları gözlemsel olarak çarpıcı biçimde bazıları da KVA (Keşifsel Veri Analizi) yöntemleri ortaya çıkarılan birtakım tespitlerimiz olmuştur. Bu bölümde sonuçlardan, çıktı performanslarından dolayı şekilde elde edilebilecek fakat ilk bakışta elde edilemeyen veri içindeki örtük bilgi (*tacit knowledge*) sunulmaya çalışılmıştır. Bazı ilginç tespitler ve bulgular aşağıda belirtilmiştir:

- Özellikle sahte haberlerin sosyal ağlardaki paylaşımında ilginç bazı detaylar görülmektedir. Sahte haberler, gerçek haberlere göre daha hızlı ve dinamik yayılım göstermektedir. Bu tarz haberler daha fazla yeniden paylaşılmakta (*retweet, share*), beğenilmekte (*like*) ya da beğenilmemektedir (*dislike*). Gerçek-zamanlı tespit çalışmalarında bu yayılım ve etkileşimler çok önemli ipuçları (kaynak tespiti, aktör/düğüm tespiti vb.) içerdiğinden Sosyal Ağ Analizi (*SNA, Social Network Analysis*) kullanılması ve sosyal bağlam kapsamındaki özniteliklere yoğunlaşılması bu bakımdan faydalı olacaktır.
- Normal kullanımında konuya ciddiyet katması ve önem atfetmesini beklediğimiz ifadelerin sahte haberlerde sıklıkla kullanıldığını ve okuyucu açısından bu ifadelerin şaşırtıcı bir şekilde anlamını yitirdiği gözlemlenmiştir. Bu örneklerden bazıları Çizelge 4.4'te sunulmuştur.
- Hiciv (*Satire*), Parodi (*Parody*), Hoax (*Şaka*) vb. mizah haberlerinin doğası gereği sahte haberlerin diğerlerinden farklı değerlendirilmesi gerekmektedir. Bu tarz haberlerin okuyucuyu aldatmaktan ziyade haber konusu veya haber aktörleri ile ilgili toplumsal saygınlığın azalması şeklinde önemli etkileri olabilir. Bazı haber sitelerinin bu tarz haberleri yaptığını açıkça beyan ettiği görülürken bu beyanı yapmayan özellikle eleştirel tarzda hiciv yapan kaynaklarda hiciv tespiti sahte haber tespitinden farklılaşmaktadır. İçerik ve söylev analizinin yanı sıra duygu analizi gibi anlamsal analizinin kullanılmasını gerektirmektedir.

Çizelge 4.4 Sahte haberlerde daha sık karşılaşılan bazı ifadeler.

Kullanılan İfade	Yazarın/Kaynağın Niyeti (N), Gerçek Durum (G)
Son Dakika!	N- Okuyucuda çok şaşırtıcı bir bilgi veya çok yeni bir haber verilecek hissi uyandırmak. G- Haber başlığı ile uyumsuz bir haberle karşılaşılması
Çok Önemli!	N- Okuyucuya haberi ciddiye alması gerektiğini hissettirmek. G- Herhangi bir haber değerinin olmaması.
İddiaya/İddialara göre	N- Okuyucunun habere daha çok inanmasını sağlamak. G- Sunulan destekleyici ifadelerin teyit edilememesi (bir temelini veya referansının olmaması).
Bunları biliyor muydunuz?	N- Yeni ve yararlı bir bilgi verildiğini vurgulamak. G- Bilimsellikten uzak ve genel-geçer olmayan bilgilerin sunulması
Sürpriz Gelişme	N- Beklenmeyen bir durum bilgisi sunarak ilgi çekmek G- Uydurma veya kurgu bir haber ile karşılaşılması.

- Özellikle sağlık ile ilgili olan sözde bilim haberleri, şehir efsaneleri ve dedikodu tarzı haberlerin belirli zamanlarda kendini tekrarladığı görülmektedir.
- Spor haberlerinde örneğin futbolda transfer dönemi haberlerinde jenerik haber formatlarının sadece içeriğinin değiştiği gözlemlenmiştir. Okuyucunun ilgisini çeken bu tarz haberlerin genellikle arka planda bir araştırma yapılmadan üretilen veya dedikodu tarzında gelişen haberler olduğunu söylemek mümkündür. Bu haberlerin Bölüm 4.8’de de görüldüğü üzere otomatik tespit açısından önemli ipuçları barındırdığı görülmektedir.
- Özellikle politik haberlerde konu, aktör ve zamanın doğru tespiti haberin gerçeklik ve sahteliği açısından önemli ipuçları içermektedir. Özellikle seçim öncesi-sonrası vb. belli dönemlerde aktörlerin merkezinde olduğu sahte ve kurgusal haberlerin çoğaldığı söylenebilir. Özellikle bu tip haberler için Varlık İsmi Tanıma (*Named Entity Recognition*) çalışmaları ile desteklenen otomatik sahte haber tespit çalışmalarının büyük katkısı olacağını değerlendirmekteyiz.
- Haber başlıkları ve metin arasındaki örüntünün (*pattern*) bazı haber tiplerinde yakalanabileceği değerlendirilmektedir. Bazı çalışmalarda da kullanıldığı gibi başlıklardan ve haber metninden yararlanılarak haberin görüşü bakış açısı tespit edilebileceği de görülmektedir.

- Sahte haberlerin daha ziyade kriz ortamlarında, karalama kampanyaları örgütlendiğinde ve toplumun doğru bilgiye erişmesinin zorlaştığı dönemlerde ortaya çıktığı bilinmektedir. Bu durum göz önüne alındığında; özellikle sosyal medyada yayılan başta Whatsapp olmak üzere mesajlaşma uygulamalarında dolaşan toplumda sansasyonel etki bırakma amacı taşıyan bilgi ve haberlerin erken tespiti, ardından hızlı bir şekilde yetkili makamlarca bilgi kirliliğinin önüne geçerek toplumun doğru haberle buluşturulması noktasında çok önemli olduğu değerlendirilmektedir.
- Görsel ve video kullanılarak servis edilen sahte haberlerde, görselleri tersine aratma (*reverse image search*) yöntemiyle orijinal kaynağını bulabileceğimiz internet hizmetleri bulunmaktadır. Bu hizmetler geliştirildikçe metin, görsel ve video içeren sahte haber tespiti alanında daha çok kullanılacağı öngörülmektedir.

Bu bölümün sonunda, Tezin sonunda verilen ve konuyu özetleyen iki EK kısaca açıklanmıştır. Bunların ilki olan EK-2’de güncel bir örnek olarak farklı olaylarda anlatımı güçlendirmek ve inandırıcılığı pekiştirmek amacıyla kullanılan bir fotoğraf sunulmuştur. 4 Ağustos 2020 tarihinde gerçekleşen Beyrut patlamasında ait olduğu belirtilen bir fotoğrafın esasen 29 Temmuz 2006 yılında Afganistan’da çekildiği gösterilmektedir.²³. Aynı fotoğraf Yemen’de çocukların yetersiz beslendiğini gösteren bir haber detayı olarak kullanılmıştır. Benzer şekilde EK-3’te teyit.org sitesi tarafından analiz edilmiş ilgi çekici bir sözde bilim haberi verilmiştir. Bu örneklerin çoğaltılması mümkündür. Tüm bu bulgu ve gözlemler özellikle önemli olaylarda ve kritik süreçlerde bu tarz haberleri yayan kaynak tespitini en hızlı şekilde bularak teyit ve doğrulamanın hızlı bir şekilde yapılabileceği sistematik ve modellerin önemini bir kez daha göstermektedir.

