
GAZİ ÜNİVERSİTESİ
Gazi
1926

AÇIK KAYNAK İSTİHBARATI (OSINT) İÇİN TÜRKÇE İÇERİK TEMELLİ KİŞİLİK ÖZELLİKLERİ TAHMİNİ

Muhammed Ali KOŞAN

DOKTORA TEZİ
BİLGİSAYAR BİLİMLERİ ANA BİLİM DALI

GAZİ ÜNİVERSİTESİ
BİLİŞİM ENSTİTÜSÜ

OCAK 2023

Muhammed Ali KOŞAN tarafından hazırlanan "AÇIK KAYNAK İSTİHBARATI (OSINT) İÇİN TÜRKÇE İÇERİK TEMELLİ KİŞİLİK ÖZELLİKLERİ TAHMİNİ" adlı tez çalışması aşağıdaki jüri tarafından OY BİRLİĞİ ile Gazi Üniversitesi Bilgisayar Bilimleri Ana Bilim Dalında DOKTORA TEZİ olarak kabul edilmiştir.

Danışman: Prof. Dr. Hacer KARACAN

Bilgisayar Mühendisliği Ana Bilim Dalı, Gazi Üniversitesi

Bu tezin, kapsam ve kalite olarak Doktora Tezi olduğunu onaylıyorum. ……..……..

İkinci Danışman: Dr. Öğr. Üyesi Burcu Ayşen ÜRGEN

Psikoloji Ana Bilim Dalı, İhsan Doğramacı Bilkent Üniversitesi

Bu tezin, kapsam ve kalite olarak Doktora Tezi olduğunu onaylıyorum. ...……

Başkan: Prof. Dr. Osman Ayhan ERDEM

Bilgisayar Mühendisliği Ana Bilim Dalı, Gazi Üniversitesi

Bu tezin, kapsam ve kalite olarak Doktora Tezi olduğunu onaylıyorum. ...……

Üye: Prof. Dr. Muhammet Ali AKCAYOL

Bilgisayar Mühendisliği Ana Bilim Dalı, Gazi Üniversitesi

Bu tezin, kapsam ve kalite olarak Doktora Tezi olduğunu onaylıyorum. ...……

Üye: Prof. Dr. Pınar KARAGÖZ

Bilgisayar Mühendisliği Ana Bilim Dalı, Orta Doğu Teknik Üniversitesi

Bu tezin, kapsam ve kalite olarak Doktora Tezi olduğunu onaylıyorum. ...……

Üye: Prof. Dr. Hasan Şakir BİLGE

Elektrik-Elektronik Mühendisliği Ana Bilim Dalı, Gazi Üniversitesi

Bu tezin, kapsam ve kalite olarak Doktora Tezi olduğunu onaylıyorum. .....……

Üye: Doç. Dr. Hacer YALIM KELEŞ

Yapay Zeka Mühendisliği Ana Bilim Dalı, Hacettepe Üniversitesi

Bu tezin, kapsam ve kalite olarak Doktora Tezi olduğunu onaylıyorum. ...……

Tez Savunma Tarihi: 03/01/2023

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

……..……

Prof. Dr. Aslıhan TÜFEKCİ

Bilişim Enstitüsü Müdürü

ETİK BEYAN

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

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

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

Muhammed Ali KOŞAN

03/01/2023

AÇIK KAYNAK İSTİHBARATI (OSINT) İÇİN TÜRKÇE İÇERİK TEMELLİ KİŞİLİK ÖZELLİKLERİ TAHMİNİ

(Doktora Tezi)

Muhammed Ali KOŞAN

GAZİ ÜNİVERSİTESİ

BİLİŞİM ENSTİTÜSÜ

Ocak 2023

ÖZET

Bu çalışmada açık kaynak istihbaratı kullanılarak sosyal medya platformlarından elde edilen içerikler üzerinden kullanıcılara ait kişilik özellikleri tahmini üzerine araştırma yapılmıştır. Araştırmanın temel çatısını oluşturmak için; açık kaynak istihbaratının yaşam döngüsü sosyal medya platformlarındaki kullanıcıların kişilik özellikleri tahminine yönelik düzenlenerek uyarlanmıştır. Bu yaşam döngüsü kullanılarak iki eş zamanlı araştırma gerçekleştirilmiştir. İlk olarak 13 negatif-pozitif anlamlı kelimeler kullanılarak elde edilen Twitter içerikleri üzerinden İngilizce içeriklerden kişilik özellikleri tahmin çalışması gerçekleştirilmiştir. Bu çalışmada hem İngilizce kişilik özellikleri veriseti hem de dengeli ve yüksek başarımlı kişilik özellikleri tahmin modeli oluşturulmuştur. Oluşturulan tahmin modeli için kurulan sistemde anlamsal veri analizi kullanılarak içerikler sosyal medya platformu baz alınarak analiz edilmiştir. Analiz edilen veriler üzerinde anlamsal yapılar kullanılarak veriye özel ön işlem adımları ile dengeli ve genellenebilir kişilik özellikleri tahmin modeli ortaya konmuştur. İkinci olarak Türkçe içerikli kişilik özellikleri veriseti oluşturmak için belirli kişilik özellikleri testleri ile oyunlaştırılmış bir web uygulama yazılımı geliştirilmiştir. Bu yazılımın tanıtım çalışmaları ile birlikte veri toplama işlemi başlatılarak kullanıcıların hem kişilik özellikleri hem de sosyal medya içerikleri toplanmıştır. Toplanan veriler incelendiğinde BFI-10 kişilik özellik test sonuçları ile kullanılabilir Twitter hesap içeriklerinin önemli bir orana sahip olduğu görülmektedir. Bundan dolayı nihai veriseti BFI-10 testine ait kişilik özellikleri sonuçları ile Twitter hesaplarından elde edilen içeriklerden oluşturulmuştur. Elde edilen verisetinde ilk olarak İngilizce veriseti üzerinde elde edilen model uyarlanmıştır. Sonrasında anlamsal ön işlem yöntemleri, sayısallaştırma yöntemleri ve derin öğrenme modelleri ile kapsamlı bir analiz yapılmıştır. Sonuç olarak Türkçe dilinin karmaşık yapısına rağmen elde edilen sonuçların dengeli, genellenebilir ve kayda değer oranda iyi sonuçlar sunduğu gözlemlenmiştir.

Bilim Kodu	:	92432
Anahtar Kelimeler	:	Yapay zekâ, Kişilik özellikleri, Kişilik tahmini, OSINT
Sayfa Adedi	:	124
Danışman	:	Prof. Dr. Hacer KARACAN
İkinci Danışman	:	Dr. Öğr. Üyesi Burcu Ayşen ÜRGEN

PREDICTION OF TURKISH CONTENT-BASED PERSONALITY TRAITS FOR OPEN SOURCE INTELLIGENCE (OSINT)

(Ph. D. Thesis)

Muhammed Ali KOŞAN

GAZİ UNIVERSITY

INSTITUTE OF INFORMATICS

January 2023

ABSTRACT

This study was conducted on the prediction of personality traits of users through the content obtained from social media platforms using open source intelligence. To create the basic framework of the research; the life cycle of open source intelligence has been adapted to predict the personality traits of users on social media platforms. Two simultaneous studies were conducted using this lifecycle. First, personality traits were predicted from English contents on Twitter obtained using 13 negative-positive words. In the study, a dataset of English personality traits was created and then a balanced and high-performance personality trait prediction model was created. The content was analyzed based on the social media platform by using semantic data analysis in the system established for the prediction model. By using semantic structures on the analyzed data, a balanced and generalizable personality trait prediction model was developed with data-specific preprocessing steps. Secondly, a gamified web application software was developed with certain personality traits tests to create a personality traits dataset with Turkish content. Along with the promotional activities of this software, data collection process was started, and both the personality traits and social media contents of the users were collected. When the collected data is examined, it is seen that the BFI-10 personality trait test results and the available Twitter account content have a significant proportion. Therefore, the final dataset was created from the personality traits results of the BFI-10 test and the content obtained from Twitter accounts. In the obtained dataset, firstly, the model obtained on the English dataset was adapted. Afterwards, a comprehensive analysis was made with semantic preprocessing methods, vectorization methods and deep learning models. As a result, it has been observed that the results obtained despite the complex structure of the Turkish language offer balanced, generalizable, and remarkably good results.

Science Code	:	92432
Key Words	:	Artificial intelligence, Personality traits, Personality prediction, OSINT
Page Number	:	124
Supervisor	:	Prof. Dr. Hacer KARACAN
Co-Supervisor	:	Assist. Prof. Dr. Burcu Ayşen ÜRGEN

TEŞEKKÜR

Lisansüstü çalışmalarım süresince bilgi birikimi, tecrübeleri ve yardımseverliği ile her zaman yoluma ışık tutan, maddi ve manevi desteğini hiçbir zaman esirgemeyen danışman hocam Prof. Dr. Hacer KARACAN'a, doktora tez çalışmam süresince yönlendirmeleri ile çalışmama değer katan ikinci danışman hocam Dr. Öğr. Üyesi Burcu Ayşen ÜRGEN'e en içten teşekkürlerimi sunarım.

Doktora tez çalışmamın tüm süreci temel alındığında çalışmama olumlu katkılar sunarak çalışmayı doğru bir zemine oturtmak için değerli eleştirilerini her zaman paylaşan ve yön veren başta Prof. Dr. M. Ali AKCAYOL hocam olmak üzere Doç. Dr. Hacer YALIM KELEŞ, Doç. Dr. Cengiz ACARTÜRK ve Dr. Öğr. Üyesi Burcu CAN hocalarıma, çalışmanın ilk aşamasında fikir ve tecrübelerini esirgemeyen meslektaşım Dr. Iqra AMEER'e teşekkürlerimi sunarım.

İş ve akademik hayatımda desteklerini hiçbir zaman esirgemeyen Gazi Üniversitesi Bilişim Enstitüsündeki yönetim, akademik ve idari kadrolardaki tüm mesai arkadaşlarıma teşekkürlerimi sunarım.

Hayatım boyunca beni desteklemekten geri durmayan, aldığım kararlarda ve yaşadığım problemlerde sürekli yanımda duran aileme teşekkürlerimi sunarım.

Son olarak beni her zaman destekleyen, bu süreçte sabır gösteren, çalışmalarımda kıymetli fikirlerini esirgemeyen, moral ve motivasyonumun kaynağı sevgili eşim Dr. Meltem KOŞAN'a ve hayat enerjim, oğlum Ömer Toprak KOŞAN'a en içten teşekkürlerimi sunarım.

İÇİNDEKİLER

Sayfa

ÖZET iv

ABSTRACT v

TEŞEKKÜR vi

İÇİNDEKİLER vii

ÇİZELGELERİN LİSTESİ xi

ŞEKİLLERİN LİSTESİ xiv

SİMGELER VE KISALTMALAR xv

1. GİRİŞ 1

2. AÇIK KAYNAK İSTİHBARATI 7

2.1. Temel İstihbarat 7

2.1.1. Yön belirleme 8

2.1.2. Toplama 8

2.1.3. İşleme 8

2.1.4. Sunum 8

2.2. İstihbarat Elde Etme Yöntemleri 9

2.2.1. İnsan istihbaratı (HUMINT) 9

2.2.2. Sinyal istihbaratı (SIGINT) 10

2.2.3. Ölçüm ve imza istihbaratı (MASINT) 10

2.2.4. Görüntü istihbaratı (IMINT) 10

2.2.5. Mekansal istihbarat (GEOINT) 10

2.2.6. Açık kaynak istihbaratı (OSINT) 11

2.3. Tez Konusu Sınırını Belirlemede Yürütülen Literatür Çalışması 15

2.3.1. Aşama 1: OSINT literatür araştırması ve çalışma alanları 16

Sayfa

2.3.2. Aşama 2: OSINT alanında kullanıcı temelli sosyal medya analizi 19

2.3.3. Aşama 3: Sosyal medya platformlarından kişilik özellikleri analizi 21

2.4. OSINT ve Kişilik Özellikleri Tahmini 23

3. KİŞİLİK ÖZELLİKLERİ TAHMİNİ 25

3.1. Tez Çalışmaları 25

3.2. 2017 Yılı Sonrası Literatürde Yapılan Çalışmalar 27

3.3. Diğer Çalışmalar 41

3.4. Motivasyon ve Sınırların Belirlenmesi 42

3.4.1. Planlama ve yönlendirme 43

3.4.2. Güvenlik 44

3.4.3. Veri toplama 45

3.4.4. İşleme 45

3.4.5. Analiz ve üretim 46

3.4.6. Yaygınlaştırma ve değerlendirme (geri bildirim) 46

4. SEMANTİK YAPILAR VE LSTM TEMELLİ SİNİR AĞLARI İLE KİŞİLİK ÖZELLİKLERİ TAHMİNİ 49

4.1. Motivasyon ve Problem Tanımı 49

4.2. Materyal ve Metod 49

4.2.1. Kapsam, veri toplama ve veri etiketleme 50

4.2.2. Tweet analizi ve ön işlem 51

4.2.3. Metinsel veriden hesaplanabilir sayısal veriye 54

4.2.4. LSTM temelli modeller 54

4.2.5. Çalışma ortamı 57

4.2.6. PAN-2015 kişilik veri seti 57

4.2.7. PAN-2015-EN ile verisetimizin karşılaştırmalı analizi 57

Sayfa

4.3. Analiz ve Tartışma 58
4.4. Değerlendirme ve Sonuç 67
5. TÜRKÇE İÇERİKLERDEN KİŞİLİK ÖZELLİKLERİ TAHMİNİ 69
5.1. Motivasyon ve Problem Tanımı 70
5.2. Türkçe Kişilik Özellikleri Veri Seti 70
5.2.1. Kullanıcılardan istenen bilgiler ve kişilik testlerinin belirlenmesi 70
5.2.2. Etik değerlendirme 82
5.2.3. Kişilik özellikleri için veri toplama sistemi 82
5.2.4. Yaygınlaştırma ve veri toplama 85
5.2.5. Toplanan verilerin analizi 85
5.2.6. Nihai veri setinin oluşturulması 86
5.3. Tahmin Modeli 89
5.3.1. Ön işlem 90
5.3.2. Ayrıştırma ve vektörleştirme 92
5.3.3. Derin öğrenme yöntemleri 93
5.3.4. Çalışma ortamı 95
5.3.5. Diğer verisetleri 95
5.4. Deney ve Tartışma 96
5.5. Değerlendirme ve Sonuç 99
6. SONUÇ 101
KAYNAKLAR 105
EKLER 117
EK-1. Veri toplama web uygulama ekran görüntüleri 1 118
EK-2. Veri toplama web uygulama ekran görüntüleri 2 119

Sayfa

EK-3. Kişilik özelliklerine göre kelime/varlık yoğunluk haritaları - NLTK 120

EK-4. Kişilik özelliklerine göre kelime/varlık yoğunluk haritaları - Zemberek 121

ÖZGEÇMİŞ 122

ÇİZELGELERİN LİSTESİ

Çizelge **Sayfa**

Çizelge 1.1. İnternet üzerinde bilgi oluşturma istatistikleri (zaman bazlı toplam – ortalama değerlerdir) 2

Çizelge 1.2. İnternet üzerinde bilgi oluşturma istatistikleri (1 saniyede ortalama değerlerdir) 2

Çizelge 1.3. Bazı sosyal medya platformlarının aktif kullanıcı sayıları 2

Çizelge 2.1. Bazı OSINT araçları 15

Çizelge 2.2. İstihbarat elde etme yöntemleri arama sorguları 16

Çizelge 2.3. İstihbarat elde etme yöntemlerinin akademik araştırma platformları üzerinde bulunan araştırma sayıları 17

Çizelge 2.4. OSINT alanında WoS üzerinde bulunan teknik ve toplam çalışma sayıları ... 17

Çizelge 2.5. OSINT alanında SD üzerinde bulunan teknik ve toplam çalışma sayıları . 18

Çizelge 2.6. OSINT alanında IX üzerinde bulunan teknik ve toplam çalışma sayıları .. 18

Çizelge 2.7. Sosyal medya ve kullanıcı kelimeleri temelli sorgular 20

Çizelge 2.8. Sosyal medya ve kullanıcı kelime varyasyonları üzerine yapılan çalışma sayıları (araştırma ve literatür çalışma sayıları / toplam çalışma sayıları).. 20

Çizelge 2.9. Kişilik özelliğine dayalı sosyal medya çalışmaları için sorgu 21

Çizelge 2.10. Kişilik özelliğine dayalı sosyal medya çalışma sayıları 22

Çizelge 3.1. 2017 ve sonrasında kişilik özellikleri tahminine yönelik çalışmaların sosyal medya platformlarına göre sayıları 28

Çizelge 3.2. 2017 ve sonrasında kişilik özellikleri tahminine yönelik çalışmaların detaylı gösterimi 30

Çizelge 4.1. Çalışmada kullanılan ana kişilik özellikleri ve alt kişilik özellikleri 51

Çizelge 4.2. Modelimizin sayısal özellikleri 55

Çizelge 4.3. Veri setimiz ve PAN-2015-EN kişilik veri setinin sayısal karşılaştırması 58

Çizelge 4.4. LSTM tabanlı model ve Doğal Dil İşleme yöntemine göre sonuçlar (RMSE) 61

Çizelge **Sayfa**

Çizelge 4.5. Emoji ve ifade ön işleme türüne göre model sonuçları (RMSE) 61

Çizelge 4.6. Etiketlerin silinip silinmediğine göre model sonuçları (RMSE) 62

Çizelge 4.7. Sürekli veya ayrık tekrarlayan verilerin silinmesiyle ilgili model sonuçları (RMSE) 63

Çizelge 4.8. Kullanıcıların tweet sayısındaki seçime göre model sonuçları (RMSE) ... 64

Çizelge 4.9. Tweet uzunluğuna (T.U.) göre model sonuçları (RMSE) 64

Çizelge 4.10. Bi-LSTM modelinin vektörleştirme yöntemine göre veri setleri üzerinde karşılaştırılması (RMSE) 66

Çizelge 5.1. Kişilik testleri ve özellikleri 72

Çizelge 5.2. BFI-10 Kişilik Testi Maddeleri 73

Çizelge 5.3. TIPI Kişilik Testi Maddeleri 74

Çizelge 5.4. MINI-IPIP Kişilik Testi Maddeleri 76

Çizelge 5.5. BFI2-S Kişilik Testi Maddeleri 78

Çizelge 5.6. BFI2 Kişilik Testi Maddeleri 79

Çizelge 5.7. Kişilik özellikleri test sonuç ekranı gösterim simgeleri listesi 84

Çizelge 5.8. Toplanan veriler - hesap bilgileri 85

Çizelge 5.9. Toplanan veriler - çözülen kişilik testi sayıları 86

Çizelge 5.10. Toplanan veriler - testi çözen kullanıcıların sosyal medya hesabı sayıları 86

Çizelge 5.11. Türkçe ve tanımsız tweetlere göre tweet sayısı temelli kullanıcı sayıları 87

Çizelge 5.12. Türkçe ve tanımsız tweetlere göre tweet ve kelime/varlık sayıları..... 88

Çizelge 5.13. Kişilik özelliklerinin PTV<=0.5 ve 0.5<PTV koşullarına göre tweet dağılımları 88

Çizelge 5.14. Kişilik özellikleri ve PTV'ye göre varlık yoğunlukları 89

Çizelge 5.15. Ön işlem yöntemleri (PPM) – açıklama listesi 91

Çizelge 5.16. Model parametreleri..... 94

Çizelge **Sayfa**

Çizelge 5.17. Oluşturulan modele göre verisetlerinin karşılaştırılması (RMSE)..... 96

Çizelge 5.18. İlk model ve son modelin karşılaştırılması (RMSE) 97

Çizelge 5.19. Derin öğrenme modelleri ile elde edilen sonuçların karşılaştırması (RMSE) 97

Çizelge 5.20. Mevcut modele göre uygulanan kelime/cümle işlemleri ile elde edilen sonuçların karşılaştırması (RMSE) 98

Çizelge 5.21. Sabit ön işlem ve Bi-LSTM modelinde farklı vektörleştirme işlemleri ile elde edilen sonuçların karşılaştırması (RMSE)..... 99

ŞEKİLLERİN LİSTESİ

Şekil **Sayfa**

Şekil 1.1. Tez çalışması grafiksel özet 5

Şekil 2.1. İstihbarat yaşam döngüsü 7

Şekil 2.2. OSINT'in diğer istihbarat toplama yöntemleri ile örnek kesişim şeması..... 11

Şekil 2.3. RAND OSINT yaşam döngüsü 13

Şekil 2.4. OSINT yaşam döngüsünün içerik türlerine göre zorluk düzeyleri 14

Şekil 2.5. Kişilik Özelliğine dayalı sosyal medya çalışma sayılarının tarihsel dağılımı 22

Şekil 3.1. Sunulan OSINT yaşam döngüsü..... 43

Şekil 4.1. Çalışmanın araştırma yöntemi - iş akışı (Grafiksel özet) 50

Şekil 4.2. Veri kümesi için ön işleme akış şeması 52

Şekil 4.3. Bir tweet için önişleme adımları (Şekil 4.2'de tweet önişleme) 53

Şekil 4.4. Modelimizin yapısal gösterimi 56

Şekil 4.5. Veri setimiz ile PAN-2015-EN kişilik veri setinin kelime bulutu ile karşılaştırılması 58

Şekil 4.6. Etiketlerin silinip silinmediğine göre model sonuçları (5081-EN, RMSE).... 62

Şekil 4.7. Veri setindeki tweet numaralarının tweet uzunluğuna göre dağılımı (5081-EN)..... 65

Şekil 4.8. Veri setindeki tweet numaralarının tweet uzunluğuna göre dağılımı (PAN-2015-EN) 66

Şekil 5.1. Araştırma metodolojisi 69

Şekil 5.2. Türkçe ve tanımsız tweetlere göre tweet sayısı temelli kullanıcı sayıları 87

Şekil 5.3. Derin öğrenme modeli 89

Şekil 5.4. Ön işlem adımları – birlikte kullanımlarda izlenmesi gereken işlem sırası gösterimi 92

Şekil 5.5. Önerilen derin öğrenme mimarisi şablonu ve kullanılan derin öğrenme gizli katmanları 94

SİMGELER VE KISALTMALAR

Bu çalışmada kullanılmış simgeler ve kısaltmalar, açıklamaları ile birlikte aşağıda sunulmuştur.

Simgeler	**Açıklamalar**
MHz	Megahertz

Kısaltmalar	**Açıklamalar**
1D	Bir Boyutlu (One-Dimensional)
A.O.	Aritmetik Ortalama
API	Uygulama Programlama Arayüzü (Application Programming Interface)
BERT	Transformatörlerden Çift Yönlü Kodlayıcı Temsilleri (Bidirectional Encoder Representations from Transformers)
BFI	Beş Faktör Kişilik Envanteri (Big Five Inventory)
BKCI-S	Book Citation Index– Science
BKT	Boşluk Karakterlerini Temizleme
BOW	Kelime Çantası (Bag-of-words)
Bi	Çift yönlü (Bidirectional)
CE	Cevap (Reply) Etiketleme
CIA	Merkezi İstihbarat Teşkilatı (Central Intelligence Agency)
CNN	Evrişimsel Sinir Ağı (Convolutional Neural Network)
COMINT	İletişim İstihbaratı (Communication Intelligence)
CPCI-S	Conference Proceedings Citation Index- Science
CTI	Siber Tehdit İstihbaratı (Cyber Threat Intelligence)
CYINT	Siber İstihbarat (Cyber Intelligence)
ÇBT	Çoklu Boşlukları Temizleme
D	Dışadönük (Extraversion)

Kısaltmalar	Açıklamalar
DA	Deneyime Açık (Openness)
DISC	Dominant, İz Bırakan, Sadık, Ciddi Kişilik Envanteri (Dominance, Influence, Steadiness, Conscientiousness Inventory)
DNN	Derin Sinir Ağı (Deep Neural Network)
DvD	Doğrulama ve Değerlendirme
EE	E-posta Etiketleme
EEB	Emoji ve Emoticonlar (İfadeler) Arası Boşluk Ekleme
EK	Etiket Kelimeleri
ELINT	Elektronik İstihbarat (Electronic Intelligence)
EMJ	Emoji Etiketleme
EMT	Emoticon (İfade) Etiketleme
ESCI	Emerging Sources Citation Index
Eİ	Emoji ve İfade (Emoticon)
FISINT	Yabancı Enstrümantasyon Sinyalleri İstihbaratı (Foreign Instrumentation Signals Intelligence)
FPR	False-Positive Rate
GB	Gigabyte
GBR	Gradyan Arttırma Regresyon (Gradient Boosting Regression)
GEOINT	Mekansal İstihbarat (Geospatial Intelligence)
GPU	Graphics Processing Unit
GRU	Geçitli Tekrarlayan Birim (Gated Recurrent Unit)
HDS	Hogan Gelişim Envanteri (Hogan Development Survey)
HE	Hashtag Etiketleme
HUMINT	İnsan İstihbaratı (Human Intelligence)
IMINT	Görüntü İstihbaratı (Imagery Intelligence)
IPIP	Uluslararası Kişilik Madde Havuzu (International Personality Item Pool)
IX	IEEE (Institute of Electrical and Electronics Engineers) Xplore

Kısaltmalar	Açıklamalar
İD	İçerik Dili
İT	İçerik Türü
KC	Kelime/Cümle Dil İşleme Yöntemleri
KNN	K-En Yakın Yokmu (K-Nearest Neighbors)
KT	Kişilik Testi
L-SVR	Doğrusal Dektek Vektör Regresyon (Linear Support Vector Regression)
LASSO	Least Absolute Shrinkage and Selection Operator
LEM	Kelime Kökleştirme – Lemmatization
LIWC	Dilbilimsel Sorgulama ve Kelime Sayımı (Linguistic Inquiry and Word Count)
LR	Lojistik Regresyon (Logistic Regression)
LSA	Latent Semantic Analysis
LSTM	Uzun-Kısa Süreli Bellek (Long-Short Term Memory)
MAE	Ortalama Mutlak Hata (Mean Absolute Error)
MASINT	Ölçüm ve İmza İstihbaratı (Measurement and Signature Intelligence)
MBTI	Myers-Briggs Kişilik Göstergesi (Myers-Briggs Type Indicator)
ME	Mention Etiketleme
MLP	Çok Katmanlı Algılayıcı (Multi-Layer Perceptron)
MSE	Ortalama Kare Hata (Mean Squared Error)
MSLE	Ortalama Kare Logaritmik Hata (Mean Squared Logarithmic Error)
MTD	Metodlar
Mini-IPIP	Mini Uluslararası Kişilik Ölçeği Havuzu Envanteri (Mini International Personality Items Pool)
N	Nevrotik (Neuroticism)
NEO-PI-R	Revize Edilmiş NEO Kişilik Envanteri (The Revised NEO (Neuroticism, Extraversion, Openness) Personality Inventory)
NLTK	Doğal Dil Araç Seti (Natural Language Toolkit)

Kısaltmalar	Açıklamalar
NRM	Normalizasyon
NT	Nesneleri Temizle
NÖK	NLTK – Önemsiz Kelimeleri Temizleme
NİT	Noktalama İşaretlerini Temizleme
OLS	Sıradan En Küçük Kareler (Ordinary Least Squares)
OSINT	Açık Kaynak İstihbaratı (Open Source Intelligence)
OWASP	Open Web Application Security Project
ÖİY	Ön İşlem Yöntemleri
ÖS	Özdenetim Sahibi (Conscientiousness)
P-SVR	Polinom Dektek Vektör Regresyon (Polynomial Support Vector Regression)
PCA	Temel Bileşen Analizi (Principal Component Analysis)
PMC	Sözde Çok Görüntülü Ortak Eğitim (Pseudo Multi-view Co-training)
POS	PoS (Metin Parçası, Part-of-Speech) Etiketleme
PPM	Ön İşlem Yöntemleri (Preprocessing Methods)
PTV	Kişilik Özelliği Değeri (Personality Trait Value)
RBF-SVR	Radyal Temel Fonksiyon - Dektek Vektör Regresyon (Radial Basis Function - Support Vector Regression)
RE	Resim Etiketleme
RET	Retweet Etiketleme
RF	Rastgele Orman (Random Forest)
RMSE	Kök Ortalama Kare Hata (Root Mean Square Error)
RMTL	Robust Multi-Task Learning
RNN	Tekrarlayan Sinir Ağı (Recurrent Neural Network)
SCI-E	Science Citation Index Expanded
SD	Science Direct
SIGINT	Sinyal İstihbaratı (Signal Intelligence)
SL	Etiketleri/Kelimeleri Sil
SLR	Basit Lojistik Regresyon (Simple Logistic Regression)

Kısaltmalar	Açıklamalar
SMP	Sosyal Medya Platformu
SOA	Second Order Attributes
SOCMINT	Sosyal Medya İstihbaratı (Social Media Intelligence)
SSD	Solid-State Drive
ST	Sayıları Temizleme
STM	Kelime Kökleştirme – Stemming
SVD	Singular Value Decomposition
SVM	Destek Vektör Makinesi (Support Vector Machine)
SVR	Dektek Vektör Regresyon (Support Vector Regression)
TAT	Tırnak/Alıntı İşaretlerini Temizleme
TF-IDF	Terim Sıklığı - Ters Doküman Sıklığı (Term Frequency – Inverse Document Frequency)
TIPI	On Maddeli Kişilik Envanteri (Ten Items Personality Inventory)
TPR	True-Positive Rate
TR/Tr	Türkçe
UE	URL Etiketleme
UND/Und	Tanımsız
URL	Uniform Resource Locator
VSD	Veri Seti Detayı
WEKA	Waikato Environment for Knowledge Analysis
WoS	Web of Science
YB	Yumuşak Başlı (Agreeableness)
YSE	Yeni Satır (\n) Etiketleme
YÖK	Yükseköğretim Kurulu
ZÖK	Zemberek – Önemsiz Kelimeleri Temizleme

1. GİRİŞ

İstihbarat tanımsal olarak; yeni toplanan veya öğrenilen bilgi, haber ve duyumların herhangi bir amaca hizmet etmek için işlenerek sunulmasıdır [1-3]. Geçmişten günümüze genel olarak askeri ve politik amaçlar için kullanılsa da günümüzde reklam ajanslarından şirketlere devletlerden örgütlere ve daha da özel alana girdiğimizde bireysel sebepler gibi birçok farklı alanda kullanılmaktadır. İstihbarat için bilginin elde edileceği bir kaynağın olması gerekmektedir. Internet'in yaygınlaşması ve teknolojinin gelişmesi ile üretilen verilerdeki artış, istihbarat çalışmalarına yeni bir boyut kazandırmıştır. Böylelikle artan veri ile birlikte bu verileri anlamlı bilgiye dönüştürerek belirli bir amaca hizmet etmek için kullanmanın diğer istihbarat faaliyetlerden daha az riskli ve daha kısa zamanlarda sonuca ulaştırdığı düşünülmektedir. Çünkü teknoloji temelsiz bir istihbarat faaliyetinde bilgiye erişim hem riskli hem de uzun süreler alabilmektedir. Sonucunda herhangi bir bilgiye ulaşamama durumu dikkate alındığında teknolojik temelli istihbarat ve özellikle Internet temelli istihbarat daha az risklidir. Internet üzerinden elde edilen bilgilerden herhangi bir anlamlı bilgi çıkarılmadığı müddetçe tekrar bilginin aranması ve ulaşılması daha olağandır ve daha hızlı gerçekleşir.

Bahsedilen ve bahsedilmeyen bir sürü avantajının yanında Internet, bize en çok elde edilecek bilginin boyutu ve çeşitliliğinde fayda sağlamaktadır. Dünyadaki Internet kullanıcı sayısı 2019 yılı sonunda 4.4 milyarı geçmekteydi ve 2021 yılı ilk yarısında 4.9 milyarı buldu. 2022 yılı Kasım ayında sunulan Internet kullanıcısı sayısı ise 5.5 milyardır. Devamında, 2019 yılında üretilen Internet trafiği 2.5 Zettabyte civarındayken 2021 yılı ilk yarısında 1.8 Zettabyte trafik üretilmiştir. Bunun yanında 2022 Kasım ayında bakılan yıllık Internet trafiği ise 4.5 Zettabyte'dır. Bununla birlikte 2019 yılı sonunda saniyede ortalama olarak 84 Terabyte trafik üretilmekteyken 2021 yılı ilk yarısında 100 Terabyte trafik üretilmiştir. Burada 2019 için yılın tamamı baz alınırken 2021 yılında ilk yarı temel alınmıştır. 2022 yılı Kasım ayında ise saniyede ortalama Internet trafiği 168 Terabyte'dır. Bu oranlar yıl bazına vurulduğunda genel olarak etkili bir değişim görülmektedir. Aynı zamanda web site sayısı 2 milyara yaklaşmaktadır. Bu bilgiler genel içeriği kapsasa da kişisel veya kurumsal bilgilerin paylaşıldığı blog içerikleri ile birlikte sosyal medya içeriklerindeki üretim oranları Çizelge 1.1 ve Çizelge 1.2'de sunulmaktadır. Görülen bilgiler kişisel veya kurumsal içerikleride kapsadığı için erişilebilir bilginin boyutlarını daha iyi ortaya koymaktadır. [4]

Çizelge 1.1. İnternet üzerinde bilgi oluşturma istatistikleri (zaman bazlı toplam - ortalama değerlerdir) [4]

İçerik	2019 Yılı	2021 Yılı İlk Yarı	2022 Yılı (Aralık Ayı Hariç)
E-posta Gönderimi	93 Trilyon	49.4 Trilyon	88 Trilyon
Google Araması	2.5 Trilyon	1.45 Trilyon	2.84 Trilyon
Blog Gönderisi	2.4 Milyar	1.40 Milyar	2.8 Milyar
Twitter Gönderisi (Tweet)	270 Milyar	149 Milyar	272 Milyar
Youtube Video İzleme	2.6 Trilyon	1.41 Trilyon	2.6 Trilyon
Instagram Resim Gönderimi	30 Milyar	16.9 Milyar	32 Milyar
Tumblr Gönderisi	51 Milyar	29.8 Milyar	58 Milyar

Çizelge 1.2. İnternet üzerinde bilgi oluşturma istatistikleri (1 saniyede ortalama değerlerdir) [4]

İçerik	2019 Yıl Sonu	2021 Yılı İlk Yarı Sonu	2022 Yılı Kasım Ayı
E-posta Gönderimi	2.8 Milyon	3 Milyon	3.2 Milyon
Google Araması	79 Bin	92.6 Bin	107.4 Bin
Twitter Gönderisi (Tweet)	8.7 Bin	9.5 Bin	10.3 Bin
Youtube Video İzleme	81 Bin	90 Bin	98.5 Bin
Instagram Resim Gönderimi	9 Yüz	1.07 Bin	1.2 Bin
Tumblr Gönderisi	1.6 Bin	1.9 Bin	2.2 Bin

Ayrıca zaman aralığına göre bazı sosyal medya platformlarının aktif kullanıcı sayıları Çizelge 1.3'de sunulmaktadır.

Çizelge 1.3. Bazı sosyal medya platformlarının aktif kullanıcı sayıları [4]

Sosyal Medya Platformu	2019 Yıl Sonu	2021 Yılı İlk Yarısı	2022 Yılı Kasım Ayı
Facebook	2.4 Milyar	2.84 Milyar	3.33 Milyar
Twitter	354 Milyon	375 Milyon	395 Milyon
Pinterest	267 Milyon	372 Milyon	497 Milyon

Sunulan içerik üretim miktarları ile aktif kullanıcı sayıları temel alındığında oluşan bilgi miktarı devasa boyutlara ulaşmaktadır.

Internet haricinde teknolojik temelli farklı istihbarat içerik ortamları bulunmaktadır. Bunların bazıları sinyaller, uydu verileri, resimler, kişisel bilgisayarlar, mobil aygıtlar, akıllı

nesneler ve robotlar olarak sıralanabilir. Her bir bilgi ortamına göre istihbarat toplamak için farklı bilgi ve yöntemler uygulanmaktadır. Fakat bunların hepsinde istihbaratın temel yaşam döngüsü işletilmektedir.

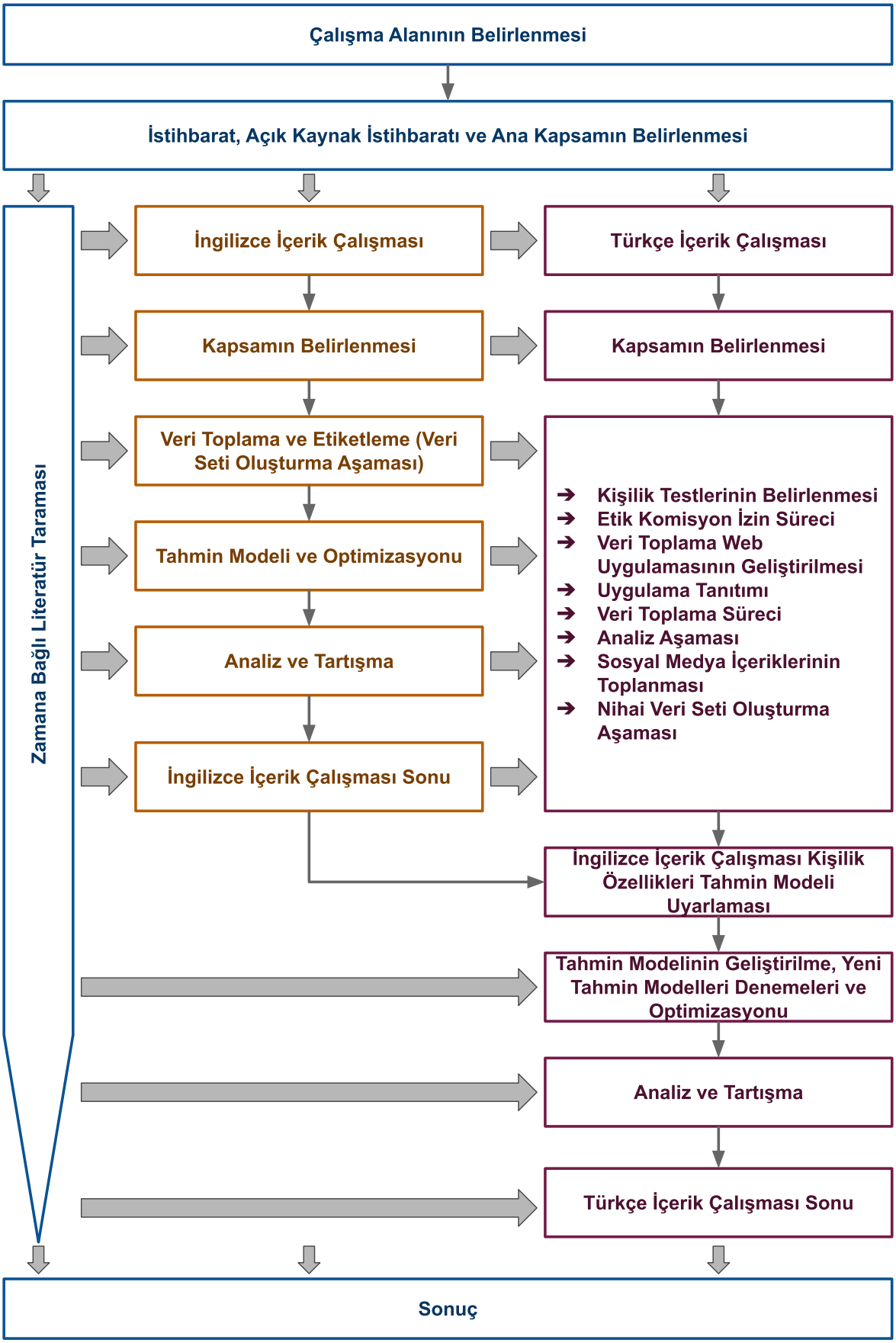
Günümüzde sosyal medya hesaplarının oluşturduğu bilgi boyutu hesaba katıldığında bu bilgilerin insanlar tarafından temel okuma ve değerlendirme yolu ile incelenmesinin mümkün olmadığı görülmektedir. Ayrıca oluşan bilginin değeri düşünüldüğünde göz ardı edilmesi de doğru bir tercih olmamaktadır. Kişisel bilgilerin istihbarat bilgisi olarak kullanılması her alanda mümkün olsa da hesaplanması en zor ve en önemli ayırt edici özelliğin kullanıcıların kişiliği olduğunu söyleyebiliriz. Bu noktada, her sosyal medya kullanıcısının kendi gerçek kişiliğine uygun içerik üretmediği kabul edildiğinden, kullanıcı kişiliği terimi daha uygun görünmektedir. Sonuç olarak paylaşımlar, kişiye özel üretilmiş bir karaktermiş gibi gözetilmelidir. Bu nedenle sosyal medya platformlarından kişilik tahmini sadece kullanıcı kişiliği hakkında bilgi verebilir. Bu tahmin, klinik bir ortamda elde edilen bilgiler kadar kesin olmasa da, belirli bir örneklem üzerinde öngörülebilirlik çalışmaları yapılabilir. Ayrıca bir kullanıcı birden fazla kişilik özelliğini birden yüksek düzeyde sağlayabilirken bazı kişilik özelliklerinde tam tersi etkiye sahip olabilir. Bu noktada Dışadönük bir bireyin Nevrotiklik düzeyinin genel olarak düşük olduğu söylenebilir. Fakat bu tam bir gerçek olarak algılanamaz. Tüm bu analizler ile birlikte sosyal medya içeriğinden kişilik özelliklerinin çıkarılması ve tahmini; çevrimiçi pazarlamada, işe alım süreçlerinde, öneri sistemlerinde, reklam ajanslarında, suç ve istihbarat profillemesinde vb. kullanılabilir [5].