4.9. Sonuç (Faz-3)

Bölüm 4’te Makine Öğrenmesi temelli yaklaşımla problem ele alınmıştır. Ayrıca Derin Öğrenme yöntemlerinden Uzun-Kısa Süreli Bellek Ağlar da deneysel uygulamaya dahil edilmiştir. Alınan sonuçlar Bölüm 3’te de öngörüldüğü üzere oldukça başarılıdır. Bu

²³ <https://eski.teyit.org/fotograflarin-beyruttan-oldugu-iddiasi/>. (Erişim tarihi: 23 Ağustos 2020)

sonular Trke dili iin sahte haber tespitini metinsel boyutta yksek bařarı oranlarının ulařılabileceđini gstermektedir. Ayrıca yine Trke iin metin sınıflandırması zerine yapılan ve nispeten yeterli olgunluk seviyesine geldiđi deđerlendirilen bazı problemlerinde metinsel erevede derin đrenme yntemleri ile daha ileri boyuta tařınabileceđini gstermesi aısından nemli bir bulgudur. Blm 5'te tm tezi kapsayacak sonu, ıkarım ve neriler daha kapsamlı Őekilde dzenlenmiřtir.

5. SONUÇLAR VE ÖNERİLER

5.1. Sonuçlar

Sahte Haber, nispeten yeni çalışılmaya başlayan bir konu olmasına rağmen, birçok disiplinin kesişim alanına giren önemli bir konu olması nedeniyle son birkaç yıl içerisinde oldukça fazla ve çeşitli akademik çalışma ortaya konmuştur. DDİ ve Yapay Zekâ disiplinleri ile doğrudan ilgili olan sayısal ve bilgisayar merkezli çözümler düşünüldüğünde, farklı dillere özgü olarak geliştirilecek projeler, kütüphaneler, uygulamalar vb. yenilikçi girişimlerle birlikte akademik çalışmaların ve pratik hayat uygulamalarının artacağını öngörmekteyiz. Bu bakımdan tezin Giriş Bölümünde de belirtildiği üzere farklı karakteristikler içeren ve köklü bir dil olan Türkçemize ait bir elin parmaklarını geçmeyecek kadar az çalışma bulunmasını, araştırmacıların dikkate alması gereken bir husus olduğunu değerlendiriyoruz. Dünyada bu alandaki gelişmelerin özellikle metin tabanlı çözümlerin dile özgü kaldığı ve kendi dilini çalışmak isteyen araştırmacılar için yeni ufuklar yaratabileceği görülmektedir. Dil kaynaklarının, kütüphanelerinin gelişimi uzun zaman almakta, kolektif bir tutum ve kümülatif çalışmalar gerektirmektedir.

Sonuç ve Öneri bölümünde, öncelikli olarak tez kapsamında yapılan çalışma değerlendirilmiş, alandaki karşılaşılabilecek zorlayıcı etkenler ve gelecek çalışmalar için bazı önemli tespitler yapılmıştır. Ardından alanda çalışmak isteyen araştırmacılara katkıda bulunmak isteyebilecekleri alt konular ve yeni araştırma alanları sunulmuştur.

Tez kapsamında, Türkçe haberlerde sahte haberin otomatik tespiti 3 temel fazda ele alınmıştır. Birinci fazda, problem tanımı yapılmış, konunun önemi ve Türkçe açısından değerlendirilmesi ele alınmış ve bütüncül bir yaklaşımla kavramın etki ve ilgi alanı belirlenmiştir. Bu faz ikinci fazda geliştirilmesi hedeflenen Türkçe Sahte Haber Sözlüğü oluşturulması için gerekli tüm ön işlemleri kapsar. Bu doğrultuda sahte ve gerçek haberleri içeren Türkçe haber derlemi oluşturulmuştur. İkinci fazda, bu derlem kullanılarak 4 farklı terim kategorisinde sahte-gerçek ton puanlamasına sahip sözlükler ortaya konmuştur. Keşifsel veri analizi kapsamında ortaya çıkan bu sözlüklerin sahte

haber tahmininde genel olarak kullanılabilir seviyede olduğu hem sözlüklerin hem de metodolojinin gelişime açık olduğu görülmüştür. Bu doğrultuda gelecek çalışmalarda derlemin büyütülmesi, sözlük kapsamının ise daraltılması ve alana özgü farklı puanlama hesaplamalarının yapılabileceği değerlendirilmektedir. Alandaki temel problemlerden biri olan sahte haber tespitinde manuel insan gücünü azaltabilecek bir karar destek sistemi olarak bu sözlüklerden faydalanılabileceği düşünülmektedir. Son fazda ise probleme makine öğrenmesi ve derin öğrenme modelleri ile yaklaşmış, ikinci fazda geliştirilen sözlük, stil işaretçileri ve özniteliklerden faydalanılmıştır. Metinlerin vektörlere çevrilerek sınıflandırma modelleri ile oluşturulan sistemin başarımları ölçümlenmiştir. Ayrıca dijital kütüphaneciliğin rolüne işaret ederek kurgulanan çatı modeli ile metnin yanı sıra ses, görsel ve video öğeleri içeren haberler içinde sahte haber tespitini kolaylaştırabileceğini öngördüğümüz evrensel bir çerçeve önerilmiştir. Bu kapsamda tezimizde deneylerle yüksek başarımları yakalayan modelimiz ile, önerdiğimiz ana çerçevenin (çatı model) sayısal çözümler açısından uygulanabilirliğini gösterilmiştir.

Tez çalışması özelinde ve genel olarak sahte haber tespiti hususunda yapılan bazı çıkarımlar aşağıda sunulmuştur:

- Bu çalışmada dahil olmakla birlikte literatürdeki sahte haberin otomatik tespitine yönelik olarak yapılan birçok çalışmada Denetimli Öğrenme modelleri kullanılmıştır. Özellikle sosyal medya platformları vb. kaynaklarda çoğu etiketsiz (*unlabelled*) verinin bulunduğu göz önüne alındığında problem denetimsiz öğrenme metodolojisi içeren farklı bakış açılarıyla ve farklı veri kaynakları ile ele alınmasının faydalı olacağı değerlendirilmektedir.
- Sözlük temelli yaklaşımda sahte-gerçeklik ton atamaları yapılırken gerçekten etkili kelimeleri daha ön plana çıkarmak için, sözcüklere geçiş sıklığı bakımından bir eşik değeri atanabileceği değerlendirilmektedir.
- Alanda yapılan çalışmaların büyük bir çoğunluğu İngilizce metinler üzerinedir. Bölüm 2.4.1’de önerdiğimiz ana çatı modelde belirttiğimiz üzere farklı dilleri birlikte ele alan çözümlere ihtiyaç bulunmaktadır.
- Sosyal ağların ve değişen iletişim teknolojilerinin etkisiyle ortaya çıkan karmaşık ve değişken ağ yapısı, dinamik çözümleri gerektirmektedir.

- Sahte haber alanında açık kaynak olarak sunulmuş referans alınabilecek kaliteli veri setlerinin azlığı da bu konuda çalışılmasını ve özellikle içerik doğrulanması anlamında etik açıdan yerinde çalışmaların yapılmasını zorlaştırmaktadır. Tez kapsamında Türkçe için titizlikle oluşturulmuş bir veri seti literatüre kazandırılmıştır.
- Sahte haber tespitinde gerçek-zamanlı tespitler yapan model ve algoritmalar geliştirmek kesin çözüm adına oldukça gerekli bir o kadar da zor bir husustur. Bunun için haberin ve kaynağının birçok bileşenini birlikte ele alabilen hibrit çözümlerin çalışılması gerekmektedir. Derin öğrenme yöntemleri sahte haber tespiti hususunda daha sık tercih edilir hale gelmiştir. Bu çalışmada elde edilen sonuçlar da göz önüne alındığında derin öğrenmenin haberin birçok bileşenini tek potada eritebilmek adına ilerleyen dönemde daha çok kullanılacağı öngörülmektedir. Öte yandan sahte haberler esas etkisini kısa bir sürede göstermektedirler. Dolayısıyla gerçek-zamanlı tespit sistemlerinin yanı sıra erken tespit sistemleri de sahte haberin potansiyel etkisini önleyebilmek açısından önemli ihtiyaç olarak ortaya çıkmaktadır.
- Tez kapsamında belirtildiği üzere çalışmaların büyük bir çoğunluğu belli alanı kapsayan ve belli tipte sahte haberlere yoğunlaşmıştır. Bu durum kullanılan teknikleri de alan ve türe özgü bırakmaktadır. Bu bakımdan alanda kapsayıcı bir çerçeve çizen genelleştirilebilir çalışmalara ihtiyaç olduğu değerlendirilmektedir.
- Bölüm 2.1.2’de çeşitli faktörlere göre kümelediğimiz sahte haberin değişik formlarının dağılım hızları ve etki potansiyelleri başlı başına çalışılabilecek bir alt konu ve hatta bir “tez çalışması” olabileceği değerlendirilmiştir.
- Kullanıcıların birden fazla sosyal medya platformunda birden fazla kullanıcı hesabı olduğu ve bazen de tek hesabın birden fazla kullanıcısı olduğu düşünüldüğünde kullanıcıların ağ davranışları ve haber yayılım ilişkilerini ortaya koymak alt problemler olarak çalışılması gereken konular arasındadır.
- Tarafsız şekilde ortaya çıkarılan çalışmaların platformlar halinde son kullanıcıya sunulması bilgi/medya okuryazarlığını daha iyi seviyelere getirerek toplumun sahte haber konusunda daha bilinçli olmasını sağlayacaktır. Aynı şekilde kurumlara özellikle gerçeklik kontrolü yapan organizasyonlara karar destek

sistemi olarak sunulması sahte haber tespit sürecini hızlandıracak ve harcanan emeği azaltacaktır.

5.2. Öneriler

Sahte Haber konusu, bir araştırma alanı olarak bir konu olmasına rağmen birçok disiplinin kesiştiği geniş bir araştırma havuzu sunmaktadır. Tüm insanlığı etkileyen neden ve sonuçları konu çok daha cazip hale getirmektedir. Tez çalışması esnasında bir kez daha teyit ettiğimiz üzere konu birçok alt problem ve çalışma alanı da barındırmaktadır. Bu bakımdan konu alt alanlara bölünüp çalışılmalıdır. Sahte haberin otomatik tespiti de bu konular içerisinde gelişime ve bilgisayar merkezli sayısal çözümlere en açık olanıdır. Tez kapsamında Türkçe Haber kaynaklarında sahte haberin otomatik tespiti konusu çalışılmıştır.

Doğal Dil İşleme ile ilgili diğer problemlerde de sıklıkla yaşandığı gibi yenilikçi çözümlerin geliştirilen dille sınırlı kalması durumu dil kaynaklarını esas alan sahte haber tespitini konusunda da kendini göstermektedir. Bu nedenle Türkçe dili için de dile özgü çalışmaların bu alanda devam ettirilmesi gerekmektedir. Birçoğu sonuç bölümünde işlenen konu ile ilgili çok sayıda alt problemin bulunması, sayısal ve sosyal disiplinlerin açısından çalışılabilecek yeni konular içermesi nedeniyle araştırmacıların konu ile ilgili çalışmasının ve katkı sunmasının çok önemli olduğunu değerlendiriyoruz.

İnsanlığın günümüzde en çok ihtiyaç duyduğu konuların başında doğru bilgiye ulaşabilme ihtiyacı gelmektedir. Bireyler, toplumlar, organizasyonlar, kurumlar ve hatta devletler için açık tehditler içeren sahte haber daha genel adıyla bilgi kirliliği maalesef bir dijital silaha dönüşmüştür. Bu bakımdan sahte haber ile mücadele bir güvenlik problemi olarak değerlendirilebilir. Bu bağlamda alandaki farklı disiplinlerde yapılacak araştırma ve geliştirme çalışmaları ülkemizin ve toplumumuzun sahte haber ve bilgi kirliliğinden korunması için önemli katkılar sağlayacaktır. Sonuç olarak dijital kütüphanecilik ekseninde önerilen çatı çerçeve çalışmasının ülkemizde çok geniş perspektifli bir araştırma grubu ile başlatılabileceğini ve bu sayede dünyada bu alanda öncü olma fırsatı yakalanabileceğini öngörmekteyiz.

6. KAYNAKLAR

- [1] D.B. Buller, J.K. Burgoon, Interpersonal deception theory, *Communication theory*, 6(3) (1996) 203-242.
- [2] W. Weir, *History's Greatest Lies*. Beverly, Massachusetts, Fair Winds Press, 2009, p. 28-41.
- [3] S. Gupta, D.B. Skillicorn, Improving a textual deception detection model, *Proceedings of the 2006 conference of the Center for Advanced Studies on Collaborative research*, 2006, p. 29-es.
- [4] C. Fuller, D. Biros, D. Twitchell, J. Burgoon, M. Adkins, An analysis of text-based deception detection tools, *AMCIS 2006 Proceedings*, (2006) 418.
- [5] J.L. Egelhofer, S. Lecheler, Fake news as a two-dimensional phenomenon: a framework and research agenda, *Annals of the International Communication Association*, 43 (2019) 97-116.
- [6] H. Allcott, M. Gentzkow, Social media and fake news in the 2016 election, *Journal of economic perspectives*, 31 (2017) 211-236.
- [7] V. Bakir, A. McStay, Fake news and the economy of emotions: Problems, causes, solutions, *Digital journalism*, 6 (2018) 154-175.
- [8] D. DiFranzo, K. Gloria-Garcia, Filter bubbles and fake news, *XRDS: Crossroads, The ACM Magazine for Students*, 23 (2017) 32-35.
- [9] A. Guess, B. Nyhan, J. Reifler, Selective exposure to misinformation: Evidence from the consumption of fake news during the 2016 US presidential campaign, *European Research Council*, 9 (2018) 4.
- [10] B.D. Horne, S. Adali, This just in: fake news packs a lot in title, uses simpler, repetitive content in text body, more similar to satire than real news, *Eleventh International AAAI Conference on Web and Social Media*, 2017.
- [11] J.L. Nelson, H. Taneja, The small, disloyal fake news audience: The role of audience availability in fake news consumption, *New media & society*, 20 (2018) 3720-3737.
- [12] G. Pennycook, D. Rand, Assessing the effect of “disputed” warnings and source salience on perceptions of fake news accuracy, *Social Science Research Network*. <https://papers.ssrn.com/sol3/papers.cfm>, (2017).
- [13] E.C. Tandoc Jr, Z.W. Lim, R. Ling, Defining “fake news” A typology of scholarly definitions, *Digital journalism*, 6 (2018) 137-153.
- [14] D.M. Lazer, M.A. Baum, Y. Benkler, A.J. Berinsky, K.M. Greenhill, F. Menczer, M.J. Metzger, B. Nyhan, G. Pennycook, D. Rothschild, The science of fake news, *Science*, 359 (2018) 1094-1096.
- [15] K. Shu, A. Sliva, S. Wang, J. Tang, H. Liu, Fake news detection on social media: A data mining perspective, *ACM SIGKDD explorations newsletter*, 19 (2017) 22-36.
- [16] N. Ruchansky, S. Seo, Y. Liu, Csi: A hybrid deep model for fake news detection, *Proceedings of the 2017 ACM on Conference on Information and Knowledge Management*, 2017, p. 797-806.