Bu tez çalışmasında 2. bölümde tez çalışmasında ana çatıyı oluşturan Açık Kaynak İstihbaratı (Open Source Intelligence - OSINT) ayrıntılı olarak sunulmaktadır. Tez çalışmasının başlangıcında temel istihbarat süreci temel alınarak OSINT'in süreç içindeki konumu ve literatürdeki çalışma alanları analiz edilmiştir. Tez çalışmasının başlangıç araştırması olduğu için belirli bir kapsam üzerinden değerlendirilen Açık Kaynak İstihbaratı başlığı devamında sunulan ana araştırma alanımız olan ve 3. Bölümde sunulan kişilik özellikleri tahmini literatür araştırması ile birlikte açık kaynak istihbarat sürecinde tez çalışmasının motivasyonu sunulmaktadır. Açık Kaynak İstihbaratı bölümünün aksine Kişilik Özellikleri Tahmini bölümü tez çalışması boyunca güncel literatür takibi yapılarak oluşturulmuş ve bu alandaki çalışmalara kapsamlı bir bakış açısı sunmaktadır. Devamında 4. bölümde İngilizce kişilik özellikleri veri seti oluşturma çalışması ve İngilizce içeriklerden

kişilik özellikleri tahmin modeli çalışması verilmektedir. İngilizce içerikli çalışmanın ana amacı Türkçe içerikli çalışma için planlanan sistem tasarımının ön bir denemesini yapmak olmuştur. Böylelikle hem İngilizce Twitter içeriklerinden elde edilen ve IBM Personality Insight ile etiketlenmiş kişilik özellikleri veri seti ve İngilizce içeriklerden kişilik özellikleri tahmin model oluşturulmuştur. Tez çalışmasının nihai amacı olan 5. bölümde Türkçe kişilik özellikleri veri seti oluşturma çalışması ve kişilik özellikleri tahmin modeli çalışması sunulmaktadır. Bu bölüm içerisinde kişilik özellikleri testlerinin belirlenmesi, etik komisyon izin süreci, veri toplama amacı ile oluşturulan web uygulaması, ilgili uygulamanın yaygınlaştırılması/tanıtılması, veri toplama süreci, toplanan verilerin analizi, sosyal medya verilerinin toplanması, nihai veri seti oluşturma aşaması, 4. bölümde oluşturulan tahmin modeli entegrasyonu, tahmin modeli optimizasyonu ve değerlendirme çalışmaları sunulmaktadır. Son olarak 6. bölümde tez çalışmasının sonucu ve gelecek yönlendirmeleri anlatılmaktadır. Tüm tez çalışması sürecinin genel gösterimi Şekil 1.1'de bağlantılı olarak verilmektedir.

Bu tez çalışması bünyesinde yürütülen alt çalışmaların sunduğu yenilik ve katkılar aşağıda sıralanmaktadır (Genel: Tüm Çalışmalar, İngilizce: İngilizce veri seti üzerinde yapılan çalışma, Türkçe: Türkçe veri seti üzerinde yapılan çalışma):

- Genel: Metinsel verilerin yapısal analizinde sunulan farklı ön işleme adımlarının detaylı değerlendirmesi yapılmıştır.
- Genel: Kişilik özelliklerini tahmin etmede literatürün kapsamlı ve kategorik analizi yapılmıştır.
- Genel: Kişilik özelliklerini tahmin etmede Twitter öznelinde sosyal medya platformlarına ait verilerin yapısal analizinin önemi ortaya konmuştur.
- İngilizce: Genellemeye uygun Twitter içeriklerine dayalı kişilik özellikleri veri seti oluşturulmuştur.
- İngilizce: Yapısal ön işleme adımlarıyla kişilik özellikleri tahmini için Dengeli Bi-LSTM modeli sunulmuştur.
- Türkçe: Türkçe içeriklerden kişilik özellikleri tahmininde karşılaştırmalı ve yapısal analiz sunulmuştur.
- Türkçe: Türkçe içeriklerden kişilik özellikleri tahmininde sosyal medya platform verilerinin Twitter öznelinde yapısal analizinin önemi ortaya konmuştur.
- Türkçe: Türkçe içerikli metinlere ait kişilik özellikleri veri seti oluşturulmuştur.
- Türkçe: Yapısal ön işlem adımları ile genelleştirilmiş derin öğrenme modelleri sunulmuştur.



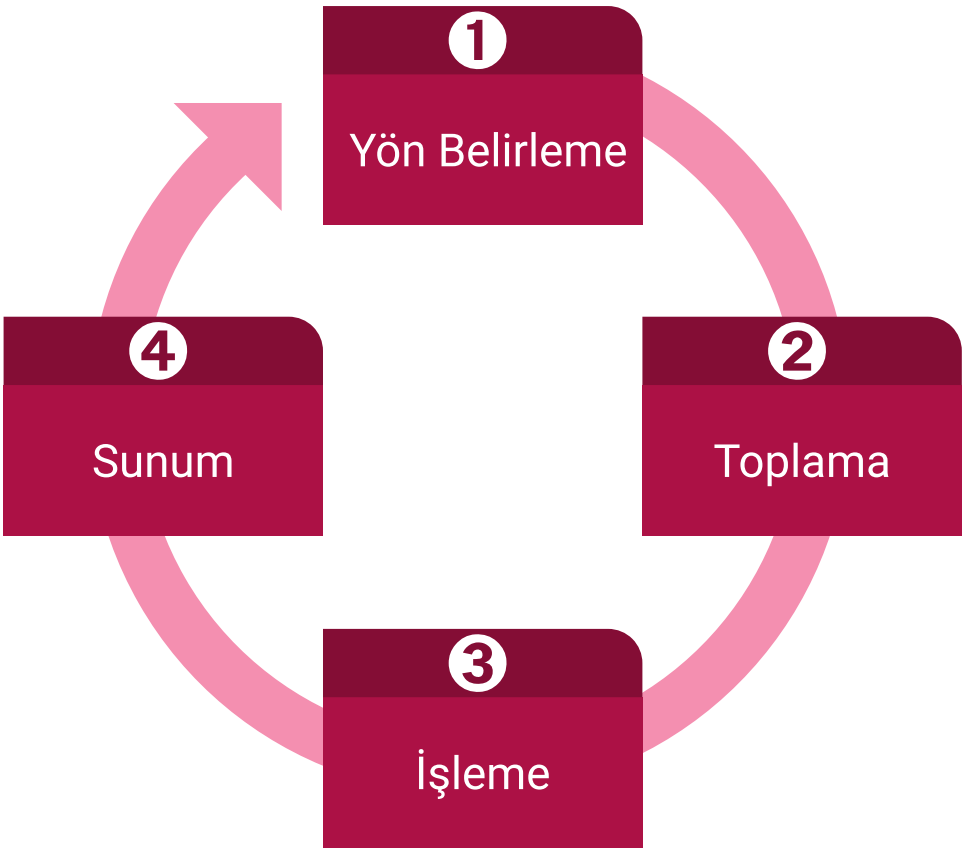
Şekil 1.1. Tez çalışması grafiksel özet

2. AÇIK KAYNAK İSTİHBARATI

Bu bölümde ilk olarak Temel İstihbarat çerçevesi tanıtılmaktadır. Sonrasında alt istihbarat alanları üzerine yapılan incelemeler sunulmakta ve alt istihbarat alanlarından OSINT detaylı olarak anlatılmaktadır.

2.1. Temel İstihbarat

İstihbarat ihtiyacının temelinde bir gereksinim yatmaktadır. Bir bilgiye olan gereksinim ve ilgili hedeflerin belirlenmesi ile yaşam döngüsü başlar. Devamında ise bu bilginin elde edilmesi, anlamlı hale getirilmesi ve kullanılması için ilgili adımlar uygulanır. İstihbarat yaşam döngüsü Şekil 2.1'de sunulmaktadır.



Şekil 2.1. İstihbarat yaşam döngüsü [6]

Her adımda uygulanan işlemler ilk adımdan son adıma kadar alt başlıklar ile ele alınmaktadır [6, 7].

2.1.1. Yön belirleme

Bu adımda ilk olarak istihbarat ihtiyacına neden olan gereksinimler ile birlikte problem tanımının yapılması gerekmektedir. İstihbarat ile ilgili gereksinim duyulan problemin tanımı yapıldıktan sonra planlama yapılmalıdır. Bu planlama evresi bilginin toplanması ile ilgili tüm planlamaları içermektedir.

2.1.2. Toplama

Bilginin toplanması ve verinin harmanlama işleminin yapıldığı adımdır. Veri toplama işlemi açık ve kapalı kaynaklardan olmak üzere iki şekilde yapılmaktadır. İlk olarak açık veri toplama işlemi her bireyin ulaşabileceği gazete, dergi, internet sitesi, radyo ve televizyon gibi kaynakları kapsamaktadır. Kapalı veri toplama ise her bireyin elde edemeyeceği ve ulaşamayacağı kapalı kaynak veya sistemlerden elde edilen veri toplama türüdür.

2.1.3. İşleme

İlgili istihbarat verilerinin toplanması sonrasında bu bilgilerin anlamlandırılarak değerli hale getirilmesi işlemidir. Toplanan her bilgi ilk adımda belirlenen gereksinimlere hizmet etmeyebilir. Bu aşamada toplanan tüm veriler belirli manipülasyon ve veri işleme adımları uygulanarak kullanılabilir hale getirilir.

2.1.4. Sunum

Bilgiler toplandıktan sonra bilgilerin düzenli hale getirilmesi ile tüm istihbarat süreci bitmemektedir. En son adım olarak ilgili bilgilerin karar vericiler ve değerlendiriciler tarafından kullanılabilmesi için düzenli şekilde raporlanmalı ve kullanılmalıdır. Bu noktada elde edilen istihbarat bilgilerinden yeni süreçler başlayabilmektedir. İstihbarat yaşam döngüsü denmesinin sebebi de budur. Her son adımdan sonra yeni gereksinimler ortaya çıkabilir ve yeni bilgilerin toplanması gerekli olabilmektedir.

2.2. İstihbarat Elde Etme Yöntemleri

İstihbarat elde etme yöntemleri birçok farklı isimlendirmelere sahip olsa da 6 ana başlık üzerinden sınıflandırılmaktadır. Bu yöntemler kendi alt alanlarında da özel olarak ayrışsa da tüm ana yöntemler ile birlikte işletilebilmektedir. Amerika Birleşik Devletlerinin Ulusal İstihbarat Ofisi tarafından yapılan sınıflandırma da yöntemler İnsan İstihbaratı (Human Intelligence - HUMINT), Sinyal İstihbaratı (Signal Intelligence - SIGINT), Ölçüm ve İmza İstihbaratı (Measurement and Signature Intelligence - MASINT), Görüntü İstihbaratı (Imagery Intelligence - IMINT), Mekansal İstihbarat (Geospatial Intelligence - GEOINT) ve son olarak Açık Kaynak İstihbaratı (Open Source Intelligence - OSINT)'ten oluşmaktadır [3]. Bunların haricinde Siber İstihbarat (Cyber Intelligence - CYINT) ve Sosyal Medya İstihbaratı (Social Media Intelligence - SOCMINT) gibi farklı istihbarat yöntemleri de bulunmaktadır. Fakat bu yöntemlerin çoğu bahsedilen 6 ana başlık içerisinde keşişim kümeleri ile tanımlanabilmektedir. Özellikle son yıllarda çoğu kaynağın dijital ortamda kontrol edilebildiği düşünüldüğünde, uluslararası ilişkilerde CYINT gibi yöntemleri kullanmak için birçok farklı çalışmanın yürütüldüğü görülmektedir [8]. Bunun haricinde mobil uygulamaların hayatımıza girmesi ve sosyal medyayı günlük yaşantımızın bir parçası yapmamızdan sonra sosyal medya kullanımı ciddi oranlara ulaşmıştır. Kullanım oranı yüksek sayılarda olan sosyal medya platformlarında oluşan veri miktarı büyük veri boyutuna eriştiğinden beri bu bilgilerin farklı amaçlarda kullanılabildiği düşünülmekte ve bu alanda birçok çalışma yapılmaya başlanmıştır. Aynı şekilde Sosyal Medya Analizi ve SOCMINT gibi farklı çalışma alanları da oluşmuştur [9]. Tabi SOCMINT yöntemi de istihbarat gereksinimine göre farklı ana istihbarat elde etme yöntemleri ile kesişmektedir.

2.2.1. İnsan istihbaratı (HUMINT)

İnsan İstihbaratı olarak isimlendirilen istihbarat elde etme yöntemi temel de teknoloji bağımsız insanlardan bire bir veya toplu şekilde elde edilen yöntem olarak sunulmaktadır [10]. Bilgi toplamanın en eski yöntemlerinden biri olarak gösterilen HUMINT, 20. yüzyılın ortalarına kadar İstihbaratın birincil bilgi elde etme yöntemiydi [3]. Teknolojinin gelişimiyle birlikte daha kolay bir şekilde HUMINT yapılabilir oldu. Özellikle telefonun gelişimi, sonrasında akıllı telefonların kullanılması, mesajlaşma ve sosyal medya gibi teknolojilerin kullanımı ile bireyler arası iletişim kolay ve hızlı hale geldi.

2.2.2. Sinyal istihbaratı (SIGINT)

Sinyal İstihbaratı denilen bu yöntem ile tüm iletişim yollarını kapsayan bilgi toplama yöntemleri kullanılsa da, yaygın olarak şifresiz iletişim kanalları üzerinden yapılan dinlemelerde kullanılmaktadır [11]. Kullanım alanına göre alt yöntemler olarak da isimlendirilen SIGINT; İletişim İstihbaratı (Communication Intelligence - COMINT), Elektronik İstihbarat (Electronic Intelligence - ELINT) ve Yabancı Enstrümantasyon Sinyalleri İstihbaratı (Foreign Instrumentation Signals Intelligence - FISINT) alt alanlara ayrılmaktadır [3].

2.2.3. Ölçüm ve imza istihbaratı (MASINT)

SIGINT ve IMINT dışındaki tüm teknik istihbarat bilgilerini kapsamaktadır. Bu istihbarat yöntemindeki verilerin sonucunda hedef ayrıt edici şekilde konumlandırılabilir, tanımlanabilir ve açıklanabilir [3]. Elektro-Optik İstihbaratı (Electro-Optical Intelligence), Radar İstihbaratı (Radar Intelligence) ve Nükleer İstihbaratı (Nuclear Intelligence) gibi alt çalışma alanları bulunmaktadır [12].

2.2.4. Görüntü istihbaratı (IMINT)

Görüntü İstihbaratı olarak isimlendirilen istihbarat elde etme yönteminde dijital veya optik araçlarla elde edilen tüm görsel materyallerin istihbaratı sunulmaktadır [3, 11]. Geçmişte elde edilen optik materyallerin dışında günümüzde gerçek zamanlı görüntü elde eden kameralardan uydulara kadar birçok kaynak ortaya çıkmaktadır. Bu kapsamda IMINT bazı noktalarda diğer İstihbarat elde etme yöntemleri ile kesişmektedir. Uydu görüntüleri ile ilgili bir istihbarat gereksiniminde GEOINT ve Internet gibi bir ortamdan elde edilen açık kaynak bir materyalden resimlerin analizinde OSINT ile birlikte kullanılmaktadır.

2.2.5. Mekansal istihbarat (GEOINT)

Coğrafik bilgilerin IMINT ile birlikte kullanımını temel alan istihbarat toplama yöntemidir [3]. İnsanların veya nesnelerin mekânsal olarak ilişkilendirilmeleri ile ilgili istihbarat çalışmalarını içermektedir [7]. Harita bilgilerinin dijital olarak erişilip özelleştirilmiş şekilde kullanılabilmesinin yanında resimlerin ve sosyal medya içeriklerindeki iletilerin konum gibi

bilgilerinden coğrafik istihbarat bilgisi çıkarılabilmektedir. Böylelikle coğrafik nesne ilişkilendirme üzerine OSINT ile ilişkili birçok çalışma yapılabilmektedir.

2.2.6. Açık kaynak istihbaratı (OSINT)

Açık Kaynak İstihbaratı açık ve gizli olan bilgilerin açık şekilde paylaşıldığı ortamlardan elde edilmesi üzerine yapılan istihbarat faaliyetlerini kapsamaktadır [13]. Internet teknolojileri ortaya çıkmadan ve yaygınlaşmadan önce Açık Kaynak İstihbaratı dergi, gazete gibi kaynaklardan elde edilen bilgilerden yapılan çalışmalar üzerine inşa edilmişti. Hatta ilk ortaya çıkışı 2. Dünya Savaşına kadar uzanmaktadır [14]. Internet teknolojilerinin ortaya çıkışı ve yaygınlaşması ile birlikte ortaya çıkan açık kaynak verilerindeki artış bu bilgi toplama yönteminin kullanım alanını da genişletmiştir. Hatta açık kaynak olarak üretilen bilgilerin çeşitliliğindeki artışla birlikte diğer istihbarat toplama yöntemleri ile kesişim oranı da artmıştır. Bu kapsamda Şekil 2.2'de örnek bir kesişim kümesi sunulmaktadır. Fakat gerçek etkileşimleri kesişim şeması ile gösterilecek düzeyde basit bir ilişki yapısına sahip değildir.



Şekil 2.2. OSINT'in diğer istihbarat toplama yöntemleri ile örnek kesişim şeması [15] (IMINT eklenmiş hali)

Bir diğer şekilde açık istihbarat kaynaklarını sıralarsak kesişim noktaları daha iyi anlaşılacaktır. Bunlar [11];

- Geleneksel medya kaynakları
- Ticari çevrimiçi prim kaynakları
- Diğer ticari niş çevrimiçi kaynakları
- Gri literatür
- Uzman olan insanlar
- Ticari görüntüler ve coğrafik bilgiler
- Internet ve WWW (e-postalar ve sesli çağrılar dahildir)

İçerik kaynaklarının geniş sunumunun yanında bilginin türüne göre dört kategoride alt türleri sınıflandırabiliriz. Bunlar Haber Medya İçeriği (News Media Content), Gri Literatür (Gray Literature), Uzun Formlu Sosyal Medya İçeriği (Long-Form Social Media Content) ve Kısa Formlu Sosyal Medya İçeriği (Short-Form Social Media Content) olarak isimlendirilmektedir. [15]

Haber Medya İçeriği: Açık olarak erişilebilir haber ve medya kaynaklarının tamamını içermektedir. Televizyon, radyo, gazete ve dergi bu içerik türünde kullanılan materyallere örnek olarak gösterilebilir.

Gri Literatür: Medya olmayan kurum ve organizasyonlardan elde edilen hem açık hem gizli içeriklerden oluşmaktadır. Ticari kurumlar, akademi, gizli yayıncılar ve ulusal hükümetler bunlara örnek gösterilebilir.

Uzun Formlu Sosyal Medya İçeriği: Bloglar, Tumblr, Reddit, Medium gibi sosyal medya içeriklerini kapsamaktadır. Bu noktada kısa formlu sosyal medya içeriklerinden içerik boyutunun büyüklüğü üzerine ayrılmaktadır.

Kısa Formlu Sosyal Medya İçeriği: Facebook, Twitter, Instagram gibi sosyal ağların içeriklerinin bulunduğu kısa boyutlu içeriklerin üretildiği tüm sosyal medya platformlarını bu başlık altında gruplanabilir.

OSINT yaşam döngüsü kaynaktan kaynağa farklı adımlarla nitelendirilmektedir. İstihbarat Çalışmaları El Kitabında sunulan OSINT yaşam döngüsü [11];

- Gereksinimlerin belirlenmesi
- Önceliklerin belirlenmesi

- Toplama
- İşleme ve Sömürü
- Analiz ve Üretim
- Güvenlik
- Yaygınlaştırma ve Değerlendirme (Geri bildirim)

Bunun yanında, CIA normal istihbarat yaşam döngüsünü tüm yöntemler üzerinde kullanmakla birlikte 5 temel adıma yaymaktadır [16]:

- Planlama ve Yönlendirme
- Toplama
- İşleme
- Analiz ve Ürünleştirme
- Yaygınlaştırma (Dağıtım)

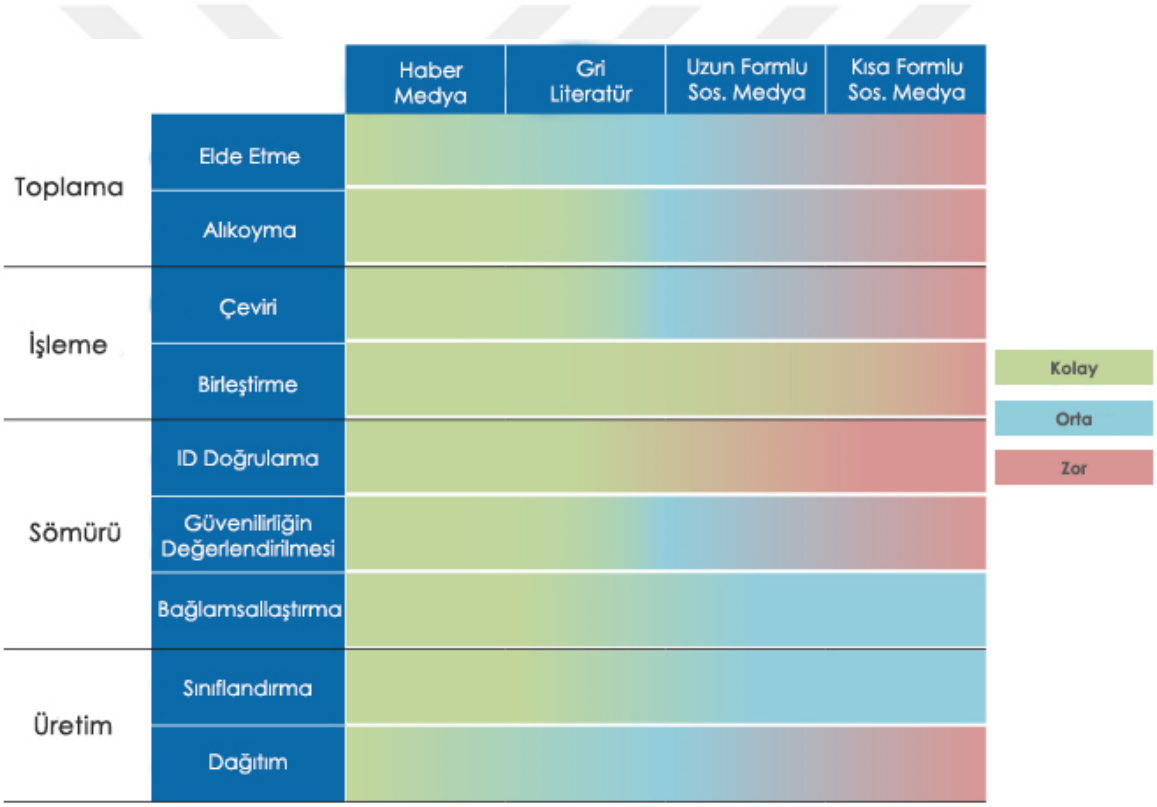
Son olarak RAND şirketi tarafından oluşturulan yaşam döngüsü Şekil 2.3'de sunulmaktadır. Bu noktada halen OSINT için bir metodolojiye ihtiyaç duyulduğu görüşü savunulmaktadır. [15]



Şekil 2.3. RAND OSINT yaşam döngüsü [15]

OSINT yaşam döngüsünün içerik türlerine göre her adımının zorluğu Şekil 2.4'te sunulmaktadır. Bu noktada hangi alanda ilerlenmesi gerektiği bilgisi üzerine bir bakış açısı sunulmaktadır.

OSINT için veri toplama adımlarında 3 farklı yöntem kullanılmaktadır. Yöntemler; el ile arama dediğimiz manuel göz ile arama, araçlar ile manuel filtreleme/arama ve son olarak dinamik ajanlar ile arama olarak sunulabilir. Bire bir ilgili içerikleri okuyarak önemli bilgilerin çıkartımı zor ve zahmetlidir. Çok fazla zaman alabilir. Ayrıca yardımcı araçlar kullanmak işleri büyük ölçüde kolaylaştırmaktadır.



Şekil 2.4. OSINT yaşam döngüsünün içerik türlerine göre zorluk düzeyleri [15]

Açık kaynak istihbaratı yaparken yardımcı olarak kullanılacak bazı örnek OSINT araçları Çizelge 2.1'de sunulmaktadır. Bunlar sadece bazı araçlar olmakla birlikte doğrudan OSINT aracı olarak kullanılan birçok araç bulunmaktadır. Sadece bu araçları kullanmak yerine, OSINT aracı olarak üretilmemiş fakat bu amaçla kullanılabilen araçlarda bulunmaktadır. Haricen, özelleştirilmiş bir problemin çözümü için devasa verilerden söz ediliyorsa bu noktada kişisel yazılmış yazılımlar oluşturmak en mantıklı yol olacaktır.

Çizelge 2.1. Bazı OSINT araçları

OSINT Araçları	Kullanım Amacı
host, dig, nslookup, quad9	DNS bilgileri üzerine araştırma ve kontrol sağlamak için.
whois	Web site, ip adresi ve ağlar üzerinde kayıt ve özlük bilgilerine erişim için.
nmap, masscan	Bir website veya ip adresi üzerinde portlar üzerine bilgi araştırması için.
google arama özellikleri	Normal aramaların dışında spesifik bilgiye erişimde özel parametreler kullanarak daha detaylı aramalar yapılmaktadır.
shodan	İnternet temelli aygıtlar üzerinde arama yapabilmek için kullanılmaktadır.
censys	Her İnternet sunucusu hakkında bilgi toplamak amacıyla kullanılmaktadır.
binaryedge	İnternet ağını tarayarak belirtilen düğümler hakkında bilgiler toplamaktadır.
securitytrails, robtex, hackertarget	Pasif dns çözümü sunmaktadır.
virustotal, greynoise, firehol ip lists	Şüpheli ip, alan adı, dosya gibi araştırmalarda ve kıyaslama yapmak amacıyla da kullanılmaktadır.
viewdns.info, whoxy, securitytrails	Dns lerde geriye dönük aramalarda kullanılabilmektedir.
spiderfoot, maltego, theharvester	Geniş çapta birçok OSINT modülü içeren araçlardır.
twint	Twitter için osint kütüphanesi ve aracı sunar.
ultimate facebook scraper	Facebook için bir sömürü aracıdır.
tweet beaver	Twitter için osint aracı sunar.

2.3. Tez Konusu Sınırını Belirlemede Yürütülen Aşamalı Literatür Çalışması

Bu bölümde belirlenen tez konusu için 2019 yılı sonuna kadar yapılmış çalışmalar baz alınarak aşamalı literatür araştırması sunulmaktadır. İlk aşama olarak literatür tarama ortamlarından istihbarat tarama yöntemleri ve tarama yöntemlerinden araştırma alanı olarak belirlenen OSINT konusuna dayalı bir araştırma yapılmıştır. İkinci aşamada ise ilk aşamada elde edilen bulgular ile araştırma sahası indirgenerek farklı anahtar kelimeler üzerine yapılan çalışmalar belirlenmiş ve incelenmiştir. Yapılan çalışmada ise Sosyal Medya üzerinde yapılan kişilik profilleme ve kişilik özelliği çıkarımı yapılan çalışmalar belirlenmiş ve incelenmiştir. Üçüncü çalışmada ise Sosyal medya üzerinden Kişilik Özellikleri analizine dayalı çalışmaların incelemesi ve diğer çalışmalar ile ilgili genel bir değerlendirme sunulmaktadır. Detaylı inceleme ise 3. Bölüm'de Kişilik Özellikleri Tahmini başlığı altında verilmektedir.

2.3.1. Aşama 1: OSINT literatür araştırması ve çalışma alanları

İstihbarat alanında yapılan çalışmaları ve yapılması muhtemel çalışmaları incelemek ve analiz etmek için bazı web içerikleri, sosyal medya hesapları takip edilmekle birlikte akademik arama ortamları da kullanılmaktadır. En çok araştırma yapılan ve kaynak bakımından birçok içeriği barındıran akademik araştırma ortamlarını WebOfScience (WoS), ScienceDirect (SD) ve IeeeXplore (IX) olarak sıralayabiliriz. Google'ın Scholar platformu da kullanılmakla birlikte çok fazla gereksiz içeriği de barındırmaktadır. Bu sebeple bu araştırma da kullanılmamıştır. İncelenen araştırma platformlarında tüm ayarlar açık şekilde kullanılmıştır. Herhangi bir filtreleme işlemi başlangıç aşamasında uygulanmamıştır.

Her istihbarat elde etme yönteminin belirtilen akademik arama motorlarında hem açık yazılışları hem de kısaltmaları kullanılarak hangi sayılarda çalışma yapıldığı irdelenmektedir. Çizelge 2.2'de arama sorguları listelenmiştir. Sonrasında ise elde edilen sonuçlar kapsamında en fazla kullanılan istihbarat elde etme yöntemi daha özelleştirilmiş arama kelimeleri kullanılarak detaylı değerlendirme yapılmaktadır.

Çizelge 2.2. İstihbarat elde etme yöntemleri arama sorguları

İstihbarat Elde Etme Yöntemi	Sorgular
HUMINT	"HUMINT" or "Human Intelligence"
SIGINT	"SIGINT" or "Signals Intelligence"
MASINT	"MASINT" or "Measurement and Signature Intelligence"
IMINT	"IMINT" or "Imagery Intelligence"
GEOINT	"GEOINT" or "Geospatial Intelligence"
OSINT	"OSINT" or "Open Source Intelligence"

İstihbarat elde etme yöntemlerine göre yapılan araştırma çalışmalarının sayıları Çizelge 2.3'de sunulmaktadır. Yapılan bir çok çalışma insan unsurunu temel aldığı için HUMINT ile ilgili elde edilen sonuçlar WoS ortamında yüksek çıkmıştır. Bunun yanında teknik olarak yapılan çalışmaların incelendiği SD ve IX ortamlarında çalışma sayısı ciddi oranda az olarak gözlemlenmiştir. Bunlar haricinde teknik akademik çalışmaların, uygulama alanındaki kolaylık sebebi ile OSINT alanında daha fazla olduğu görülmektedir.

Çizelge 2.3. İstihbarat elde etme yöntemlerinin akademik araştırma platformları üzerinde bulunan araştırma sayıları

İstihbarat Elde Etme Yöntemi	WoS	SD	IX
HUMINT	1942	92	10
SIGINT	190	119	17
MASINT	17	28	7
IMINT	60	35	3
GEOINT	154	42	7
OSINT	216	153	44

Yapılan incelemelerde çalışmaların ilgili ve ilgisiz olma durumları göz ardı edilmiştir. Bunun ile ilgili çalışma sonraki aşamada detaylı olarak yapılacaktır.

OSINT alanı ile ilgili yapılan incelemelerde WoS platformu üzerinde çekirdek koleksiyonlar dan teknik olanların ayrımı Çizelge 2.4'de sunulmaktadır. Diğer alanlar ile ilgili bu ayrım yapıldığında (özellikle HUMINT üzerine) yapılan çalışmaların sayısının daha da azaldığı görülmektedir. OSINT üzerine yapılan çalışmaların 139 tanesi bilimsel indeksler üzerinde taranan çalışmaları 77 tanesi ise sosyal içerikli indeksler üzerinde taranan çalışmaları kapsamaktadır. Özellikle bilimsel değeri yüksek SCI-E indeksli çalışmaların sayısı 17 adettir. İlgili indekslerin açıklamaları aşağıda sunulmaktadır.

- SCI-E: Science Citation Index Expanded
- CPCI-S: Conference Proceedings Citation Index- Science
- BKCI-S: Book Citation Index– Science
- ESCI: Emerging Sources Citation Index

Çizelge 2.4. OSINT alanında WoS üzerinde bulunan teknik ve toplam çalışma sayıları

Koleksiyonlar	OSINT-WoS
SCI-E	17
CPCI-S	89
BKCI-S	12
ESCI	21
Bilimsel Toplam	139
Sosyal Bilimler Toplam	77
Genel Toplam	216

SCI-E kapsamında elde edilen çalışmaların başlık ve özetlerinde en çok geçen kelimelere bakıldığında ilk olarak Siber Tehdit İstihbaratı (Cyber Threat Intelligence - CTI) ve Sosyal içerikler göze batmaktadır.
Aynı şekilde WoS üzerinde ESCI indeksli OSINT çalışmalarının başlık ve özet bilgilerinde en çok kelimelere bakıldığında dikkat çeken kelime grubu sosyal medya analizidir (social, media, analysis).

WoS ile benzer şekilde SD'de aratılan OSINT içeriklerinin türüne göre çalışma toplamları Çizelge 2.5'de sunulmaktadır.

Çizelge 2.5. OSINT alanında SD üzerinde bulunan teknik ve toplam çalışma sayıları

Koleksiyonlar	**OSINT-SD**
Literatür Çalışması	4
Araştırma Çalışması	39
Kitap Kritiği	1
Kitap Bölümü	70
Editöryal	3
Kısa İletişim	2
Diğer	34
Genel Toplam	153

Bakılan çalışmalardan 39 adedinin araştırma makalesi olduğu görülmektedir. Bunların dışında yapılan çalışmalar genel olarak literatürsel bilgi veya tanıtım içermektedir. Elde edilen 39 çalışmanın anahtar kelimeler, başlık ve özetleri temel alındığında çalışmaların büyük çoğunluğu Sosyal Ağ Analizi üzerinedir.

İncelenen son akademik çalışma arama platformu IX'de ise OSINT alanında aratılan çalışmaların türüne göre ayrımı Çizelge 2.6'da sunulmaktadır. Genel olarak yapılan çalışmaların türü konferanslarda sunulan içeriklerden oluşmaktadır.

Çizelge 2.6. OSINT alanında IX üzerinde bulunan teknik ve toplam çalışma sayıları

Koleksiyonlar	**OSINT-IX**
Konferans	42
Dergi/Magazin	1

Çizelge 2.6. (devam) OSINT alanında IX üzerinde bulunan teknik ve toplam çalışma sayıları

Koleksiyonlar	OSINT-IX
Ders İçeriği	1
Toplam	44

Elde edilen tüm çalışmaların başlık, özet ve anahtar kelime bilgilerinden en fazla kullanılan kelimelere bakıldığında diğer araştırma platformu sonuçları ile benzer çıkmaktadır. Bu noktada yapılan çalışmaların çoğunun sosyal medya analizine dayalı olduğu görülmektedir.

Araştırılan üç farklı platformda WoS için SCI-E ve ESCI, SD için Araştırma Çalışmaları ve IX için tüm çalışmaların çıkartılan başlık, özet ve anahtar kelime bilgileri temel alındığında OSINT alanında Sosyal Medya ile ilgili çalışmaların yoğunlaşmış olduğudur. Bunun sebebi olarak erişim kaynağı verilerinin çokluğu ve rahat erişilebilir olması söylenebilir. Son olarak SD ve IX platformlarının sunduğu anahtar kelime bilgileri üzerinde yapılan incelemede ise temel kelimeler (information, intelligence, OSINT) haricinde dikkat çeken kelime grupları "social, media, analysis" (sosyal, medya, analiz) ve "cyber, threat, analysis" (siber, tehdit, analiz) kelime gruplarının dikkat çektiği gözlemlenmiştir. Yapılan çalışmalar bireysel olarak incelendiğinde de OSINT için ana çalışma alanları genel olarak sosyal medya analizi ve siber tehdit analizidir.

2.3.2. Aşama 2: OSINT alanında kullanıcı temelli sosyal medya analizi

Önceki bölümde tespit edilen OSINT çalışma alanlarında en çok bahsedilen konulardan biri sosyal medya analizidir. Sosyal medya üzerinde ise bir istihbarat çalışması yapılabilmesi için istihbarat toplanacak hedefin belirlenmesi gerekmektedir. Bu hedef bir kullanıcı, kurum, kuruluş, organizasyon, dernek vb. olabilmektedir. Fakat sosyal medya ortamlarının genel olarak kullanıcı temelli olduğu düşünüldüğünde hedefinde kullanıcı temelli olarak seçilmesinin daha doğru olacağı öngörülmektedir. Bu bağlamda, literatürde yapılan inceleme sosyal medya analizi ile birlikte kullanıcı temelli olarak yapılmıştır.

İlk olarak, 3 araştırma ortamında (WoS, SD ve IX) Çizelge 2.7'de sunulan sorgular kullanılarak aramalar yapılmıştır. Sorgularda kullanılan anahtar kelimeler kullanıcı ve sosyal ağ analizi üzerine yapılmış çalışmaların genel olarak ilgi durumu ve anahtar kelimelerine

bakılarak belirlenmiştir. Bu noktada ilgili kelimelerin varyasyonları arama ortamları tarafından tamamlandığından diğer kullanımlar sorgulara dahil edilmemiştir.

Yapılan aramalar başlık, özet ve anahtar kelime alanlarına bakılarak yapılmış ve sonucunda kullanıcı görünüşü (user profile) ve kullanıcı kişiliği (user personality) temelli sosyal ağlarda yapılan çalışma sayıları Çizelge 2.8'de sunulmuştur. Araştırma ortamlarında yapılan aramalar bilimsel değeri daha yüksek görülen araştırma ve literatür çalışmaları özelinde filtrelenmiştir. Bu noktada yapılan filtrelemelerde, WoS araştırma ortamında SCI-E ve ESCI, SD araştırma ortamında Araştırma ve Literatür makaleleri, IX araştırma ortamında Dergi ve Magazin kategorisinde Araştırma ve Literatür makaleleri temel alınmıştır.

Çizelge 2.7. Sosyal medya ve kullanıcı kelimeleri temelli sorgular

Sorgu No	**Sorgu**
1	("user profiling" OR "user profile") AND ("social media" OR "social network")
2	("user personality") AND ("social media" OR "social network")

Çizelge 2.8. Sosyal medya ve kullanıcı kelime varyasyonları üzerine yapılan çalışma sayıları (araştırma ve literatür çalışma sayıları / toplam çalışma sayıları)

Sorgu No	**WoS**	**SD**	**IX**
1	114/359	103/105	23/265
2	5/21	16/16	2/17

Araştırma ortamlarından elde edilen sonuçlarda kesişimler olmakla birlikte, bulunan akademik çalışmalardan yalnızca bir çalışma literatür değerlendirmesi kategorisinde sınıflandırılmaktadır. Ana amacı twitter üzerinde yapılan duygu analizi çalışmalarının analizi ve değerlendirmesi olan çalışmada, içerdiği literatür çalışmalarında odaklanılan alan, kullanılan yöntem, veri seti uzayı, uygulanan hedef ülke ve doğrulama parametreleri temelinde değerlendirme yapılmaktadır [17].

Arama sonucunda listelenen çalışmaların atıf yapılma sayılarına göre sıralanması sonucu iki sorgu ve arama platformları için ayrı olmak üzere en fazla atıf alan beşer çalışma genel olarak incelenmiştir. Arama platformlarındaki sonuçlardaki kesişimler hesaba katılarak elde edilen çalışma sayısı kullanıcı kişiliği için 11, kullanıcı görünüşü için 13 çalışma olarak elde edilmiştir.

Kullanıcı görünüşü ile ilgili elde edilen çalışmalar dikkate alındığında genel çalışma alanları; tavsiye sistemleri [18, 19], çıkarım saldırıları [20], isim belirsizliği problemleri [21], market analizi [22], gizli grupların tespiti (folksonomi) [23, 24], yaş benzerlik grupları [25], mekânsal dağılım ve hareketlilik [26], anonim profil eşleştirme [27], kullanıcı özellik çıkarımı [28], içerik çıkartımı/farkındalığı [29], hassas bilgi tespiti [30] olarak gözlemlenmiştir.

Sorgu 2 ile elde edilen en fazla atıf alan çalışmalar incelendiğinde genel çalışma alanlarının; akıllı şebekelerde kişiliğin rolü [31], görüntü kullanımına dayalı kişilik analizi [32], bilişsel profil çıkarımı [33], 5 büyük-model çıkarımı [34, 35], tavsiye sistemleri [36], kişisel özellik ile kişilik ilişkisi [37], sosyal ağlar arası kişilik ilişkisi [38], kişilik ile davranış ilişkisi [39, 40], video/vlog üzerinden kişilik analizi [41] olarak ayrıştığı görülmüştür.

2.3.3. Aşama 3: Sosyal medya platformlarından kişilik özellikleri analizi

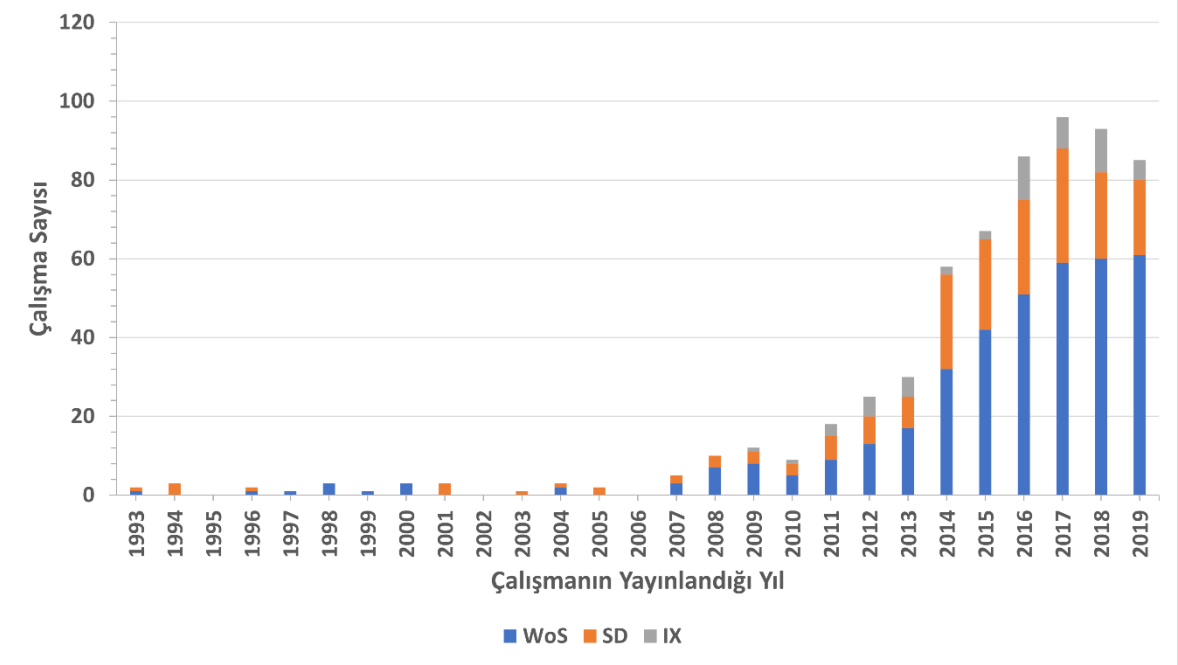
Bu aşamada önceki aşamalarda incelenen çalışmalarda kişiliği sosyal medya çalışmalarında terminolojik olarak tanımlayan anahtar kelimeler üzerinden bir araştırma yapılarak elde edilen çalışmalar incelenmiştir. Elde edilen anahtar kelimeler sosyal medya (social media), sosyal ağ (social network), kişilik (personality), kişilik özellikleri (personality traits) ve tahmin (predict)'dir. Kişilik ve kişilik özellikleri ayrımında kişilik ile ilgili aramalarda 'personality traits' anahtar kelimesinin geçmediği alakalı çalışmaların varlığı sebebi ile kapsam kişilik kelimesine göre sınırlandırılmıştır. WoS için SCI-E, CPCI-S, BKCI-S ve ESCI koleksiyonlarında arama yapılmıştır. Çizelge 2.9'da kullanılan sorgunun üç farklı tarama dizininde aranması sonucu elde edilen çalışma sayıları Şekil 2.5 ve Çizelge 2.10'da sunulmaktadır. Sunulan bilgilere bakıldığında yapılan çalışmaların özellikle son 6 yılda çoğunlukta olduğu görülmektedir. Ayrıca son 3 yıl yapılan çalışmaların dağılımı tüm zamanlarda yapılan çalışmaların neredeyse yarısını oluşturmaktadır.