- [17] J.S. Adams, *Internet Journalism and Fake News*, Cavendish Square Publishing, LLC, **2018**.
- [18] A. Taub, M. Fisher, Where countries are tinderboxes and Facebook is a match, *New York Times*, 21 (**2018**).
- [19] A. Douai, Global and Arab media in the post-truth era: Globalization, authoritarianism and fake news, *IEMed: Mediterranean Yearbook*, (**2019**) 124-132.
- [20] I. Khaldarova, M. Pantti, Fake news: The narrative battle over the Ukrainian conflict, *Journalism practice*, 10 (**2016**) 891-901.
- [21] K. Starbird, J. Maddock, M. Orand, P. Achterman, R.M. Mason, Rumors, false flags, and digital vigilantes: Misinformation on twitter after the 2013 boston marathon bombing, *IConference 2014 Proceedings*, (**2014**).
- [22] P. Barbera, Explaining the spread of misinformation on social media: evidence from the 2016 US presidential election, *Symposium: Fake News and the Politics of Misinformation. APSA*, **2018**.
- [23] J. Gillin, How Pizzagate went from fake news to a real problem for a DC business, *PolitiFact*, (**2016**).
- [24] V. Narayanan, P.N. Howard, B. Kollanyi, M. Elswah, Russian involvement and junk news during Brexit, *The computational propaganda project. Algorithms, automation and digital politics*. <https://comprop.oii.ox.ac.uk/research/working-papers/russia-and-brexit>, (**2017**).
- [25] E. Ferrara, Disinformation and social bot operations in the run up to the 2017 French presidential election, *arXiv preprint arXiv:1707.00086*, (**2017**).
- [26] C. Ireton, J. Posetti, *Journalism, fake news & disinformation: handbook for journalism education and training*, UNESCO Publishing, **2018**.
- [27] U. Mertoğlu, H. Sever, B. Genç, Savunmada Yenilikçi bir Dijital Dönüşüm Alanı: Sahte Haber Tespit Modeli, 9. Savunma Teknolojileri Kongresi, 27-29 Haziran, ODTÜ, Ankara, **2018**.
- [28] H. Rashkin, E. Choi, J.Y. Jang, S. Volkova, Y. Choi, Truth of varying shades: Analyzing language in fake news and political fact-checking, *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, **2017**, p. 2931-2937.
- [29] P. Meel, D.K. Vishwakarma, Fake news, rumor, information pollution in social media and web: A contemporary survey of state-of-the-arts, challenges and opportunities, *Expert Systems with Applications*, (**2019**) 112986.
- [30] A.K. Chaudhry, D. Baker, P. Thun-Hohenstein, Stance detection for the fake news challenge: identifying textual relationships with deep neural nets, *CS224n: Natural Language Processing with Deep Learning*, (**2017**).
- [31] R. Masood, A. Aker, *The Fake News Challenge: Stance Detection using Traditional Machine Learning Approaches*, *KMIS*, **2018**, p. 126-133.
- [32] E. Tacchini, G. Ballarin, M.L. Della Vedova, S. Moret, L. de Alfaro, Some like it hoax: Automated fake news detection in social networks, *arXiv preprint arXiv:1704.07506*, (**2017**).
- [33] S. Vosoughi, D. Roy, S. Aral, The spread of true and false news online, *Science*, 359 (**2018**) 1146-1151.

- [34] J. Mandalios, RADAR: An approach for helping students evaluate Internet sources, *Journal of Information Science*, 39 (2013) 470-478.
- [35] S. Blakeslee, The CRAAP test, *LOEX Quarterly*, 31 (2004) 4.
- [36] V.L. Rubin, N. Conroy, Discerning truth from deception: Human judgments and automation efforts, *First Monday*, 17 (2012).
- [37] V. Pérez-Rosas, B. Kleinberg, A. Lefevre, R. Mihalcea, Automatic detection of fake news, arXiv preprint arXiv:1708.07104, (2017).
- [38] J.P. Friesen, T.H. Campbell, A.C. Kay, The psychological advantage of unfalsifiability: The appeal of untestable religious and political ideologies, *Journal of personality and social psychology*, 108 (2015) 515.
- [39] T. Pavleska, A. Školkay, B. Zankova, N. Ribeiro, A. Bechmann, Performance analysis of fact-checking organizations and initiatives in Europe: a critical overview of online platforms fighting fake news, *Social media and convergence*, 29 (2018).
- [40] Y. Lahlou, S. El Fkihi, R. Faizi, Automatic detection of fake news on online platforms: A survey, 2019 1st International Conference on Smart Systems and Data Science (ICSSD), IEEE, 2019, p. 1-4.
- [41] C. Janze, M. Risius, Automatic Detection of Fake News on Social Media Platforms, *PACIS*, 2017, p. 261.
- [42] T. Granskogen, Automatic Detection of Fake News in Social Media using Contextual Information, *NTNU*, 2018.
- [43] L. Graves, F. Cherubini, The rise of fact-checking sites in Europe, 2016.
- [44] S. Zannettou, M. Sirivianos, J. Blackburn, N. Kourtellis, The web of false information: Rumors, fake news, hoaxes, clickbait, and various other shenanigans, *Journal of Data and Information Quality (JDIQ)*, 11 (2019) 1-37.
- [45] T. Gröndahl, N. Asokan, Text Analysis in Adversarial Settings: Does Deception Leave a Stylistic Trace?, *ACM Computing Surveys (CSUR)*, 52 (2019) 1-36.
- [46] N.J. Conroy, V.L. Rubin, Y. Chen, Automatic deception detection: Methods for finding fake news, *Proceedings of the Association for Information Science and Technology*, 52 (2015) 1-4.
- [47] V.L. Rubin, Deception detection and rumor debunking for social media, *The SAGE Handbook of Social Media Research Methods*, Sage, 2017, p. 342.
- [48] I. Inuwa-Dutse, M. Liptrott, I. Korkontzelos, Detection of spam-posting accounts on Twitter, *Neurocomputing*, 315 (2018) 496-511.
- [49] D.H. Fusilier, M. Montes-y-Gómez, P. Rosso, R.G. Cabrera, Detecting positive and negative deceptive opinions using PU-learning, *Information processing & management*, 51 (2015) 433-443.
- [50] D.M. Markowitz, J.T. Hancock, Linguistic obfuscation in fraudulent science, *Journal of Language and Social Psychology*, 35 (2016) 435-445.
- [51] S. Feng, R. Banerjee, Y. Choi, Syntactic stylometry for deception detection, *Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, 2012, p. 171-175.

- [52] H. Zhang, Z. Fan, J. Zheng, Q. Liu, An improving deception detection method in computer-mediated communication, *Journal of Networks*, 7 (2012) 1811.
- [53] M. Ott, C. Cardie, J.T. Hancock, Negative deceptive opinion spam, *Proceedings of the 2013 conference of the north american chapter of the association for computational linguistics: human language technologies*, 2013, p. 497-501.
- [54] V.L. Rubin, T. Lukoianova, Truth and deception at the rhetorical structure level, *Journal of the Association for Information Science and Technology*, 66 (2015) 905-917.
- [55] V.L. Rubin, N.J. Conroy, Y. Chen, Towards news verification: Deception detection methods for news discourse, *Hawaii International Conference on System Sciences*, 2015, p. 5-8.
- [56] S. Poria, E. Cambria, D. Hazarika, P. Vij, A deeper look into sarcastic tweets using deep convolutional neural networks, *arXiv preprint arXiv:1610.08815*, (2016).
- [57] D. Küçük, N. Arıcı, Doğal Dil İşlemede Derin Öğrenme Uygulamaları Üzerine bir Literatür Çalışması, *Uluslararası Yönetim Bilişim Sistemleri ve Bilgisayar Bilimleri Dergisi*, 2 (2018) 76-86.
- [58] J. Allen, *Natural Language Understanding*, Benjamin/Cummings Publishing Company, USA, 1995.
- [59] E.D. Liddy, *Natural Language Processing. Encyclopedia of Library and Information Science*, M. Drake (Eds), Vol.2, New York, 2001.
- [60] M.Z. Kurdi, *Natural language processing and computational linguistics 2: Semantics, Discourse and Applications*, John Wiley & Sons, New Jersey, USA, 2017.
- [61] S.E. Şeker, Doğal Dil İşleme (Natural Language Processing), *YBS Ansiklopedi*, 2 (2015) 14-22.
- [62] L. Doyuran, Medyatik Bir Çalışma Alanı Olarak Eleştirel Söylem Çözümlemesi (Televizyon Dizileri Örneğinde), *Erciyes İletişim Dergisi*, 5 (2018) 301-323.
- [63] E. Adalı, Doğal Dil İşleme, *Türkiye Bilişim Vakfı Bilgisayar Bilimleri ve Mühendisliği Dergisi*, 5 (2012) 1.
- [64] J. Ashok, S. Suppiah, J. VijiPriya, A review on significance of sub fields in artificial intelligence, *International Journal of Latest Trends in Engineering and Technology (IJLTET)*, 6 (2016) 542-548.
- [65] M. Bates, R.J. Bobrow, R.M. Weischedel, *Critical Challenges for Natural Language Processing*, Cambridge University Press, USA, 3-34, 1993.
- [66] S. Hellmann, S. Auer, Towards Web-Scale Collaborative Knowledge Extraction, *The People's Web Meets NLP*, Springer, Berlin, 2013, p. 287-313.
- [67] J. Hirschberg, C.D. Manning, *Advances in Natural Language Processing*, Science, 349 (2015) 261-266.
- [68] D. Jurafsky, *Speech & Language Processing*, Pearson Education India, 2000.
- [69] V.W. Feng, G. Hirst, Detecting deceptive opinions with profile compatibility, *Proceedings of the Sixth International Joint Conference on Natural Language Processing*, 2013, p. 338-346.

- [70] A.M. Braşoveanu, R. Andonie, Semantic fake news detection: a machine learning perspective, International Work-Conference on Artificial Neural Networks, Springer, **2019**, p. 656-667.
- [71] M. Tsikerdekis, S. Zeadally, Online deception in social media, Communications of the ACM, **57** (2014) 72-80.
- [72] J. Zhang, L. Cui, Y. Fu, F.B. Gouza, Fake news detection with deep diffusive network model, arXiv preprint arXiv:1805.08751, (2018).
- [73] K. Oflazer, Türkçe ve Doğal Dil İşleme, Türkiye Bilişim Vakfı Bilgisayar Bilimleri ve Mühendisliği Dergisi, **5** (2012).
- [74] K. Oflazer, M. Saraçlar, Turkish and its challenges for language and speech processing, Turkish Natural Language Processing, K. Oflazer, M. Saraçlar (Eds), Springer, Cham, Switzerland, **2018**, p. 1-19.
- [75] A.A. Akın, M.D. Akın, Zemberek, an open source nlp framework for turkic languages, Structure, **10** (2007) 1-5.
- [76] O. Bilgin, Ö. Çetinoğlu, K. Oflazer, Building a wordnet for Turkish, Romanian Journal of Information Science and Technology, **7** (2004) 163-172.
- [77] K. Oflazer, B. Say, D.Z. Hakkani-Tür, G. Tür, Building a Turkish Treebank, Invited chapter in Building and Exploiting Syntactically-annotated Corpora, Anne Abeille Editor, Kluwer Academic Publishers, **2003**.
- [78] M. Kutlu, I. Cicekli, A hybrid morphological disambiguation system for turkish, Proceedings of the Sixth International Joint Conference on Natural Language Processing, **2013**, p. 1230-1236.
- [79] H. Sak, T. Güngör, M. Saraçlar, Morphological disambiguation of Turkish text with perceptron algorithm, International Conference on Intelligent Text Processing and Computational Linguistics, Springer, **2007**, p. 107-118.
- [80] O. Bilgin, Frequency effects in the processing of morphologically complex Turkish words, Master's thesis, Boğaziçi University, İstanbul, **2016**.
- [81] E. Akyürek, E. Dayanık, D. Yuret, Morphological analysis using a sequence decoder, Transactions of the Association for Computational Linguistics, **7** (2019) 567-579.
- [82] K. Oflazer, Two-level description of Turkish morphology, Literary and linguistic computing, **9** (1994) 137-148.
- [83] H. Sak, T. Güngör, M. Saraçlar, Resources for Turkish morphological processing, Language resources and evaluation, **45** (2011) 249-261.
- [84] J. Hankamer, Morphological parsing and the lexicon, Lexical representation and process, MIT Press, **1989**, p. 392-408.
- [85] E. Arisoy, D. Can, S. Parlak, H. Sak, M. Saraçlar, Turkish broadcast news transcription and retrieval, IEEE Transactions on Audio, Speech, and Language Processing, **17** (2009) 874-883.
- [86] C. Burcu, LSTM Ağları ile Türkçe Kök Bulma, Bilişim Teknolojileri Dergisi, **12** (2019) 183-193.
- [87] E. Ünür, Geleneksel Medya'nın Sosyal Medyayı Nasıl Kullandığına İlişkin Bir İnceleme: Dizilerin Twitter Kullanımı, Turkish Journal of Social Research/Türkiye Sosyal Araştırmalar Dergisi, **20** (2016).

- [88] N. Nic, R. Fletcher, A. Kalogeropoulos, D.A. Levy, R.K. Nielsen, Reuters institute digital news report 2018, Reuters Institute for the Study of Journalism, **2018**.
- [89] M. Deuze, The web and its journalisms: considering the consequences of different types of newsmedia online, *New Media & Society*, 5 (2003) 203-230.
- [90] C.S. Sütçü, Değişen haber algısı: Kullanıcılar sosyal medyada haberleri nasıl değerlendiriyorlar?, *Global Media Journal: Turkish Edition*, 6 (2015).
- [91] S. Kattimani, P. Kumbargoudar, D.S. Gobbur, Training of the library professionals in digital era: key issues, (2006).
- [92] L.S. Connaway, H. Julien, M. Seadle, A. Kasprak, Digital literacy in the era of fake news: Key roles for information professionals, *Proceedings of the Association for Information Science and Technology*, 54 (2017) 554-555.
- [93] M. Sullivan, Libraries and Fake News: What's the Problem? What's the Plan?, *Communications in Information Literacy*, 13 (2019) 91-113.
- [94] W. Finley, B. McGowan, J. Kluever, Fake news: An opportunity for real librarianship, *ILA reporter*, 35 (2017) 8-12.
- [95] A.L. Association, Resolution on access to accurate information, Retrieved February, 10 (2017) 2018.
- [96] N. Rochlin, Fake news: belief in post-truth, *Library Hi Tech*, 35 (2017) 386-392.
- [97] C. Wardle, H. Derakhshan, Information disorder: Toward an interdisciplinary framework for research and policy making, Council of Europe report, 27 (2017).
- [98] A. Neely-Sardon, M. Tignor, Focus on the facts: A news and information literacy instructional program, *The Reference Librarian*, 59 (2018) 108-121.
- [99] L. Jacobson, The smell test: In the era of fake news, librarians are our best hope, *School Library Journal*, 63 (2017) 24-29.
- [100] A. Cruse, *Meaning in Language: An Introduction to Semantics and Pragmatics*, Oxford University Press UK, (2011).
- [101] V.L. Rubin, N.J. Conroy, Challenges in automated deception detection in computer-mediated communication, *Proceedings of the American Society for Information Science and Technology*, 48 (2011) 1-4.
- [102] J. Bachenko, E. Fitzpatrick, M. Schonwetter, Verification and implementation of language-based deception indicators in civil and criminal narratives, *Proceedings of the 22nd International Conference on Computational Linguistics (COLING 2008)*, 2008, p. 41-48.
- [103] V.L. Rubin, N. Conroy, Y. Chen, S. Cornwell, Fake news or truth? using satirical cues to detect potentially misleading news, *Proceedings of the second workshop on computational approaches to deception detection*, 2016, p. 7-17.
- [104] K. Leetaru, P.A. Schrodt, *Gdelt: Global data on events, location, and tone, 1979–2012*, ISA Annual Convention, Citeseer, San Francisco, California, USA, 2013, p. 1-49.
- [105] GDELT, <https://blog.gdeltproject.org/> (Erişim tarihi: **19 Ocak 2020**).
- [106] F. Sağlam, *Otomatik Duygu Sözlüğü Geliştirilmesi ve Haberlerin Duygu Analizi*, Doktora Tezi, Hacettepe Üniversitesi, Ankara, 2019.