Çizelge 2.9. Kişilik özelliğine dayalı sosyal medya çalışmaları için sorgu

Sorgu
(((((('social media') OR 'social network') AND 'personality') AND 'predict'))

Çizelge 2.10. Kişilik özelliğine dayalı sosyal medya çalışma sayıları

Tarih Aralığı	WoS	SD	IX
2017-2019 sonu	180	74	24
2020 öncesi	379	189	54



Şekil 2.5. Kişilik Özelliğine dayalı sosyal medya çalışma sayılarının tarihsel dağılımı

2020 yılı öncesi son 3 yılda yapılan çalışmaların çoğunluğu oluşturması ve çalışma kapsamını güncel çalışmalardan oluşturma isteğimizden dolayı literatür incelemesi son 3 yıl için sınırlandırılmıştır. İlk inceleme için Türkçe içerikler üzerinde yapılan çalışmaların belirlenmesi çalışmanın geleceği açısından önemlidir. Bu sebeple, toplanan çalışmalar "turkish" or "türkçe" anahtar kelimeleri ile filtrelenmiş ve elde edilen 20 çalışma incelenmiştir. İncelenen çalışmaların kapsam bakımından akademik ve öğrenci performans analizi, internet, teknoloji, sosyal medya ve akıllı telefon bağımlılığı vb. üzerine yapılan çalışmalar olduğu, kişilik ölçeği anket sonuçlarına göre yapılan sosyal bilimler çalışmaları olduğu ve sosyal medya üzerinden kişilik analizi ile alakalı olmadığı görülmüştür. Bu nedenle incelenen çalışmalar belirlenen tez çalışması ile alakalı olmadığından literatür havuzundan çıkarılmıştır. Ayrıca geri kalan çalışmalarda sosyal medya kullanılarak yapılan anketler ile alkol kullanımı, siber zorbalık, bitki/hayvan kişiliği, yeme alışkanlığı, trolleme, bağımlılık gibi parametreler üzerinden incelenerek tez çalışması ile ilişkili olmayan çalışmalar literatür havuzumuzdan silinmiştir. Sonuç olarak içerik ve kapsam temel

alındığında konumuz ile ilişkili olarak kişilik özellikleri tahminine dayalı çalışmaların detaylı sunumu Bölüm 3'de sunulmaktadır.

2.4. OSINT ve Kişilik Özellikleri Tahmini

İnternet ortamında herkese açık (public) şekilde paylaşılan içerikler OSINT'in çalışam alanlarından bir tanesidir. İnsanlar genel olarak kullandıkları sosyal medya platformlarından çoğunlukla herkese açık görünebilecek şekilde içerik üretimi yapmaktadırlar. Herkese açık şekilde öznel hareket ve fikirleri içeren paylaşımlar kişiler veya kullanıcı personaları hakkında farklı ipuçları verebilir. Buradaki asıl olay bilinmeyeni elde etme ve öğrenme isteğidir. Bilinmeyen bir bilginin elde edilme faaliyeti OSINT'in her açıdan çalışma alanıdır. Özetle İnternet ortamı veya daha sınırlı bir alan içinde sosyal medya platformlarından kişiler/kullanıcılar hakkında bir durumu elde etme/tahmin etme faaliyeti açık kaynak istihbarat faaliyeti olarak tanımlanabilir. Bu açıdan incelenen OSINT yaklaşımı ile kişilik özellikleri tahminine yönelik literatürsel analiz Bölüm 3'de detaylı olarak sunulmaktadır.

3. KİŞİLİK ÖZELLİKLERİ TAHMİNİ

Bu bölümde kişilik özellikleri tahmini üzerine literatürde yapılan çalışmaların kapsamlı analizi sunulmaktadır. Alt başlıklarda ilk olarak, Türkiye'de kişilik özelliklerine dayalı yapılmış tez çalışmaları incelenmektedir. Devamında son 5 yılda yapılmış çalışmaların incelemesi yapılmaktadır. Sonrasında diğer yıllarda yapılmış fakat etki düzeyi ve kapsamı geniş olan çalışmalar incelenmektedir. Son olarak, tüm literatür çalışmalarının incelenerek tez motivasyonu ve sınırlarının belirlenmesi amacı ile yapılan değerlendirme sunulmaktadır.

3.1. Tez Çalışmaları

YÖK Tez Tarama sistemi üzerinden yapılan araştırmalarda elde edilen çalışmalar incelendiğinde kişilik özelliklerinin kullanımına yönelik farklı çalışma alanları görülmektedir. Buna ek olarak kişilik özellikleri tahminine dayalı birkaç çalışma tespit edilmiştir. Genel araştırma bulguları bu başlık altında sırası ile sunulmaktadır.

Kişilik özelliklerinin mobil uygulama kullanımı arasındaki ilişkiyi araştıran çalışmada [42] mobil ve web ortamında birbirini tamamlayıcı iki deney gerçekleştirilmiştir. Bu uygulamalar yardımı ile kullanıcıların mobil uygulama tavsiyelerini dikkate alma durumları ile kişilik özellikleri arasındaki ilişki araştırılmıştır. Sonuç olarak kullanıcıların mobil uygulama tercihleri ile kişilik özellikleri arasında belirli ilişkilerin olduğu tespit edilmiştir. Özellikle cinsiyete göre ve uygulama kategorisi temelli pozitif ve negatif ilişkili olma durumları her kişilik özelliğine göre detaylı olarak sunulmuştur.

Çevrimiçi çalışma gruplarının başarısı üzerinde kişilik özelliklerinin etkisini araştıran çalışmada [43] BFI-10 kişilik ölçeği öğrenci olan katılımcılara uygulanmıştır. Devamında öğrenciler kişilik özellikleri temel alınarak 35 alt gruba ayrıştırılmıştır. İlgili gruplara online bir sistem üzerinden basit bir aktivite yapmaları istenmiş ve kişilik özelliklerine göre kullanıcıların grup başarısı üzerindeki etkileri incelenmiştir. Çalışmanın sonucunda kişilik özelliklerinin grup çalışmasındaki başarının üzerinde etkili olduğu gözlemlenmiştir. Bunun yanında bir diğer araştırmada [44] kişiliğin takım çalışması üzerindeki etkisi ortaya konmuştur. Böylelikle işe alım süreci, takım değerlendirmeleri veya takım oluşturma süreçlerinde kişilik özelliklerinin önemi ortaya konmaktadır.

Yazılımların son kullanıcı üzerindeki etkisi genellikle kullanılabilirlik temelinde değerlendirilmektedir. Yapılan alt çalışma alanlarından biri de kişilik özelliklerinin kullanılabilirlik üzerindeki etkisidir. Kişilik özellikleri ile öznel kullanılabilirlik değerlendirmesi arasındaki ilişkiyi inceleyen tez çalışması [45] bu noktada örnek gösterilebilir. İlgili araştırma kullanılabilirlik ile kişilik özellikleri arasında ilişki olduğunu ortaya koymuştur.

İnsanların kişilik özelliklerini tahmin etme kelime türleri ve dil biliminin önemini araştıran çalışmada [46] farklı analizler ortaya konmuştur. Fakat ilgili çalışmanın bu tez çalışması için önemi dil mekanikleri ve kelimelere göre kişilik üzerindeki etkinin ortaya konmuş olmasıdır.

Oyun sektöründe en önemli konu kullanıcının oyun üzerinde daha fazla vakit geçirmesidir. Bu bağlamda kullanıcıların oyun üzerindeki oynama sürelerini arttırmak için kullanılan karakter ve oyun içi aksiyon türlerinin kullanıcıların kişiliğine göre değişkenlik gösterebileceğini ortaya koyan çalışma [47] incelenebilir.

Suça yatkınlık veya suçta etkili olabilecek kişilik özellikleri üzerine yapılan araştırmalar [48, 49] bulunmaktadır. İlgili araştırmalar incelendiğinde kişilik özellikleri ile suça yatkınlık arasında belirli bir ilişki olduğu görülmektedir. Bunun yanında insanlar hakkında genel bilgi ediniminin en kolay yapılacağı alan sosyal medya platformlarıdır. Sosyal medya platformlarının kişilik özelliklerini yansıtıp yansıtmadığı üzerine yapılan araştırmada [50] 47 kişi ile çalışılmıştır. Araştırma ile kişilerin son 50 paylaşımı üzerinden %80'e yakın tahmin edilebilirlik ortaya konmuştur. Bunun haricinden kişilik özelliklerinin sosyal medya kullanımı üzerindeki etkisi üzerine de çalışmalar [51-53] bulunmaktadır.

Sosyal medya platformları genel olarak değerlendirildiğinde 2 farklı içerik türüne sahiptir. Bu içerik türleri görsel materyal ve metinsel içerikleri kapsamaktadır. İçerik türleri dışında ayrıca paylaşım özellikleride analizlerde kullanılabilmektedir. Görsel materyaller üzerinden yapılan kişilik özellikleri tahmin çalışmalarında [54, 55] yüz, çevre, ses, çevriyazı ve duygu durumu gibi materyal üzerinden elde edilecek özelliklere göre kişilik tahmininin yapılabileceği ve kişilik özelliklerinin tahminine ek olarak aldatma tespitinin yapılabileceği [56] ortaya konmaktadır. Bunun haricinden Facebook ortamında paylaşılan içeriklerin özelliklerinin kişilik özellikleri üzerindeki etkisi ve kişilik özelliklerinin tahminine yönelik bir tez çalışması [57] bulunmaktadır. İlgili çalışmaya göre paylaşımı yapan kullanıcının

paylaşım düzeyi arttırkça kişilik özellikleri tahmini daha iyi sonuçlar vermektedir. Buna ek olarak bir diğer bulgu, kullanıcının arkadaşlarına ait kişilik özellikleri dikkate alındığında başarımın daha iyi sonuçlar verdiğidir.

3.2. 2017 Yılı Sonrası Literatürde Yapılan Çalışmalar

Çalışmaların büyük çoğunluğunun son 5 yılda yapılmış olması, mevcut çalışmalardan çalışmanın kapsamını oluşturma isteği ile birleştiğinden, literatür taraması son 5 yılda yapılmış çalışmalarla sınırlandırılmıştır. İncelenen çalışmaların büyük çoğunluğunu kapsam olarak akademik ve öğrenci performans analizi, internet, teknoloji, sosyal medya, akıllı telefon bağımlılığı ve kişilik ölçeği anket sonuçlarına dayalı sosyal bilimler vb. konularda yapılan çalışmalar oluşturmaktadır. Bu çalışmalar, sosyal medya platformlarından tahmin edilen kişilik özellikleri ile ilgili değildir. Bu nedenle incelenen ve kapsam dışı olduğu belirlenen çalışmalar literatür havuzundan çıkarılmıştır. Geri kalan çalışmalar içerisinde alkol kullanımı, siber zorbalık, bitki/hayvan kişiliği, beslenme alışkanlıkları, trolleme, bağımlılık gibi sosyal medya üzerinde (anket uygulama, test çözme vb.) sosyal bilimler temelli çalışmalar bulunmaktadır. Son olarak, çalışmanın kapsamı ile ilgili olmayan çalışmalar bu parametreler incelenerek literatür havuzundan silinmiştir.

2017 sonrasında yapılan çalışmalarda kişilik tahmininde Facebook, Twitter, Instagram, Sina Weibo, Youtube, Google+ gibi sosyal medya platformları kullanılmıştır. Facebook verileri üzerinde yapılan çalışmaların büyük çoğunluğu bir süre önce aktif olarak dağıtılan myPersonality veri setini [58] kullanıyor ve şu anda sadece 250 kullanıcı verisi herkese açık olarak paylaşılmaktadır. Çalışmalarda kullanılan veri setleri çoğunlukla İngilizce yazılmıştır. Beş Faktör Kişilik Envanteri (Big Five Inventory - BFI), yaygın kabul görmesi nedeniyle en yaygın kullanılan kişilik ölçeğidir. Her çalışmada kullanılan veri ön işleme yöntemleri farklılık göstermektedir. Ancak tahmin modelleri için temel istatistiksel yöntemlere ek olarak yapay zeka alanında kullanılan sınıflandırma, regresyon ve kümeleme yöntemleri kullanılmaktadır.

İçerik ve kapsam dikkate alınarak 2017 ve sonrasında elde edilen çalışmaların ayrıntılı bir sunumu kategorik özelliklerine göre bu bölüm içinde sırası ile verilmektedir. İlk olarak sosyal medya platformlarına göre çalışmaların sayıları Çizelge 3.1’de sunulmaktadır.

Çizelge 3.1. 2017 ve sonrasında kişilik özellikleri tahminine yönelik çalışmaların sosyal medya platformlarına göre sayıları

Sosyal Medya Platformu	Çalışmalar	Çalışma Sayısı
Facebook	[59-79]	21
Twitter	[77, 80-88]	10
Instagram	[77, 89, 90]	3
Sina Weibo	[91-98]	8
Çoklu-Platform	[77, 99-103]	6

İncelenen çalışmalarda noktalama işareti temizleme, kök çıkarımı, farklı dillerin dönüştürülmesi gibi birçok farklı ön işleme yöntemi kullanılmaktadır. Ayrıca çalışmalarda kullanılan içerik dilleri çoğunlukla İngilizce olmakla birlikte diğer diller İspanyolca, Almanca, Felemenkçe, İtalyanca, Çince, Portekizce, Fransızca, Arapça, Malayca, Türkçe, Bahasa ve Japonca'dır. Kişilik özellikleri testleri arasında en çok kullanılanı Big-Five olmakla birlikte DISC, NEO-PI-R, HDS, MBTI vb. testler de kullanılmaktadır. Tahmin yöntemleri için sınıflandırma, regresyon ve kümeleme yöntemleri istatistik ve yapay zekada farklı kombinasyonlarda kullanılmaktadır. Kullanılan tahmin yöntemine göre çeşitli test ve değerlendirme yöntemleri (Doğruluk, f-ölçümü, MAE, RMSE vb.) kullanılmıştır.

Facebook platformunun çalışmalarda çoğunluk oluşturmasındaki en büyük etken myPersonality veri setidir. Bu çalışmaların detaylı incelemesi Çizelge 3.2'de sunulmaktadır. Çizelge 3.2'de sunulan özellikler sırası ile Sosyal Medya Platformu (SMP), Kişilik Testi (KT), İçerik Dili (İD), İçerik Türü (İT), Veri Seti Detayı (VSD), Ön İşlem Yöntemleri (ÖİY), Metodlar (MTD), Doğrulama ve Değerlendirme Metrikleri (DvD) ve çalışma sonuçları olarak sunulmaktadır. Ayrıca kullanılan ön işlem yöntemlerinde gösterimi daha anlaşılır hale getirmek amacı ile ön işlem yöntemlerine verilen kodlar aşağıda sunulmaktadır.

ÖİY-a. Noktalama işaretlerinin silinmesi
ÖİY-b. Sayıların silinmesi
ÖİY-c. Metinin küçük harfe dönüştürülmesi
ÖİY-d. Önemsiz kelimelerin (stop words) temizlenmesi
ÖİY-e. Fazla boşlukların temizlenmesi
ÖİY-f. Kök çıkarımı (stemming, lemmatization)
ÖİY-g. Kelime sayısı
ÖİY-h. LIWC özellikleri
ÖİY-i. İçerik sınıflandırması
ÖİY-j. İçerikten bilgi çıkarımı
ÖİY-k. Görüntüden özellik çıkarımı
ÖİY-l. Resimden yüz tespiti

ÖİY-m. Metin olmayan nesneleri filtreleme
ÖİY-n. Metinden özellik çıkarımı
ÖİY-o. Öznitelik seçimi
ÖİY-p. Farklı dillerin dönüşümü
ÖİY-q. Özel karakterlerin silinmesi
ÖİY-r. Kelime yazım düzeltmesi (normalization)
ÖİY-s. Vektörleştirme
ÖİY-t. Metin olmayan nesneleri anlamlı metine dönüştürme
ÖİY-u. PoS (Part of Speech) etiketleme
ÖİY-v. Ayrıştırma/Jetonlaştırma (tokenization)
ÖİY-w. Kelime türü çıkarımı (fiil, nesne vb.)
ÖİY-x. N-grams
ÖİY-y. Emojilerin ve emoticonların anlamlı metine dönüştürme
ÖİY-z. Pozitif-negatif etiketleme

Çizelge 3.2. 2017 ve sonrasında kişilik özellikleri tahminine yönelik çalışmaların detaylı gösterimi

#	Veri Seti Özellikleri	Model	Sonuç
[80]	**SMP:** Twitter **KT:** DISC **İD:** İngilizce **İT:** Tweet **VSD:** Anahtar kelimelere dayalı toplanan milyonlarca tweet	**ÖİY:** a, b, c, d, e, f **MTD:** Kelime sıklığı ve kelime bulutu ağırlığı hesaplanarak, en baskın kelimeleri bulmak için kişilik özellikleri analiz edildi. **DvD:** Her özellik için en az 20 kez hesaplanır.	Her özellik için bulunan en etkili kelimeler sırası ile D: get, new, I: feelings, social, S: help, open ve C:clear, details kelimeleridir.
[59]	**SMP:** Facebook **KT:** Extraversion (E), neuroticism (N) ve shyness (S) **İD:** Malayca ve İngilzice **İT:** Post **VSD:** 170 kullanıcı ve 12573 veri	**ÖİY:** g, h **MTD:** Özellik seçimi Pearson Korelasyonu kullanılarak yapılmış ve ardından WEKA üzerinde ZeroR, Hoeffding tree, Naif Bayes, Logistic MultiClass, J48 ve OneR yöntemleri uygulanmıştır. **DvD:** 10-kat çapraz doğrulama, doğruluk (accuracy), f-ölçüm (f-measure)	Extraversion için Hoeffding tree yöntemi ile doğruluk değeri 67.81, f-ölçüm değeri 0.66 elde edilmiştir. Neuroticism için Naif Bayes ile doğruluk değeri 58.26 ve Shyness için J48 ile doğruluk değeri 70 olarak elde edilmiştir.
[99]	**SMP:** Facebook, Twitter, Youtube ve Makaleler **KT:** NEO-PI-R, Big-Five özellikleri ve Myers-Briggs kişilik türleri **İD:** İngilizce, İspanyolca, Almanca, Flemenkçe, İtalyanca, Portekizce, Fransızca ve Arapça **İT:** Post, Tweet, Vlog, Makale **VSD:** 9918 Facebook gönderisi, 404 vlog ve 7818 Twitter kullanıcısı	**ÖİY:** g **MTD:** SVD (Singular Value Decomposition) ile önerilen yöntem **DvD:** Monte Carlo çapraz doğrulama, doğruluk, f-ölçüm, en iyi oran seçimi (best choice of ratio)	F1-ölçümüne göre Facebook 0.82, Youtube 0.81 ve Makaleler 0.73'dir. Twitter için f-1 ölçüm değerleri dillere göre sırası ile; İngilizce 0.9, Almanca 0.76, İtalyanca 0.9, İspanyolca 0.72, Fransızca 0.86, Portekizce 0.85, Flemenkçe 0.85 ve Arapça 0.76'dır.
[81]	**SMP:** Twitter **KT:** Big-Five özellikleri **İD:** İngilizce **İT:** Tweet, Profil Resmi **VSD:** 54784 Twitter kullanıcısı	**ÖİY:** i, j, k, l **MTD:** Face++ ve EmoVu kullanılarak resimlerdeki yüzlerden bilgi çıkarımı yapılmıştır. Ayrıca, metin tabanlı tweet verilerinden yaş gibi kullanıcıya ilişkin bilgiler tahmin edilmiş ve görüntülerden elde edilen bilgilerle Pearson Korlasyonu kullanılarak analiz edilmiştir. Doğrusal Regresyon (Linear Regression)) ile Elastik Net Düzenlileştirme (Elastic Net Regularization) yöntemi kullanılarak profil resminden kişilik tahmini yapılabileceği önerilmiştir. **DvD:** Pearson Korelasyonu	r>:135

Çizelge 3.2. (devam) 2017 ve sonrasında kişilik özellikleri tahminine yönelik çalışmaların detaylı gösterimi

#	Veri Seti Özellikleri	Model	Sonuç
[82]	**SMP:** Twitter **KT:** NEO-PI-R **İD:** - **İT:** Beğenilen ve gönderilen resimler, Tweet **VSD:** 4000 kullanıcının 1.5 milyon resim verisi	**ÖİY:** k, l **MTD:** Pearson Korelasyonu ile çıkarılan özniteliklerin etkileri hesaplanmaktadır. Doğrusal Regresyon ve Elastik Net Düzenlileştirme yöntemleri kullanılarak resimlerden kişilik tahmini yapılabileceği önerilmiştir. **DvD:** VGG-Net ve Imagga tarafından çıkarılan öznitelikler önerilen öznitelik çıkarımı ile karşılaştırılmıştır. Doğruluk için RMSE ve performans için Pearson Korelasyonu kullanılmaktadır. 10-kat çapraz doğrulama.	Conscientiousness, agreeableness ve neuroticism için resim özelliklerinden daha iyi sonuç elde edilirken openness ve extraversion özellikleri için en iyi sonuçlar metin içeriklerinden elde edilmiştir.
[91]	**SMP:** Sina Weibo **KT:** Big-Five Berkeley-44 **İD:** Çince **İT:** Mikroblog **VSD:** 994 kullanıcı	**ÖİY:** m, n **MTD:** Çok görevli öğrenme (Multi-task learning) ve Sağlam çok görevli öğrenme (Robust multi-task learning - RMTL) yöntemleri önerilmiştir. Önerilen yöntemler Naif Bayes, Lojistik Regresyon, RepTree ve Random Forest yöntemleri karşılaştırılmıştır. **DvD:** Kesinlik (precision), duyarlılık (recall) ve f-ölçüm	f-ölçüm kriterine göre RMTL, test edilen tüm yöntemlerden daha iyi sonuç vermiştir.
[60]	**SMP:** Facebook **KT:** Big-Five **İD:** - **İT:** Beğeni **VSD:** myPersonality veri setinden 92255 kullanıcı verisi	**ÖİY:** - **MTD:** LASSO algoritması **DvD:** MSE, Pearson Korelasyonu	Kişilik özellikleri MSE değeri için O: 0.024, C: 0.030, E: 0.038, A: 0.030 ve N: 0.038. Pearson Korelasyonu için O: 0.38, C: 0.29, E: 0.34, A: 0.22 ve N: 0.27'dir.
[92]	**SMP:** Sina Weibo **KT:** Big-Five **İD:** Çince **İT:** Mikroblog **VSD:** -	**ÖİY:** n, o **MTD:** Kullanıcı tarafından sağlanan bilgilerden özellik seçimi için PCA ve Korelasyon analizi kullanılmıştır. Daha sonra seçilen öznitelikler ile Çoklu Regresyon Modeli (Multiple Regression Model), Gri Tahmin Modeli (Gray Prediction Model) ve Çoklu Görev Modeli (Multi-tasking Model) uygulanmıştır. **DvD:** MAE	Gri Tahmin Modeli kullanılarak MAE ile 0.1-0.2 aralığında daha iyi sonuçlar elde edilmiştir.

Çizelge 3.2. (devam) 2017 ve sonrasında kişilik özellikleri tahminine yönelik çalışmaların detaylı gösterimi

#	Veri Seti Özellikleri	Model	Sonuç
[93]	**SMP:** Sina Weibo **KT:** Big-Five (conscientiousness, extroversion ve agreeableness) **İD:** Çince **İT:** Mikroblog **VSD:** 968854 post	**ÖİY:** - **MTD:** İlk deneyde metindeki duygu, anlam ve içerik türüne göre kural temelli bir kişilik tahmini yaklaşımı önerilmiştir. Önerilen yöntem ile Naif Bayes, SVM, Lojistik Regresyon ve Karar Ağacı yöntemleri karşılaştırılmıştır. İkinci deneyde önerilen kişilik temelli duygu sınıflandırma yöntemi sunulmuştur. **DvD:** Kesinlik, duyarlılık ve f-ölçüm	İlk deneyde elde edilen sonuçlar incelendiğinde önerilen yöntemin kesinlik açısından daha iyi olduğu ve diğer yöntemlerin ise duyarlılık açısından daha iyi sonuç verdiği gözlemlenmiştir.
[61]	**SMP:** Facebook **KT:** Big-Five **İD:** İngilizce **İT:** Post **VSD:** myPersonality veri setinden 250 kullanıcıdan 10000 gönderi alınmış. İkinci veri seti olarak applymagicsauce.com'dan 150 kullanıcı etiketlenmiştir.	**ÖİY:** a, b, c, d, e, f, p **MTD:** Derin öğrenme yöntemleri (MLP, LSTM, GRU ve CNN) ve geleneksel makine öğrenme yöntemleri (Naif Bayes, SVM, Lojistik Regresyon, Gradient Boosting ve Lineer Diskriminant Analizi (Linear Discriminant Analysis) yöntemleri karşılaştırılmaktadır. **DvD:** Doğruluk, 10-kat çapraz doğrulama	myPersonality veri setinde en iyi sonuçlar MLP ile elde edilmiş olup ikinci veri setinde MLP ve CNN yöntemleri ile elde edilmiştir.
[62]	**SMP:** Facebook **KT:** Big-Five **İD:** İngilizce **İT:** Post **VSD:** myPersonality, 250 kullanıcı	**ÖİY:** a, c, d, e, o, q, r, s, t, u, v **MTD:** TF-IDF ile vektörleştirme sonrası PCA ile öznitelik seçimi yapılmıştır. Sonrasında KNN ve SVM sınıflandırma yöntemleri kullanılmıştır. **DvD:** KNN için k değeri, kesinlik, duyarlılık, f-ölçüm	KNN yöntemi için en iyi k değeri her kişilik özelliğinde farklı olarak tespit edilmiştir. SVM yönteminde kişilik özelliklerinin tahminine dayalı başarım PCA kullanılıp/kullanılmama durumuna göre değişiklik göstermektedir.
[100]	**SMP:** Facebook, Twitter, Google+ ve diğerleri **KT:** Big-Five **İD:** İngilizce **İT:** Metinsel gönderiler **VSD:** myPersonality ve diğerleri	**ÖİY:** a, b, c, d, e, f, o, s, u, v **MTD:** Multinomial Naif Bayes, KNN ve SVM yöntemleri kullanılmıştır. **DvD:** -	Önerilen yöntemlerle yapılan tahmin süreçlerinde her yöntemin farklı durumlarda farklı başarı oranları göstereceği sunulmuştur. Bu nedenle, duruma göre Çoğunluk Oylama Yöntemi (Majority Voting Method) ile en doğru sonucun alınabileceği sunulmuştur.

Çizelge 3.2. (devam) 2017 ve sonrasında kişilik özellikleri tahminine yönelik çalışmaların detaylı gösterimi

#	Veri Seti Özellikleri	Model	Sonuç
[63]	**SMP:** Facebook **KT:** Hogan Development Survey (HDS) **İD:** İngilizce **İT:** Post **VSD:** 51712 kullanıcı	**ÖİY:** h, n, w **MTD:** Sıradan En Küçük Kareler (Ordinary Least Squares - OLS) ve Lasso algoritmaları kullanılmıştır. **DvD:** k-kat çapraz doğrulama, en küçük (min), en büyük (max), ortalama (mean), standart sapma (standard deviation), skew, kurtosis, standart hata (standard error)	Excitable, Dutiful ve Bold özelliklerinin dil ile güçlü bir ilişkisi olduğu ve Cautious, Colorful ve Leisurely özelliklerinin dil ile daha zayıf bir ilişkisi olduğu tespit edilmiştir.
[89]	**SMP:** Instagram **KT:** Big-Five-44 **İD:** İngilizce **İT:** İçerik özellikleri ve Resim **VSD:** 193 kullanıcıdan 54962 resim	**ÖİY:** k, n **MTD:** Karşılaştırma için ZeroR yönteminin yanı sıra M5' kuralları, Random Forest ve Radial Basis Function Network yöntemleri kullanılmıştır. **DvD:** 10-kat çapraz doğrulama, ortalama karekök hatası (Root Mean Square Error - RMSE)	Elde edilen sonuçlara bakıldığında hem metinde hem de resimde farklı kişilik özelliklerinde farklı başarı oranları gözlemlenmiştir.
[64]	**SMP:** Facebook **KT:** Big-Five **İD:** İngilizce **İT:** Post **VSD:** myPersonality	**ÖİY:** d, s, u, v, w **MTD:** Anlamsal benzerlik ölçümü için WordNet kullanılmıştır. Girdiler 3 farklı vektör olarak oluşturulmuş ve her vektör belirtilen yöntemle test edilmiştir. Vektörler vec1 (İsim, Sıfat, Fiil ve Zamir), vec2 (Tüm İsimler), vec3 (Duygusal etkiye sahip isimler – NRC duygu sözlüğü) vektörleri, WordNet'te bulunan JCN ve Yol Uzunluğu (Path Length) benzerlik ölçüm yöntemleri kullanılarak karşılaştırılmıştır. **DvD:** Doğruluk, keskinlik, duyarlılık ve f-ölçüm	Tüm kişilik özellikleri esas alındığında JCN ölçüm yönteminin genel olarak daha başarılı olduğu görülmüştür. Ayrıca vec1 olarak oluşturulan vektörün doğruluk (0,64) ve f-ölçüm (0,65) değerleri ile diğer vektörlere göre daha iyi olduğu görülmüştür.
[65]	**SMP:** Facebook **KT:** Big-Five **İD:** İngilizce **İT:** Post **VSD:** myPersonality, 115872 kullanıcı	**ÖİY:** d, h, m, n, s **MTD:** LDA ve LIWC özellikleri ile farklı veri setleri oluşturulmuştur. Bu veri setleri Lineer Regresyon, Destek Vektör Regresyon (Support Vector Regression - SVR) (Linear (L-SVR), Polinom (P-SVR) ve Radyal Temel Fonksiyon (RBF-SVR)) yöntemleri kullanılarak test edilmiştir. **DvD:** MSE	Deneylerde P-SVR ve RBF-SVR'nin diğer yöntemlere göre genel olarak daha başarılı olduğu belirlenmiştir.

Çizelge 3.2. (devam) 2017 ve sonrasında kişilik özellikleri tahminine yönelik çalışmaların detaylı gösterimi

#	Veri Seti Özellikleri	Model	Sonuç
[90]	**SMP:** Instagram **KT:** Big-Five **İD:** - **İT:** Resim **VSD:** 179 üniversite öğrencisinin hesabından 25394 resim	**ÖİY:** k, l **MTD:** Pearson Korelasyonu **DvD:** RMSE	Sonuçlar, Extraversion kişilik özelliğinin cinsiyetle, Extraversion, Agreeableness ve Openness ölçeklerinin ise resimlerdeki yüzlerin sayısı ve duygu durumlarıyla ilişkili olduğunu göstermiştir.
[66]	**SMP:** Facebook **KT:** Big-Five **İD:** İngilizce **İT:** Post **VSD:** myPersonality	**ÖİY:** h, n, s, v **MTD:** KNN, Naif Bayes ve Karar Ağacı yöntemleri 3 farklı özellik seti (1:4 genel özellik, 2:1 + psikolojik özellikler, 3:1 + stil özellikleri) oluşturularak kullanılmıştır. Tüm özelliklerin tartışıldığı veri kümesine ek olarak, Pearson korelasyon ve Parçacık Sürü Optimizasyonu (Particle Swarm Optimization) yöntemleri kullanılarak elde edilen 3 veri kümesi üzerinde testler yapılmıştır. **DvD:** f-ölçüm, 10-kat çapraz doğrulama	Özellik türlerine göre yapılan deneylerde psikolojik ve biçimsel özellikleri içeren özellik setleri ile her ölçekte farklı başarı oranları elde edilmiştir. Ayrıca, öznitelik seçimine dayalı analizde Parçacık Sürü Optimizasyonu ile belirlenen öznitelik setinin büyük başarı oranları ile daha iyi olduğu belirlenmiştir.
[83]	**SMP:** Twitter **KT:** MBTI **İD:** - **İT:** Tweet **VSD:** 450 twitter kullanıcısından 1 milyonun üzerinde tweet toplanmıştır	**ÖİY:** h, n **MTD:** Twitter verileri ile kişilik testi sonuçları arasında Pearson korelasyon analizi yapıldı. Daha sonra elde edilen öznitelikler ile oluşturulan veri setinde Yapay Sinir Ağı ve Lojistik Regresyon yöntemleri kullanılmıştır. **DvD:** Doğruluk	Twitter verileri ile birçok kişilik testi özelliği arasındaki ilişkinin yüksek olduğu belirlendi. Sınıflandırma yöntemlerinin test edilmesi sonucunda doğruluk oranları farklı ölçekler için %51 ile %59 arasında değişmiştir.
[67]	**SMP:** Facebook **KT:** Big-Five **İD:** İngilizce **İT:** Post, kullanıcı bilgisi (ağ boyutu, aradalık, yoğunluk, aracılık ve geçişlilik) (network size, betweenness, density, brokerage ve transitivity) **VSD:** myPersonality	**ÖİY:** h, m, o, s, v **MTD:** Özellik seçimi için SNA, LIWC, SPLICE ve bunların kombinasyonları kullanılmıştır. Elde edilen 4 farklı özellik seti ile XGBoost, Lineer Regresyon, Gradient Boost ve SVM yöntemleri kullanılmıştır. **DvD:** Pearson Korelasyonu, Doğruluk	4 farklı özellik seti üzerinde yapılan testlerde XGBoost, diğer yöntemlere göre daha yüksek bir başarı oranı göstermiştir. En yüksek başarı oranı SNA ile elde edilen öznitelik setinde görülmüştür.

Çizelge 3.2. (devam) 2017 ve sonrasında kişilik özellikleri tahminine yönelik çalışmaların detaylı gösterimi

#	Veri Seti Özellikleri	Model	Sonuç
[68]	**SMP:** Facebook **KT:** Big-Five **İD:** English **İT:** Post **VSD:** myPersonality, 115872 kullanıcı	**ÖİY:** e, o, r, s, v **MTD:** İstatistiksel Dil özellikleri ve Doküman vektörü (Doc2Vec) ile birlikte AttRCNN ile elde edilen özellik vektörü, RCNN-CNN ve CNN'den elde edilen vektörlerden oluşan özellik kümeleri kombinasyonları Support Vector Regression (SVR), Gradient Boosting Regression (GBR), Random Forest ve MLP yöntemleri ile test edilmiştir. **DvD:** MAE, 5-kat çapraz doğrulama	Yapılan testler sonucunda AttRCNN ile elde edilen özellik ve kombinasyonları ile en iyi başarımlar gözlemlenmiştir. Tahmin yöntemlerinde ise en iyi başarım SVR ve GBR ile elde edilmiştir.
[69]	**SMP:** Facebook **KT:** Big-Five **İD:** İngilizce **İT:** Post **VSD:** myPersonality	**ÖİY:** e, f, h, n **MTD:** Veri setinden özellik çıkarımında psiko dil özellikleri ve dil özelliklerine dayalı özellik kümeleri oluşturulmuştur. Bu özellik kümeleri üzerinde Naif Bayes, Karar Ağacı, Random Forest, Lineer Regresyon (Simple Logistic Regression - LogitBoost) ve SVM (Sequential Minimal Optimization) yöntemleri kullanılmıştır. **DvD:** Keskinlik, duyarlılık, f-ölçüm, 10-kat çapraz doğrulama	Uygulanan tahmin yöntemlerinden Naif Bayes psiko dil özellikleri ile en iyi başarımı göstermiştir.
[70]	**SMP:** Facebook **KT:** Big-Five **İD:** İngilizce **İT:** Post **VSD:** myPersonality veri seti ve 9856 etiketsiz kullanıcı veri seti	**ÖİY:** h, n, s, v **MTD:** Büyük etiketsiz verilerdeki avantajından dolayı Pseudo Multi-view Co-training (PMC) yöntemi önerilmiştir. İlk deneyde LIWC ile elde edilen özellikler ile PMC yöntemi eğitilmiştir. İkinci deneyde ise hem LIWC hemde unigram ile elde edilen özellikler ile PMC yöntemi uygulanmıştır. **DvD:** Keskinlik, duyarlılık, f-ölçüm	Kişilik özelliği tahmininde ortalama f-ölçüm değeri Deney 1'de 0.66 ve Deney 2'de 0.67'dir. Önerilen yöntem, karşılaştırılan diğer yöntemlerden tüm ölçüt değerlerinde daha iyi sonuç göstermiştir.
[84]	**SMP:** Twitter **KT:** Big-Five **İD:** Endonezya Dili **İT:** Varlık miktarları (tweet, takipçi, takip edilen, retweet, cevap ve hashtag) **VSD:** 5 kişilik özelliğine dayalı olarak Yüksek: H ve Düşük: L etiketli yayınlanmamış el yazmasından 508 twitter verisi.	**ÖİY:** j, n, x **MTD:** KNN, J48, Random Forest, SVM, Naif Bayes **DvD:** Duyarlılık, keskinlik, f-ölçüm	En iyi f-ölçüm değeri 0.744 olarak Random Forest yöntemi ile elde edilmiştir.

Çizelge 3.2. (devam) 2017 ve sonrasında kişilik özellikleri tahminine yönelik çalışmaların detaylı gösterimi

#	Veri Seti Özellikleri	Model	Sonuç
[85]	**SMP:** Twitter **KT:** Big-Five **İD:** Türkçe **İT:** Tweet **VSD:** 40 gönüllüye ait veriler ground data olarak 2000 rastgele kullanıcı ise kümeleme testi amacı ile kullanılmaktadır.	**ÖİY:** a, f, m, o, q, s, u **MTD:** K-means, Agglomerative Kümeleme **DvD:** Hata oranı (error rate), siluet katsayısı (silhouette coefficient)	K-means kümelemenin Word2Vec ile en iyi siluet katsayı skorlarına sahip olduğu söylenmektedir. En dengeli sonuçlar ise k=2 değeri ile elde edilmiştir.
[71]	**SMP:** Facebook **KT:** Big-Five **İD:** İngilizce **İT:** Post **VSD:** myPersonality	**ÖİY:** a, b, e, f, g, j, n, w **MTD:** Naif Bayes, Decision Tree, Random Forest, Simple Logistic Regression (SLR), SVM. Öznitelik Seçim Yöntemleri olarak: Pearson Correlation Coefficient, Corelation-based feature subset, information gain, symmetric uncertainly ve chi-squared **DvD:** Doğruluk	En iyi doğruluk oranı 5 kişilik özelliği ortalaması olarak sırası ile SVM %59.6, SLR %58.28 ve RF %57.38 olarak sunulmuştur. Pearson Korelasyon temelli öznitelik seçimi ile Naif Bayes ile %61.89 ile %70.08 arasında elde edilmiştir.
[101]	**SMP:** Twitter, Facebook ve Youtube **KT:** Big-Five **İD:** İngilizce **İT:** Metin içerikleri **VSD:** myPersonality-500 kullanıcı, Youtube_Personality – 400 kayıt ve PAN2015_İngilizce veri seti 294 kullanıcı	**ÖİY:** h, j, n, y, z **MTD:** Personality2vec (önerilen yöntem), adawalk, node2vec, deepwalk, AttRCNN-CNN, CNN, Mairesse, Doc2Vec ve Random Guess **DvD:** MAE	Tüm veri kümelerinde en iyi MAE değeri önerilen yöntem personality2vec ile elde edilmiştir. myPersonality için 0.4852-0.5971, Youtube_Personality için 0.5824-0.7213 ve PAN2015_İngilizce için 0.1007-0.1425 olarak elde edilmiştir.
[94]	**SMP:** Sina Weibo **KT:** NEO-FF-60 **İD:** Çince **İT:** Mikroblog **VSD:** 3.3G mikroblog metin havuzuna sahip veri kümesi ve 400 katılımcılı Sina Weibo veri kümesi	**ÖİY:** h, j, n, s **MTD:** Logistic Regression, SVM, Random Forest. Öznitelik Kategorisi: CLIWC, BOW, LSA+SOA, Personality Lexicon (PL), PL_Hownet, PL_Tongyici, PL_ITH **DvD:** f-ölçüm	En iyi f-ölçüm değeri ortalama olarak PL_ITH öznitelik kategorisinde Random Forest yöntemi ile 0.704 olarak elde edilmiştir.

Çizelge 3.2. (devam) 2017 ve sonrasında kişilik özellikleri tahminine yönelik çalışmaların detaylı gösterimi

#	Veri Seti Özellikleri	Model	Sonuç
[102]	**SMP:** Sosyal medya platformları **KT:** MBTI **İD:** İngilizce **İT:** Metin temelli içerik **VSD:** 8675 kayıttan son 50 gönderi ile 422845 kayıt içeren MBTI veri seti	**ÖİY:** d, f, s, v **MTD:** XGboost (önerilen yöntem), NB, KNN, Decision Tree, Random Forest, MLP, SVM, MNB, Logistic Regression, SGD Veri setinde yeniden örnekleme yöntemleri (resampling methods) kullanılmıştır. **DvD:** Doğruluk, keskinlik, duyarlılık, f-ölçüm	Xgboost yönteminin yapılan tüm testlerde çok iyi değerlere sahip olduğu görülmüştür. Özellikle yeniden örnekleme yöntemleri ile elde edilen sonuçlar XGboost ile %100'e yakın sonuçlar sunmaktadır.
[72]	**SMP:** Facebook **KT:** Big-Five **İD:** İngilizce **İT:** Post ve Beğeni **VSD:** myPersonality veri seti, 22000 kullanıcı 3183816 post ve 301105 beğeni	**ÖİY:** s, t **MTD:** Lojistik Regresyon, Support Vector Regression, Derin Sinir Ağı (DNN) kullanılmaktadır. Girdi olarak Doküman Vektörü (DV) ve Beğeni Vektörü (LV) olarak ele alınır. **DvD:** MAE	En iyi MAE değerleri DV+LV ile DNN yöntemi ile 5 kişilik özelliğinden 0.5038-0.-6354 arasında sonuç vermektedir.
[103]	**SMP:** Youtube, Facebook, Twitter **KT:** Big-Five **İD:** İngilizce **İT:** Metin temelli içerik **VSD:** Youtube Personality, myPersonality, PAN2015 ve OpenPsychometrics	**ÖİY:** s **MTD:** AdaWalk, node2vec, Deep2Walk, 2CLSTMs, doc2vec, Kampman, mairesse, TFIDF, Wei, random guess **DvD:** RMSE	4 veri seti üzerinde yapılan denemelerde en iyi RMSE değerleri genel olarak AdaWalk ile elde edilmiştir.
[95, 96]	**SMP:** Sina Weibo **KT:** Dört kısa cevaplı soru (Four short-answer questions), Proaktif Kişilik Anketi (Proactive Personality Questionnaire) **İD:** Çince **İT:** Mikroblog **VSD:** 901 katılımcıdan 4955 post ve Weibo metin veri seti	**ÖİY:** d, j, n, o **MTD:** SVM, XGboost, KNN, NB, LR **DvD:** Doğruluk, f-ölçüm, SEN, SPE, PPV, NPV, AUC	En iyi sonuçlar her iki veri kümesinin kullanımı ile elde edilmiştir. SVM ve NB yöntemleri en iyi sonuçlara sahip olmuştur.

Çizelge 3.2. (devam) 2017 ve sonrasında kişilik özellikleri tahminine yönelik çalışmaların detaylı gösterimi

#	Veri Seti Özellikleri	Model	Sonuç
[73]	**SMP:** Facebook **KT:** Big-Five **İD:** İngilizce **İT:** Beğeni ve beğeninin yapıldığı varlık türü **VSD:** myPersonality, 8530 beğeni ile 19700 Facebook kullanıcı verisi	**ÖİY:** t **MTD:** LR, AttRNN, BiGRU **DvD:** Accuracy	The best results were obtained with the AttRNN method with an accuracy rate of 0.29-0.48.
[74]	**SMP:** Facebook **KT:** Big-Five **İD:** İngilizce **İT:** Post **VSD:** 6893 eğitim veri seti, 3024 test veri seti	**ÖİY:** s **MTD:** Bayesian Network, Random Forest, SVM, Attention-based LSTM **DvD:** Duyarlılık, keskinlik, f-ölçüm	En iyi f-ölçüm oranı Attention-based LSTM yöntemi ile %72.2'dir.
[97]	**SMP:** Sina Weibo **KT:** DISC **İD:** Çince **İT:** Mikroblog **VSD:** 198 kullanıcı, her kullanıcının sadece 30 sayfası	**ÖİY:** n, o **MTD:** NB, Bayes Net, KNN (k=1,3,5), J48, LibSVM **DvD:** 10-kat çapraz doğrulama, duyarlılık, keskinlik, f-ölçüm, MVV, ROC, PRC, doğruluk	En iyi f-ölçüm oranı KNN (k=1) ile %80,5'dir. Diğer değerlendirme kriterlerinde ise en iyi sonuçlar KNN yöntemi ile elde edilmiştir.
[75]	**SMP:** Facebook **KT:** Big-Five **İD:** İngilizce **İT:** Beğeni, etiket, güncellemeler, olaylar, gruplar, arkadaşlar, doğum günü, yaş, cinsiyet, ilişki durumu, ilgiler, zaman dilimi ve ağ boyutu **VSD:** 7438 kullanıcı, myPersonality	**ÖİY:** o, s **MTD:** Yapay Sinir Ağı (Artificial Neural Network) **DvD:** 10-kat çapraz doğrulama, doğruluk	10 kat çapraz doğrulama olan ve olmayan iki farklı şema oluşturulmuştur. En iyi sonuçlar 10 kat çapraz doğrulama kullanılarak elde edilmiştir.

Çizelge 3.2. (devam) 2017 ve sonrasında kişilik özellikleri tahminine yönelik çalışmaların detaylı gösterimi

#	Veri Seti Özellikleri	Model	Sonuç
[76]	**SMP:** Facebook **KT:** Big-Five **İD:** İngilizce **İT:** Post ve Post Kategorisi **VSD:** 100 kullanıcı, 38 kategoride sınıflandırılmış 27182 post	**ÖİY:** a, b, c, d, i, j, n, o, q, s **MTD:** Linear Support Vector Classification (LSVC), Logistic Regression (LR), The multinomial Naive Bayes classifier (MNB), Random Forest Classifier (RFC) **DvD:** Doğruluk, %80 eğitim veri seti, %20 test veri seti	En iyi başarım LSVC ile %89.37 accuracy değeri olarak elde edilmiştir.
[77]	**SMP:** Facebook, Twitter, Instagram **KT:** Big-Five **İD:** İngilizce ve Bahasa **İT:** Post ve Tweet **VSD:** myPersonality (250 kullanıcı), Twitter (502 kullanıcı, 46238 tweet)	**ÖİY:** c, d, e, f, h, j, q, t, y **MTD:** Term Frequency & Inverse Gravity Moment (TF-IGM) ile Binary-Partitioning Transformer (BPT) yöntemi **DvD:** f1-ölçüm, doğruluk	LIWC ve SPLICE ile metodlar karşılaştırılmış. Facebook için en iyi sonuç BPT+TF-IGM ile f1-score 0.762 ve accuracy %78.34 olarak elde edilmiştir. Twitter için en iyi sonuç yine önerilen yöntem ile f1-score 0.783 ve accuracy %79.67 olarak elde edilmiştir. Instagram için en iyi sonuç f1-score 0.821 ve accuracy %86.84 olarak elde edilmiş.
[86]	**SMP:** Twitter **KT:** Big-Five, Anxiety ve Avoidance **İD:** İngilizce **İT:** Tweet ve Takipçiler **VSD:** 229 kullanıcı, 350000 tweet	**ÖİY:** a, b, d, f, j, m, n, o, q, u, v, x **MTD:** Random Forest, Porposed Holistic Model and Independent Model **DvD:** RMSE, Siluet Skoru (Silhouette Score)	Önerilen holistic modelin en iyi özelliklerle RMSE değeri 0.192 olarak elde edilmiştir.
[87]	**SMP:** Twitter **KT:** MBTI **İD:** Japonca **İT:** Tweet **VSD:** -	**ÖİY:** i, n **MTD:** BERT, BERT-CLS, BERT-AVG, BoW-SVD, DistilBERT, VM, SMOTE, ADASYN, borderline SMOTE, SVM-SMOTE, SMOTE-ENN **DvD:** Keskinlik, duyarlılık, f1-ölçüm doğruluk, macro-f1, weighted-f1	Önerilen yöntem ile kullanıcı kişilik tahmini %70 accuracy olarak tahmin edilebilir.

Çizelge 3.2. (devam) 2017 ve sonrasında kişilik özellikleri tahminine yönelik çalışmaların detaylı gösterimi

#	Veri Seti Özellikleri	Model	Sonuç
[88]	**SMP:** Twitter **KT:** MBTI **İD:** İngilizce **İT:** Tweet **VSD:** (MBTI) Myers-Briggs Personality Type Dataset, 8675 kayıt	**ÖİY:** a, b, c, d, h, j, m, n, q, r, s **MTD:** Linear Support Vector, Support Vector, Logistic Regression, Random Forest, Decision Tree ve Multinomial Naive Bayes **DvD:** MSE, doğruluk, f1-score	Linear Support Vector ile en iyi accuracy 0.661095 olarak elde edilmiştir. Referans alınan 2017 makalesinde Logistic regression ile 0.6659 başarım elde edilmiş.
[78]	**SMP:** Facebook **KT:** Big-Five **İD:** İngilizce **İT:** Post **VSD:** myPersonality	**ÖİY:** d, f, h, j, m, n, o, s **MTD:** Önerilen yöntemler ve geleneksel yöntemler: KNN, NB, DT **DvD:** 10-kat çapraz doğrulama, duyarlılık, keskinlik, f1-ölçüm	Psikolojik karakteristik özellik seti ile LIWC'den daha iyi başarımlar elde edilmiş olup, geleneksel modellerde başarım (normal veya stil özellikleri ile) en iyi KNN'de, çalışmada önerilen modellerde başarım (normal özellikler ile) NB'de. KNN ve NB'de stil özellikleri ile uygulama yapıldığında başarım diğer özellik veri kümelerine göre artmakta.
[98]	**SMP:** Sina Weibo **KT:** Big-Five **İD:** Çince **İT:** Mikroblog **VSD:** microblogPCU veri seti (432 kullanıcı, 40000 post), toplanan veri seti (187 kullanıcı, 13709 post)	**ÖİY:** j, n, s **MTD:** BERT, Decision Tree, AdaBoost, Linear Discriminant Analysis (LDA), Baggin classifier, Random Forest, Fully Connection Network (FC), MLP, SVM, Logit **DvD:** K-means, Earth Mover (EM) Mesafesi, Jensen-Shannon (JC) Mesafesi, Duyarlılık, Keskinlik, F1-ölçüm, Doğruluk	İlk deneyde BERT kullanılarak EM distance ile yapılan kümelemede yüksek başarım gözlendi. İkinci deneyde BERT ile tüm metriklerde en yüksek performans FC ile elde edildi. Üçüncü deneyde en iyi başarım MLP de gözlemlendi.
[79]	**SMP:** Facebook **KT:** Big-Five **İD:** İngilizce **İT:** Post **VSD:** myPersonality (250 kullanıcı, 9917 post)	**ÖİY:** a, c, f, m, r, s, y **MTD:** Attention based BiLSTM Neural Network, SVM, RF, KNN, MLP with LIWC, Doc2vec, BERT, Basic/Word Level/Sentence Level **DvD:** Doğruluk, f1-ölçüm	En iyi başarımlar Word Level/Sentence Level Attention BiLSTM metodu ile elde edilmiştir. Word Level için accuracy 0.64, f1-score 0.67 ve Sentence Level için accuracy 0.68, f1-score 0.65'dir.

3.3. Diğer Çalışmalar

Literatürde kişilik özellikleri tahminine yönelik çalışmalar yıllardır üzerinde çalışılan konulardandır. 2017 öncesi yapılan çalışmaları inceleyen ve meta analizini sunan çalışmada [104] IPIP, Big-Five ve Big Five Inventory 10 kişilik özellikleri testleri kullanıldığı sunulmaktadır. Facebook sosyal medya platformu yoğunlukla kullanılmakla birlikte Twitter, Sina Weibo ve Instagram da çalışılan sosyal medya platformları arasındadır. Çalışmaların yarısı myPersonality veri setini kullanan çalışmalardır. Bunun dışında kullanıcıların sosyal medya üzerindeki aktivite, resim, demografi, dil ve beğeni gibi dijital ayak izleri kullanılmaktadır.

PAN, dijital metin adli analizi ve stilometri üzerine paylaşımlı görevlerin işletildiği ve WEBIS tarafından organize edilen bilimsel etkinlik serisidir. 2015 yılında yapılan bir paylaşımlı görevde 4 farklı dilde (İngilizce, Flemenkçe, İtalyanca ve İspanyolca) twitter verilerinden oluşturulan bir kişilik özellikleri veri seti kullanılmıştır [105]. Bu veri seti eğitim ve test verisetleri olarak ayrı sunulmuştur. İngilizce, Flemenkçe, İtalyanca ve İspanyolca dillerinde eğitim veri seti için sırasıyla 152, 34, 38 ve 100 kullanıcı verisi, test veri seti için sırasıyla 142, 32, 36 ve 88 kullanıcı verisi oluşturulmuştur. Big-five ölçeğine dayalı olarak ilgili kişilik özellikleri -0.5 ile 0.5 arasında değerlere sahiptir. Kişilik özellikleri tahmininde başarım ölçütü olarak da RMSE kullanılması istenmiş ve nihai değer olarak ortalama RMSE değeri alınmıştır. Çalışmaya katkı sağlayan her grup farklı başarımlar elde etmekle birlikte en iyi başarım 3 dil için tek grup tarafından elde edilmiştir. İlgili çalışmada Second Order Attributes (SOA) ve Latent Semantic Analysis (LSA)'nın birleşimden oluşmaktadır [106]. Bag-of-words (BOW), SOA ve LSA yöntemleri yazarların önerdiği yöntem ile karşılaştırılmıştır. Elde edilen sonuçlara göre önerilen yöntemin daha iyi değerlere sahip olduğu sunulmuştur.

Literatürdeki resim temelli çalışmalara örnek olarak Jia Xu ve ark. [107], 2D ve 2.5D profil resimleri üzerinde S-NNPP (Kişilik Tahmini için Yumuşak Eşik Tabanlı Sinir Ağı - Soft Threshold-Based Neural Network for Personality Prediction) yöntemini kullanarak bir kişilik özelliği tahmin yöntemi sunmuştur. Karşılaştırma yöntemi olarak Geleneksel BP, MobileNetV2 ve ResNeSt50 yöntemleri kullanılmıştır. Değerlendirme kriterleri olarak TPR, FPR, F-measure ve ROC Curve yöntemlerini kullanmışlardır. Çalışmada en iyi performans S-NNPP yöntemi kullanılarak 2.5D profil resimleri ile elde edilmiştir.

3.4. Motivasyon ve Sınırların Belirlenmesi

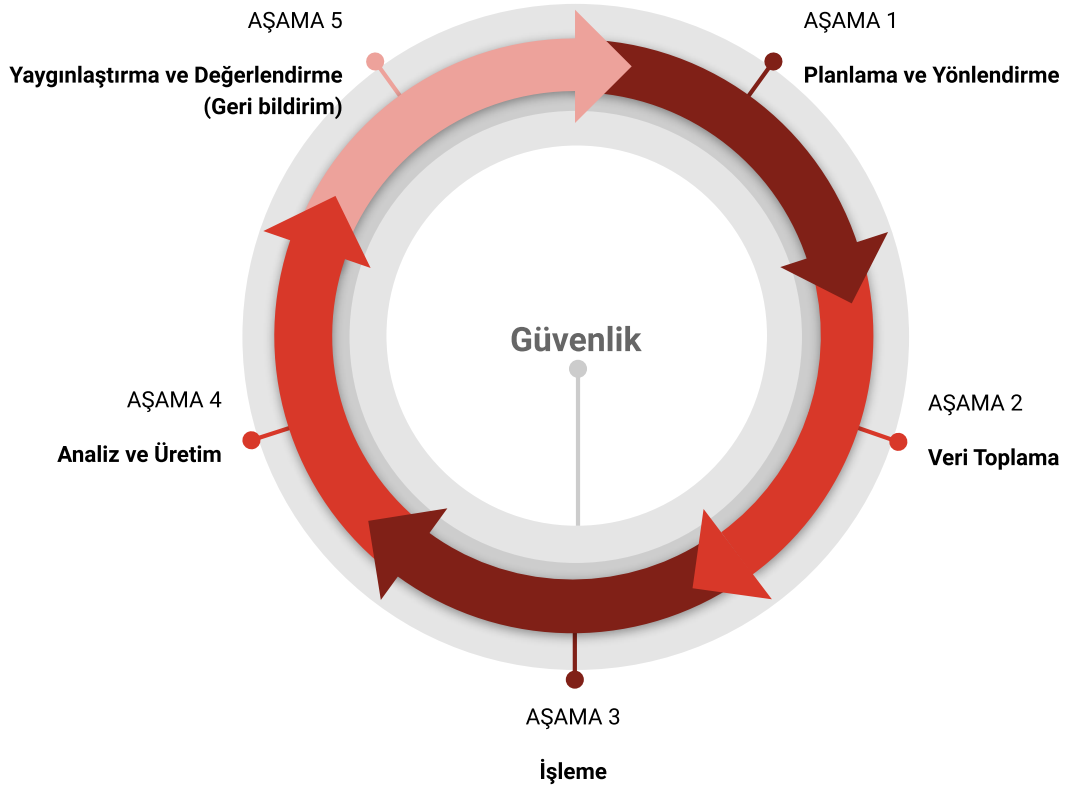
Kişilik özelliklerinin sosyal medya içeriklerinden çıkarımı işe alım süreçlerinde, danışmanlık merkezlerinde, çevrimiçi pazarlamada, kurumsal kullanımda, psikolojik profil oluşturmada, psikolojik kişiye göre tasarım ve içerik sunumunda, öneri sistemlerinde, antisosyallik ve suça yatkınlığın ön değerlendirmesinin yapılabileceği tüm ortamlar için kullanılabilir. [5]

İncelenen tüm çalışmalardan hareketle her bir sosyal medya ortamı için yapılan yapısal analiz ile ön işleme ile oluşturulan eğitim ve test veri setlerinin çoğunlukla diğer sosyal medya ortamları için uygun olmadığı tespit edilmiştir. Bu noktada her bir sosyal medya ortamının kendi içindeki yapısı analiz edilerek ve yapısal olarak anlamlı/anlamsız tüm oluşumlar belirlenerek en genel yapısal dönüşümlerin gerekliliği belirlenmiştir. Ayrıca veri ön işlemedeki çeşitliliğin model eğitimindeki farklılıklara etkisi tam olarak vurgulanmamış ve bu bağlamda karşılaştırmalı analiz eksikliği tespit edilmiştir. Aynı zamanda, yapılan araştırmalar neticesinde bulunan çalışmalar arasında Türkçe dilindeki içeriklerden kişilik tahmininin oluşturulan yeni bir veri seti ile yapılmadığı görülmektedir. Aynı zamanda Türkçe kişilik özellikleri tahminine dayalı bir veri seti de bulunmamıştır. Türkçe yapısı gereği matematiksel çözümlemesi zor olan bir dil olduğundan ön işlem adımlarında problemler oluşabilmektedir. Burdan yola çıkarak Türkçe içeriklerden oluşan bir kişilik veri seti oluşturulması ve bu veri seti üzerinden önişlem ve eğitim yolu ile kişilik tahmini çalışması yapılmıştır. Yapılan analizlerin daha iyi başarımlar göstereceği öngörülmekle birlikte, en kötü başarım ölçütü literatürde sunulan diğer dil içeriklerinden oluşan tahmin oranlarına yakın bir başarı sağlamak olmuştur.

Tez çalışmasında temel amacımız Türkçe içeriklerden kişilik özellikleri tahmini olsa da, veri toplama aşamasında geçen süreyi değerlendirmek için hem süreci yeniden modellemek hem de İngilizce veri setlerine ulaşım zorluğu nedeni ile İngilizce içeriklerden kişilik özellikleri tahminine yönelik bir çalışma yapmak olmuştur. Diğer bir deyiş ile amacımız hem Türkçe içeriklerden kişilik özellikleri tahminine yönelik farklı bir bakış açısı ortaya koymak hem de İngilizce içeriklerden kişilik özellikleri tahminine yönelik yeni bir veri seti ve yeni bir model sistem tasarımı sunmaktır.

Tez çalışması süresince yapılan işlemler farklı kaynaklardan elde edilen OSINT yaşam döngüleri temel alınarak yeni bir yaşam döngüsü üzerinden gelecek çalışmalara kaynak olması amacı ile sunulmuştur. Bu yaşam döngüsü Şekil 3.1'de ve döngüye ait alt başlıklar aşağıda sunulmaktadır:

- Planlama ve Yönlendirme
- Güvenlik
- Veri Toplama
- İşleme
- Analiz ve Üretim
- Yaygınlaştırma ve Değerlendirme (Geri bildirim)



Şekil 3.1. Sunulan OSINT yaşam döngüsü

3.4.1. Planlama ve yönlendirme

İlk olarak problem tanımının açık olması gerekmektedir. Ana araştırma problemi Türkçe içerik ortamlarından kişilik tahmini yapılabileceğidir. Toplanacak bilgilerin hangi ortamlardan toplanacağı belirlenmeli ve toplanacak bilgilerin türleri tanımlanmalıdır. Ayrıca kullanıcılar üzerinden ölçülecek kişilik modeli, alt ölçütler ve kişilik testleri belirlenerek bu testlerin türkçe uyarlamaları elde edilmelidir.

IPIP-120 ve IPIP-300 kişilik testleri ile daha fazla kişilik özelliği tahmin edilmekle birlikte soru sayıları fazla olduğundan kullanıcıların tercih etmeyebilecekleri öngörülmektedir. Bunun yanısıra BFI-10 (Big-Five Inventory), TIPI ve MINI-IPIP testleri soru sayısı bakımından daha az olmakla birlikte 5 ana ölçeğe göre kişiliği tahmin etmek amacı ile kullanılabilmektedir. Bu bilgilerden yola çıkarak kullanıcılara sunulacak anketlerde çeşitlilik sunulması ve en fazla veri toplanan kişilik envanteri üzerinden çalışmaya devam edilmesi daha anlamlı olacaktır. Ek olarak kullanıcıların aktif olarak kullandıkları ve içerik ürettikleri sosyal medya, blog gibi ortamların çeşitliliği düşünüldüğünde ana problem ortamının elde edilecek veri miktarındaki anlamlı ve işlenebilir veri çokluğuna göre şekillenmesi daha doğru bir karar olarak düşünülmektedir.

Yapılan tartışma üzerinden gereksinim olarak belirlenen kişilik envanterleri seçilerek ilgili dile özgü test yapısı elde edilmelidir. Bu testlerin oyunlaştırma kullanılarak kullanıcılara sunulması ve bu etkileşimle yayılımının gerçekleştirilmesi, katılımcı sayısının arttırılması amacı ile önemlidir. Bu sebeple, tez çalışmasında Türkçe çevrimleri ile oyunlaştırma kullanılarak sunulan bir web uygulaması geliştirilmiş ve yayına sunulmuştur.

3.4.2. Güvenlik

Toplanan kullanıcı bilgilerinin hangi amaçla kullanılacağı açık ve anlaşılır düzeyde anketlerin doldurulmasından önce sunulmalıdır. Ayrıca kullanıcı profillerinden sadece herkese açık verilerin toplanacağı belirtilmelidir. Uygulamanın online olarak yönetimsel ekranlarının olmaması temel güvenlik prensipleri açısından önemlidir. Yönetimsel işlemler gerektiği zaman ilgili veri tabanını belirli süreli güvenli bir sistem üzerinden erişilerek yapılmalıdır. İlgili işlemler bittikten sonra erişim ayarları tekrar kapatılmalıdır. Bunun yanında toplanacak bilgiler kişilik özelliği bakımından anonimleştirilmiş kullanıcı verileri ile uygulama sunucusunda barındırılmalıdır. Sosyal medya bilgileri ise belirli aralıklarla kişisel bilgisayar (erişimi güvenli ve dışarıya kapalı) üzerinden çekilerek etiketleme işlemleri yapılmalıdır. Uygulama ve uygulama sunucusu en az OWASP Top-10 güvenlik açıklarına göre kapsamlı şekilde incelenerek test edilmelidir. Uygulamanın tüm bileşenleri mümkün mertebe birbirinden ayrıştırılmış farklı erişim noktalarında ve istemci tarafından erişimin gerekmediği bölümlerinin sadece ilgili sunucu erişimine açık şekilde konfigüre edilmesi gerekmektedir.

3.4.3. Veri toplama

İlk olarak belirlenen kişilik testleri (BFI, IPIP-120 vb.) oyunlaştırılarak bir web uygulaması oluşturulmuştur. Oluşturulan bu web uygulaması ilgi çekici bir web sitesi üzerinden kullanıcılara sunulmuştur. Ayrıca kullanıcı sözleşmesinde bilgilerin kullanımı ile ilgili her tür yönelim sağlanmıştır. Sistem, kullanıcıların test işlemini yapmadan önce belirli kişisel özellik bilgilerini (cinsiyet, yaş, eğitim durumu, yaşadığı şehir, memleket vb.) dolduracakları şekilde tasarlanmıştır. Her kullanıcıya ait eşsiz bir özet değeri verilmiş ve kullanıcı bilgileri bu benzersiz özet değeri ile tutulmuştur. Ayrıca kullanıcının kullandığı sosyal medya ortamları, blog gibi hesaplarının adresleri istenmiştir. Alınan içerik ortamları bilgilerinin veri toplama aşaması, kişisel bilgisayar üzerinden yapılarak elde edilen bilgiler işlenmeden önce daha güvenli ortamlarda tutulmuştur.

Kullanıcı temelli içerik oluşturan ortamlardan veri çekilmesi API veya Crawler kullanılarak yapılmaktadır. Sosyal medya platformlarının kendi API'leri ile hizmet verdiği bilinmekle birlikte son bir yıldır katı kurallar ile API verilmesi zorlaşmıştır. Özellikle Twitter API için kişilik araştırmaları yasaklı listededir. Bunun yanında araştırmanın ana olgusu herkese açık veriler olduğundan, API kullanılması çok önemli bir getiri sağlamayacaktır. Twitter üzerinden veri alımı için herhangi bir kullanıcı girişi de gerekmemektedir. Facebook ise yaptığı güncelleme ile Graph API'de sunduğu kullanıcı gönderilerini kapattığını duyurmuştur. Bunun yanında kullanıcı girişi yapmadan kullanıcıların herkese açık profillerindeki gönderilerde görüntülenememektedir. Veri toplanması için Facebook kullanıcısı oluşturularak herkese açık veriler üzerinden Crawl işlemi yapılabilmektedir.

3.4.4. İşleme

Elde edilen verilerin öncelikle doğru şekilde elde edildiği ispatlanmalıdır. Bunun için sürekli aynı yönelim üzerinden yapılan test işaretlemeleri çalışma dışı bırakılmalıdır. Ayrıca çalışmada kullanılabilmesi için sosyal medya hesaplarında anlamlandırılabilir içeriklerin olup olmadığına bakılmalıdır. Bu noktada yapılan analizler sonucu hangi bilgilerin çıkarılması gerektiğine karar verilebilmektedir. Son olarak işlenebilir halde tutulan bilgilerin tamamında analiz edilmeden önce belirli önişlem adımları uygulanmalıdır. Bu önişlem adımları literatürde belirlenen ve ayrıştırılan ön işlem adımları ile yapılabilmektedir. Bunun

yanı sıra zamanla literatürde gözlemlenecek farklı ön işlem adımları da çalışma süresince yapılan çalışmada test edilmelidir.

3.4.5. Analiz ve üretim

Bu adım İşleme adımı ile beraber düşünülmelidir. Analiz yapılırken kullanılacak Yapay Zeka yöntemlerinde farklı ön işlem adımları ile farklı modeller denenebilmektedir. Modellerin başarımı farklı seçilimlere göre değiştiğinden dolayı önişlem ve analiz süreci istenen başarım oranı elde edilene kadar sürekli devam edecek bir döngü içindedir. Yapılacak/Yapılan analizlerde kullanılacak yöntemlerin seçimi, literatür ve Derin Öğrenme yönelimi dikkate alınarak seçilmeli ve denenecek Yapay Zeka yöntemleri ve bunların hibrit örnekleri üzerinden gerçekleştirilmelidir.

Yapay Zeka yöntemleri üzerinde kullanılacak veriler 3 farklı bölümde kullanılmaktadır. Eğitim, doğrulama ve test verisetleri üzerinden başarımlar test edilmektedir. Genel olarak sınıflandırma ve regresyon analizi üzerinden çalışmalar yapılmakla birlikte bunların başarım ölçütleri farklılık göstermektedir. Sınıflandırma için kullanılan yaygın başarım ölçütleri accuracy, precision, recall, f-measure ve roc-area'dır. Regresyon için ise MAE, MSE, MSLE ve RMSE gibi yöntemler kullanılabilmektedir. Başarım hem eğitim verilerinin işlenmesindeki zamansal akışta görülebilmekte hemde bu süreç içerisinde doğrulama ve test verisetlerinin eğitilen model ile ne düzeyde eşleştiği ölçülebilmektedir. Türkçe içerikler üzerine yapılan çalışmalar/verisetleri tespit edilemediğinden dolayı çıkarılan başarım sonuçları literatürdeki diğer çalışmalar ile kıyaslanarak analiz edilmektedir.

Başarım oranı yeterli düzeyde görülen modellerden kullanıcı bilgisi verilen ve bu kullanıcı bilgisi üzerinden kişilik tahmininde bulunan bir sistem tasarımı gerçekleştirilmektedir.

3.4.6. Yaygınlaştırma ve değerlendirme (geri bildirim)

Tüm süreç boyunca elde edilen bulguların karar verici ve değerlendiriciler tarafından değerlendirilmesi gerekmektedir. Bir bildiri veya makale sürecinde bu editör ve hakemler olmakla birlikte tez sürecinde jürilerdir. Ayrıca konu uzmanları görüş sorularak yapılan geri bildirimler ile ilgili değerlendirme süreci tamamlanabilmektedir. Geri bildirimlerin, çalışmanın yeterli başarımı göstermemesi, isterlerini karşılamaması veya yeni isterlerin

ortaya çıkabileceği durumları da mevcuttur. Bu durumda problem tanımı varolan yöntem üzerinde analiz edilerek ilgili sürece dönüş yapılıp yeni yönelimlerde çalışma devam ettirilmelidir.

4. SEMANTİK YAPILAR VE LSTM TEMELLİ SİNİR AĞLARI İLE KİŞİLİK ÖZELLİKLERİ TAHMİNİ

Bu bölümde sunulan çalışma ile, Twitter üzerinden sosyal medya platformlarından alınan belirli bir örneklemden oluşturulan veri seti, IBM Personality Insight [108-110] kullanılarak Big Five Modeline göre etiketlenmiştir. Daha sonra tweet verilerinin yapısal analizi sunulmuş ve metin ön işleme aşamasındaki önemli işlem adımları çoklu seçim şeklinde verilmiştir. Ardından, farklı ön işleme ve LSTM tabanlı yapay sinir ağı model kombinasyonları kullanılarak Derin Öğrenme eğitimi ve test analizleri gerçekleştirilmiştir. Son olarak Twitter verileri üzerinde yapısal analiz, ön işleme modeli ve derin öğrenme modeli ile karşılaştırmalı bir analiz yapılarak literatüre (hem akademi hem de bu amaçla çalışan kamu/özel sektör için) katkı sağlanması amaçlanmıştır. Çalışmanın motivasyonu, problem tanımı ve diğer detayları aşamalı olarak alt başlıklarda sunulmaktadır.

4.1. Motivasyon ve Problem Tanımı

İlk olarak kişilik özelliklerinin tahmininde yaygın kullanımı ve geliştirilen araçların çeşitliliği sebeplerinden dolayı İngilizce içerikli materyaller üzerinde çalışılmaya karar verilmiştir. Devamında, İngilizce kişilik özelliklerinin etiketli olduğu veri setlerine erişim sağlamaya çalışırken hem veri setlerine erişim zorluğu hem de erişilebilen veri setlerinin yetersizliği sebebi ile yüksek boyutlu veri seti oluşturma ihtiyacı ortaya çıkmıştır. Veri setini oluşturmak için belirlenen temel problemler sosyal medya platformu ve bu platformdan alınacak metinsel verilerin kişilik özellikleri temelinde nasıl etiketleneceğidir. Bunun yanında literatürde farklı ön işlem kombinasyonları ile kişilik özellikleri tahmin modeli denemeleri belirli sınırlar içerisindedir. Bu çalışma ile hem İngilizce kişilik özellikleri veri seti oluşturulması hem de dengeli ve genellenebilir bir tahmin modeli çalışması yapılmıştır.

4.2. Materyal ve Metod

Bu bölüm, tweet verilerinden kişilik özellikleri tahmin modelini oluşturmak için kullanılacak veri seti ile birlikte önerilen ön işleme ve tahmin modelini açıklamaktadır. Araştırmanın tüm iş akışı Şekil 4.1'de sunulan grafik özette genel hatlarıyla sunulmuştur. İlk olarak, kapsam belirlendikten sonra toplanan tweet verileri ile birlikte IBM Personality

Insight hizmeti ile kişilik özellikleri etiketlenmiştir. Daha sonra ön işleme, LSTM tabanlı sinir ağı modeli ve analiz sonuçlarının değerlendirilmesi üzerine bir döngü sunulmaktadır. Bu aşamada her bir süreç adımı ve model parametresi değerlendirilerek sonuçların analizi ile oluşturulan modelin en iyi sonuçlarını elde etmek hedeflenmiştir. İlgili testlerin yeterli düzeyde yapıldığı göz önünde bulundurularak, PAN-2015 Kişilik Veri Seti, modeller arası ön işleme ve analiz sonuçları ile karşılaştırılmıştır.



Şekil 4.1. Çalışmanın araştırma yöntemi - iş akışı (Grafiksel özet)

4.2.1. Kapsam, veri toplama ve veri etiketleme

Kullanıcılara çeşitlilik sağlamak amacıyla sunulan hem olumlu hem de olumsuz anlamda 13 kelime [111] kullanılarak, 30 Kasım 2020 tarihinde 04:07 ile 04:21 (GMT+3) arasında her kelime için 1000 tweet toplanmıştır. Toplamda 13000 tweet yazan 12101 ayrı kullanıcı elde edilmiştir. Bu kullanıcı listesinden herkese açık paylaşım yapan 11984 kullanıcının en son 3200 tweet kaydı alınmıştır. Daha sonra İngilizce içeriği olmayanlar elenerek ve IBM Personality Insight hizmetinin sınırlamaları ile elde edilen kullanıcılar azaltılmıştır. Ayrıca maksimum 100 tweet'e sahip kullanıcılar ve toplam tweet verisi maksimum 1200 kelime olan kullanıcılar temizlenmiştir. Son olarak 5081 kullanıcı verisi ile toplam 11769203 tweet içeriğine sahip bir veri seti elde edilmiştir. Bu veri kümesinde, bir tweet başka bir tweet'e yanıtsa, [REPLY] etiketiyle işaretlenmiştir.

5081 kullanıcının tweet verileri IBM Personality Insight hizmetinin girdi formatına dönüştürülmüş ve hizmet çıktısı olarak 35 farklı kişilik özelliğinin sonuçları toplanmıştır. Elde edilen kişilik özellikleri, her kullanıcı için sınıf etiketi olarak kullanılmıştır. İlgili kişilik özellikleri 5 ana başlık altında toplanmıştır. Bunlar Deneyime Açık (Openness, DA), Özdenetim Sahibi (Conscientiousness, ÖS), Dışadönük (Extraversion, D), Yumuşak Başlı (Agreeableness, YB) ve Nevrotik (Neuroticism, N)'dir. Diğer 30 kişilik özelliği ise her bir ana başlık için 6 adet olmak üzere ana kişilik özelliklerini belirlemede kullanılan alt kişilik özellikleridir. Tüm kişilik özellikleri grup temelli olarak Çizgelge 4.1'de sunulmaktadır. Her kişilik özelliğinin 0 ile 1 arasında ondalık değeri vardır. Bu değer, tweet bilgisinin kişilik özelliğinin yüzdesini temsil etmektedir. Çalışma ortamı kısıtı, verilerin boyutu, PAN-2015 veri seti kişilik özellikleri etiketleri ile uyumluluğu ve ana kişilik özelliği kullanımı temel alınarak deneysel çalışma için 5 ana kişilik özelliğinin kullanılmasının yeterli olacağı düşünülmüştür.

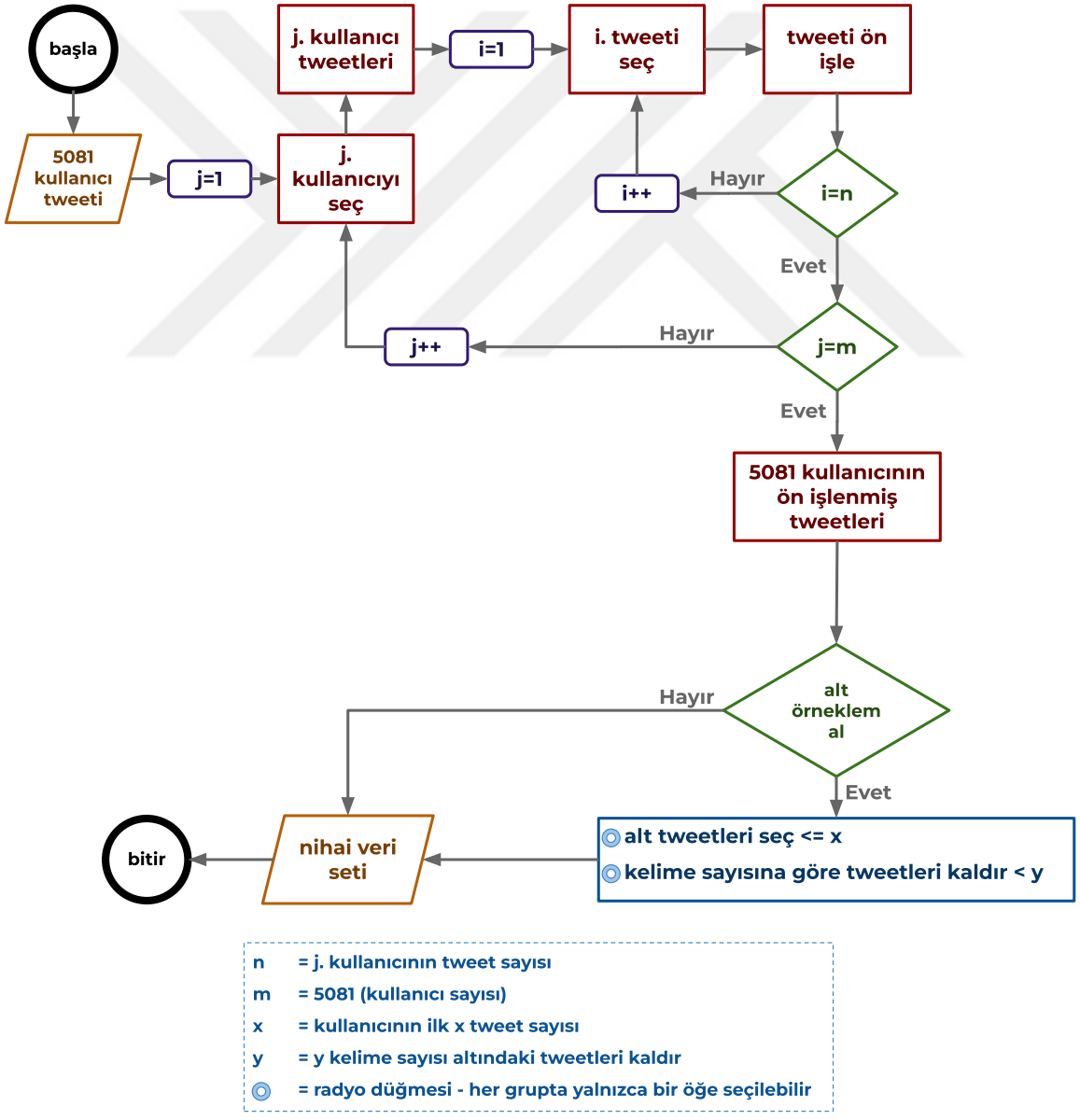
Çizelge 4.1. Çalışmada kullanılan ana kişilik özellikleri ve alt kişilik özellikleri

Ana Kişilik Özelliği	**Openness** (Deneyime Açık)	**Conscientiousness** (Özdenetim Sahibi)	**Extraversion** (Dışadönük)	**Agreeableness** (Yumuşak Başlı)	**Neuroticism** (Nevrotik)
Alt Kişilik Özelliği	Adventurousness	Achievement striving	Activity level	Altruism	Anger
	Artistic interests	Cautiousness	Assertiveness	Cooperation	Anxiety
	Emotionality	Dutifulness	Cheerfulness	Modesty	Depression
	Imagination	Orderliness	Excitement seeking	Morality	İmmoderation
	Intellect	Self discipline	Friendliness	Sympathy	Self consciousness
	Liberalism	Self efficacy	Gregariousness	Trust	Vulnerability

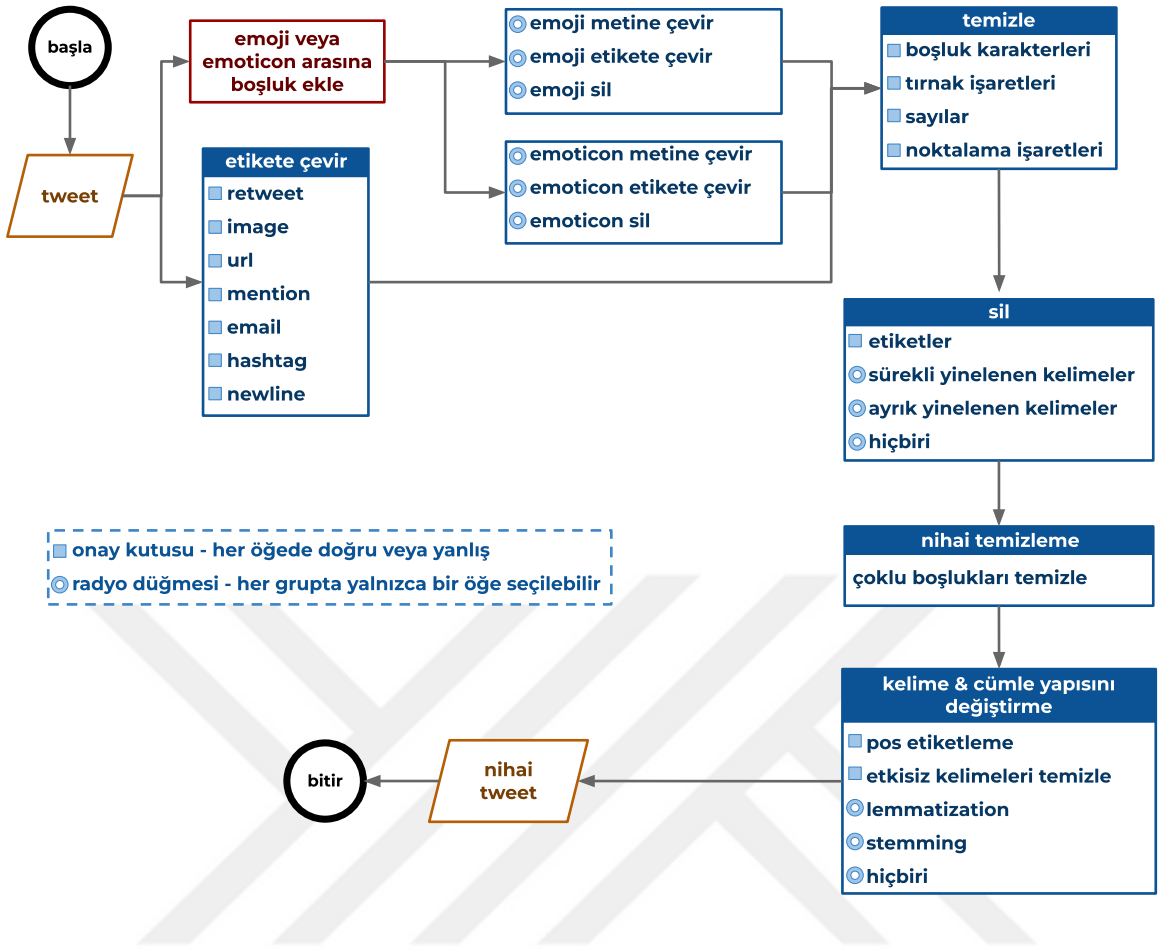
4.2.2. Tweet analizi ve ön işlem

Bir tweet, birçok varlık ve metnin bir kombinasyonu olabilir. Bu varlıklar tweet içeriğine göre büyük farklılıklar gösterdiğinden, genelleme için ortak bir payda bulmayı da zorlaştırabilmektedir. Örneğin, bir tweet'te paylaşılan bir söz ile başka bir tweet'te paylaşılan bir söz arasında bir kesişme noktası bulmak kolay olmayabilmektedir. Ayrıca ilişki kurmayı zorlaştıran bu varlıklar, kelime dağarcığı kelime listesini devasa boyutlara taşıyacağı için işlem gücü açısından bizi içinden çıkılmaz bir konuma getirebilmektedir. Bu noktada yapılan işlemleri sıralayacak olursak; varlıkları kategori etiketlerine (MENTION, HASHTAG,

URL, IMAGE, EMAIL, RETWEET, NEWLINE), hem anlam hem de kategori etiketlerine (emoji, emoticon) dönüştüren, yorumlanamayan karakterleri (boşluk karakterleri, tırnak işaretleri, sayılar ve noktalama işaretleri), etiketli verileri (etiketler), benzer sözcükleri (sürekli ve ayrık yinelenen sözcükler) temizleme ve veri kümelerinde doğal dil işleme adımlarının (metin parçası etiketleme, etkisiz kelimeleri temizleme, basitleştirme ve kök çıkarma) uygulanmasıdır. Bu ön işleme adımlarının birbiri ile birlikte kullanılabilmesi başarı oranı üzerinde önemlidir. Şekil 4.2 ile tüm veri kümesinde hangi işlem adımlarının uygulanacağını ve Şekil 4.3'te bir tweet'e hangi önişlem adımlarının hangi sıra ile uygulanabileceği gösterilmektedir.