- [107] M. Egele, G. Stringhini, C. Kruegel, G. Vigna, Towards detecting compromised accounts on social networks, *IEEE Transactions on Dependable and Secure Computing*, 14 (2015) 447-460.
- [108] K. Shu, D. Mahudeswaran, H. Liu, FakeNewsTracker: a tool for fake news collection, detection, and visualization, *Computational and Mathematical Organization Theory*, 25 (2019) 60-71.
- [109] M. AlRubaian, M. Al-Qurishi, M. Al-Rakhami, M.M. Hassan, A. Alamri, Credfinder: A real-time tweets credibility assessing system, 2016 *IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*, IEEE, 2016, p. 1406-1409.
- [110] N. Hassan, F. Arslan, C. Li, M. Tremayne, Toward automated fact-checking: Detecting check-worthy factual claims by ClaimBuster, *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2017, p. 1803-1812.
- [111] U. Mertoğlu, B. Genç, H. Sever, F. Sağlam, Auto-Tagging Model for Turkish News, *International Ankara Conference on Scientific Researches*, Ankara, 2019, p. 615-623.
- [112] X.-Y. Liu, J. Wu, Z.-H. Zhou, Exploratory undersampling for class-imbalance learning, *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, 39 (2008) 539-550.
- [113] G.K. Zipf, *Human behavior and the principle of least effort: An introduction to human ecology*, Ravenio Books, 2016.
- [114] Zipf yasası, https://tr.wikipedia.org/wiki/Zipf_yasası (Erişim tarihi: 7 Nisan 2020).
- [115] B. Copeland, *Artificial Intelligence: Definition, Examples, and Applications*, *Encyclopedia Britannica*, <https://www.britannica.com/technology/artificial-intelligence>, (Erişim tarihi: 20 Mayıs 2020).
- [116] P. Gambus, S.L. Shafer, *Artificial intelligence for everyone*, *Anesthesiology: The Journal of the American Society of Anesthesiologists*, 128 (2018) 431-433.
- [117] M. Mohri, A. Rostamizadeh, A. Talwalkar, *Foundations of Machine Learning*, MIT Press, London, England, 2018.
- [118] C.M. Bishop, *Pattern Recognition and Machine Learning*, Springer, New York, 2006.
- [119] E. Alpaydın, *Introduction to machine learning*, 4th ed., MIT Press, 2020.
- [120] F. Monti, F. Frasca, D. Eynard, D. Mannion, M.M. Bronstein, Fake news detection on social media using geometric deep learning, *arXiv preprint arXiv:1902.06673*, (2019).
- [121] F. Qian, C. Gong, K. Sharma, Y. Liu, Neural User Response Generator: Fake News Detection with Collective User Intelligence, *IJCAI*, 2018, p. 3834-3840.
- [122] Z. Jin, J. Cao, H. Guo, Y. Zhang, J. Luo, Multimodal fusion with recurrent neural networks for rumor detection on microblogs, *Proceedings of the 25th ACM international conference on Multimedia*, 2017, p. 795-816.
- [123] C. Guo, J. Cao, X. Zhang, K. Shu, M. Yu, Exploiting emotions for fake news detection on social media, *arXiv preprint arXiv:1903.01728*, (2019).

- [124] S. Esmaeilzadeh, G.X. Peh, A. Xu, Neural abstractive text summarization and fake news detection, arXiv preprint arXiv:1904.00788, **(2019)**.
- [125] L. Borges, B. Martins, P. Calado, Combining similarity features and deep representation learning for stance detection in the context of checking fake news, *Journal of Data and Information Quality (JDIQ)*, 11 **(2019)** 1-26.
- [126] B. Rath, W. Gao, J. Ma, J. Srivastava, From retweet to believability: Utilizing trust to identify rumor spreaders on twitter, *Proceedings of the 2017 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining 2017*, **2017**, p. 179-186.
- [127] A. Roy, K. Basak, A. Ekbal, P. Bhattacharyya, A deep ensemble framework for fake news detection and classification, arXiv preprint arXiv:1811.04670, **(2018)**.
- [128] S. Sadiq, N. Wagner, M.-L. Shyu, D. Feaster, High Dimensional Latent Space Variational AutoEncoders for Fake News Detection, *2019 IEEE Conference on Multimedia Information Processing and Retrieval (MIPR)*, IEEE, **2019**, p. 437-442.
- [129] C. Conforti, M.T. Pilehvar, N. Collier, Towards automatic fake news detection: cross-level stance detection in news articles, *Proceedings of the First Workshop on Fact Extraction and VERification (FEVER)*, **2018**, p. 40-49.
- [130] T. Mikolov, I. Sutskever, K. Chen, G.S. Corrado, J. Dean, Distributed representations of words and phrases and their compositionality, *Advances in neural information processing systems (NIPS 2013)*, **2013**, p. 3111-3119.
- [131] T. Mikolov, K. Chen, G. Corrado, J. Dean, Efficient estimation of word representations in vector space, arXiv preprint arXiv:1301.3781, **(2013)**.
- [132] T. Robinson, Disaster tweet classification using parts-of-speech tags: a domain adaptation approach, Master Thesis, Kansas State University, Manhattan Kansas, **2016**.
- [133] O. Kaynar, H. Arslan, Y. Görmez, Y.E. Işık, Makine Öğrenmesi ve Öznitelik Seçim Yöntemleriyle Saldırı Tespiti, *Bilişim Teknolojileri Dergisi*, 11 **(2018)** 175-185.
- [134] B. Bezirci, A.E. Yılmaz, Metinlerin okunabilirliğinin ölçülmesi üzerine bir yazılım kütüphanesi ve Türkçe için yeni bir okunabilirlik ölçütü, *Dokuz Eylül Üniversitesi Mühendislik Fakültesi Fen ve Mühendislik Dergisi*, 12 **(2010)** 49-62.
- [135] G. Robert, *The technique of clear writing*, (Revised Edition), New York: McGraw Hill, **(1968)**.
- [136] E. Ateşman, Türkçede okunabilirliğin ölçülmesi, *Dil Dergisi*, 58 **(1997)**.
- [137] E. Kartal, Z. Özen, Dengesiz Veri Setlerinde Sınıflandırma, *Mühendislikte Yapay Zeka ve Uygulamaları*, Sakarya Üniversitesi Kütüphanesi Yayınevi, Sakarya, **(2017)** 109-131.
- [138] N.V. Chawla, K.W. Bowyer, L.O. Hall, W.P. Kegelmeyer, SMOTE: synthetic minority over-sampling technique, *Journal of artificial intelligence research*, 16 **(2002)** 321-357.
- [139] G.C. Cawley, N.L. Talbot, On over-fitting in model selection and subsequent selection bias in performance evaluation, *The Journal of Machine Learning Research*, 11 **(2010)** 2079-2107.
- [140] L.E. Peterson, K-nearest Neighbor, *Scholarpedia*, 4 **(2009)** 1883.