Şekil 4.2. Veri kümesi için ön işleme akış şeması



Şekil 4.3. Bir tweet için önişleme adımları (Şekil 4.2'de tweet önişleme)

Şekil 4.3'de sunulan sil ve temizle işlemlerinin farkı anlamlı ve anlamsız karakter/kelime grupları ile alakalıdır. Temizleme işlemi anlamsız karakter ve varlıkların temizlenmesi işlemini ifade etmektedir. Silme işlemi ise anlamlı varlık ve kelimelerin ilgili metinden silinmesi işlemini tanımlamaktadır.

Onay Kutusu olarak tanımlanan alanlar tek tek seçilebilirken, Radyo Düğmesi olarak tanımlanan alanlar sadece grup içinde seçim olarak sunulabilir.

Tüm veri kümesine bağlı olarak, her kullanıcı için tweet sayısı değişir. Bu şekilde, her kullanıcının tahmin modeli üzerindeki etkisi, aynı maksimum sayıda tweet seçilerek analiz edilebilir (Şekil 4.2'deki alt tweet'leri seçin). Başka bir deyişle, kullanıcının seçilen maksimum tweet sayısından daha az tweet'i varsa, kullanıcının tüm tweet'leri alınır. Ayrıca kelime sayısına göre tweet'lerin model üzerindeki etkileri araştırılabilir (Şekil 4.2'de kelime sayısına göre tweet'leri kaldırın).

4.2.3. Metinsel veriden hesaplanabilir sayısal veriye

Metin verilerine belirli matematiksel işlemleri uygulamadan önce, metni sayısallaştırmak gerekir. Derin öğrenme yöntemlerinde kullanılan sayısallaştırma yöntemleri çeşitli olmakla birlikte bu çalışmada iki sayısallaştırma yöntemi kullanılmıştır. Bunlardan ilki, yaygın olarak kullanılan ve ilgili metindeki her bir kelimeye sıklığına göre karşılık gelen sayısal bir ifadeye dönüşen bir kelime havuzu oluşturularak gerçekleştirilmiştir. İkinci yöntem ise son yıllarda popülerlik kazanan CBOW ve Skip-Gram yöntemleri ile alt örnekleme yaparak vektörleştirme sağlayan FastText [112] yöntemidir. Bu aşamada dikkat edilmesi gereken asıl nokta, sözlük oluşturma aşamasında sadece eğitim veri setinin kullanılması gerektiğidir. Test veri setinin kullanılması tüm çıktıların başarı oranını artırsa da genelleme yapmamaktadır. Eğitim veri setimiz ve PAN-2015 eğitim veri seti kullanılarak FastText yöntemi için 300 boyutlu vektörlere sahip 0.025 alfa ile 3 pencereli 1 gram(ngram) FastText modeli her veri seti için ayrı ayrı eğitilmiştir. Tüm kelimeler sayısallaştırılmadan önce küçük harfe dönüştürülmüştür.

4.2.4. LSTM temelli modeller

Uzun Kısa Süreli Bellek (Long Short-Term Memory - LSTM) [113, 114], Derin öğrenme yöntemlerinin ve Tekrarlayan Sinir Ağı'nın (Recurrent Neural Network - RNN) en çok kullanılan varyantlarından biridir. Bunun temel nedeni birçok problemin çözümünde göstermiş olduğu başarı oranlarıdır.

Önerilen yapay sinir ağı modelinde öncelikle alınan tweet verilerine ayrıştırma/jetonlaştırma işlemi (tokenization) uygulanarak bir sözlük oluşturulmuştur. Ayrıştırma işlemi boşluk temelli ayrıştırma kullanılarak sağlanmıştır. Sözlükteki her tweet için vektörleştirme uygulanmış ve vektörleştirilmiş tweet bilgisi girdi olarak modele sunulmuştur. Vektörleştirme işleminin nihai hali her sayısallaştırma yöntemi için nihai olarak Gömme Katmanında gerçekleştirilmektedir. Gömme katmanında her metin dizisinin maksimum kelime boyutunda vektörel gösterimleri oluşturulmaktadır. Bu gösterimler FastText modeli için FastText modelinden elde edilen ağırlık vektörlerinin Gömme Katmanına ağırlık olarak atanması ile özelleştirilmiştir. Gömme Katmanı temel parametreleri maksimum kelime sayısı, gömme boyutu ve girdi uzunluğu olarak sunulmaktadır. Ayrıca her girdinin uzunluğunu sabitlemek için Dizi/Cümle uzunluğuna göre cümle sonu dolgulama işlemi (post

padding) yapılmıştır. Nihai olarak her cümle benzerlik ve yönelimi ayarlamak için yoğunluk vektörüne çevirilerek derin öğrenme modeli içinde başarım ve genelleştirmenin arttırılması sağlanmıştır.

Daha sonra hiper parametreleri ile en iyi sonucu elde etmek için LSTM modeli optimize edilmiş ve kişilik özellikleri çıktılarının orantılı değerleri verilmiştir. LSTM modellerini optimize ederken; parti boyutu (batch size), dönem sayısı (epoch), gizli katman sayısı, farklı gizli katman türleri (+2. gizli katmanda LSTM veya Dense), katmanların birim sayıları (unit size), aktivasyon fonksiyonları (swish, sigmoid, tanh vb.), öğrenme oranları (learning rates), bırakmalar (dropouts, tekrarlayan (recurrent) veya sıralı (sequential)) ve l1/l2 düzenlileştirme (regularization) vb. kullanılmıştır. LSTM modellerinin hiperparametreleri optimize edilirken bir parametrenin ideal değeri bulunurken diğer parametreler sabit tutulmuştur. Ek olarak, aşırı öğrenmeyi (overfitting) önlemek için bırakma, düzenlileştirme ve min delta değişikliğinin kontrol edilmesi ile öğrenme dengelenmektedir. Modelde, aktivasyon fonksiyonu olarak Swish aktivasyon fonksiyonu kullanılmıştır ve yaygın olarak kullanılan aktivasyon fonksiyonlarından daha iyi performans elde edilmiştir [115]. Ayrıca, optimizasyon fonksiyonu olarak Adam [116] (Öğrenme oranı 0,001 ve 0,0001), kayıp fonksiyonu olarak Ortalama Kare Hata (Mean Squared Error - MSE) ve değerlendirme metriği olarak Kök Ortalama Kare Hata (Root Mean Squared Error - RMSE) kullanılmıştır. Tüm modellerde hem normal hem de tekrarlayan bırakmalar 0,3 olarak alınmıştır.

LSTM modelinin birçok farklı çeşidi vardır. Bu çalışmada, temel LSTM modeli ile Çift Yönlü LSTM (Bidirectional LSTM, Bi-LSTM) [117] modeli kullanılmıştır. Bi-LSTM'nin normal LSTM'den tek farkı girdiyi çift yönlü değerlendirerek modeli oluşturmasıdır.

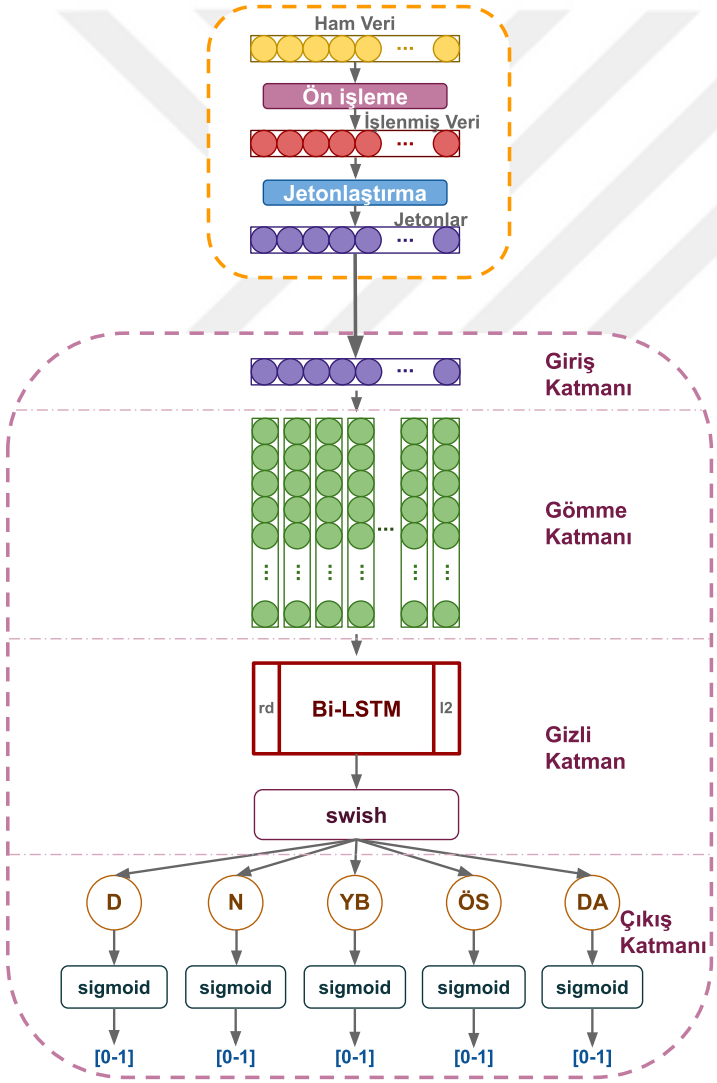
Modelimizin model özellikleri Çizelge 4.2'de ve yapısal gösterimi Şekil 4.4'te sunulmuştur.

Çizelge 4.2. Modelimizin sayısal özellikleri

Özellik	**Değer**
Dizi/Cümle uzunluğu (Sequence length)	48
Gömme boyutu (Embedding dimension)	96
Parti boyutu (Batch size)	1024
Dönem sayısı (Epoch)	max: 1000 ve doğrulama kaybı (validation loss) ile minimum delta 0.0001 ise işlem tamamlanır

Çizelge 4.2. (devam) Modelimizin sayısal özellikleri

Özellik	Değer
LSTM – birim	256
LSTM – tekrarlayan bırakma oranı	%30
LSMT – l2 aktivite düzenlileştirme değeri	0.001
LSTM – aktivasyon fonksiyonu	Swish
Çıkış – aktivasyon fonksiyonu	Sigmoid
Kayıp fonksiyonu (Loss function)	MSE
Optimizasyon fonksiyonu	Adam (lr: 0.0001, b1: 0.9, b2: 0.999, e:1e-7)
Değerlendirme fonksiyonu	RMSE



Şekil 4.4. Modelimizin yapısal gösterimi (rd: tekrarlayan bırakma, l2: l2 düzenlileştirici)

4.2.5. Çalışma ortamı

Çalışma ortamı donanım ve yazılım olarak iki farklı şekilde sunulmaktadır. Bu ortamlar her biri ayrı paragraf olarak açıklanmaktadır.

Donanım olarak AMD Ryzen 9 3900X 12 Çekirdekli işlemci, 64 GB (4*16 GB) 3600 MHz Ram, 500 GB (3400 okuma / 2500 yazma) SSD ve Nvidia Chipset RTX 2080Ti 11 GB ekran kartı ile çalışma ortamı kullanıldı. Özellikle donanımlar arası veri iletimi ve işlenmesinde uyum ve denge gözetilmelidir.

Yazılım açısından ise Windows 10 işletim sistemi üzerinde Python 3.8.3 programlama dili kullanılarak CUDA 10.2 üzerinde TensorFlow 2.4 ile model testleri yapılmıştır. Ön işleme adımlarında emoji ve emoticonların anlamlandırılması için temel python kodlarına ek olarak emoji ve emot kütüphaneleri kullanılmıştır. Ayrıca, Konuşma Bölümü Etiketleme (Part of Speech Tagging, POS-Tagging), Önemsiz Kelimelerin Temizlenmesi (Stop Words) ve Kökleştirme (Lemmatization ve Stemming) yöntemlerinin bir kısmı için NLTK 3.5 kütüphanesi kullanılmıştır. Son olarak FastText için Gensim 3.8.3 kütüphanesinden faydalanılmıştır.

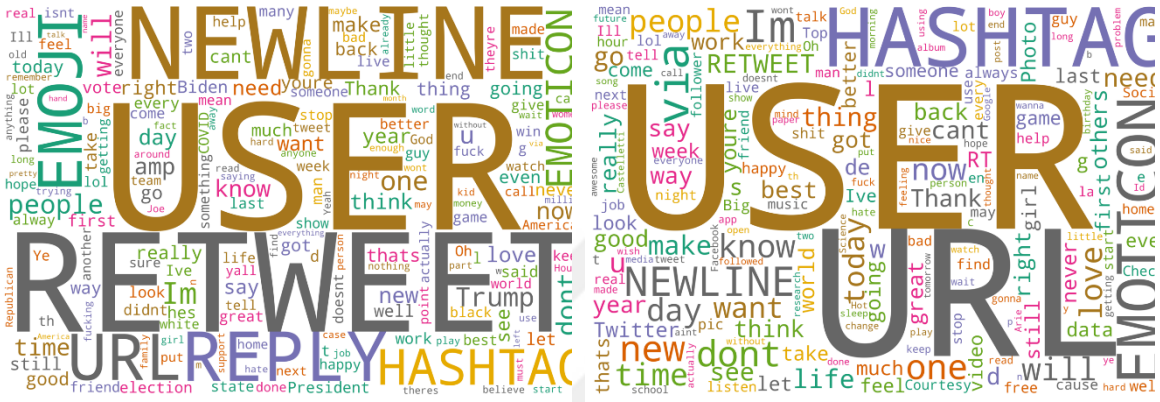
4.2.6. PAN-2015 kişilik veri seti

PAN-2015 Kişilik Veri Seti, 4 dilden oluşan eğitim ve test veri setlerine ayrılmıştır, ancak bu çalışmada sadece İngilizce veri setleri kullanılmıştır. İngilizce eğitim veri seti için 152 kullanıcı bilgisi ve İngilizce test veri seti için 142 kullanıcı bilgisi bulunmaktadır. İki veri setinde toplam 27344 tweet bulunmaktadır. Bu çalışmada modeller üzerinde kıyaslama oluşturduğumuz veri seti ve PAN-2015-EN Kişilik Veri Seti kullanılarak yapılmaktadır.

4.2.7. PAN-2015-EN ile veri setimizin karşılaştırmalı analizi

Veri setimizi oluştururken temel amacımız erişilebilir veri setlerindeki veri miktarının az olması, varlık ve kelime çeşitliliğinin azlığıydı. Aynı zamanda karmaşık modellerle daha fazla veriye ihtiyaç duyulmuştur. Ayrıca yeni oluşturulan veri setinde temel kişilik özelliklerini ifade eden alt kişilik özellikleri bulunmaktadır. Alt kişilik özelliklerinin başarı oranı, ana kişilik özelliklerinin ana değerini belirlediğinden, çalışmayı karmaşıklaştırmamak

için değerlendirme ve analize yer verilmemiştir. Yeni bir veri seti oluşturmak için anketlerden ve klinik çalışmalardan elde edilecek sonuç sayısının sınırlı olacağı bilinmektedir. Bunun nedeni; hem ilgili verileri sunacak kişi sayısının azlığı hem de katılan her kişinin sosyal medya hesaplarında aktif olarak paylaşım yapmamasıdır. İki veri setinin sayısal bilgileri Çizelge 4.3'te, varlık ve kelime yoğunluğuna göre kelime bulutları ise Şekil 4.5'te sunulmuştur.



(a) **(b)**

Şekil 4.5. Veri setimiz ile PAN-2015-EN kişilik veri setinin kelime bulutu ile karşılaştırılması. (a) Veri setimizin kelime bulutu, (b) PAN-2015-EN kişilik veri setinin kelime bulutu

Çizelge 4.3. Veri setimiz ve PAN-2015-EN kişilik veri setinin sayısal karşılaştırması

Bilgi	Bizim veri setimiz	PAN-2015-EN veri seti
Kullanıcı sayısı	5081	294 (154 eğitim and 142 test)
Tweet sayısı	11769202	27343
Maksimum tweet uzunluğu	890	80
Minimum tweet uzunluğu	1	1
En yaygın varlık	Kullanıcı (User)	Kullanıcı (User)

4.3. Analiz ve Tartışma

Veri setimiz (5081-EN), Büyük Beşli Envanteri (Big Five Inventory) modeline dayalı IBM Personality Insights hizmeti ile 13 olumlu/olumsuz kelime ile tweet atan ve 0-1 arasında değerlerle etiketlenen 5081 kullanıcıya ait 11769202 tweet bilgisini içermektedir. Bu veri setinin içerik analizi dinamik yöntemlerle yapılmış ve içeriğindeki varlıklara ve yapılara

göre belirli ön işleme adımları belirlenmiştir. Bu ön işleme adımlarının farklı kombinasyonları farklı sonuçlar üretebilir. Örneğin, hashtag kelime öbeklerini [HASHTAG] etiketlerine dönüştürme devre dışı bırakıldığında ve noktalama işaretleri temizleme etkinleştirildiğinde, hashtag'ler normal kelimeler olarak kabul edilir. Ayrıca resim bilgisi içeren 638 tweet, URL bilgisi içeren 3275304 tweet, hesap etiketleme (mention) bilgisi içeren 9280208 tweet, e-posta bilgisi içeren 2384 tweet, hashtag bilgisi içeren 1200813 tweet ve \n (yeni satır) bilgisi içeren 1996337 tweet bulunmaktadır.

Gösterimlerdeki ön işleme adımlarını daha kolay ifade etmek için kodlar aşağıda tanımlanmıştır. Kodlar; Etiket Kelimeleri (EK), Emoji ve İfade (Eİ), Nesneleri Temizle (NT), Etiketleri/Kelimeleri Sil (SL) ve Kelime/Cümleyi (KC) içerir.

- EK1: Retweet Etiketleme [RETWEET]
- EK2: Resim Etiketleme [IMAGE]
- EK3: URL Etiketleme [URL]
- EK4: Mention Etiketleme [USER]
- EK5: E-posta Etiketleme [EMAIL]
- EK6: Hashtag Etiketleme [HASHTAG]
- EK7: Yeni Satır (\n) Etiketleme [NEWLINE]
- Eİ1: Emojiyi Anlamlı Metine Dönüştürme
- Eİ2: Emoji Etiketleme [EMOJI]
- Eİ3: Emojiyi Silme
- Eİ4: İfadeyi Anlamlı Metine Dönüştürme
- Eİ5: İfade Etiketleme [EMOTICON]
- Eİ6: İfadeyi Silme
- NT1: Boşluk Karakterlerini Temizleme
- NT2: Tırnak/Alıntı İşaretlerini Temizleme
- NT3: Sayıları Temizleme
- NT4: Noktalama İşaretlerini Temizleme
- SL1: Etiketleri Silme ([URL], [USER] etc.)
- SL2: Sürekli Tekrar Eden Kelimeleri Temizleme
- SL3: Ayrık Tekrar Eden Kelimeleri Temizleme
- KC1: PoS Etiketleme
- KC2: Önemsiz Kelimeleri Temizleme
- KC3: Kelime Kökleştirme (Lemmatization)
- KC4: Kelime Kökleştirme (Stemming)

Model ve ön işleme özelliklerine göre farklı kombinasyonlar kullanılarak bir analiz yapılmış ve en iyi sonuçların elde edilmesi ve sonuçlardaki farklılıkların analiz edilmesi için birçok çeşitte modeller üzerinde hiperparametre dengeleme işlemleri gerçekleştirilmiştir. 5081 kullanıcılı 11769202 tweet kaydından oluşan veri setimizde, veri setinin %60'ı LSTM model eğitimi, %20'si doğrulama ve %20'si test için kullanılmıştır. PAN-2015-EN veri setinin

eğitim ve test veri setleri ayrı ayrı sunulduğu için eğitim veri setinin %25'i doğrulama veri seti olarak kullanılmıştır. Değerlendirme metriği olarak RMSE (en küçük değer en iyi sonuç) kullanılmıştır.

EK1-EK7 ve NT1-NT4 arasında tüm ön işlemlerin seçilmesi, anlamsız bilgiler bulunduğundan ve tweetler arasında herhangi bir korelasyon bulunmadığından önemlidir. Bu 11 ön işlem, standart ön işleme adımları olarak adlandırılır ve sonraki tüm testlerde kullanılmıştır. İlk olarak kelimelere ve cümlelere uygulanan Doğal Dil İşleme (NLP) yöntemlerinin performansları modeller üzerinde değerlendirilmiştir. Standart ön işleme adımlarına ek olarak, KC1 ve KC4 arasındaki tüm ön işleme adımlarının aynı özelliklere sahip LSTM ve Çift Yönlü LSTM (Bi-LSTM) modelleri üzerindeki etkileri Çizelge 4.4'de gösterilmiştir. Elde edilen RMSE değerleri her bir kişilik özelliği için farklı başarı oranları gösterdiğinden dolayı modelin genel başarı oranını değerlendirmek için Aritmetik Ortalama (A.O.) kullanılmıştır. Ayrıca metinsel ifadelerin vektörleştirilmesinde ilk olarak standart frekans tabanlı vektörleştirme işlemi uygulanmıştır.

Çizelge 4.4'de sunulan LSTM tipine ve NLP ön işlemine göre yapılan analiz sonucunda Bi-LSTM'nin temel LSTM'ye göre daha iyi sonuçlar verdiği gözlemlenmiştir. Ayrıca Bi-LSTM modelinde herhangi bir işlem yapılmadığında (Yok) ve KC2 ve KC3 işlemleri yapıldığında da benzer sonuçlar elde edilmiştir. KC2 ve KC3 ön işleme ile (Yok) başarı oranlarında küçük bir fark yaratmasa da sözlüğün boyutu küçültülebilir ve ayırt edici kelimeler üzerinde daha fazla konsantrasyon elde edilebilir. Testlerin bu aşamasından sonra Bi-LSTM kullanılmaktadır. Ancak KC2, KC3 ve Yok olmak üzere farklı seçeneklerde NLP ön işleme yöntemleri kullanılabilir.

Standart ön işleme adımlarına ek olarak, önerilen modellerde emoji ve ifade başarı oranlarının ön işlem türüne göre farklılıkları Çizelge 4.5'de sunulmuştur.

Emoji ve ifadelerin değerlendirildiği model, Bi-LSTM üzerindeki standart ön işleme adımlarına ek olarak KC2 ve KC3 ön işleme adımlarını kullanmaktadır. Bu ön işleme adımlarında metinsel ifadelere ve etiket ifadelerine dönüştürülerek uygulanan modellerin sonuçları her kişilik özelliğinde farklılık gösterse de aritmetik olarak birbirine çok yakın sonuçlara sahiptir. Elde edilen sonuçlarda emoji ve ifadeleri metinsel ifadelere dönüştürerek

sözlük kapasitesini ve işleme karmaşıklığını artırmak yerine emoji ve emoticon'ları etikete dönüştürmek daha uygun olacaktır.

Çizelge 4.4. LSTM tabanlı model ve Doğal Dil İşleme yöntemine göre sonuçlar (RMSE)

Veriseti	Kişilik özellikleri	KC1		KC2		KC3		KC4		Yok	
		LSTM	Bi-LSTM	LSTM	Bi-LSTM	LSTM	Bi-LSTM	LSTM	Bi-LSTM	LSTM	Bi-LSTM
5081-EN	D	0.1602	0.1590	0.1573	0.1574	0.1585	0.1571	0.1575	0.1590	0.1602	0.1573
	N	0.1817	0.1821	0.1821	0.1811	0.1857	0.1817	0.1825	0.1835	0.1907	0.1815
	YB	0.2133	0.2119	0.2088	0.2086	0.2089	0.2078	0.2092	0.2083	0.2091	0.2077
	ÖS	0.2341	0.2329	0.2307	0.2316	0.2303	0.2300	0.2297	0.2314	0.2306	0.2298
	DA	0.1652	0.1654	0.1644	0.1637	0.1654	0.1655	0.1648	0.1649	0.1694	0.1634
	A.O.	0.1909	0.1902	0.1886	**0.1884**	0.1897	**0.1884**	0.1887	0.1894	0.1920	**0.1879**
PAN-2015-EN	D	0.1611	0.1595	0.1607	0.1601	0.1600	0.1602	0.1601	0.1600	0.1607	0.1599
	N	0.2277	0.2271	0.2272	0.2266	0.2266	0.2264	0.2265	0.2267	0.2271	0.2274
	YB	0.1521	0.1504	0.1507	0.1496	0.1505	0.1490	0.1507	0.1492	0.1506	0.1495
	ÖS	0.1501	0.1471	0.1482	0.1477	0.1477	0.1484	0.1466	0.1479	0.1477	0.1478
	DA	0.1624	0.1612	0.1608	0.1609	0.1613	0.1601	0.1614	0.1602	0.1607	0.1596
	A.O.	0.1707	0.1691	0.1695	**0.1690**	0.1692	**0.1688**	0.1691	**0.1688**	0.1693	**0.1688**

Çizelge 4.5. Emoji ve ifade ön işleme türüne göre model sonuçları (RMSE)

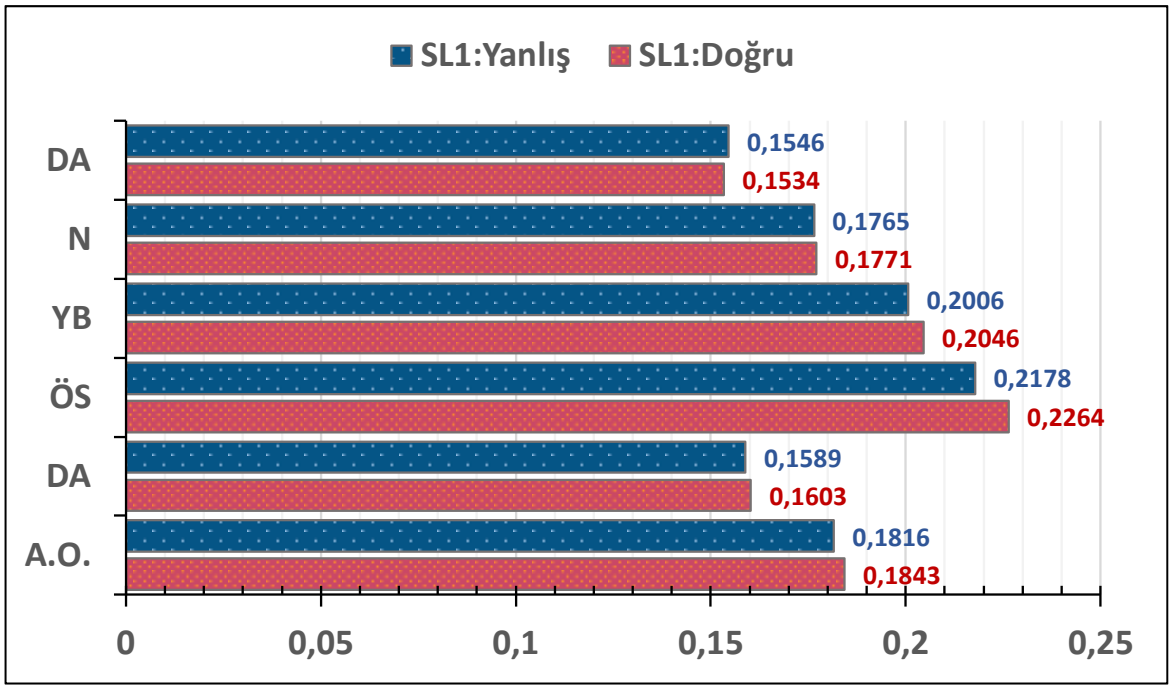
Kişilik özellikleri	5081-EN			PAN-2015-EN		
	Eİ1 & Eİ4	Eİ2 & Eİ5	Eİ3 & Eİ6	Eİ1 & Eİ4	Eİ2 & Eİ5	Eİ3 & Eİ6
D	0.1518	0.1521	0.1550	0.1614	0.1593	0.1610
N	0.1732	0.1749	0.1757	0.2267	0.2261	0.2292
YB	0.1979	0.1975	0.1991	0.1493	0.1491	0.1497
ÖS	0.2163	0.2155	0.2174	0.1486	0.1473	0.1485
DA	0.1574	0.1571	0.1588	0.1597	0.1605	0.1593
A.O.	**0.1793**	**0.1794**	0.1812	0.1691	**0.1685**	0.1695

Model parametreleri aynı özelliklerle uyarlanmış ve standart özelliklere Eİ2&Eİ5 eklenmiştir. Bu test sürecinden sonra tüm modellerde bu ön işlemler de kullanılmıştır. Çizelge 4.6 ve Şekil 4.6, etiketlerin silinip silinmemesine bağlı olarak, ön işleme adımlarındaki etiketleme ifadelerinin dönüşümlerdeki başarı oranı üzerindeki etkisini değerlendirmektedir.

Çizelge 4.6. Etiketlerin silinip silinmediğine göre model sonuçları (RMSE)

Kişilik özellikleri	**5081-EN**		**PAN-2015-EN**	
	SL1=True	**SL1=False**	**SL1=True**	**SL1=False**
D	0.1534	0.1546	0.1600	0.1599
N	0.1771	0.1765	0.2268	0.2272
YB	0.2046	0.2006	0.1491	0.1491
ÖS	0.2264	0.2178	0.1470	0.1473
DA	0.1603	0.1589	0.1604	0.1598
A.O.	0.1843	**0.1816**	**0.1687**	**0.1687**

Dışadönük kişilik özelliğinde elde edilen RMSE değeri SL1=Doğru durumunda daha iyi olmasına rağmen, SL1=Yanlış durumunda diğer 4 kişilik özelliğinin daha iyi sonuçlar verdiği görülmektedir. Tüm kişilik özellikleri ortalama olarak değerlendirildiğinde, model başarı oranı açısından SL1 ön işleme adımını gerçekleştirmemenin (SL1=Yanlış) daha uygun olacağı sonucuna varılmıştır. Bu test sonucunda SL1 ön işleme adımı, sonraki model denemeleri için standart ön işleme adımlarına SL1=Yanlış olarak eklenmiştir. Çizelge 4.6'da elde edilen sonuçların Çizelge 4.5'de elde edilen sonuçlardan düşük olmasının nedeni, etiketlenmesi gereken alanlardan kaynaklanmaktadır.



Şekil 4.6. Etiketlerin silinip silinmediğine göre model sonuçları (5081-EN, RMSE)

Devamında ön işlem türlerine göre kelime ve etiket tekrarlarının modeller üzerindeki etkileri değerlendirilmiştir. Bu aşamada, SL2 ve SL3 ön işlem adımlarının uygulanıp uygulanmadığına bağlı olarak aynı model ve standart ön işleme adımları ile elde edilen sonuçlar Çizelge 4.7'de sunulmuştur.

Çizelge 4.7. Sürekli veya ayrık tekrarlayan verilerin silinmesiyle ilgili model sonuçları (RMSE)

Kişilik özellikleri	**5081-EN**			**PAN-2015-EN**		
	SL2	**SL3**	**Yok**	**SL2**	**SL3**	**Yok**
D	0.1510	0.1518	0.1515	0.1608	0.1586	0.1599
N	0.1733	0.1732	0.1727	0.2292	0.2263	0.2270
YB	0.1968	0.1979	0.1974	0.1499	0.1502	0.1499
ÖS	0.2147	0.2163	0.2152	0.1486	0.1479	0.1480
DA	0.1563	0.1574	0.1564	0.1601	0.1610	0.1603
A.O.	**0.1784**	0.1793	**0.1786**	0.1697	**0.1688**	**0.1690**

SL2 ve SL3 önişlemlerinde tekrarlanan kelimelerin olmaması durumunda sonuçlar aritmetik ortalama açısından birbirine yakın çıkmıştır. Ancak SL2 veya SL3 ile işlenecek olan ardışık tekrarların azaltılması ve cümle uzunluklarının kısaltılmasının hem işlem gücü açısından kazanç sağlayacağı hem de modellerdeki karmaşıklığın artmasına yardımcı olacağı düşünülmektedir.

5081 kullanıcının farklı sayıda tweet kaydı ve her tweet'in farklı uzunlukları bulunmaktadır. Kullanıcı tweet'lerine göre belirli bir orana sahip tweet'lerin seçimi ile Bi-LSTM sonuçları Çizelge 4.8'de, tweet'lerdeki kelime sayısına göre Bi-LSTM üzerindeki kısıtlamaların sonuçları Çizelge 4.9'da sunulmuştur.

Kullanıcılardan gelen tweet sayıları farklı yönlerden alınıp sonuç dağılımını eşitlemeye çalışsak da veri ne kadar fazlaysa başarı da o kadar yüksek olmaktadır. Dolayısıyla kullanıcının tweet bilgilerinin tamamının elde edilmesinin başarı oranı açısından önemli olduğu sonucuna varılmıştır.

Çizelge 4.8. Kullanıcıların tweet sayısındaki seçime göre model sonuçları (RMSE)

Kişilik özellikleri	**Seçilen Tweet Sayısı**				
	250	**500**	**1000**	**2000**	**3200**
D	0.1669	0.1626	0.1574	0.1527	0.1510
N	0.1936	0.1882	0.1819	0.1760	0.1733
YB	0.2128	0.2093	0.2044	0.1990	0.1968
ÖS	0.2314	0.2254	0.2203	0.2155	0.2147
DA	0.1741	0.1705	0.1657	0.1607	0.1563
A.O.	0.1957	0.1912	0.1859	0.1807	**0.1784**

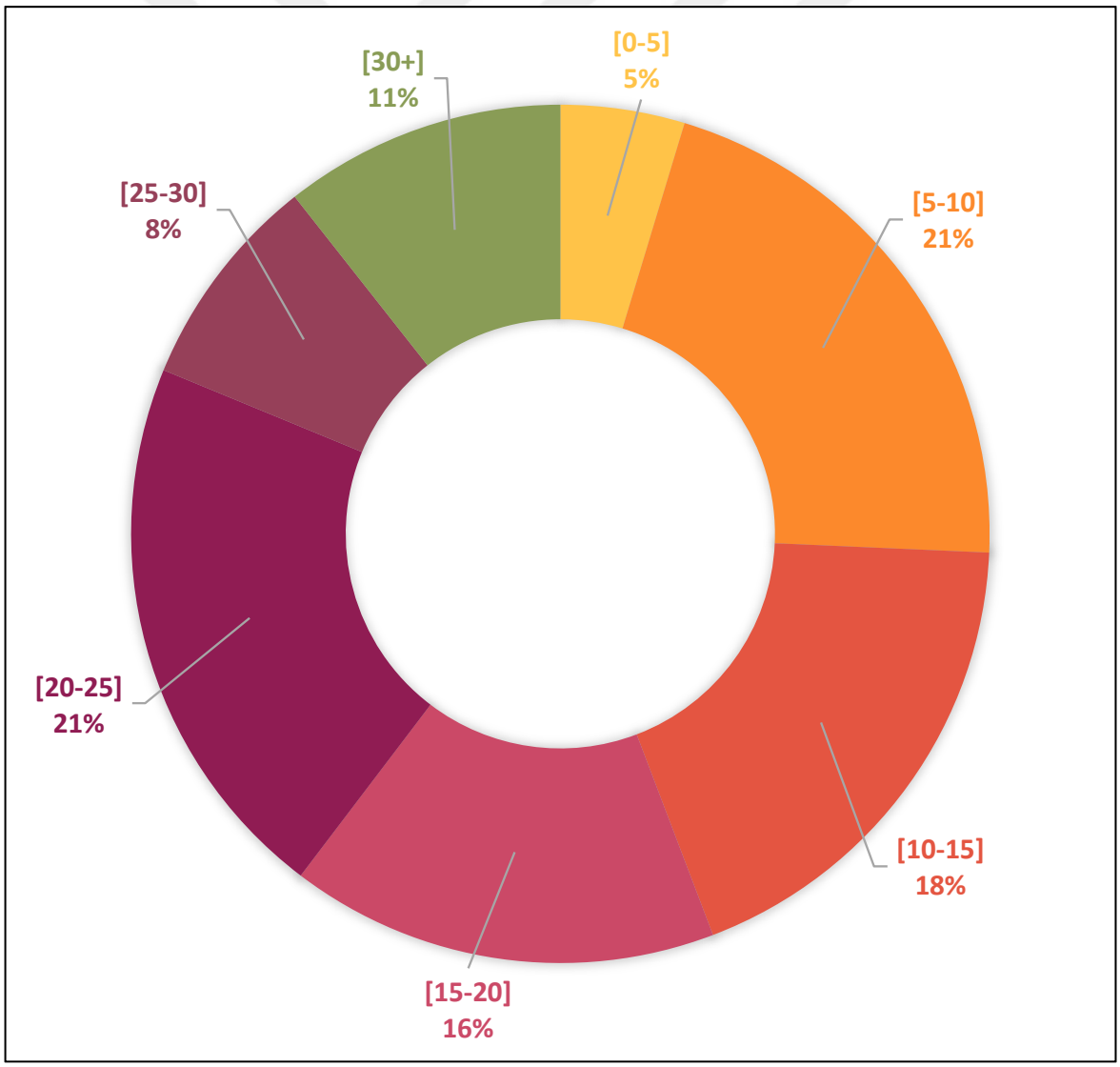
Çizelge 4.9. Tweet uzunluğuna (T.U.) göre model sonuçları (RMSE)

Veriseti	**Kişilik özellikleri**	**Tweet Uzunluğu**						
		0	**5**	**10**	**15**	**20**	**25**	**30**
5081-EN	D	0.1510	0.1491	0.1472	0.1434	0.1422	0.1465	0.1500
	N	0.1733	0.1707	0.1654	0.1594	0.1577	0.1618	0.1622
	YB	0.1968	0.1953	0.1925	0.1897	0.1889	0.1886	0.1824
	ÖS	0.2147	0.2121	0.2129	0.2109	0.2112	0.2097	0.2018
	DA	0.1563	0.1542	0.1449	0.1373	0.1332	0.1321	0.1217
	A.O.	0.1784	0.1762	0.1725	0.1681	**0.1666**	0.1677	**0.1636**
PAN-2015-EN	D	0.1604	0.1599	0.1597	0.1627	0.1627	0.1640	0.1421
	N	0.2264	0.2252	0.2278	0.2338	0.2422	0.2527	0.2518
	YB	0.1492	0.1503	0.1504	0.1587	0.1655	0.1698	0.1683
	ÖS	0.1481	0.1478	0.1472	0.1451	0.1482	0.1616	0.1787
	DA	0.1611	0.1612	0.1597	0.1610	0.1578	0.2084	0.1598
	A.O.	**0.1690**	**0.1689**	**0.1690**	0.1722	0.1753	0.1913	0.1801

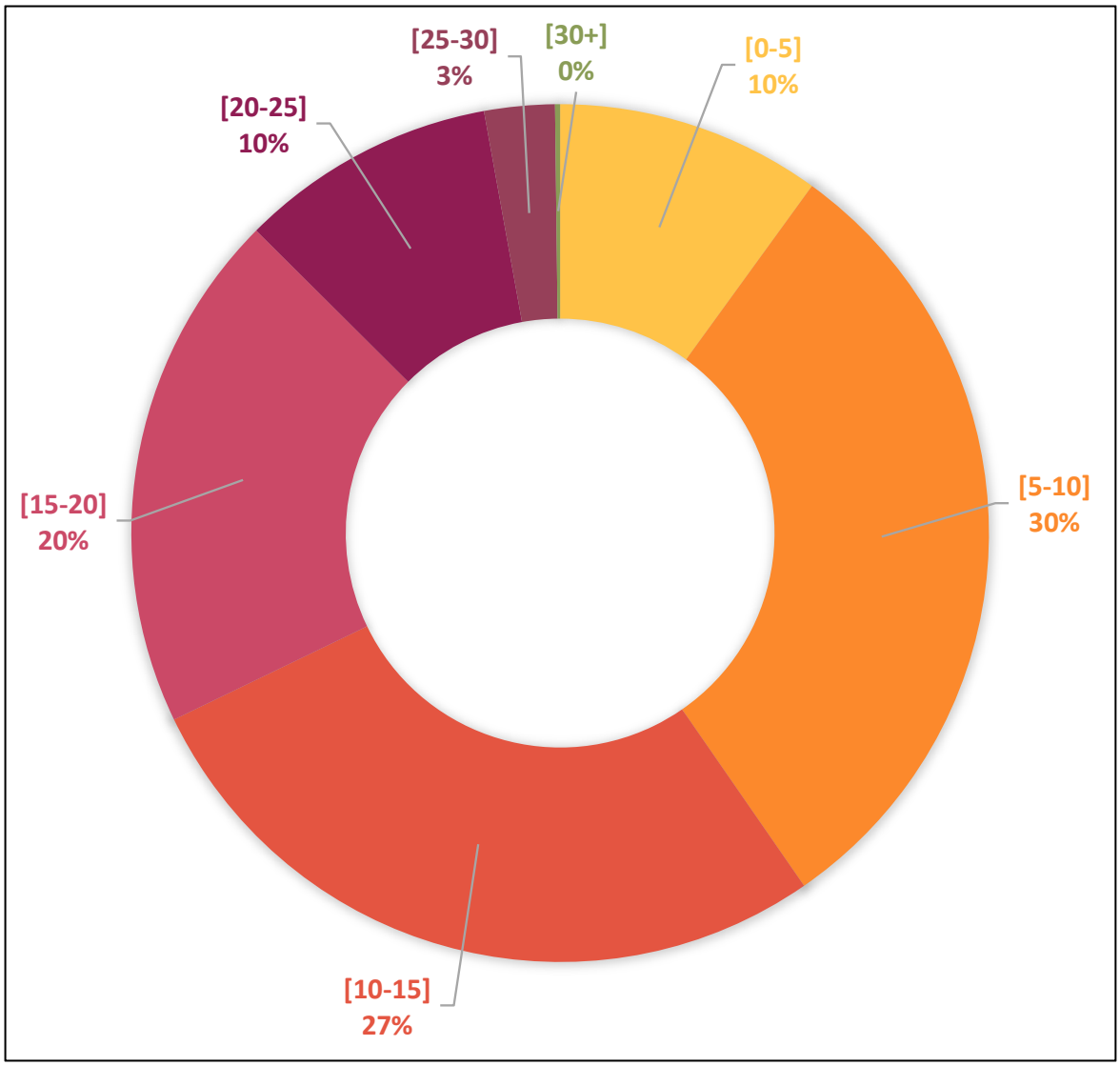
Tweetlerin uzunluklarına göre Çizelge 4.9'daki sonuçlardan önce tweetlerin uzunluklarına göre tweet sayılarının dağılımının bilinmesi gerekmektedir. Şekil 4.7'de veri setimizde bulunan tweet sayısının tweet uzunluğuna göre dağılımı görülmektedir. Şekil 4.7 ve Çizelge 4.9'daki sonuçlar birlikte değerlendirildiğinde, tweet uzunluğunun artmasıyla başarı oranını belirginleştirmenin etkisi görülmektedir. Ancak Şekil 4.7'deki dağılıma göre 20 veya daha fazla tweet seçilerek veri setinin büyük bir kısmı elenmiştir. Bu noktada modellerin genelleme üzerindeki etkisinin düşük olacağı düşünülmüştür. %50'nin üzerindeki seçim, en

iyi dağılımı elde etmek için her veri seti için ideal bir dağılım olarak görülebilir. Veri setimiz için 15 kelime ve üzeri uzunlukta tweetler seçime dahil edilmiştir. Ayrıca Şekil 4.8'de sunulan PAN-2015-EN veri setinin tweet uzunluk dağılımında, veri setinin %50 üzerinde olduğu için 10 kelime ve daha uzun olan tweetler seçime dahil edilmiştir. PAN-2015-EN veri setinde 30 ve üzeri kelime içeren 52 kayıt bulunduğundan dağılımda görülmemektedir.