- [141] S. Hatko, k-Nearest Neighbour Classification of Datasets with a Family of Distances, Department of Mathematics and Statistics, Master Thesis, Ottawa University, Canada, **2015**.
- [142] V. Vapnik, The nature of statistical learning theory, Springer science & business media, New York, **2013**.
- [143] V. Vapnik, Estimation of dependences based on empirical data, Springer Science & Business Media, New York, **2006**.
- [144] C.A. Mertler, R.V. Reinhart, Advanced and multivariate statistical methods: Practical application and interpretation, Taylor & Francis, **2016**.
- [145] J.F. Hair, W.C. Black, B.J. Babin, R.E. Anderson, R.L. Tatham, Multivariate data analysis 6th Edition, Pearson Prentice Hall. New Jersey. humans: Critique and reformulation. Journal of Abnormal Psychology, 87 (**2006**) 49-74.
- [146] Ö. Çokluk, Lojistik regresyon analizi: Kavram ve uygulama, Kuram ve Uygulamada Eğitim Bilimleri, 10 (**2010**) 1357-1407.
- [147] M. Kantardzic, Data mining: concepts, models, methods, and algorithms, John Wiley & Sons, **2011**.
- [148] C. Aydın, Makine Öğrenmesi Algoritmaları Kullanılarak İtfaiye İstasyonu İhtiyacının Sınıflandırılması, Avrupa Bilim ve Teknoloji Dergisi, 14 (**2018**) 169-175.
- [149] A. Şeker, B. Diri, H.H. Balık, Derin öğrenme yöntemleri ve uygulamaları hakkında bir inceleme, Gazi Mühendislik Bilimleri Dergisi, 3 (**2017**) 47-64.
- [150] O. Ajao, D. Bhowmik, S. Zargari, Fake news identification on twitter with hybrid cnn and rnn models, Proceedings of the 9th international conference on social media and society, **2018**, p. 226-230.
- [151] D. Khattar, J.S. Goud, M. Gupta, V. Varma, Mvae: Multimodal variational autoencoder for fake news detection, The World Wide Web Conference, **2019**, p. 2915-2921.
- [152] S. Girgis, E. Amer, M. Gadallah, Deep learning algorithms for detecting fake news in online text, 2018 13th International Conference on Computer Engineering and Systems (ICCES), IEEE, **2018**, p. 93-97.
- [153] R.S. Olson, W. La Cava, Z. Mustahsan, A. Varik, J.H. Moore, Data-driven advice for applying machine learning to bioinformatics problems, arXiv preprint arXiv:1708.05070, (**2017**).
- [154] J. Bergstra, Y. Bengio, Random search for hyper-parameter optimization, The Journal of Machine Learning Research, 13 (**2012**) 281-305.

EKLER

EK 1 – Veri seti oluşturulurken yararlanılan yaygın kullanıma sahip haber siteleri

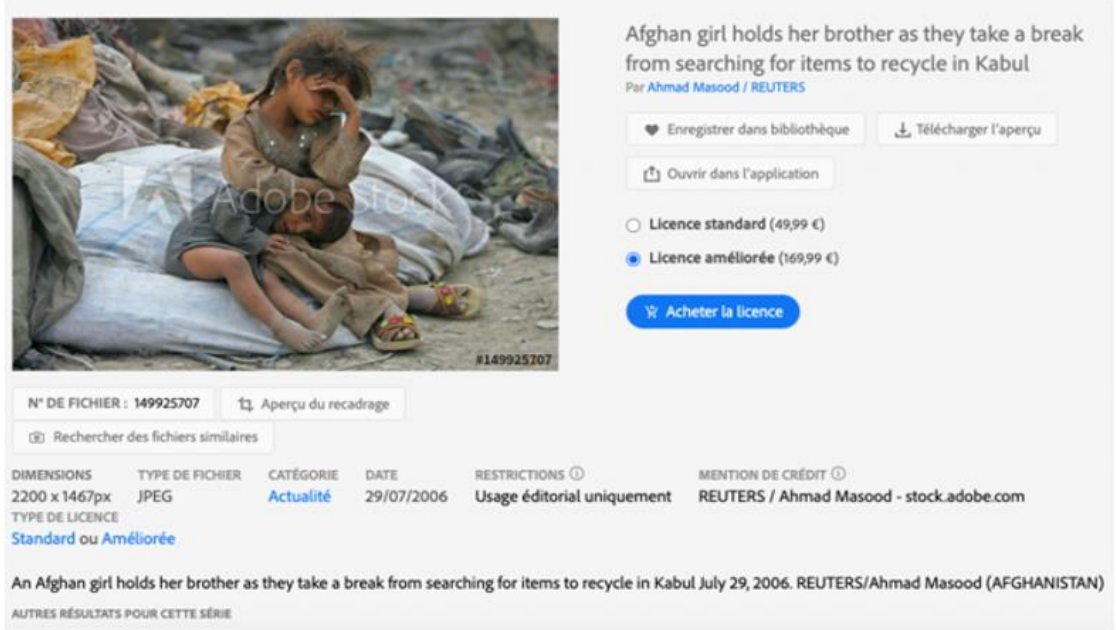
SNo	Kaynak Adresi (URL)
1	https://www.aksam.com.tr/
2	https://www.cumhuriyet.com.tr/
3	https://www.zaytung.com/ ²⁴
4	http://www.gazetevatan.com/
5	https://www.hurriyet.com.tr/
6	https://www.milliyet.com.tr/
7	https://www.haberturk.com/
8	https://www.ntv.com.tr/
9	https://www.posta.com.tr/
10	https://www.takvim.com.tr/
11	https://www.sozcu.com.tr/
12	https://www.star.com.tr/
13	https://www.sabah.com.tr/
14	https://www.cnnturk.com/
15	https://www.dogrula.org/ ²⁵ ,
16	https://www.aydinlik.com.tr/
17	https://www.yenicaggazetesi.com.tr/
18	https://gazeteyenigun.com.tr/
19	https://www.yeniasir.com.tr/
20	www.diken.com.tr/
21	https://www.takvim.com.tr/
22	https://www.turkgun.com/
23	https://www.haber365.com.tr/
24	https://onedio.com/
25	https://www.webtekno.com/
26	https://www.yeniakit.com.tr/
27	https://www.fotomac.com.tr/
28	https://www.ulusal.com.tr/
29	https://www.aspor.com.tr/
30	https://www.gazeteoku.com/
31	https://tr.sputniknews.com/
32	https://teyit.org/
33	https://www.ahaber.com.tr/
34	https://www.mynet.com/
35	https://www.yenisafak.com/
36	https://www.ensonhaber.com
37	https://t24.com.tr/

Yukarıdaki kaynaklardan haber teyiti, çapraz doğrulama, içerik kontrolü, veri seti oluşturma, benzer haber tespiti vb. amaçlı faydalanılmıştır. Bu kaynaklar dışında facebook, twitter paylaşımlarının yanı sıra sosyal medyada viral olan kaynağı belirsiz bazı haberlerde veri setinde yerini almıştır.

²⁴ www.zaytung.com alay ve hiciv amaçlı uydurma haberler yapan web sitesidir.

²⁵ <https://www.dogrula.org/> sitesi teyit.org benzeri haber analizleri yapan web sitesidir.