PAN-2015-EN veri seti ve veri setimiz üzerinde FastText yöntemi ile ön işleme ve vektörleştirme yönteminin karşılaştırmalı analizi Çizelge 4.10'da sunulmuştur. Her modelde en iyi performansı elde ettiğimiz Bi-LSTM yöntemi gizli katmanda tek bir model olarak ele alınmıştır. Elde edilen sonuçlara göre nihai modelimizin FastText + BiLSTM modeline göre daha iyi performans gösterdiği ve genellemeye daha uygun olduğu görülmektedir.



Şekil 4.7. Veri setindeki tweet numaralarının tweet uzunluğuna göre dağılımı (5081-EN)



Şekil 4.8. Veri setindeki tweet numaralarının tweet uzunluğuna göre dağılımı (PAN-2015-EN)

Çizelge 4.10. Bi-LSTM modelinin vektörleştirme yöntemine göre veri setleri üzerinde karşılaştırılması (RMSE)

Kişilik Özellikleri	**5081-EN**		**PAN-2015-EN**	
	Nihai Modelimiz	**FastText**	**Nihai Modelimiz**	**FastText**
D	0.1434	0.2388	0.1590	0.1879
N	0.1594	0.3432	0.2262	0.2566
YB	0.1897	0.2529	0.1509	0.1772
ÖS	0.2109	0.3677	0.1470	0.1773
DA	0.1373	0.2148	0.1605	0.1670
A.O.	**0.1681**	0.2834	**0.1687**	0.1932

4.4. Değerlendirme ve Sonuç

İlk olarak anlamsız varlıkları anlamlı hale getirme işlemi uygulanmıştır. Daha sonra NLP yöntemleri ve LSTM türleri kullanılarak Bi-LSTM'nin KC2, KC3 ve Yok türleri ile tahmin başarı oranının yüksek olduğu görülmüştür. Daha sonra Emoji ve İfadelerin başarı oranına etkisi araştırılmış ve en iyi performansın Eİ1&Eİ4 ve Eİ2&Eİ5 seçimleri ile elde edildiği ancak derlem boyutundan dolayı en iyi seçimin Eİ2&Eİ5 olacağı ortaya konulmuştur. Daha sonra tüm etiketlerin (SL1) temizlenmesinin başarı oranını artırmadığı tespit edilmiştir. Tekrarlayan kayıtların (SL2, SL3 ve Yok) temizlenmesinde SL2 ve SL3 yöntemlerinin işleme performansı ve karmaşıklık açısından faydalı olacağı sunulmuştur. Ardından her bir kullanıcının içeriğinin tamamının kullanılmasının tahmin başarı oranında etkili olduğu görülmüştür. Daha sonra içerik bazlı kelime sayımlarının başarı oranına etkisi incelenmiş ve dengeli çeşitlilik elde etmek için veri setinin yarısından fazlasının seçilmesinin başarı oranını olumlu etkilediği sonucuna varılmıştır. Son olarak modelimiz ve FastText+Bi-LSTM yöntemimiz hem 5081-EN hem de PAN-2015-EN veri seti ile karşılaştırılmıştır. Veri setimiz ve modelimiz kullanılarak daha iyi bir tahmin başarı oranı elde edildiği belirlenmiştir.

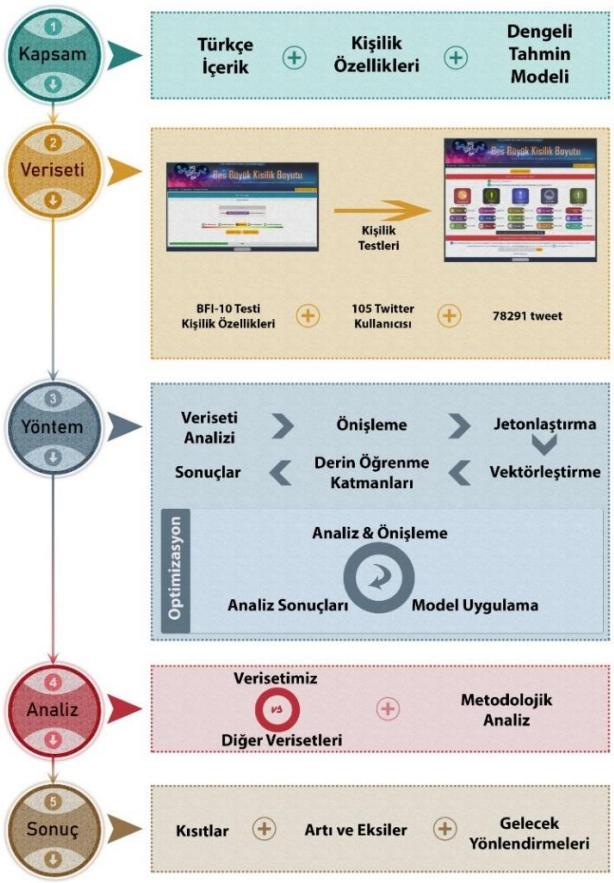
Yapısal analiz ile ardışık denemelerin herhangi bir platformda metin tabanlı içerikler için kişilik özelliklerini tahmin etme başarısını etkileyeceği görülmektedir. Platformda yer alan bilgi türleri doğru bir şekilde çıkarılarak başarı oranlarını tahmin etmedeki etkisi araştırılmalı, sonuçlar doğru veri formatı ve başarı oranı açısından değerlendirilmelidir.

Bir örneklemdeki verilere dayalı olarak kişilik tahmini yapabilmek için ilgili örneklemdeki verilerin yapısal analizinin yapılması gerekmektedir. Bu çalışmada bir Twitter örneği üzerinde yapısal analiz yapılmış, Tweet bilgilerindeki varlıklar ve metinsel farklılıklar oluşturacak yapılar incelenmiştir. Ön işleme adımları ile genellemeyi optimum seviyede tutmak amaçlanmıştır. Oluşturulan nihai model ile birlikte bu çalışmanın en önemli noktalarından biri de oluşturulan veri setidir. Yüksek boyutlu ve ideal olarak dağıtılan veri setinin başarı oranının PAN-2015-EN veri setine göre dengeli bir başarı oranına sahip olduğu gözlemlenmiştir. Aslında bazı kişilik özelliklerinde daha iyi sonuçlar elde edilse de tüm sonuçlar birlikte değerlendirildiğinde aritmetik ortalamaya göre sonuçlar birbirine yakın çıkmıştır.

Bu çalışmada, Twitter verileri üzerinde genellemeyi optimize etmek için gerekli ön işleme adımları ile başarı oranını artıran LSTM modellerinde uygulanan hiperparametre ayarlaması ile başarı oranını en ideal seviyede tutmayı amaçlanmıştır. Ancak deneyler sonucunda gözlemlenen en büyük sorun, işlenen veri boyutu ve model karmaşıklığından kaynaklanan donanım kısıtlamaları olmuştur. Bu noktada donanım kısıtlamalarının izin verdiği kadar sınırlar zorlanarak başarı oranı optimum seviyeye getirilmiştir. Çalışmanın bir diğer sınırlılığı ise Twitter verilerinden kişilik analizi için diğer veri setlerine ulaşılamamasıdır. Bu nedenle veri setimizin ileride yapılacak çalışmalar için faydalı bir kaynak olacağı düşünülmektedir. Literatüre katkısının yanı sıra, sistemimizin kolluk kuvvetlerinin istihbarat faaliyetleri, şirket veya kurumların işe alım süreçleri, reklam ajanslarının hedef kitlelerinin durum analizi, reklam ajanslarının hedef kitlelerinin durum analizi ve psikoloji alanında hizmet veren ortamlarda ön değerlendirme vb. gibi alanlarda kullanılabileceğine ve fayda sağlayacağına inanıyoruz.

5. TÜRKÇE İÇERİKLERDEN KİŞİLİK ÖZELLİKLERİ TAHMİNİ

Çalışmanın literatür taraması sonrası araştırma metodolojisini oluşturmak amacı ile bir sistematik akış belirlenmiştir. Bu akış şeması Şekil 5.1'de sunulmaktadır. Araştırma akışında ilk olarak literatüre katkısı düşünelerek kapsamın sınırlarının çizilmesi gereklidir. Bu sınırlar motivasyon ve problem tanımının yapıldığı ilk bölümde sunulmaktadır. Sonrasında veri toplama işlemleri için Etik Komisyon izinleri toplanmış ve paralelinde Kişilik Özelliklerinin kullanıcılar tarafından doldurulması ve teşvik etmek amacı ile oyunlaştırma temelli bir web uygulaması geliştirilmiştir. Bu web uygulamasının yayınlanmasıyla birlikte uygulamanın tanıtımı hem resmi hem de İnternet ortamındaki iletişim ve yayın kanalları ile yapılmıştır. Veri toplama işlemi tamamlandıktan sonra elde edilen verilerin analizi ve kullanılabilir veri seti oluşturma çalışmaları yapılmıştır. Nihai veri seti elde edildikten sonra anlamsal ön işlem yöntemleriyle birlikte farklı derin öğrenme yöntemleri ve vektörleştirme yöntemleri kullanılarak denemeler yapılmıştır. Son olarak ilgili sonuçlar analiz edilerek karşılaştırmalı analiz sunulmaktadır. İlgili adımlar sonraki başlıklarda detaylandırılmaktadır.



Şekil 5.1. Araştırma metodolojisi

5.1. Motivasyon ve Problem Tanımı

Literatür bütün olarak incelendiğinde her veri ortamı için farklı yapısal analizler ve ön işlem yollarının işlenmesi farklı sonuçlar doğurmaktadır. Bunun yanında her sosyal medya platformunda farklı veri türleri ve varlıklar bulunmaktadır. Bu noktada her sosyal medya platformunun ayrı yapısal analize tabi tutulması ve ön işlem adımlarının platformun yapısına göre tasarlanması gerekmektedir. Özellikle belirli bir seviyeye ulaşmış yapay zeka modellerinde başarımı arttırmak ve genelleştirmeye yaklaşmak için veri merkezli yaklaşıma dayalı hareket etmek daha fazla önem arz etmektedir [118].

Bu noktada Türkçe içerikli kişilik özellikli veri seti eksikliği ve Türkçe içerik temelli kişilik özellikleri tahmin modeli eksikliği görülmektedir. Yapılan çalışmalar ise içerikten kişilik özelliği tahminine dayalı değildir. Ayrıca bu tahmin modeli için dengeli ve genelleştirilebilir bir metodoloji oluşturmanın gerçek kullanım alanlarında da katkı sağlayacağı düşüncesi ile anlamsal yapı analizi kullanılarak içerik türüne göre veri merkezli bir yaklaşımın ortaya çıkarılması uygun bulunmuştur. Bu çalışma ile Türkçe içerik temelli sosyal medya platformlarından kişilik özellikleri tahminine yönelik Türkçe kişilik özellikleri veri seti oluşturulması ve Türkçe içeriklerden kişilik özellikleri tahminine dayalı genelleştirilmiş bir sistem tasarımı sunulması amaçlanmıştır.

5.2. Türkçe Kişilik Özellikleri Veri Seti

Türkçe içerikli kişilik özellikleri veri seti oluşturmak için ilk olarak kullanıcılara uygulanacak kişilik testleri belirlenmiştir. Devamında uygulanacak testler ile ilgili Etik Değerlendirme ve Etik Komisyon izinleri alınmıştır. Belirlenen kişilik testleri ile oyunlaştırma temelinde bir web uygulaması geliştirilerek Türkiye genelinde tanıtım çalışmaları yapılmıştır. Veri toplama işlemi tamamlandıktan sonra elde edilen veriler analiz edilerek nihai veri seti oluşturma işlemi tamamlanmıştır. Bu işlem adımları alt başlıklarda açıklanmaktadır.

5.2.1. Kullanıcılardan istenen bilgiler ve kişilik testlerinin belirlenmesi

Elde edilecek veri kümesinin dağılımını demografik olarak göstermek amacı ile kullanıcılardan istenecek ama doldurulması zorunlu tutulmayacak demografik bilgiler

belirlenmiştir. En çok kullanılan kişilik özellikleri boyutları Büyük Beşli (Big Five) olarak isimlendirilmektedir. Seçilen kişilik özelliği testleri hem Büyük Beşli'ye göre oluşturulmuş hem de doğrulanmış testlerdir. Bunlar, Büyük Beşli Envanteri 10 (Big Five Inventory 10 – BFI10), On Maddeli Kişilik Envanteri (Ten Items Personality Inventory – TIPI), Mini Uluslararası Kişilik Ölçeği Havuzu Envanteri (Mini International Personality Items Pool – Mini-IPIP), Büyük Beşli Envanteri 2-Küçük (Big Five Inventory 2 Small – BFI2-S), Büyük Beşli Envanteri 2 (Big Five Inventory 2 – BFI2)'dir. Sırası ile testlerde sunulan madde sayıları 10, 10, 20, 30 ve 60'dır. BFI2-S testi BFI2 testinin madde sayısı küçültülerek tekrar revize edilmiş halidir.

Tüm kişilik özelliği testleri karşılama metni ve sunulan maddelerin referans olduğu bir cümleye sahiptir. Maddeler bu referans alınan cümlede ilgili yere konularak kişinin kendini yada bir başkasını tanımlamasına yönelik oluşturulmuştur. Tüm testler bir başka özne içinde doldurulabilmektedir.

Kullanıcılardan kullanıcı hesabı tanımlaması için kullanıcı adı, şifre ve e-posta bilgileri zorunlu olarak istenmektedir. Ayrıca oluşturulan "Aydınlatma Metni ve Kullanıcı Sözleşmesi" nin okunup onaylanması zorunlu tutulmuştur. Twitter, Facebook ve Instagram hesaplarından herkese açık olarak paylaşım yapanlardan en az bir tanesini girmeleri istenmiştir. Hem araştırma sonucunu hem de verilerinden elde edilecek çıktılar hakkinde bilgi edinmek isteyen kullanıcılardan isteyenler için gerekli seçenekler oluşturulmuştur.

- Cinsiyet
- Yaş
- Eğitim durumu
- Medeni Hali
- Memleket
- Yaşanılan şehir
- Twitter, Facebook, Instagram, Youtube ve Blog adresleri
- Araştırma sonucunda araştırma çıktıları hakkında bilgi almak istenip istenmediği bilgisi

Beş kişilik ölçeği temelinde belirlenen 5 farklı kişilik özelliği testi Çizelge 5.1'de sunulmaktadır. Kullanıcıların hangi teste yoğun ilgi göstereceklerini anlamak amacıyla ve birden fazla testin sonucu ile kişilik özelliğini netleştirme düşüncesi ile birden fazla kişilik testi belirlenmiştir. Ayrıca kullanıcılara tanıdıkları tarafından değerlendirilme imkânı da sunulmuştur.

Kişilik testleri ön bilgilendirme amacı ile 10 kişiye uygulanmıştır. Ve bu uygulama ile ortalama test çözme süreleri elde edilmiştir. Elde edilen test çözme süreleri kişilik testlerinin yayınlandığı web uygulamasında ön bilgilendirme amacı ile sunulmuştur. Ortalama test çözme süreleri sırasıyla 41, 53, 82, 112 ve 230 saniyedir.

Çizelge 5.1. Kişilik testleri ve özellikleri

Kişilik Testi	**Referans**	**Soru Sayısı**	**Ölçek**
BFI-10	[119, 120]	10	5
TIPI	[121, 122]	10	7
MINI-IPIP	[123, 124]	20	5
BFI2-S	[125, 126]	30	5
BFI2	[125, 126]	60	5

BFI-10 Kişilik Testi

Test literatürde incelenen bir çok çalışmada kullanılmakta birlikte, ilk olarak 2007 yılında yapılan bir çalışmada [119] kullanılmıştır. Kişilik testinin ingilizce içeriği [119] ve Türkçe çevirimleri farklı çalışmalardan [120] elde edilmiştir. Aşağıda sunulan giriş cümleleri testin sunumunda kullanılmaktadır.

- İngilizce Giriş Cümlesi: How well do the following statements describe your personality?
- Türkçe Giriş Cümlesi: Aşağıdaki ifadeler kişiliğinizi ne kadar iyi tanımlamaktadır?
- İngilizce Cümle Yapısı: I see myself as someone who …
- Türkçe Cümle Yapısı: Kendimi … biri olarak görüyorum.

Ayrıca cümle yapısı ise maddelerin tamamlayıcı cümlesi olarak sunulmaktadır. Her madde cümle yapısında bulunan "..." işareti yerine konularak tanımlayıcı olmakta ve her defasında aynı cümleyi okumak yerine önemli olan maddenin belirleyici olması sağlanmaktadır. Kişilik Özelliği testine ait maddeler Çizelge 5.2'de İngilizce ve Türkçe karşılıkları olarak verilmektedir.

Her madde için verilecek cevaplar Hiç katılmıyorum (Disagree strongly) 1 puan, Biraz katılmıyorum (Disagree a little) 2 puan, Kararsızım (Neither agree nor disagree) 3 puan, Biraz katılıyorum (Agree a little) 4 puan, Tamamen katılıyorum (Agree strongly) 5 puan şeklinde isimlendirilmiş ve puanlandırılmıştır.

Test ile elde edilen Kişilik Özelliklikleri ise 5 farklı boyut olarak: Dışadönüklük (Extraversion), Yumuşak Başlılık (Agreeableness), Özdenetim/Sorumluluk (Conscientiousness), Nevrotiklik (Neuroticism), Deneyime Açıklık (Openness)'dır. 1(R) ve 6 maddeler Dışadönüklük, 2 ve 7(R) maddeler Yumuşak Başlılık, 3(R) ve 8 maddeler Özdenetim/Sorumluluk, 4(R) ve 9 maddeler Nevrotiklik, 5(R) ve 10 maddeler Deneyime Açıklık boyutlarına etki etmektedir. R harfi ile işaretlenen maddeler puan olarak işaretlemenin tersine sahiptir. Örnek olarak Hiç katılmıyorum seçeneği işaretlenmiş normal madde 1 puanken R olarak işaretlenmiş madde 5 puandır. Kişilik Özelliklerinin türkçe çevrimleri ise ilgili kaynaklardan [120, 124] elde edilmiştir.

Çizelge 5.2. BFI-10 Kişilik Testi Maddeleri

#	**İngilizce**	**Türkçe**
1	… is reserved.	… içine kapanık ...
2	… is generally trusting.	… genellikle güvenilir …
3	… tends to be lazy.	… yavaş hareket etme eğiliminde …
4	… is relaxed, handles stress well.	… rahat ve stresle başa çıkabilen …
5	… has few artistic interests.	… çok az sanatsal ilgisi olan …
6	… is outgoing, sociable.	… dışa dönük, sosyal …
7	… tends to find fault with others.	… başkalarının hatasını bulma eğiliminde …
8	… does a thorough job.	… bir işi tam yapacak …
9	… gets nervous easily.	… kolay sinirlenen …
10	… has an active imagination.	… yaratıcı …

TIPI Kişilik Testi:

Kişilik testinin ingilizce içeriği [121] ilgili çalışmadan ve Türkçe çevirimleri 2013 yılında yapılan bir çalışmadan [122] elde edilmiştir. Aşağıda sunulan giriş cümleleri testin sunumunda kullanılmaktadır.

- İngilizce Giriş Cümlesi: Here are a number of personality traits that may or may not apply to you. Please write a number next to each statement to indicate the extent to which you agree or disagree with that statement. You should rate the extent to which the pair of traits applies to you, even if one characteristic applies more strongly than the other.
- Türkçe Giriş Cümlesi: Aşağıda sizi tanımlayan ya da tanımlamayan birçok kişilik özelliği bulunmaktadır. Lütfen her bir ifadenin yanına, o ifadenin size tanımlama düzeyini dikkate alarak, o ifadeye katılıp katılmadığınızı belirtmek için 1 ile 7 arasında bir rakam yazın. İfadelerde sizi en çok tanımlayan özelliği dikkate alarak, uygun gördüğünüz rakamı yazın.

- İngilizce Cümle Yapısı: I see myself as …
- Türkçe Cümle Yapısı: Kendimi … olarak görüyorum.

Ayrıca cümle yapısı ise maddelerin tamamlayıcı cümlesi olarak sunulmaktadır. Her madde cümle yapısında bulunan "..." işareti yerine konularak tanımlayıcı olmakta ve her defasında aynı cümleyi okumak yerine önemli olan maddenin belirleyici olması sağlanmaktadır. Kişilik Özelliği testine ait maddeler Çizelge 5.3'de İngilizce ve Türkçe karşılıkları olarak verilmektedir.

Çizelge 5.3. TIPI Kişilik Testi Maddeleri

#	İngilizce	Türkçe
1	Extraverted, enthusiastic.	Dışa dönük, istekli
2	Critical, quarrelsome.	Eleştirel, kavgacı
3	Dependable, self-disciplined.	Güvenilir, öz-disiplinli
4	Anxious, easily upset.	Kaygılı, kolaylıkla hayal kırıklığına uğrayan
5	Open to new experiences, complex.	Yeni yaşantılara açık, karmaşık
6	Reserved, quiet.	Çekingen, sessiz
7	Sympathetic, warm.	Sempatik, sıcak
8	Disorganized, careless.	Altüst olmuş, dikkatsiz
9	Calm, emotionally stable.	Sakin, duygusal olarak dengeli
10	Conventional, uncreative.	Geleneksel, yaratıcı olmayan

Her madde için verilecek cevaplar Hiç katılmıyorum (Disagree strongly) 1 puan, Kısmen katılmıyorum (Disagree moderately) 2 puan, Biraz katılmıyorum (Disagree a little) 3 puan, Kararsızım (Neither agree nor disagree) 4 puan, Biraz katılıyorum (Agree a little) 5 puan, Kısmen katılıyorum (Agree moderately), Tamamen Katılıyorum (Agree strongly) 7 puan şeklinde isimlendirilmiş ve puanlandırılmıştır.

Test ile elde edilen Kişilik Özelliklikleri ise 5 farklı boyut olarak: Dışadönüklük (Extraversion), Yumuşak Başlılık (Agreeableness), Özdenetim/Sorumluluk (Conscientiousness), Duygusal Denge/Nevrotiklik (Emotional Stability/Neuroticism), Deneyime Açıklık (Openness)'dır. 1 ve 6(R) maddeler Dışadönüklük, 2(R) ve 7 maddeler Yumuşak Başlılık, 3 ve 8(R) maddeler Özdenetim/Sorumluluk, 4(R) ve 9 maddeler Nevrotiklik, 5 ve 10(R) maddeler Deneyime Açıklık boyutlarına etki etmektedir. R harfi ile işaretlenen maddeler puan olarak işaretlemenin tersine sahiptir. Örnek olarak Hiç

katılmıyorum seçeneği işaretlenmiş normal madde 1 puanken R olarak işaretlenmiş madde 7 puandır. Kişilik Özelliklerinin türkçe çevrimleri ise ilgili kaynaklardan [122, 124] elde edilmiştir.

MINI-IPIP Kişilik Testi

Kişilik testinin ingilizce içeriği [123] ilgili çalışmadan elde edilmiştir. Uluslararası Kişilik Maddeleri Havuzu (International Personality Items Pool - IPIP) tarafından oluşturulan havuzda yer alan kişilik testinin her maddesi bir koda sahiptir. Çizelge 5.4'te sunulan maddelerin listesi IPIP sitesinde [127] maddeler halinde bulunmaktadır. Tüm havuzun belirli bir kısmında bulunan maddelerin ise Türkçe çevirimleri Oya SOMER tarafından [128] yapılmış ve IPIP organizasyon sitesinde sunulmaktadır. Aşağıda sunulan giriş cümleleri testin sunumunda kullanılmaktadır.

- İngilizce Giriş Cümlesi: How much do you agree with each statement about you as you generally are now, not as you wish to be in the future?
- Türkçe Giriş Cümlesi: Lütfen olmak istediğiniz kişi olarak değil, şu anda olduğunuz kişiye göre maddeleri işaretleyiniz.
- İngilizce Cümle Yapısı: In general, I …
- Türkçe Cümle Yapısı: Genelde, Ben …

Ayrıca cümle yapısı ise maddelerin tamamlayıcı cümlesi olarak sunulmaktadır. Her madde cümle yapısında bulunan ... işareti yerine konularak tanımlayıcı olmakta ve her defasında aynı cümleyi okumak yerine önemli olan maddenin belirleyici olması sağlanmaktadır. Kişilik Özelliği testine ait maddeler Çizelge 5.4'te İngilizce ve Türkçe karşılıkları olarak verilmektedir.

Her madde için verilecek cevaplar Hiç katılmıyorum (Strongly Disagree) 1 puan, Katılmıyorum (Somewhat Disagree) 2 puan, Kararsızım (Neither Agree nor Disagree) 3 puan, Katılıyorum (Somewhat Agree) 4 puan, Tamamen katılıyorum (Strongly Agree) 5 puan şeklinde isimlendirilmiş ve puanlandırılmıştır.

Çizelge 5.4. MINI-IPIP Kişilik Testi Maddeleri

#	İngilizce	IPIP Kodu	Türkçe
1	Am the life of the party	H34, X94	Toplantıların gözdesiyimdir.
2	Sympathize with others' feelings	H1130	Başkalarının duygularını anlayıp paylaşırım.
3	Get chores done right away	E119	İşlerin doğru yapılmasını sağlarım.
4	Have frequent mood swings	E92	Ruhsal dengem sık bozulur.
5	Have a vivid imagination	X14	**Canlı bir hayal gücüne sahibim.
6	Don't talk a lot	X56	*Çok konuşmam.
7	Am not interested in other people's problems	X227	*Başka insanların problemleriyle ilgilenmem.
8	Often forget to put things back in their proper place	X82	*Eşyaları sık sık yerine koymayı unuturum.
9	Am relaxed most of the time	E141	Genelde rahatımdır.
10	Am not interested in abstract ideas	X228	Soyut fikirlerle ilgilenmem.
11	Talk to a lot of different people at parties	X83	*Toplantılarda değişik insanlarla konuşabilirim.
12	Feel others' emotions	E136	Başkalarının duygularını hissederim.
13	Like order	X118	*Sipariş vermeyi severim.
14	Get upset easily	X95	Moralim çabuk bozulur.
15	Have difficulty understanding abstract ideas	X176	Soyut fikirleri kavramakta zorlanırım.
16	Keep in the background	H154, X34	Arka planda kalmayı tercih ederim.
17	Am not really interested in others	X165	Başkalarıyla pek ilgilenmem.
18	Make a mess of things	H1467	İşleri karmakarışık yaparım.
19	Seldom feel blue	X156	Nadiren kendimi sıkıntılı hissederim.
20	Do not have a good imagination	X272	Hayal gücüm kuvvetli değildir.

* Türkçe kod listesinde bulunmadığından dolayı el ile Türkçeye çevirisi yapılmıştır.
** Türkçe kod listesinde yanlış yazıldığından (tam tersi olarak ifade edilmiştir) düzeltilmiş hali eklenmiştir.

Test ile elde edilen Kişilik Özelliklikleri ise 5 farklı boyut olarak: Dışadönüklük (Extraversion), Yumuşak Başlılık (Agreeableness), Özdenetim/Sorumluluk (Conscientiousness), Nevrotiklik (Neuroticism), Deneyime Açıklık (Openness)'dır. 1, 6(R), 11 ve 16(R) maddeler Dışadönüklük, 2, 7(R), 12 ve 17(R) maddeler Yumuşak Başlılık, 3, 8(R), 13 ve 18(R) maddeler Özdenetim/Sorumluluk, 4, 9(R), 14 ve 19(R) maddeler Nevrotiklik, 5, 10(R), 15(R) ve 20(R) maddeler Deneyime Açıklık boyutlarına etki etmektedir. R harfi ile işaretlenen maddeler puan olarak işaretlemenin tersine sahiptir. Örnek olarak Hiç katılmıyorum seçeneği işaretlenmiş normal madde 1 puanken R olarak işaretlenmiş madde 5 puandır. Kişilik Özelliklerinin türkçe çevrimleri ise ilgili kaynaktan [124] elde edilmiştir.

BFI2-S ve BFI2 Kişilik Testleri

Test ilk olarak 2017 yılında yapılan bir çalışmada [125] kullanılmış ve ingilizce test içerikleri ilgili çalışmadan elde edilmiştir. Yapılan çalışmada BFI2 için belirlenen 60 madde [126], kısaltılmış form BFI2-S için 30 madde [129] ve BFI2-XS 15 madde [130] olarak tasarlanmıştır. BFI2-XS'in kişilik özelliklerini ölçmede tercih edilmemesi yazarlar tarafından önerilmiştir. Türkçe çevirimleri ise aynı kuruma ait web sayfasından [131] elde edilmiştir. Aşağıda sunulan giriş cümleleri testin sunumunda kullanılmaktadır.

- İngilizce Giriş Cümlesi: Here are a number of characteristics that may or may not apply to you. For example, do you agree that you are someone who likes to spend time with others? Please write a number next to each statement to indicate the extent to which you agree or disagree with that statement.
- Türkçe Giriş Cümlesi: Aşağıda sizi kısmen tanımlayan (ya da pek tanımlayamayan) birtakım özellikler sunulmaktadır. Örneğin, başkaları ile zaman geçirmekten hoşlanan birisi olduğunuzu düşünüyor musunuz? Lütfen aşağıda verilen özelliklerin sizi ne oranda yansıttığını ya da yansıtmadığını belirtmek için sizi en iyi tanımlayan ifadeyi işaretleyiniz.
- İngilizce Cümle Yapısı: I am someone who …
- Türkçe Cümle Yapısı: Kendimi … biri olarak görüyorum.

Ayrıca cümle yapısı ise maddelerin tamamlayıcı cümlesi olarak sunulmaktadır. Her madde cümle yapısında bulunan ... işareti yerine konularak tanımlayıcı olmakta ve her defasında aynı cümleyi okumak yerine önemli olan maddenin belirleyici olması sağlanmaktadır. BFI2-S Kişilik Özelliği testine ait maddeler Çizelge 5.5'te BFI2 Kişilik Özelliği testine ait maddeler Çizelge 5.6'da İngilizce ve Türkçe karşılıkları olarak verilmektedir. Her iki test için temel bilgiler ortak olduğundan ayrı bölümler olarak sunulmamıştır. Aynı bölüm içinde testlerin maddeleri ve hangi maddenin hangi boyuta etki ettiği ayrı bilgiler olarak sunulmaktadır.

Önceki bölümlerde sunulan 3 farklı test, 5 farklı kişilik boyutunu temel almaktadır. BFI2 testleri ise bu ana boyutların (domains) altında alt boyutlarda (facets) kişilik özelliklerini belirlemektedir. Her ana boyut altında 3 alt boyut bulunmaktadır. Böylelikle toplamda 5 ana boyut ve 15 alt boyut sunulmaktadır.

Çizelge 5.5. BFI2-S Kişilik Testi Maddeleri

#	İngilizce	Türkçe
1	Tends to be quiet.	Sessiz olmaya eğilimli
2	Is compassionate, has a soft heart.	Şefkatli, yumuşak kalpli
3	Tends to be disorganized.	Dağınık olma eğiliminde
4	Worries a lot.	Çok endişelenen
5	Is fascinated by art, music, or literature.	Sanat, müzik ya da edebiyatla çok ilgili
6	Is dominant, acts as a leader.	Baskın, lider gibi davranan
7	Is sometimes rude to others.	Zaman zaman başkalarına kaba davranan
8	Has difficulty getting started on tasks.	İşe başlamakta zorlanan
9	Tends to feel depressed, blue.	Depresif, hüzünlü hissetmeye eğilimli
10	Has little interest in abstract ideas.	Soyut konulara az ilgi duyan
11	Is full of energy.	Enerji dolu
12	Assumes the best about people.	Başkaları hakkında hep iyi düşünen
13	Is reliable, can always be counted on.	Sözünde duran, başkalarının güvenebildiği
14	Is emotionally stable, not easily upset.	Duygusal olarak dengeli, keyfi kolay kaçmayan
15	Is original, comes up with new ideas.	Özgün, yeni fikirler üreten
16	Is outgoing, sociable.	Dışadönük, sosyal
17	Can be cold and uncaring.	Soğuk ve başkalarını umursamayan
18	Keeps things neat and tidy.	Etrafını temiz ve derli toplu tutan
19	Is relaxed, handles stress well.	Rahat, stresle baş edebilen
20	Has few artistic interests.	Sanatsal ilgileri az olan
21	Prefers to have others take charge.	Kararları başkalarının vermesini tercih eden
22	Is respectful, treats others with respect.	Saygılı, başkalarına saygılı davranan
23	Is persistent, works until the task is finished.	Kolay vazgeçmeyen, işin sonunu getiren
24	Feels secure, comfortable with self.	Güvenli, kendiyle barışık
25	Is complex, a deep thinker.	Çok yönlü, derin düşünen
26	Is less active than other people.	Başkalarından daha az hareketli
27	Tends to find fault with others.	Başkalarında hata arama eğiliminde olan
28	Can be somewhat careless.	Biraz özensiz olabilen
29	Is temperamental, gets emotional easily.	Değişken mizaçlı, çabuk sinirlenen
30	Has little creativity.	Yaratıcı yönü zayıf olan

Çizelge 5.6. BFI2 Kişilik Testi Maddeleri

#	İngilizce	Türkçe
1	Is outgoing, sociable.	Dışadönük, sosyal
2	Is compassionate, has a soft heart.	Şefkatli, yumuşak kalpli
3	Tends to be disorganized.	Dağınık olma eğiliminde
4	Is relaxed, handles stress well.	Rahat, stresle baş edebilen
5	Has few artistic interests.	Sanatsal ilgileri az olan
6	Has an assertive personality.	Atılgan, girişken
7	Is respectful, treats others with respect.	Saygılı, başkalarına saygılı davranan
8	Tends to be lazy.	Tembelliğe eğilimli
9	Stays optimistic after experiencing a setback.	Bir aksilik yaşadığında iyimserliğini koruyan
10	Is curious about many different things.	Farklı birçok şeye merak duyan
11	Rarely feels excited or eager.	Nadiren heyecanlanan ya da heveslenen
12	Tends to find fault with others.	Başkalarında hata arama eğiliminde olan
13	Is dependable, steady.	Güvenilir, istikrarlı
14	Is moody, has up and down mood swings.	Dakikası dakikasına uymayan, ruh hali inişli çıkışlı
15	Is inventive, finds clever ways to do things.	Yaratıcı, bir işi yapmanın akıllıca yöntemlerini bulan
16	Tends to be quiet.	Sessiz olmaya eğilimli
17	Feels little sympathy for others.	Başkalarının halinden pek anlamayan
18	Is systematic, likes to keep things in order.	Sistemli, her şeyin düzenli olmasını seven
19	Can be tense.	Gergin olabilen
20	Is fascinated by art, music, or literature.	Sanat, müzik ya da edebiyatla çok ilgili
21	Is dominant, acts as a leader.	Baskın, lider gibi davranan
22	Starts arguments with others.	Başkaları ile tartışma başlatan
23	Has difficulty getting started on tasks.	İşe başlamakta zorlanan
24	Feels secure, comfortable with self.	Güvenli, kendiyle barışık
25	Avoids intellectual, philosophical discussions.	Entelektüel, felsefi tartışmalardan kaçınan
26	Is less active than other people.	Başkalarından daha az hareketli
27	Has a forgiving nature.	Affedici bir yapısı olan
28	Can be somewhat careless.	Biraz özensiz olabilen
29	Is emotionally stable, not easily upset.	Duygusal olarak dengeli, keyfi kolay kaçmayan
30	Has little creativity.	Yaratıcı yönü zayıf olan
31	Is sometimes shy, introverted.	Bazen utangaç, içe dönük
32	Is helpful and unselfish with others.	Yardımsever, bencil olmayan
33	Keeps things neat and tidy.	Etrafını temiz ve derli toplu tutan
34	Worries a lot.	Çok endişelenen
35	Values art and beauty.	Sanata ve estetiğe değer veren

Çizelge 5.6. (devam) BFI2 Kişilik Testi Maddeleri

#	İngilizce	Türkçe
36	Finds it hard to influence people.	Başkalarını etkilemede zorlanan
37	Is sometimes rude to others.	Zaman zaman başkalarına kaba davranan
38	Is efficient, gets things done.	Verimli, iş bitiren
39	Often feels sad.	Sıkça üzgün hisseden
40	Is complex, a deep thinker.	Çok yönlü, derin düşünen
41	Is full of energy.	Enerji dolu
42	Is suspicious of others' intentions.	Başkalarının iyi niyetinden şüphe eden
43	Is reliable, can always be counted on.	Sözünde duran, başkalarının güvenebildiği
44	Keeps their emotions under control.	Duygularını kontrol altında tutan
45	Has difficulty imagining things.	Zihinde canlandırma yapmada zorlanan
46	Is talkative.	Konuşkan
47	Can be cold and uncaring.	Soğuk ve başkalarını umursamayan
48	Leaves a mess, doesn't clean up.	Arkasını toplamayan, dağınık bırakan
49	Rarely feels anxious or afraid.	Nadiren kaygılanan ya da korkan
50	Thinks poetry and plays are boring.	Şiir ve tiyatroyu sıkıcı bulan
51	Prefers to have others take charge.	Kararları başkalarının vermesini tercih eden
52	Is polite, courteous to others.	Kibar, başkalarına nezaketle yaklaşan
53	Is persistent, works until the task is finished.	Kolay vazgeçmeyen, işin sonunu getiren
54	Tends to feel depressed, blue.	Depresif, hüzünlü hissetmeye eğilimli
55	Has little interest in abstract ideas.	Soyut konulara az ilgi duyan
56	Shows a lot of enthusiasm.	Coşku dolu
57	Assumes the best about people.	Başkaları hakkında hep iyi düşünen
58	Sometimes behaves irresponsibly.	Bazen sorumsuzca davranan
59	Is temperamental, gets emotional easily.	Değişken mizaçlı, çabuk sinirlenen
60	Is original, comes up with new ideas.	Özgün, yeni fikirler üreten

Her madde için verilecek cevaplar Hiç katılmıyorum (Disagree strongly) 1 puan, Biraz katılmıyorum (Disagree a little) 2 puan, Kararsızım (Neither agree nor disagree) 3 puan, Biraz katılıyorum (Agree a little) 4 puan, Tamamen katılıyorum (Agree strongly) 5 puan şeklinde isimlendirilmiş ve puanlandırılmıştır.

Test ile elde edilen Kişilik Özelliklikleri ile maddelerin etkileri, 5 ana boyut ve 15 alt boyut olarak aşağıda sunulmaktadır. R harfi ile işaretlenen maddeler puan olarak işaretlemenin tersine sahiptir. Örnek olarak Hiç katılmıyorum seçeneği işaretlenmiş normal madde 1 puanken R olarak işaretlenmiş madde 5 puandır. Ana boyutların puan ise alt boyutların

ortalaması ile elde edilmektedir. Kişilik Özelliklerinin türkçe çevrimleri ise ilgili kaynaklardan [124, 131] bazı düzeltmeler yapılarak elde edilmiştir.

BFI2-S testine ait boyut-madde ilişkisi aşağıda maddeler olarak sunulmaktadır.

- Dışadönüklük (Extraversion)
    - Sosyallik (Sociability): 1(R), 16
    - Kendine Güven (Assertiveness): 6, 21(R)
    - Enerji Seviyesi (Energy Level): 11, 26(R)
- Yumuşak Başlılık (Agreeableness)
    - Merhamet (Compassion): 2, 17(R)
    - Saygılılık (Respectfulness): 7(R), 22
    - Güven (Trust): 12, 27(R)
- Özdenetim/Sorumluluk (Conscientiousness)
    - Düzenli Olma (Organization): 3(R), 18
    - Verimlilik (Productiveness): 8(R), 23
    - Sorumluluk (Responsibility): 13, 28(R)
- Nevrotiklik/Olumsuz Duygu (Neuroticism/Negative Emotion)
    - Anksiyete (Anxiety): 4, 19(R)
    - Depresyon (Depression): 9, 24(R)
    - Duygusal Değişkenlik (Emotional Volatility): 14(R), 29
- Deneyime Açıklık (Openness)
    - Estetik Hassasiyet (Aesthetic Sensitivity): 5, 20(R)
    - Entelektüel Merak (Intellectual Curiosity): 10(R), 25
    - Yaratıcı Hayal Gücü (Creative Imagination): 15, 30(R)

BFI2 testine ait boyut-madde ilişkisi aşağıda maddeler olarak sunulmaktadır.

- Dışadönüklük (Extraversion)
    - Sosyallik (Sociability): 1, 16(R), 31(R), 46
    - Kendine Güven (Assertiveness): 6, 21, 36(R), 51(R)
    - Enerji Seviyesi (Energy Level): 11(R), 26(R), 41, 56
- Yumuşak Başlılık (Agreeableness)
    - Merhamet (Compassion): 2, 17(R), 32, 47(R)
    - Saygılılık (Respectfulness): 7, 22(R), 37(R), 52
    - Güven (Trust): 12(R), 27, 42(R), 57
- Özdenetim/Sorumluluk (Conscientiousness)
    - Düzenli Olma (Organization): 3(R), 18, 33, 48(R)
    - Verimlilik (Productiveness): 8(R), 23(R), 38, 53
    - Sorumluluk (Responsibility): 13, 28(R), 43, 58(R)
- Nevrotiklik/Olumsuz Duygu (Neuroticism/Negative Emotion)
    - Anksiyete (Anxiety): 4(R), 19, 34, 49(R)
    - Depresyon (Depression): 9(R), 24(R), 39, 54
    - Duygusal Değişkenlik (Emotional Volatility): 14, 29(R), 44(R), 59
- Deneyime Açıklık (Openness)
    - Estetik Hassasiyet (Aesthetic Sensitivity): 5(R), 20, 35, 50(R)
    - Entelektüel Merak (Intellectual Curiosity): 10, 25(R), 40, 55(R)
    - Yaratıcı Hayal Gücü (Creative Imagination): 15, 30(R), 45(R), 60

5.2.2. Etik değerlendirme

Veri toplama işlemi için Gazi Üniversitesi Etik Komisyonuna başvuruda bulunulmuştur. 23.07.2020 tarih ve E.78156 sayılı yazı ile başvuru onaylanmıştır. Pandemi nedeni ile veri toplamada yeterli düzeye ulaşılamadığı düşünüldüğü için ek uzatma talebinde bulunulmuştur. Ek uzatma talebi 26.01.2021 tarihinde yapılan toplantı ile 6 ay olarak onaylanmıştır.

Elde edilen kişilik verileri ve sosyal medya platformlarından toplanan veriler tersine tespiti engelleyecek şekilde anonimleştirilmiştir. Ayrıca ilgili verilerden oluşturulan veri seti sadece üniversite kurumlarından gelecek resmi e-posta başvurusu ile kullanılacak şekilde anonim düzeyde dağıtılmaktadır.

5.2.3. Kişilik özellikleri için veri toplama sistemi

Veri toplama sistemi web uygulaması olarak herhangi bir cihazdan erişilebilecek şekilde Responsive olarak geliştirilmiştir. Ayrıca kullanıcıların uyguladıkları testlere her zaman erişebilecekleri, kendilerini değerlendirmeleri için yakınlarına istekte bulunabilecekleri ve sosyal medyada paylaşabilecekleri icon setleri ile uygulama zenginleştirilmiştir. Web uygulamaya ait ana ekran, test çözüm ekranı ve sonuç ekranları (bfi-10 ve bfi-2) sırasıyla EK-1'de Şekil 1.1, Şekil 1.2 ve EK-2'de Şekil 2.1, Şekil 2.2'de sunulmaktadır.

Web uygulamanın detaylı özellikleri aşağıda sunulmaktadır:

- Ana sayfadan tüm testler erişilebilir durumda sunulmaktadır.
- Responsive tasarım ile mobil cihazlardan görünüm ve test çözümü yapılabilmektedir.
- Başlangıçta kullanıcı kaydolup giriş yaptıktan sonra testleri çözebiliyorken sonrasında kaydolmadan giriş yapabilme özellikleri eklenmiştir.
- Her testin çözümü öncesi test ve çalışma hakkında bilgilendirme ve onay bölümü bulunmaktadır. Sistem kullanıcı sözleşmeyi onayladıktan sonra test çözüm işlemi başlayacak şekilde tasarlanmıştır.
- Kullanıcı test çözüm işlem sürecindeyken her soru EK-1'de Şekil 1.2'deki gibi belirgin yapıda sunulmaktadır. Önceki ve Sonraki sorulara geçişler hemen seçim alanlarının altında sunulmaktadır.
- Seçimler bittikten sonra tüm cevapların kontrolü ve değerlendirmeyi tamamla süreç ekranı gelmektedir. Tüm cevaplar verilmeden süreç tamamlama buttonu aktif olmamaktadır.
- Test çözüm işlemi tamamlandıktan sonra Değerlendirmeyi Tamamla buttonuna tıklandıktan sonra Test Sonuç ekranı (EK-2'de Şekil 2.1 ve Şekil 2.2) görünmektedir.

Tasarlanan kişilik özellikleri görselleri Çizelge 5.7'de sunulmaktadır. (BFI2 ve BFI2-S testleri hariç sadece 5 ana kişilik özelliği temel alınır)

- Test Sonuç ekranında kullanıcı kişilik testi sonuçlarını görebilmekte ve bunlardan herhangi birini sosyal medya hesabında (twitter ve facebook) paylaşabilmektedir. İlgili paylaşıma entegre olarak sistemin yayınlaştırmasıda sağlanmaktadır.
- Test sonuç ekranının en alt bölümünde herhangi bir yakınınızın sizi tanımlaması için ilgili testi onun e-posta adresine gönderebilmektesiniz. Sizi tanımlaması için bir yakınınızı ekledikten sonra ilgili e-posta adresine bir link ile birlikte e-posta gönderilmektedir. E-postada yer alan link ile yakınınız ilgili testi kullanarak sizin tanımlamanızı yapabilmektedir.
- Yakınınız sizi tanımlayıcı testi çözdükten sonra ilgili sonuçlara aynı sayfada eklediğiniz bölümden erişebilmektesiniz.
- Çözmüş olduğunuz tüm testler ve oturumları sisteme kullanıcı girişi yapılarak sonrasında da erişilebilecek şekilde tasarlanmıştır.
- Kullanıcı tercihlerini sistem tamamen kapanana kadar değiştirebilecek şekilde tasarlanmış ve bu konuda kullanıcılar bilgilendirilmiştir.

Web uygulamanın geliştirme ve yayınlanma sürecinde dikkat edilen hususlar aşağıda sunulmaktadır:

- Web uygulama yazılımsal olarak NodeJS altyapısı üzerine kurulmuştur.
- Ön yüz geliştirme aşaması React ve PrimeReact teknolojilerin kullanımına ek olarak kişisel olarak geliştirdiğimiz komponentler ile geliştirilmiştir.
- Arka plan geliştirme süreci ExpressJS kullanılarak RestApi ile tamamlanmıştır.
- Veritabanı olarak PostgreSQL kullanılmıştır.
- Geliştirme sürecinde kullanıcı kaydı, test yönetimi ve kullanıcının kendi çözdüğü testlerin gösterimi dışında herhangi bir yönetimsel ekran geliştirmesi yapılmamıştır.
- Ön yüz uygulaması özel paketleme yöntemleri kullanılarak yayınlanmıştır.
- 3 farklı bulut sunucu üzerinde ön yüz, api ve veri tabanı olarak yayınlanma işlemleri gerçekleştirilmiştir.
- Veri tabanına sadece api'nin bulunduğu sunucudan erişim ayarlaması yapılmıştır.
- Yönetimsel süreçler haricinde dosya ve veri tabanına erişim yetkilendirmeleri kapalı tutulmuştur.
- Veritabanı sunucusunda belirli sürelerde yetkilendirmeler açılarak veri tabanı kopyası yerel bilgisayara alınarak üzerinde çalışılmıştır.

Çizelge 5.7. Kişilik özellikleri test sonuç ekranı gösterim simgeleri listesi

Ana Kişilik Özellikleri	Alt Kişilik Özellikleri		
			
			
			
			
			

5.2.4. Yaygınlaştırma ve veri toplama

Web uygulamanın yayılımı için yapılan işlemler aşağıda sıralanmıştır.

- Türkiye'de bulunan tüm üniversite ve bakanlıklara resmi yazı yazıldı
- Kişisel ve farklı sosyal medya hesaplarından (Twitter, Linkedin, Facebook, Whatsapp, Instagram vb.) bildirimler yapıldı
- Üniversitelerde birkaç öğrenci grubuna tanıtımlar gerçekleştirildi
- Mail listeleri ve forumlarda tanıtımlar yapıldı
- Akademik ve özel e-posta listeleri oluşturularak spam listelere girmeyecek düzeyde toplu e-postalar atıldı

Veri toplama işlemi Ağustos 2021 itibari ile Etik Komisyon izni bittiğinde sonlandırılmıştır.

5.2.5. Toplanan verilerin analizi

Toplanan veriler tüm kullanıcı hesapları ve eklenen sosyal medya hesapları bilgileri Çizelge 5.8'de, çözülen test sayıları Çizelge 5.9'da ve testi çözen ve verilerine erişebildiğimiz (herkese açık - public) kullanıcıların sosyal medya hesapları Çizelge 5.10'da sunulmaktadır. 3 sosyal ağdan (Facebook, Twitter, Instagram) birine sahip olup herhangi bir kişilik testini çözen anlamlı kullanıcı sayısı 259'dur.

Çizelge 5.8. Toplanan veriler - hesap bilgileri

Parametre	Değer
Toplam kullanıcı	899
Test çözen kullanıcı	813
Test çözmeyen kullanıcı	86
Twitter hesabı	253
Facebook hesabı	255
Instagram hesabı	449
Youtube hesabı	4
Blog hesabı	4

Çizelge 5.9. Toplanan veriler - çözülen kişilik testi sayıları

BFI-10	TIPI	MINI-IPI	BFI-2	BFI-2S
762	407	370	276	274

Çizelge 5.10. Toplanan veriler - testi çözen kullanıcıların sosyal medya hesabı sayıları

Twitter	Facebook	Instagram
123	77	59

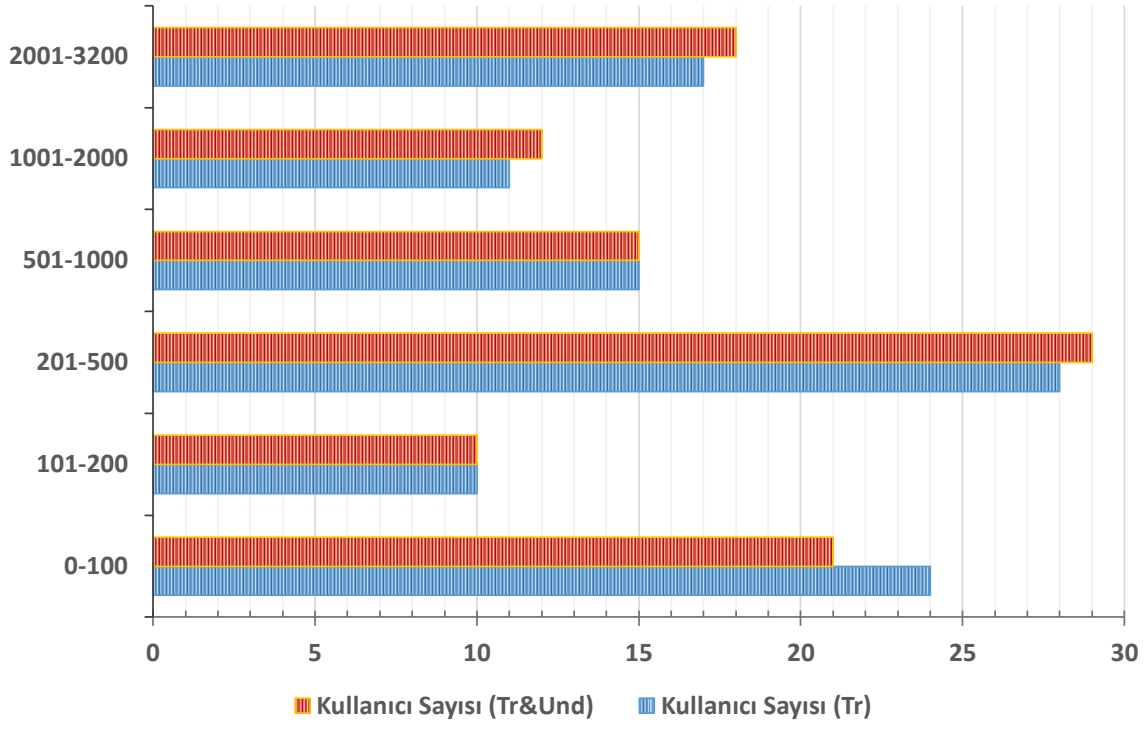
Çizelge 5.10'da sunulan verilerden Instagram üzerinde yapılan paylaşımlar hem düşük seviyede hem de fotoğraflar için yapılan açıklama gibi bilgilerin etiketler üzerine yoğunlaştığı görülmektedir. Facebook üzerinde yapılan paylaşımlar ise genellikle başka grupların gönderilerinin paylaşımı ve fotoğraf ve bilgi güncelleme üzerinedir. Metinsel veriler üzerinde yapılacak analizde Twitter için yapılmış test çözümleri değerlendirilerek 123 kullanıcıdan 113 kullanıcının tweet bilgileri toplanabilmiştir. Bu 113 kullanıcının 105 tanesi BFI-10 testini çözmüş fakat geri kalan 8 tanesi diğer testlerden birinin çözümünü gerçekleştirmiştir. Bu noktada tüm çözümlerin aynı testin ürünü olacak şekilde değerlendirilmesinin daha uygun olacağı düşünelerek BFI-10 testini çözen ve anlamlı twitter hesabı bulunan 105 kullanıcıdan oluşan bir veri seti oluşturulmuştur.

5.2.6. Nihai veri setinin oluşturulması

Toplanan kullanıcı bilgileri üzerinden yapılan incelemelerde Twitter kullanıcıları farklı diller üzerinde de paylaşımlar yapmaktadırlar. Bunun yanında birden fazla dilin karışımı veya hashtag, mention, url vb paylaşımlarda da Twitter tarafından Tanımsız (undefined - und) tanımlaması ile tweet bilgileri işaretlenmektedir. Türkçe (Tr) ve Und olarak işaretlenmiş tweet bilgilerine göre veri seti istatistikleri Çizelge 5.11 ve Şekil 5.2'de sunulmaktadır. Elde edilen bilgilere göre dağılım farklı tweet sayılarına göre birbirine yakın şekilde görülmektedir.

Çizelge 5.11. Türkçe ve tanımsız tweetlere göre tweet sayısı temelli kullanıcı sayıları

Tweet Sayısı Aralığı	Kullanıcı Sayısı (Tr)	Kullanıcı Sayısı (Tr&Und)
0-100	24	21
101-200	10	10
201-500	28	29
501-1000	15	15
1001-2000	11	12
2001-3200	17	18
Toplam	105	105



Şekil 5.2. Türkçe ve tanımsız tweetlere göre tweet sayısı temelli kullanıcı sayıları

105 Twitter kullanıcısından toplanan tweet bilgileri incelendiğinde toplam 78291 tweet bulunmaktadır. Bununla birlikte 1.104.699 kelime/varlık (eşsiz değildir) sayısında veri içermektedir. Türkçe ve Tanımsız tweetler ele alındığında Çizelge 5.12'de sunulan bilgiler elde edilmektedir. Tr ve Und olmayan tüm tweet verileri veri kümesinden temizlenmiştir.

Çizelge 5.12. Türkçe ve tanımsız tweetlere göre tweet ve kelime/varlık sayıları

	Tweet Sayısı (Tr)	**Tweet Sayısı (Tr&Und)**
Tweet sayısı	73896	78291
Kullanıcı ortalama tweet sayısı	798	840
Seçilen dilde yayınlanan tweetlerin kelime/varlık sayısı	1091745	1104699
Seçilen dilde yayınlanan tweetlerin kelime/varlık sayısının kullanıcı bazlı ortalaması	13027	13167

Oluşturulan nihai veri seti tr&und olarak kullanılmaktadır. Her kişilik özelliği değeri (Personality Trait Value - PTV) 0 ile 1 arasındadır. Veri setimizde her kişilik özelliğinin PTV<=0.5 (eşit ve altında) ve 0.5<PTV (üstünde) koşullarına göre tweet dağılımları Çizelge 5.13'te sunulmaktadır. Dağılımda genel olarak denge gözlemlenmektedir. Sadece YB kişilik özelliğinde 0.5 değeri ve üstü tweet yoğunluğu görülmektedir.

Çizelge 5.13. Kişilik özelliklerinin PTV<=0.5 ve 0.5<PTV koşullarına göre tweet dağılımları

Koşul	**D**	**N**	**YB**	**ÖS**	**DA**
0.5<PTV	43747	39766	81298	64357	51255
PTV<=0.5	47458	51439	9907	26848	39950

Veri setinin kelime ve varlık yoğunluğunu görmek için veri seti üzerinde standart ön işlem adımları ile birlikte Zemberek ve NLTK Stop Words'lerine göre temizleme işlemi uygulanarak farklı dağılımlar gözlemlenmiştir. İlgili dağılımlar NLTK için EK-3'de ve Zemberek için EK-4'de sunulmaktadır. EK-3 ve EK-4 birlikte incelendiğinde Stop Words listelerine göre farklı kelime yoğunlukları oluşmaktadır. Bunun yanında kişilik özelliklerine ve dağılımına göre varlık/kelime yoğunluklarının bazı durumlarda ayırt edici olduğu gözlenmektedir. Varlık yoğunluklarının PTV değerine göre gösterimleri Çizelge 5.14'te sunulmaktadır. Kullanılan önemsiz kelimeler içinde ilgili varlık kelimeleri bulunmadığından yoğunlukta bir fark oluşturmamaktadır. Kelime yoğunluklarındaki değişimler ise EK-3 ve EK-4 kelime bulutlarında görülmektedir.

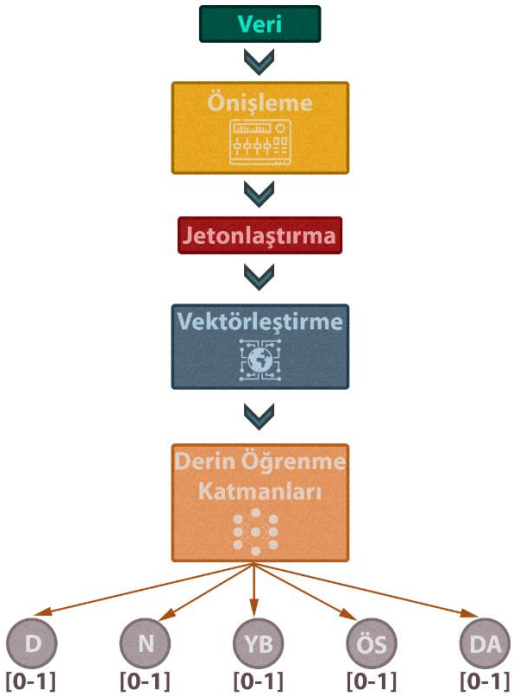
Çizelge 5.14. Kişilik özellikleri ve PTV'ye göre varlık yoğunlukları (soldan sağa büyükten küçüğe)

Kişilik özellikleri	**PTV Değeri (Değer: 0.5)**	
	Eşit veya daha küçük	**Daha büyük**
Dışadönük	E3, E2, E9, E5, E4, E1, E7, E8, E6	E3, E2, E5, E7, E9, E1, E4, E8, E6
Nevrotik	E3, E9, E2, E5, E1, E4, E7, E8, E6	E3, E2, E9, E5, E1, E4, E7, E8, E6
Yumuşak Başlı	E2, E9, E5, E3, E1, E4, E7, E8, E6	E3, E2, E9, E5, E1, E4, E7, E8, E6
Özdenetim Sahibi	E3, E9, E2, E1, E4, E5, E7, E8, E6	E3, E2, E5, E9, E7, E4, E1, E8, E6
Deneyime Açık	E3, E2, E9, E5, E1, E4, E7, E8, E6	E3, E2, E9, E5, E1, E7, E4, E8, E6

* E1:REPLY, E2:RETWEET, E3:USER, E4:HASHTAG, E5:URL, E6:IMAGE, E7:EMOJI, E8:EMOTICON, E9: NEWLINE

5.3. Tahmin Modeli

Veri seti elde edildikten sonra derin öğrenme tahmin modeli için standart bir sistematik şema üzerinden farklı yaklaşımların kombinasyonları ile en uygun derin öğrenme modelini oluşturmak için bir şablon ortaya konmuştur. Bu şablon genel hatları ile Şekil 5.3'te sunulmaktadır. Detayları ise alt başlıklarda açıklanmaktadır.



Şekil 5.3. Derin öğrenme modeli

5.3.1. Ön işlem

İngilizce verisetleri üzerinde yapılan önişlem denemelerinden yola çıkarak aşağıdaki ön işlem adımları standart olarak tahmin modelinde uygulanmıştır. Bu ön işlem adımları (Bölüm 4):

- Cevap (Reply) Etiketleme (CE)
- Retweet Etiketleme (RET)
- Resim Etiketleme (RE)
- URL Etiketleme (UE)
- Mention Etiketleme (ME)
- E-posta Etiketleme (EE)
- Hashtag Etiketleme (HE)
- Yeni Satır (\n) Etiketleme (YSE)
- Emoji ve Emoticonlar Arası Boşluk Ekleme (EEB)
- Emoji Etiketleme (EMJ)
- Emoticon Etiketleme (EMT)
- Boşluk Karakterlerini Temizleme (BKT)
- Tırnak/Alıntı İşaretlerini Temizleme (TAT)
- Sayıları Temizleme (ST)
- Noktalama İşaretlerini Temizleme (NİT)
- Çoklu Boşlukları Temizleme (ÇBT)

Bunun dışında dil farklılığından dolayı dile özgü yöntemler ile denemeleri yapılan ön işlem adımları aşağıda sıralanmaktadır:

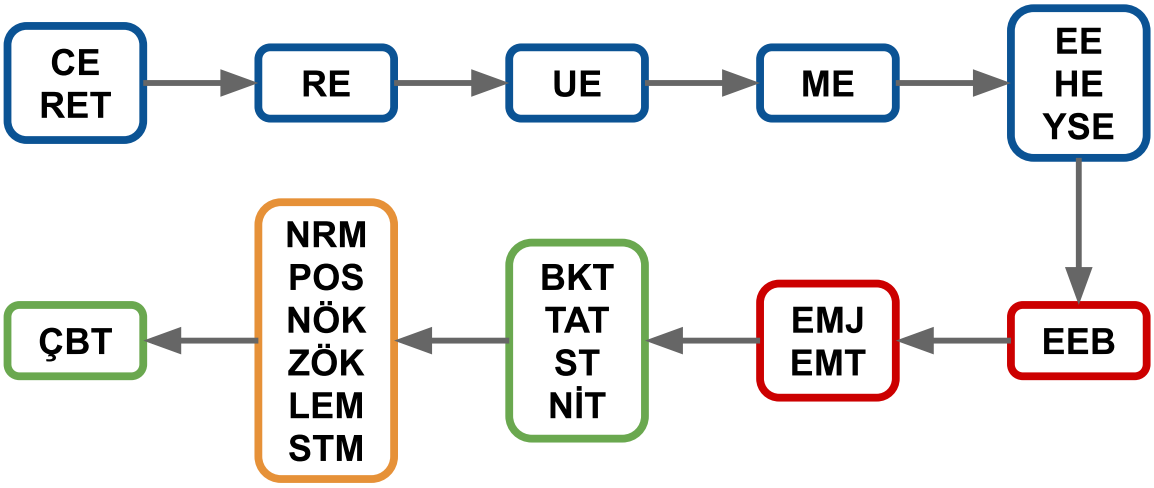
- Normalizasyon (NRM)
- PoS Etiketleme (POS)
- NLTK – Önemsiz Kelimeleri Temizleme (NÖK)
- Zemberek – Önemsiz Kelimeleri Temizleme (ZÖK)
- Kelime Kökleştirme – Lemmatization (LEM)
- Kelime Kökleştirme – Stemming (STM)

Önemsiz kelimelerin temizlenmesi hem NLTK hem de Zemberek stop-word listelerine göre farklı denemelere tabi tutulmaktadır. Ön işlem yöntemlerine ait işlem açıklama listesi Çizelge 5.15’te sunulmaktadır.

Çizelge 5.15. Ön işlem yöntemleri (PPM) – açıklama listesi

PPM	**Açıklama**
CE	Eğer ilgili gönderi bir başka gönderiye cevap olarak verilmiş ise [REPLY] etiketi eklenir.
RET	Eğer ilgili gönderi bir başka gönderiden replika edilmiş ise [RETWEET] etiketi eklenir.
RE	Eğer ilgili gönderi bir resim içeriyorsa ilgili resim bilgisi yerine [IMAGE] etiketi eklenir.
UE	Eğer ilgili gönderi bir URL içeriyorsa ilgili URL bilgisi yerine [URL] etiketi eklenir.
ME	Eğer ilgili gönderi bir başka hesabı refere ediyorsa (Örn: @user) ilgili kayıt yerine [USER] etiketi eklenir.
EE	Eğer ilgili gönderi bir e-posta adresi içeriyorsa ilgili e-posta adres bilgisi yerine [EMAIL] etiketi eklenir.
HE	Eğer ilgili gönderi bir hashtag içeriyorsa ilgili hashtag yerine [HASHTAG] etiketi eklenir.
YSE	Eğer ilgili gönderi bir yeni satır (\n) bilgisi içeriyorsa ilgili tanımlama yerine [NEWLINE] etiketi eklenir.
EEB	Emoji ve emoticonlar arasında boşluklar oluşturularak ilgili emoji ve emoticonlar üzerindeki işlemler kolaylaştırılır. EJT ve ETT için kolaylaştırma adımı olarak da kullanılmaktadır.
EMJ	Eğer ilgili gönderi bir emoji içeriyorsa ilgili emoji yerine [EMOJI] etiketi eklenir.
EMT	Eğer ilgili gönderi bir emoticon içeriyorsa ilgili emoticon yerine [EMOTICON] etiketi eklenir.
BKT	Eğer ilgili gönderi temel space işlemi haricinde bir boşluk tanımlaması içeriyorsa (\t \n \r \f \v) ilgili tanımlama yerine temel space karakteri eklenir.
TAT	Eğer ilgili gönderi tırnak işaretleri içeriyorsa ilgili işaretler yerine temel space karakteri eklenir. Tırnak işaretleri sonraki içeriklerle bitişik olabildiğinden dolayı ilk olarak temel space karakteri eklenerek ayrıştırma korunur. Sonrasında fazla boşluklar DMS ile temizlenir.
ST	Eğer ilgili gönderi rakam içeriyorsa ilgili rakamlar temizlenir.
NİT	Eğer ilgili gönderi noktalama işaretleri içeriyorsa ilgili noktalama işaretleri temizlenir.
ÇBT	Tokenization işleminin sağlıklı yapılması için hem normal içerikte bulunan hem de farklı ön işlem adımları ile oluşan fazla space karakterleri temizlenir.
NRM	Eğer ilgili gönderide yazım yanlışı olan kelimeler var ise ilgili kelimelerin doğru hali eklenir.
POS	Cümlenin öğelerine göre her kelimenin cümle öğresi bilgisi ilgili kelimeye eklenerek yeni cümle yapısı oluşturulur.
NÖK	NLTK kütüphanesinde bulunan Türkçe Stop Words listesine göre cümledeki önemsiz kelimeler temizlenir.
ZÖK	Zemberek kütüphanesinde bulunan Türkçe Stop Words listesine göre cümledeki önemsiz kelimeler temizlenir.
LEM	Kelimelerin kökleri bulunarak kökler ile cümle oluşturulur.
STM	Kelimelerin ön ve son eklerinden temizlenerek elde edilen yeni kelimelerle cümle oluşturulur.

Sunulan ön işlem adımları birlikte kullanımlarda bazı problemler oluşturabilir. Örneğin resimler bağlantı olarak içerikde bulunduğundan dolayı resimler ile ilgili işlemler URL filtreleme işlemlerinden önce uygulanmalıdır. Bu mantıkla oluşturulan ve seçime bağlı şekilde kullanılabilecek ön işlem adımları akışı Şekil 5.4'te sunulmaktadır.



Şekil 5.4. Ön işlem adımları – birlikte kullanımlarda izlenmesi gereken işlem sırası gösterimi

5.3.2. Ayrıştırma ve vektörleştirme

Her metinsel ifadenin matematiksel olarak işlenebilmesi için sayısal bir ifadeye dönüştürülmesi gerekmektedir. Metinsel ifade, ilk olarak boşluk karakteri (white space) temelli ayrıştırma/jetonlaştırma (tokenization) yöntemi uygulanarak ayırt edici ifadelere bölünür. Sonrasında belirli sayısallaştırma yöntemleri ile ilgili ifade birbirlerinden ayrıt edilecek sayısal ifadelere dönüştürülmektedir. Bu çalışmada eğitim veri setinden oluşturulan sözlük kullanılarak temel sayısallaştırma yöntemi ile birlikte BERTurk (Turkish BERT)[132], Electra[132, 133], FastText[112, 134], Doc2Vec[135, 136] ve Entity Frequency sayısallaştırma yöntemleri uygulanmıştır.

Entity Frequency sayısallaştırma yöntemi bir tweet içerisinde bulunabilecek genel varlıkların boyutuna göre oluşturulan bir sayısallaştırma yöntemi olarak kullanılmıştır. Bu yöntemi denememizdeki amacımız sadece varlıkların ilgili kayıt üzerindeki etki düzeylerinin kişilik tahmini modeli üzerindeki başarıma katkısı olup olmadığını anlamaktı. Entity Frequency sayısallaştırma yöntemi için REPLY, RETWEET, USER, HASHTAG, URL, IMAGE, EMAIL, NEWLINE, EMOJI ve EMOTICON (ifade) varlıkları sırası ile bir

dizi oluşturmaktadır. İlgili varlık içerik üzerinde hangi frekans düzeyinde geçiyorsa ona göre dizi oluşturma işlemi yapılmaktadır.

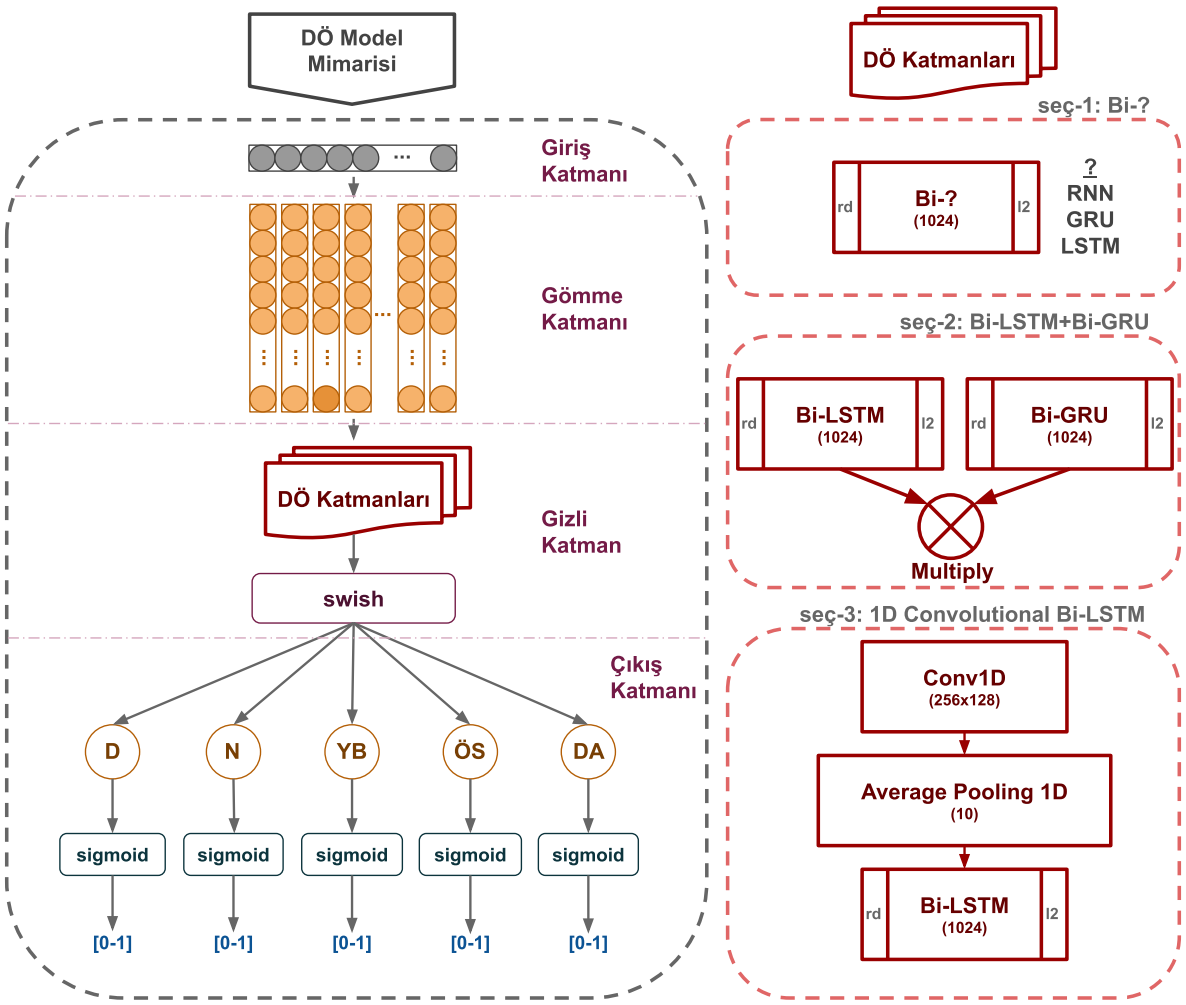
Sözlükteki her tweet için vektörleştirme uygulanmış ve vektörleştirilmiş tweet bilgisi girdi olarak modele sunulmuştur. Vektörleştirme işleminin nihai hali her sayısallaştırma yöntemi için nihai olarak Gömme Katmanında gerçekleştirilmektedir. Gömme katmanında her metin dizisinin maksimum kelime boyutunda vektörel gösterimleri oluşturulmaktadır. Bu gösterimler FastText, Doc2Vec vb. modelleri için ilgili model eğitiminden elde edilen ağırlık vektörlerinin Gömme Katmanına ağırlık olarak atanması ile özelleştirilmiştir. Gömme Katmanı temel parametreleri maksimum kelime sayısı, gömme boyutu ve girdi uzunluğu olarak sunulmaktadır. Ayrıca her girdinin uzunluğunu sabitlemek için Dizi/Cümle uzunluğuna göre cümle sonu dolgulama işlemi (post padding) yapılmıştır. Nihai olarak her cümle benzerlik ve yönelimi ayarlamak için yoğunluk vektörüne çevirilerek derin öğrenme modeli içinde başarım ve genelleştirmenin arttırılması sağlanmıştır.

5.3.3. Derin öğrenme yöntemleri

Çift yönlü hesaplama ile girdi olarak aldığı vektörleri daha iyi sonuçlar ile eğiten Bidirectional modeller tek yönlü modellerden daha iyi başarım göstermiştir. Bu sebeple bu çalışma bünyesinde RNN[137], GRU[138], LSTM[113, 117, 138] ve 1D Convolutional LSTM[138, 139] mimarilerinin Bidirectional kullanımları üzerine çalışılmıştır. Bu çalışmada kullanılan derin öğrenme mimarisi ve derin öğrenme katmanlarına göre varyasyonların gösterimi Şekil 5.5'te ve kullanılan yöntemlere ait parametre değerleri Çizelge 5.16'da sunulmaktadır.

Derin öğrenme modeli oluşturma çalışmalarında ilk olarak İngilizce veri seti üzerinde uygulanan Bi-LSTM modeli (Bölüm 4) uygulanmıştır. Başarımı arttırmak için hiper parametreler optimize edilmiştir. Modeller optimize edilirken epoch sayısı, katmanların birim sayısı, batch boyutu, farklı gizli katman sayıları, aktivasyon fonksiyonları (sigmoid, tanh, swish vb.), öğrenme oranı, dropout (sequential or recurrent), l1/l2 regularizatioın vb. parametreleri kullanılmıştır. Öğrenmenin dengeli şekilde (overfitting ve underfitting olmaması için) yapılması için dropout, regularization ve min-delta değişim kontrolü işlemleri yapılmıştır. Gizli katmanda Swish aktivasyon fonksiyonu [115] kullanılmıştır. Burada yapılan denemeler ve önceki çalışmaların etkisi bulunmaktadır. Ek olarak

değerlendirme metriği olarak Root Mean Squared Error (RMSE), kayıp fonksiyonu olarak Mean Squared Error (MSE) ve optimizasyon fonksiyonu olarak Adam kullanılmıştır.



Şekil 5.5. Önerilen derin öğrenme mimarisi şablonu ve kullanılan derin öğrenme (DÖ) gizli katmanları

Çizelge 5.16. Model parametreleri

Parametre	Değer
Dizi/Cümle uzunluğu (Sequence length)	32
Gömme boyutu (Embedding dimension)	256
Parti boyutu (Batch size)	128
Dönem sayısı (Epoch)	max: 1000 and if the minimum delta is 0.0001 by validation loss, the process is finished
Conv1D boyutu	256 x 128 (filter size x kernel size)

Çizelge 5.16. (devam) Model parametreleri

Parametre	Değer
Pool boyutu	10
Gizli Katman – aktivasyon fonksiyonu	swish
Gizli Katman – birim	1024
l2 aktivite düzenlileştirme değeri	0.001
Tekrarlayan bırakma oranı	%30 (1D Convolutional Bi-LSTM), %90 (diğerleri)
Çıkış – aktivasyon fonksiyonu	sigmoid
Kayıp fonksiyonu (Loss function)	MSE
Optimizasyon fonksiyonu	Adam [116] (lr: 0.0001, b1: 0.9, b2: 0.999, e:1e-7)
Değerlendirme fonksiyonu	RMSE

5.3.4. Çalışma ortamı

İlk olarak çalışma ortamını donanımsal altyapı olarak ifade etmek gerekirse;

- AMD Ryzen 9 3900X 12-Core Processor
- 64 GB (4x16 GB) 3600 MHz Ram
- 500 GB SSD (3400 okuma/2500 yazma)
- Nvidia Chipset RTX 2080Ti 11GB GPU

Donanımlar arasında verimli bir iletişim için donanımlar arası okuma, yazma ve iletim ile ilgili dengenin gözetilmesi gerekmektedir.

Yazılım temelli çalışma ortamında ise Windows 10 işletim sistemi üzerinde Tensorflow 2.4 için Cuda 10.2 ve Python 3.8.3 kullanılmaktadır. Ön işlem adımları ve vektörleştirme yaklaşımları için temel python işlemlerine ek olarak NLTK 3.5, Gensim 3.8.3, Zemberek-grpc 0.16.1 kütüphaneleri kullanılmaktadır.

5.3.5. Diğer verisetleri

Bu çalışma bünyesinde oluşturulan veri setini karşılaştırmak için Türkçe içerikli kişilik özellikleri veri seti bulunamadığından dolayı literatürde kullanılan PAN-2015-EN [106] ve

5081-EN (Bölüm 4'de oluşturulan veri seti) veri seti karşılaştırma amacı ile kullanılmaktadır. Dil yapısı dikkate alındığında uygulanan ön işlem adımları ve vektörleştirmede ortaya çıkacak farklı dizilimlerden dolayı eğitimdeki eğilim yakın olmayacaktır.

5.4. Deney ve Tartışma

Türkçe içerikli oluşturulan veri setimiz (105-TR) üzerinde PAN-2015-EN ve 5081-EN verisetlerinin kıyaslandığı derin öğrenme modeli ilk olarak uygulanmış olup sonuçlar Çizelge 5.17'de sunulmaktadır. RMSE değerleri incelendiğinde dil farklılığı temel alındığında dile göre optimize edilmiş modelin Türkçe içerikler üzerinde aynı etkiyi göstermediği görülmektedir.