EK 2 – Görsellerle desteklenen teyitsiz sahte haber örneği (Örnek-1)



Afghan girl holds her brother as they take a break from searching for items to recycle in Kabul
Par Ahmad Masood / REUTERS

Enregistrer dans bibliothèque Télécharger l'aperçu

Ouvrir dans l'application

Licence standard (49,99 €)
Licence améliorée (169,99 €)

Acheter la licence

N° DE FICHIER : 149925707 Aperçu du recadrage

Rechercher des fichiers similaires

DIMENSIONS	TYPE DE FICHIER	CATÉGORIE	DATE	RESTRICTIONS	MENTION DE CRÉDIT
2200 x 1467px	JPEG	Actualité	29/07/2006	Usage éditorial uniquement	REUTERS / Ahmad Masood - stock.adobe.com

TYPE DE LICENCE
Standard ou Améliorée

An Afghan girl holds her brother as they take a break from searching for items to recycle in Kabul July 29, 2006. REUTERS/Ahmad Masood (AFGHANISTAN)

AUTRES RÉSULTATS POUR CETTE SÉRIE

a. Fotoğraf, Afganistan'ın Kabil şehrinde 2006 yılında çekilmiştir.

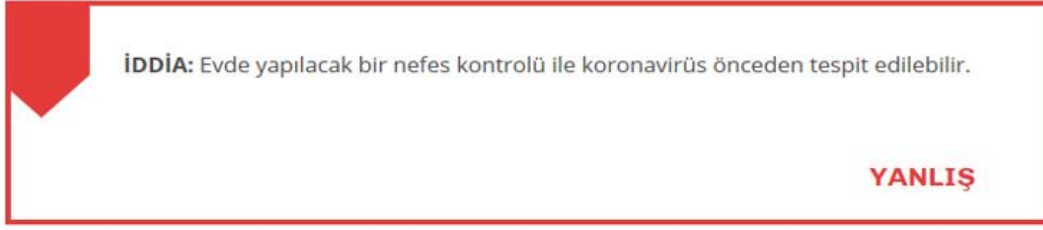


b. Yukarıdaki görselde, 1 ve 2 numaralı fotoğraflar Lübnan-Beyrut patlaması (2020) ile ilgilidir. Fakat 3 numaralı fotoğraf ilgisizdir. Sosyal medyada yoğun bir şekilde olayla ilişkilendirildiği görülmektedir.

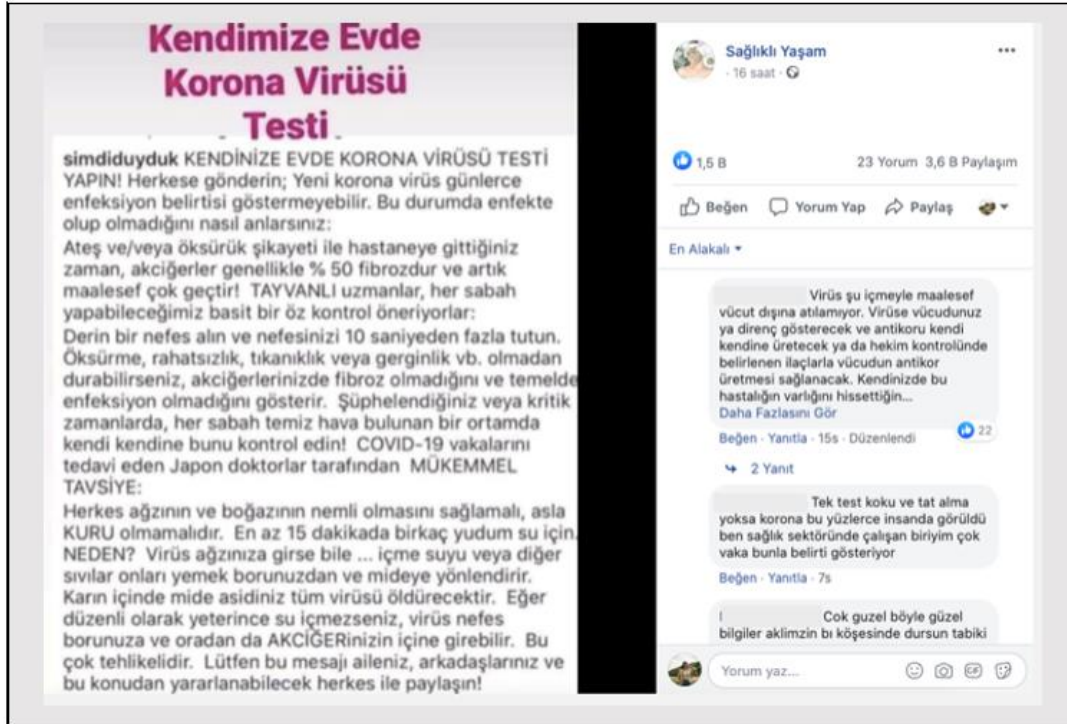
c. Fotoğraf, Yemen'deki çocukların yetersiz beslenmesi ve kıtlığı anlatan haber görsellerinde yerli ve yabancı yaygın ve kaynaklar tarafından kullanılmıştır.

EK-2'de benzer görsellerle desteklenen teyit edilmemiş haber ve paylaşım örnekleri görülmektedir. Sosyal medyada sıklıkla karşılaşılan bu tarz içeriklerin bazıları bilinçsizce oluşturulurken bazılarının ise propaganda, karalama maksadıyla bilinçli oluşturulduğu anlaşılmaktadır. Her iki durumda da toplumda karamsarlık yaratma etkisi bulunan bu tarz haber ve paylaşımlar özellikle örnekteki gibi kriz zamanlarında ortaya çıkmaktadır.

EK 3 – Sahte bilgiler içeren bilim haberleri ve paylaşımları (Örnek-2)



Covid-19 etkisini gösterdiği ilk zamanlardan beri [sosyal medyada](#), evde yapılan nefes kontrolü ile koronavirüsün önceden tespit edilebileceği iddia ediliyor. Bu iddialardan biri 9 Ağustos 2020'de Facebook'ta paylaşıldı ve yaklaşık 3 bin 300 kez beğenildi.



EK-3'te, teyit.org sitesi tarafından ayrıntılı analizi yapılarak "Yanlış" olarak etiketlenen bu görselde, COVID-19 salgını ile ilgili yapılan bir bilgi paylaşımı görülmektedir²⁶. Bu tarz paylaşımlar oldukça popüler olmakla birlikte genellikle zararsız görülmektedirler. Fakat özellikle sağlık ile ilgili yanlış uygulamaların geri döndürülemez etkileri olabileceği düşünüldüğünde art niyetli olmasa bile sahte bir haberlerin ciddiye alınması gerektiği bilinmelidir. Herhangi bir temele dayanmayan sosyal medya ve mesajlaşma uygulamalarında sık olarak karşılaşılmaktadır.

²⁶ <https://eski.teyit.org/nefesinizi-10-saniye-tutarak-koronavirus-testi-yapamazsiniz/> (Erişim tarihi: 20 Ağustos 2020)

EK 4 - Tezden TüretilmiŖ Yayınlar

U. Mertođlu, B. Genç, Automated Fake News Detection in the Age of Digital Libraries, Information Technologies and Libraries, **2020** (Accepted:14.08.2020, will be published: in December Issue).

EK 5 - Tezden Türetilmiş Bildiriler

U. Mertoğlu, H. Sever, B. Genç, Savunmada Yenilikçi bir Dijital Dönüşüm Alanı: Sahte Haber Tespit Modeli, SAVTEK 2018, 9. Savunma Teknolojileri Kongresi, 27-29 Haziran, ODTÜ, Ankara, **2018**.

U. Mertoğlu, B Genç, H Sever, F Sağlam, Auto-Tagging Model for Turkish News, International Ankara Conference on Scientific Researches, **2019**, p. 615-623.

U. Mertoğlu, B Genç, H Sever, Text-based fake news detection via machine learning, International Conference on Artificial Intelligence and Applied Mathematics in Engineering ICAIAME (2020), Antalya (will be held in October **2020**)