Çizelge 5.17. Oluşturulan modele göre verisetlerinin karşılaştırılması (RMSE)

Veri seti	**Kişilik özellikleri**					**A.O.**
	D	**N**	**YB**	**ÖS**	**DA**	
5081-EN-IBM	0.1434	0.1594	0.1897	0.2109	0.1373	0.1681
PAN-2015	0.1590	0.2262	0.1509	0.1470	0.1605	0.1687
105-TR	0.3153	0.2660	0.1693	0.2132	0.2333	0.2394

Dil yapısı ve kullanım çeşitliliği farklılığı temel alınarak model üzerinde hiper parametre kestirimi bir parametre sabit tutulup seçilen parametrenin oynatılması ile ayarlanmaya çalışılmıştır. Ayrıca model karmaşıklığı ve hibrit çözümler üzerine de denemeler yapılmıştır. İlk model ile yapılan kestirimde elde edilen son modele ait hiperparametreler ve model yapı değişikliği ile elde edilen sonuçlar Çizelge 5.18'de sunulmaktadır. İlk Model İngilizce veri setinde oluşturulan model yapısıdır. Son Model ise hiper parametre ve model yapısında yapılan ayarlamalar ile Türkçe veri seti üzerinde elde edilen yeni sonuçları temsil etmektedir.

Oluşturulan model yapısının veri setinde kullanılan dil yapısı ve oluşturulan vektör yapılarına göre başarımlarda farklılık gösterebilmektedir. Bundan dolayı ilk olarak derin öğrenme mimarisine göre yapılan farklı model denemeleri 105-TR veri seti üzerinde uygulanmış ve en uygun derin öğrenme modeli bulunmaya çalışılmıştır. Sabit ön işlem ve

vektörleştirme işleminde derin öğrenme modeline göre karşılaştırmalı sonuçlar Çizelge 5.19'da sunulmaktadır.

Çizelge 5.18. İlk model ve son modelin karşılaştırılması (RMSE)

Model	**Kişilik özellikleri**					**A.O.**
	D	**N**	**YB**	**ÖS**	**DA**	
İlk Model	0.3153	0.2660	0.1693	0.2132	0.2333	0.2394
Son Model	0.2972	0.2495	0.1590	0.1995	0.2135	**0.2237**

Çizelge 5.19. Derin öğrenme modelleri ile elde edilen sonuçların karşılaştırması (RMSE)

Model	**Kişilik özellikleri**					**A.O.**
	D	**N**	**YB**	**ÖS**	**DA**	
Bi-RNN	0.3358	0.2684	0.3316	0.2773	0.2829	0.2992
Bi-GRU	0.2933	0.2490	0.1607	0.2000	0.2154	**0.2237**
Bi-LSTM	0.2972	0.2495	0.1590	0.1995	0.2135	**0.2237**
Bi-LSTM + Bi-GRU	0.3019	0.2601	0.1615	0.2028	0.2202	0.2293
Convolutional Bi-LSTM	0.3168	0.2628	0.1641	0.2085	0.2209	0.2346

Elde edilen sonuçlara göre Bi-GRU ve Bi-LSTM 1024 birimlik modellerde ortalama başarım aynı olmakla birlikte Bi-LSTM ile elde edilen sonuçlarda 3 farklı kişilik özelliği için en iyi başarımlar gözlemlenmektedir. Bu noktada model seçiminde genel başarım ve ayrık başarım düşünüldüğünde Bi-LSTM modelinin kullanımı daha mantıklı görünmektedir.

Derin öğrenme modeli, hibrit yaklaşımlar, hiper parametre kestirimi gibi denemelere ek olarak oluşturulan derin öğrenme model şablonundaki önişlem adımlarına da eklemeler yapılmıştır. Yanlış yazımların modelin başarımı üzerinde etkisi için ilgili tweetler üzerinde Normalization işlemine ek olarak Part-of-Speech Tagging, Stemming, Lemmatization ve Clear Stop Words işlemleri uygulanarak model başarımı üzerindeki etkisine bakılmıştır. Temel önişlemler sabit kalmak koşulu ile kelime/cümle yapısı ile ilgili nlp işlemleri ve Bi-LSTM modeli ile elde edilen sonuçlar Çizelge 5.20'de sunulmaktadır.

Çizelge 5.20. Mevcut modele göre uygulanan kelime/cümle işlemleri ile elde edilen sonuçların karşılaştırması (RMSE)

Kelime/Cümle İşlemi	**Kişilik özellikleri**					**A.O.**
	D	**N**	**YB**	**ÖS**	**DA**	
Hiçbiri	0.2972	0.2495	0.1590	0.1995	0.2135	**0.2237**
Normalizasyon	0.2959	0.2559	0.1610	0.2032	0.2181	0.2266
PoS Etiketleme	0.2957	0.2521	0.1604	0.1993	0.2178	0.2251
NLTK – Önemsiz Kelimeleri Temizleme	0.2982	0.2478	0.1606	0.1992	0.2167	**0.2245**
Zemberek – Önemsiz Kelimeleri Temizleme	0.2922	0.2509	0.1606	0.1995	0.2157	**0.2239**
Lemmatization	0.2965	0.2540	0.1619	0.2023	0.2200	0.2269
Stemming	0.2953	0.2482	0.1591	0.1981	0.2140	**0.2229**

Çizelge 5.20'ye göre herhangi bir kelime/cümle işlemi uygulanmadığındaki başarı ile Clear Stop Words ve Stemming işlemleri ile uygulandığındaki başarı birbirine yakın durumdadır. Kelime havuzlarını minimize etme durumlarında Clear Stop Words ve Stemming işlemleri uygulanarak kelime havuzu indirgeme işlemi yapılabilir.

Son olarak Bi-LSTM modeli sabit kalmak koşulu ile vektörleştirme işlemlerinde farklı yaklaşımlar kullanılmıştır. Bu yaklaşımlar ile yapılan denemeler Çizelge 5.21'de sunulmaktadır. Vektörleştirme adımı hazır olan modeller de dahil olmak üzere oluşturduğumuz eğitim veri setine göre düzenlenmiştir.

Çizelge 5.21 incelendiğinde standart vektörleştirme yöntemi dediğimiz eğitim veri setinden oluşturulan sözlük ile oluşturulan kelime frekanslarından vektörleştirme işleminin başarımının yüksek olduğu gözlenmektedir. Bunun yanında BERTurk'e ait olan 32K ve 128K sözlüklere sahip olan modeller ile yapılan denemelerde de yine başarımların farklı kişilik özelliklerinde yüksek elde edildiği gözlemlenmiştir. Genelleştirme temel alındığında BERTurk ile oluşturulan modellerde daha yüksek kapasiteli sözlük işletildiğinden daha iyi sonuç alınacağı düşünülmektedir. Bunun yanında model performansı temel alındığında Standart vektörleştirmenin daha az sözlük ve sözlük frekanslarındaki skalanın daha az olmasından dolayı daha az işlem kapasitesi ile işlem yapabileceği görülmektedir.

Çizelge 5.21. Sabit ön işlem ve Bi-LSTM modelinde farklı vektörleştirme işlemleri ile elde edilen sonuçların karşılaştırması (RMSE)

Vektörleştirme Yöntemi	**Kişilik özellikleri**					**A.O.**
	D	**N**	**YB**	**ÖS**	**DA**	
Standart	0.2972	0.2495	0.1590	0.1995	0.2135	**0.2237**
BERTurk – 32K uncased	0.2922	0.2498	0.1603	0.2028	0.2149	**0.2240**
BERTurk – 128K uncased	0.2946	0.2485	0.1578	0.2009	0.2143	**0.2232**
Doc2Vec	0.3202	0.2677	0.1690	0.2182	0.2348	0.2420
Electra	0.2958	0.2529	0.1607	0.2008	0.2171	0.2254
Entity Frequency	0.3047	0.2629	0.1653	0.2069	0.2322	0.2344
FastText – wiki.tr	0.3284	0.2681	0.1702	0.2249	0.2398	0.2463
FastText – cc.tr.300	0.3659	0.2758	0.1709	0.2403	0.2415	0.2589

5.5. Değerlendirme ve Sonuç

Herhangi bir dili temel alarak elde edilen içeriklerden kullanıcıların kişilik özelliklerini tahmin etmek için, dilin yapısı düşünülerek ilgili içeriğin analizi ve model mimarisinin oluşturulması gerekmektedir. Diller arasındaki yapısal farklılıklar nedeni ile oluşacak sayısal tanımlar da farklı özelliklerde olacaktır. Bu sebeple her dil için ayrı bir model mimarisinin düşünülmesi gerekmektedir.

Bu çalışma ile Türkçe içeriklerden kişilik özellikleri tahmini için ilk olarak yaygın sosyal medya platformları düşünülmekle birlikte elde edilen nihai veri seti ile Twitter içerikleri ile kişilik özellikleri ilişkilendirilebilmiştir. İngilizce verisetleri üzerinde başarım gösteren modeli Türkçe içeriklerden oluşturulan veri seti üzerinde uygulayarak diller arasındaki yapısal işlem farklılıklarını gözlemlenmiştir. Bunun yanında farklı ön işlem adımları ile ne gibi sonuçlar elde edilebileceğini görülmüştür. Devamında farklı derin öğrenme mimarilerinin veri seti üzerindeki başarım etkisi incelenerek Bi-LSTM'in başarım üzerindeki etkisi bir daha ortaya konmaktadır. Son olarak vektörleştirme adımında uygulanan farklı yöntemler ile Standart ve BERTurk yöntemlerinin iyi başarım gösterdikleri gözlemlenmektedir.

Her mimari ve ara işlemler farklı verisetleri ve platforma dayalı verilere göre farklılık gösterebilmektedir. Bu çalışmada oluşturulan ve kullanılan veri seti Twitter üzerinde oluşturulan Tweet yapısını temel almaktadır. Tweetler belirli karakter kısıtı ve içeriklerindeki varlık çeşitliliği çeşitli sayıdadır. Aynı şekilde farklı platformlarda daha farklı varlıkların olabileceği bilinmektedir. Bu sebeple oluşturulacak her modelin temel alacağı platforma göre tekrar yapılandırılması önem arzetmektedir.

Bu çalışma ile bu alanda yapılacak akademik çalışmalar ile birlikte kamu ve özel kurumlarda insanlar/kullanıcılar hakkında ön bir değerlendirmeyi temel alacak tüm alanlara katkı sağlanması amaçlanmıştır.

6. SONUÇ

Bir bilgiye olan ihtiyaçtan kaynaklı ortaya çıkan istihbarat gereksinimi, aslında hayatımızın her anını etkileyebilmektedir. Günümüzde hükümetler, reklam ajansları, firmalar ve hatta kişisel ihtiyaçlar için birçok farklı istihbarat gerektiren durum oluşmakta ve bunun için kişisel veya profesyonel çabalara başvurulmaktadır. Özellikle Internetin hatta daha da özelleştirirsek sosyal medyanın kullanımının hayatımızın vazgeçilmez bir parçası olmasını da düşünürsek ciddi bir bilgi üretim ve yayım ortamı da bulunmaktadır. Bu noktada geriye kalan yalnızca başarıya götürecek adımlarla istenilen bilginin birçok bilginin içerisinden çıkartılıp kullanılmasıdır. Başarıya ulaştıran adımlardan kasıt, çalışmamızda sunulan OSINT hayat döngüsünün bir problem için doğru şekilde doğru araçlar ve yöntemlerle uygulanmasıdır. Her problemin oluşturduğu veri uzayı ve çıkarım yöntemi başarımda farklı etkiler ortaya çıkarttığından dolayı hem özel sektörde hem de akademik ortamda problemlere özgü modeller geliştirilmekte ve araştırılmaktadır. Sosyal medya alanında kullanıcı kişiliğine dayalı kişilik özelliği tahmini çalışmalarında başarım oranının yeterince iyi düzeylerde, dengeli veya genelleştirilebilir olmadığı görülmektedir.

Bu tez çalışmasında ilk olarak genel istihbarat türleri üzerinden araştırmalar yapılmış ve OSINT'in önemi ortaya konmuştur. OSINT'in literatürdeki çalışma alanları üzerinde yoğunlaşılan çalışma alanlarından sosyal medya platformlarının önemi görülmüştür. Sosyal medya platformları üzerinde kullanıcıların temel profil özelliklerine ek olarak ilk bakışta görülmeyen özelliklerinin tespitinin önemli olduğu görülmüş ve kullanım alanları önem düzeyleri belirtilerek sunulmuştur.

OSINT yöntemleri kullanılarak kişilik özellikleri tespiti çalışma alanının önemini ortaya koyan çalışmamızda literatürde kişilik özellikleri tahmini üzerine yapılan çalışmalar kategorik özellikler temel alınarak analiz edilerek detaylı olarak çalışmamızda sunulmuştur. Böylelikle bu çalışma alanında yapılan ve yapılacak olan çalışmalara önemli bir bakış açısı kazandırılmıştır. Kişilik özellikleri tahminine yönelik İngilizce çalışmalarda kullanılan verisetlerine erişim problemleri, ilgili verisetlerinin yeterli boyutta örnekleme oluşturacak düzeyde olmadığı ve genelleştirilebilir düzeyde çeşitli içeriğe sahip olmadığı gerekçeleri ile tez çalışmamızın kapsamlarından biri belirlenmiştir. Devamında Türkçe içeriklerden kişilik tahminine yönelik çalışmaların eksikliğinin tespiti, yeterli düzeyde çalışmaların bu kapsama

yönelmediği gerekçesi ve Türkçe içerikli kişilik özelliği veri seti oluşturma çabası ile dengeli ve başarım oranı yüksek tahmin modeli oluşturma düşüncesi ile tez çalışmamızın ana kapsamı şekillendirilmiştir. Bunlara ek olarak OSINT yaşam döngüsünün önemli bölümleri tekrar şekillendirilerek revize edilmiş bir yaşam döngüsü ile çalışmanın ana hatları belirlenmiştir.

İngilizce içerikli kişilik özellikleri tahmini çalışması için öncelikle 13 pozitif-negatif anlama gelebilen kelimeler üzerinden paylaşım yapan kullanıcıların belirli bir örneklemi üzerin kullanıcı tweet verileri çekilmiş olup, sonrasında bu içeriklerin etikleme işlemleri IBM Personality Insight servisi üzerinden etiketlenmiş ve kapsamlı bir İngilizce kişilik özellikleri veri seti oluşturulmuştur. Bu veri seti üzerinden yapılan anlamsal analiz ile birlikte uygulanan ön işlem yöntemleri kullanılarak kişilik özellikleri tahmin modeli tasarımı yapılmış ve ön işlem adımlarına göre tahmin modeli genelleştirme ve optimizasyon işlemleri uygulanmıştır. Sonuç olarak elde edilen model ile hem oluşturduğumuz veri seti hem de literatürde kullanılan PAN-2015-EN veri seti üzerinde genelleştirilebilir, dengeli ve başarımı yüksek bir tahmin modeli ortaya konmuştur.

Türkçe içeriklerden kişilik özellikleri tahmini çalışması daha uzun bir süreçte tamamlanabilmiştir. Bundaki ana etkenler; veri seti oluşturma aşamasında uyguladığımız kişilik testlerinin belirlenmesi, veri toplama web uygulamasının geliştirilmesi, web uygulamanın tanıtım çalışmaları ve pandemi sürecindeki doğal etkenler, yapılan kişilik testleri ve kişilerin sosyal medya hesaplarının kapsamlı analizi ve nihai veri seti oluşturma sürecindeki analiz süreçleridir. Çok çeşitli veri toplamadaki ana amacımız birden fazla sosyal medya platformu ile birden fazla kişilik testi ile birlikte daha kapsamlı bir veri seti ve daha uygun bir tahmin modeli çalışması olmasıydı. Fakat bu süreçte hem doldurulan kişilik testlerindeki çeşitliliğin azlığı hem katılım sağlayan kullanıcıların sosyal medya hesaplarını paylaşmadaki kaygıları hem de paylaşılan sosyal medya hesaplarındaki içerik yetersizlikleri nedeni ile BFI-10 kişilik testini çözen ve Twitter hesabında anlamlı içerik barındıran kullanıcılardan Türkçe içerikli kişilik özellikleri veri seti oluşturulmuştur. Sonrasında elde edilen Türkçe veri seti üzerinde İngilizce veri seti üzerinde uygulanmış tahmin modeli entegrasyonu gerçekleştirilmiştir. Dil yapısı farklılıkları, elde edilen sözlük farklılıkları ve bazı ön işlem yöntemlerinde uygulanan metodların etkisi nedeni ile birebir aynı başarım oranı elde edilememiştir. Bu sebeple ilgili model üzerinde tekrar optimizasyon çalışmaları yapılarak farklı yöntemler ile birlikte daha iyi bir model oluşturulmuş ve aşamalı analizi

sunulmuştur. Böylelikle Türkçe içerikli kişilik özellikleri veri seti ile birlikte Türkçe içerikli kişilik özellikleri tahmin modeli oluşturma amacımız nihayete ermiştir.

Tüm tez çalışması incelendiğinde, bu tez çalışması bünyesinde yürütülen alt çalışmaların sunduğu yenilik ve katkılar aşağıda sunulmaktadır:

- Genel olarak;
  - Metinsel verilerin yapısal analizinde sunulan farklı ön işleme adımlarının detaylı değerlendirmesi yapılmıştır.
  - Kişilik özelliklerini tahmin etmede literatürün kapsamlı ve kategorik analizi yapılmıştır.
  - Kişilik özelliklerini tahmin etmede Twitter öznelinde sosyal medya platformlarına ait verilerin yapısal analizinin önemi ortaya konmuştur.
- İngilizce içerik üzerine;
  - Genellemeye uygun Twitter içeriklerine dayalı kişilik özellikleri veri seti oluşturulmuştur.
  - Yapısal ön işleme adımlarıyla kişilik özellikleri tahmini için Dengeli Bi-LSTM modeli sunulmuştur.
- Türkçe içerik üzerine;
  - Türkçe içeriklerden kişilik özellikleri tahmininde karşılaştırmalı ve yapısal analiz sunulmuştur.
  - Türkçe içeriklerden kişilik özellikleri tahmininde sosyal medya platform verilerinin Twitter öznelinde yapısal analizinin önemi ortaya konmuştur.
  - Türkçe içerikli metinlere ait kişilik özellikleri veri seti oluşturulmuştur.
  - Yapısal ön işlem adımları ile genelleştirilmiş derin öğrenme modelleri sunulmuştur.

Yapılan çalışmalarımızda toplanan verinin boyutu anlamlı ve dengeli olarak arttırıldıkça başarım oranının arttığı görülmektedir. Kişilik testleriyle kullanılabilir ve anlamlı içerik metinlerinden kişilik özellikleri veri setini devasa boyutlarda oluşturmanın maliyeti bu çalışmanın ana kısıtlarından bir tanesidir. Her ne kadar yeterli düzeyde bir veri seti oluşturulsa da daha yüksek boyutta dengeli bir veri seti oluşturmanın başarım ve genelleştirme üzerinde etkisi yadsınamaz bir gerçektir. Diğer bir kısıt ise çalışma ortamından kaynaklanmaktadır. Donanımsal çalışma ortamında spesifik konfigürasyonlar ile tüm süreçler yönetildiğinden ve hassas veriler üzerinde çalışıldığından dolayı büyük ölçekli bulut servisler kullanılamadı. Kişisel çalışma ortamımızdaki donanım kısıtlarından dolayı çok daha karmaşık modeller üzerinde deneme yapılamaması diğer bir kısıt olarak sunulabilir.

Gelecek çalışmalarda gelişen ve geliştirilen farklı tahmin modellerinin hibrit yaklaşımları üzerine farklı model denemeleri yapılabilir. Bunun yanında veri merkezli yaklaşım temelinde farklı sosyal medya platformlarındaki veri şablonları üzerinde farklı veri işleme

modelleri oluşturulabilir. Ayrıca dile göre değişkenlik gösteren farklı ayrıştırma/jetonlaştırma yöntemlerinin (tokenization methods) kişilik özelliği tahmin modelleri üzerindeki etkisi incelenebilir. Ek olarak sosyal medya platformlarında bulunan varlıklar (resim, web sayfası vb.) derinlemesine incelenerek içerikte yapılan değişikliklerle başarımdaki etkileri gözlemlenebilir. Bu noktada bir resim varlığının içeriği, renkler, renk tonları, kategorik şekiller örnek gösterilebilir. Ayrıca bir web sayfasının içeriğini temsil eden anahtar kelimelerin ilgili içerik ile ilgisi üzerinden tahmin modeli başarımı üzerine denemeler yapılabilir. Son olarak Varlık İsmi Tanıma (Name Entity Recognition) ile etiketlenmiş önemli isim varlıkları ile tweetler arası ilişkilerin model başarımı ve genelleştirmesi üzerindeki etkileri çalışılabilir.

KAYNAKLAR

1. İnternet: Tdk. Türk Dil Kurumu. İstihbarat Kelime Anlamı. URL: http://www.tdk.gov.tr/index.php?option=com_gts&arama=gts&guid=TDK.GTS.5c272c32beb8e3.27975555, Son Erişim Tarihi: 25.06.2021.

2. İnternet: Od. Oxford Dictionary. Intelligence Word Mean. URL: https://en.oxforddictionaries.com/definition/intelligence, Son Erişim Tarihi: 25.06.2021.

3. İnternet: Office of the Director of National Intelligence. What is Intelligence? URL: https://www.dni.gov/index.php/what-we-do/what-is-intelligence, Son Erişim Tarihi: 25.06.2021.

4. İnternet: Internet Live Stats. Internet Live Stats. URL: http://www.internetlivestats.com/, Son Erişim Tarihi: 26.11.2022.

5. Dandannavar, P. S., Mangalwede, S. R., and Kulkarni, P. M. (21-22 Dec. 2018). *Social Media Text - A Source for Personality Prediction*. Paper presented at 2018 International Conference on Computational Techniques, Electronics and Mechanical Systems (CTEMS).

6. Prunckun, H. (2014). *Scientific methods of inquiry for intelligence analysis*. Rowman & Littlefield.

7. Keleştemur, A. (2015). *Siber İstihbarat*. Level Yayınevi.

8. Gomes De Assis, C. (2017). The new era of information as power and the field of Cyber Intelligence. *Urvio, Revista Latinoamericana de Estudios de Seguridad.* 20. 94-109.

9. Omand, D., Bartlett, J., and Miller, C. (2012). Introducing social media intelligence (SOCMINT). *Intelligence and National Security.* 27(6). 801-823.

10. Stottlemyre, S. A. (2015). HUMINT, OSINT, or something new? Defining crowdsourced intelligence. *International Journal of Intelligence and CounterIntelligence.* 28(3). 578-589.

11. Johnson, L. K. (2007). *Handbook of intelligence studies*. Routledge.

12. Aid, M. M. (2013). Measurement and Signature Intelligence.

13. Yeboah-Ofori, A., and Brimicombe, A. (2018). Cyber Intelligence & OSINT: Developing Mitigation Techniques Against Cybercrime Threats on Social Media A Systematic Review July 2017. *International Journal of Cyber-Security and Digital Forensics.* 7(1). 87-99.

14. Glassman, M., and Kang, M. J. (2012). Intelligence in the internet age: The emergence and evolution of Open Source Intelligence (OSINT). *Computers in Human Behavior.* 28(2). 673-682.

15. Williams, H. J., and Blum, I. (2018). Defining Second Generation Open Source Intelligence (OSINT) for the Defense Enterprise; RAND Corporation Santa Monica United States.

16. İnternet: CIA. The Intelligence Cycle. URL: https://www.cia.gov/kids-page/6-12th-grade/who-we-are-what-we-do/the-intelligence-cycle.html, Son Erişim Tarihi: 02.01.2019.

17. Khan, N., and Khan, M. N. A. (2018). A Review of Opinion Mining in Twitter Streams. *International Journal of Next-Generation Computing.* 9(1). 66-72.

18. Zhou, X. J., Xu, Y., Li, Y. F., Josang, A., and Cox, C. (2012). The state-of-the-art in personalized recommender systems for social networking. *Artificial Intelligence Review.* 37(2). 119-132.

19. Yin, H. Z., Cui, B., Chen, L., Hu, Z. T., and Zhang, C. Q. (2015). Modeling Location-Based User Rating Profiles for Personalized Recommendation. *Acm Transactions on Knowledge Discovery from Data.* 9(3).

20. Cai, Z., He, Z., Guan, X., and Li, Y. (2018). Collective Data-Sanitization for Preventing Sensitive Information Inference Attacks in Social Networks. *IEEE Transactions on Dependable and Secure Computing.* 15(4). 577-590.

21. Tang, J., Yao, L. M., Zhang, D., and Zhang, J. (2010). A Combination Approach to Web User Profiling. *Acm Transactions on Knowledge Discovery from Data.* 5(1).

22. Ikeda, K., Hattori, G., Ono, C., Asoh, H., and Higashino, T. (2013). Twitter user profiling based on text and community mining for market analysis. *Knowledge-Based Systems.* 51. 35-47.

23. Xie, H., Li, Q., Mao, X., Li, X., Cai, Y., and Rao, Y. (2014). Community-aware user profile enrichment in folksonomy. *Neural Networks.* 58. 111-121.

24. Xie, H., Li, X., Wang, T., Lau, R. Y. K., Wong, T.-L., Chen, L., Wang, F. L., and Li, Q. (2016). Incorporating sentiment into tag-based user profiles and resource profiles for personalized search in folksonomy. *Information Processing & Management.* 52(1). 61-72.

25. Pfeil, U., Arjan, R., and Zaphiris, P. (2009). Age differences in online social networking – A study of user profiles and the social capital divide among teenagers and older users in MySpace. *Computers in Human Behavior.* 25(3). 643-654.

26. Luo, F., Cao, G., Mulligan, K., and Li, X. (2016). Explore spatiotemporal and demographic characteristics of human mobility via Twitter: A case study of Chicago. *Applied Geography.* 70. 11-25.

27. Liang, X., Li, X., Zhang, K., Lu, R., Lin, X., and Shen, X. S. (2013). Fully Anonymous Profile Matching in Mobile Social Networks. *IEEE Journal on Selected Areas in Communications.* 31(9). 641-655.

28. Fang, Q., Sang, J., Xu, C., and Hossain, M. S. (2015). Relational User Attribute Inference in Social Media. *IEEE Transactions on Multimedia.* 17(7). 1031-1044.

29. Roussaki, I., Kalatzis, N., Liampotis, N., Kosmides, P., Anagnostou, M., Doolin, K., Jennings, E., Bouloudis, Y., and Xynogalas, S. (2012). Context-awareness in wireless and mobile computing revisited to embrace social networking. *IEEE Communications Magazine.* 50(6). 74-81.

30. Zheng, X., Cai, Z., Yu, J., Wang, C., and Li, Y. (2017). Follow But No Track: Privacy Preserved Profile Publishing in Cyber-Physical Social Systems. *IEEE Internet of Things Journal.* 4(6). 1868-1878.

31. Cassidy, A., Strube, M., and Nehorai, A. (2015). A Framework for Exploring Social Network and Personality-Based Predictors of Smart Grid Diffusion. *IEEE Transactions on Smart Grid.* 6(3). 1314-1322.

32. Samani, Z. R., Guntuku, S. C., Moghaddam, M. E., Preatiuc-Pietro, D., and Ungar, L. H. (2018). Cross-platform and cross-interaction study of user personality based on images on Twitter and Flickr. *Plos One.* 13(7).

33. Antoniou, A. (2017). Social network profiling for cultural heritage: combining data from direct and indirect approaches. *Social Network Analysis and Mining.* 7(1).

34. Souri, A., Hosseinpour, S., and Rahmani, A. M. (2018). Personality classification based on profiles of social networks' users and the five-factor model of personality. *Human-Centric Computing and Information Sciences.* 8.

35. Ortigosa, A., Carro, R. M., and Quiroga, J. I. (2014). Predicting user personality by mining social interactions in Facebook. *Journal of Computer and System Sciences.* 80(1). 57-71.

36. Song, H., and Moon, N. (2019). Eye-tracking and social behavior preference-based recommendation system. *Journal of Supercomputing.* 75(4). 1990-2006.

37. Correa, T., Hinsley, A. W., and De Zúñiga, H. G. (2010). Who interacts on the Web?: The intersection of users' personality and social media use. *Computers in Human Behavior.* 26(2). 247-253.

38. Hughes, D. J., Rowe, M., Batey, M., and Lee, A. (2012). A tale of two sites: Twitter vs. Facebook and the personality predictors of social media usage. *Computers in Human Behavior.* 28(2). 561-569.

39. Shen, J., Brdiczka, O., and Liu, J. (2015). A study of Facebook behavior: What does it tell about your Neuroticism and Extraversion? *Computers in Human Behavior.* 45. 32-38.

40. Winter, S., Neubaum, G., Eimler, S. C., Gordon, V., Theil, J., Herrmann, J., Meinert, J., and Krämer, N. C. (2014). Another brick in the Facebook wall – How personality traits relate to the content of status updates. *Computers in Human Behavior*. 34. 194-202.

41. Biel, J., and Gatica-Perez, D. (2013). The YouTube Lens: Crowdsourced Personality Impressions and Audiovisual Analysis of Vlogs. *IEEE Transactions on Multimedia.* 15(1). 41-55.

42. Ünal, P. (2015). *An Analysis on User Profiles and Usage Preferences for Mobile Application Recommendations*. Yayımlanmamış Doktora Tezi, Orta Doğu Teknik Üniversitesi Bilişim Sistemleri Ana Bilim Dalı.

43. Küçüközer, Ş. (2013). *Analyzing the effects of personality traits on the success of online study groups*. Yayımlanmamış Yüksek Lisans Tezi, Orta Doğu Teknik Üniversitesi Bilişim Sistemleri Ana Bilim Dalı.

44. Hafizoğlu, F. M. (2010). *The effect of trust and team personality on multiagent teamwork performance*. Yayımlanmamış Yüksek Lisans Tezi, Boğaziçi Üniversitesi Sistem ve Kontrol Mühendisliği Ana Bilim Dalı.

45. Alizadeh, L. (2019). *Assessing system usability according to different user personalities*. Yayımlanmamış Yüksek Lisans Tezi, İstanbul Aydın Üniversitesi İşletme Ana Bilim Dalı.

46. Demİray, Ç. K. (2015). *Change in usage of cognitive words, affect words, discourse markers, and first person pronoun position among clients in psychotherapy process*. Yayımlanmamış Doktora Tezi, Orta Doğu Teknik Üniversitesi Psikoloji Ana Bilim Dalı.

47. Kavum, M. F. (2021). *Player modeling by using unsupervised learning methods*. Yayımlanmamış Yüksek Lisans Tezi, İstanbul Teknik Üniversitesi Oyun ve Etkileşim Teknolojileri Ana Bilim Dalı.

48. Erel, Ö. (2021). *Şiddet içeren ve içermeyen suç davranışının risk ve koruyucu faktörler açısından incelenmesi*. Yayımlanmamış Doktora Tezi, İstanbul Üniversitesi-Cerrahpaşa Sosyal Bilimler Ana Bilim Dalı.

49. Aral, F. (1997). *Suçta etken olabilecek kişilik boyutlarının incelenmesi*. Yayımlanmamış Doktora Tezi, İstanbul Üniversitesi Sosyal Bilimler Ana Bilim Dalı.

50. Günay, A. (2019). *Sosyal medya kişilik özelliklerini yansıtır mı? İnsan kaynakları perspektifinde twitter üzerinden bir söylem analizi*. Yayımlanmamış Doktora Tezi, Süleyman Demirel Üniversitesi İşletme Ana Bilim Dalı.

51. Koçak, E. (2017). *The relationship between personality traits and behaviors of instagram users vs. non-users*. Yayımlanmamış Yüksek Lisans Tezi, Boğaziçi Üniversitesi Yönetim Bilişim Sistemleri Ana Bilim Dalı.

52. Akbaş, M. C. (2019). *The relationship between snapchat users' motivation, attitude and personality and the application's content and features*. Yayımlanmamış Yüksek Lisans Tezi, Boğaziçi Üniversitesi Yönetim Bilişim Sistemleri Ana Bilim Dalı.

53. Yaşar, N. N. (2015). *Türkiye'deki bireylerin kişilik özellikleri, facebook kullanımları ve boş zaman aktiveleri*. Yayımlanmamış Yüksek Lisans Tezi, İstanbul Teknik Üniversitesi Endüstri Mühendisliği Ana Bilim Dalı.

54. Aslan, S. (2019). *Multimodal video-based personality recognition using long short-term memory and convolutional neural networks*. Yayımlanmamış Yüksek Lisans Tezi, İhsan Doğramacı Bilkent Üniversitesi Bilgisayar Mühendisliği Ana Bilim Dalı.

55. Taş, E. (2019). *Prediction of personality traits from videos by using machine learning algorithms*. Yayımlanmamış Yüksek Lisans Tezi, İstanbul Teknik Üniversitesi Bilgisayar Mühendisliği Ana Bilim Dalı.

56. Mandira, B. (2021). *Personality-aware deception detection from behavioral cues*. Yayımlanmamış Yüksek Lisans Tezi, İhsan Doğramacı Bilkent Üniversitesi Bilgisayar Mühendisliği Ana Bilim Dalı.

57. Ateş, Ü. (2014). *Inference of personality using social media profiles*. Yayımlanmamış Yüksek Lisans Tezi, Orta Doğu Teknik Üniversitesi Bilişim Sistemleri Ana Bilim Dalı.

58. Kosinski, M., Stillwell, D., and Graepel, T. (2013). Private traits and attributes are predictable from digital records of human behavior. *Proceedings of the national academy of sciences.* 110(15). 5802-5805.

59. Ahmad, Z., Lutfi, S. L., Kushan, A. L., and Yixing, R. T. (2017). Personality Prediction of Malaysian Facebook Users: Cultural Preferences and Features Variation. *Advanced Science Letters.* 23(8). 7900-7903.

60. Laleh, A., and Shahram, R. (18-21 Dec. 2017). *Analyzing Facebook Activities for Personality Recognition*. Paper presented at 2017 16th IEEE International Conference on Machine Learning and Applications (ICMLA).

61. Tandera, T., Hendro, Suhartono, D., Wongso, R., and Prasetio, Y. L. (2017). Personality Prediction System from Facebook Users. *Procedia Computer Science.* 116. 604-611.

62. Vaidhya, M., Shrestha, B., Sainju, B., Khaniya, K., and Shakya, A. (15-18 Nov. 2017). *Personality Traits Analysis from Facebook Data*. Paper presented at 2017 21st International Computer Science and Engineering Conference (ICSEC).

63. Akhtar, R., Winsborough, D., Ort, U., Johnson, A., and Chamorro-Premuzic, T. (2018). Detecting the dark side of personality using social media status updates. *Personality and Individual Differences.* 132. 90-97.

64. Hassanein, M., Hussein, W., Rady, S., and Gharib, T. F. (18-19 Dec. 2018). *Predicting Personality Traits from Social Media using Text Semantics*. Paper presented at 2018 13th International Conference on Computer Engineering and Systems (ICCES).

65. Howlader, P., Pal, K. K., Cuzzocrea, A., Kumar, S. D. M., and Assoc Comp, M. (2018). *Predicting Facebook-Users' Personality based on Status and Linguistic Features via Flexible Regression Analysis Techniques*. New York, Assoc Computing Machinery.

66. Mao, Y., Zhang, D., Wu, C., Zheng, K., and Wang, X. (7-10 Dec. 2018). *Feature Analysis and Optimisation for Computational Personality Recognition*. Paper presented at 2018 IEEE 4th International Conference on Computer and Communications (ICCC).

67. Tadesse, M. M., Lin, H., Xu, B., and Yang, L. (2018). Personality Predictions Based on User Behavior on the Facebook Social Media Platform. *IEEE Access.* 6. 61959-61969.

68. Xue, D., Wu, L. F., Hong, Z., Guo, S. Z., Gao, L., Wu, Z. Y., Zhong, X. F., and Sun, J. S. (2018). Deep learning-based personality recognition from text posts of online social networks. *Applied Intelligence.* 48(11). 4232-4246.

69. Marouf, A. A., Hasan, M. K., and Mahmud, H. (7-9 Feb. 2019). *Identifying Neuroticism from User Generated Content of Social Media based on Psycholinguistic Cues*. Paper presented at 2019 International Conference on Electrical, Computer and Communication Engineering (ECCE).

70. Zheng, H. C., Wu, C. H., and Assoc Comp, M. (2019). *Predicting Personality Using Facebook Status Based on Semi-supervised Learning*. New York, Assoc Computing Machinery.

71. Al Marouf, A., Hasan, M. K., and Mahmud, H. (2020). Comparative Analysis of Feature Selection Algorithms for Computational Personality Prediction From Social Media. *Ieee Transactions on Computational Social Systems.* 7(3). 587-599.

72. Sun, J. S., Tian, Z. Q., Fu, Y. L., Geng, J., and Liu, C. L. (2020). Digital twins in human understanding: a deep learning-based method to recognize personality traits. *International Journal of Computer Integrated Manufacturing.* 14.

73. Wang, S., Cui, L., Liu, L., Lu, X., and Li, Q. *Personality Traits Prediction Based on Users' Digital Footprints in Social Networks via Attention RNN*. Paper presented at 2020 IEEE International Conference on Services Computing (SCC).

74. Zhao, J. H., Zeng, D. L., Xiao, Y. J., Che, L. P., and Wang, M. J. (2020). User personality prediction based on topic preference and sentiment analysis using LSTM model. *Pattern Recognition Letters.* 138. 397-402.

75. Başaran, S., and Ejimogu, O. H. (2021). A Neural Network Approach for Predicting Personality From Facebook Data. *SAGE Open.* 11(3). 21582440211032156.

76. Bakry, M. R., Nasr, M. M., and Alsheref, F. K. (2022). Personality Classification Model of Social Network Profiles based on their Activities and Contents. *International Journal of Advanced Computer Science and Applications.* 13(7). 16-21.

77. Kamalesh, M. D., and Bharathi, B. (2022). Personality prediction model for social media using machine learning Technique. *Computers & Electrical Engineering.* 100. 12.

78. Yang, B. (2022). Analysis Model of Personality and Psychological Characteristics of Network Users under High-Pressure Working Environment. *Security and Communication Networks.* 2022. 10.

79. Zhou, L. X., Zhang, Z. Y., Zhao, L. J., and Yang, P. L. (2022). Attention-based BiLSTM models for personality recognition from user-generated content. *Information Sciences.* 596. 460-471.

80. Ahmad, N., and Siddique, J. Personality Assessment using Twitter Tweets. *Procedia Computer Science*. Zannimerk, C., Frydman, C., Toro, C., Hicks, Y., Howlett, R. J. ve Jain, L. C.s: Zannimerk, C., Frydman, C., Toro, C., Hicks, Y., Howlett, R. J. ve Jain, L. C. Knowledge-Based and Intelligent Information & Engineering Systems, 1964-1973, Amsterdam, Elsevier Science Bv, 2017.

81. Bhatti, S. K., Muneer, A., Lali, M. I., Gull, M., and Din, S. M. U. (2017). *Personality Analysis of the USA Public Using Twitter Profile Pictures*. New York, Ieee.

82. Guntuku, S. C., Lin, W. S., Carpenter, J., Ng, W. K., Ungar, L. H., Preotiuc-Pietro, D., and Acm. (2017). *Studying Personality through the Content of Posted and Liked Images on Twitter*. New York, Assoc Computing Machinery.

83. Raje, M. S., and Singh, A. Personality Detection by Analysis of Twitter Profiles. *Advances in Intelligent Systems and Computing*. Abraham, A., Cherukuri, A. K., Madureira, A. M. ve Muda, A. K.s: Abraham, A., Cherukuri, A. K., Madureira, A. M. ve Muda, A. K. Proceedings of the Eighth International Conference on Soft Computing and Pattern Recognition, 667-675, Cham, Springer International Publishing Ag, 2018.

84. Jeremy, N. H., Prasetyo, C., and Suhartono, D. (2019). Identifying Personality Traits for Indonesian User from Twitter Dataset. *International Journal of Fuzzy Logic and Intelligent Systems.* 19(4). 283-289.

85. Tutaysalgir, E., Karagoz, P., and Toroslu, I. H. *Clustering based personality prediction on turkish tweets*. Paper presented at 2019 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM).

86. Karanatsiou, D., Sermpezis, P., Gruda, D., Kafetsios, K., Dimitriadis, I., and Vakali, A. (2022). My Tweets Bring All the Traits to the Yard: Predicting Personality and Relational Traits in Online Social Networks. *Acm Transactions on the Web.* 16(2). 26.

87. Matsumoto, K., Kishima, R., Tsuchiya, S., Hirobayashi, T., Yoshida, M., and Kita, K. (2022). Relationship Between Personality Patterns and Harmfulness: Analysis and Prediction Based on Sentence Embedding. *International Journal of Information Technology and Web Engineering.* 17(1). 24.

88. Rathi, S., Verma, J. P., Jain, R., Nayyar, A., and Thakur, N. (2022). Psychometric profiling of individuals using Twitter profiles: A psychological Natural Language Processing based approach. *Concurrency and Computation-Practice & Experience.* 34(19). 19.

89. Ferwerda, B., Tkalcic, M., and Acm. (2018). *Predicting Users' Personality from Instagram Pictures: Using Visual and/or Content Features?* New York, Assoc Computing Machinery.

90. Kim, Y., and Kim, J. H. (2018). Using computer vision techniques on Instagram to link users' personalities and genders to the features of their photos: An exploratory study. *Information Processing & Management.* 54(6). 1101-1114.

91. Huang, S. G., Zheng, J. H., Xue, D., and Zhao, N. Predicting Big-Five Personality for Micro-blog Based on Robust Multi-task Learning. *Communications in Computer and Information Science*. Zou, B., Li, M., Wang, H., Song, X., Xie, W. ve Lu, Z.s: Zou, B., Li, M., Wang, H., Song, X., Xie, W. ve Lu, Z. Data Science, Pt 1, 486-499, Berlin, Springer-Verlag Berlin, 2017.

92. Li, C., Wan, J., and Wang, B. (13-16 Oct. 2017). *Personality Prediction of Social Network Users*. Paper presented at 2017 16th International Symposium on Distributed Computing and Applications to Business, Engineering and Science (DCABES).

93. Lin, J., Mao, W., and Zeng, D. D. (2017). Personality-based refinement for sentiment classification in microblog. *Knowledge-Based Systems.* 132. 204-214.

94. Han, S. Q., Huang, H. L., and Tang, Y. Q. (2020). Knowledge of words: An interpretable approach for personality recognition from social media. *Knowledge-Based Systems.* 194. 20.

95. Wang, P., Yan, Y., Si, Y. D., Zhu, G. C., Zhan, X. P., Wang, J., and Pan, R. S. (2020). Classification of Proactive Personality: Text Mining Based on Weibo Text and Short-Answer Questions Text. *Ieee Access.* 8. 97370-97382.

96. Wang, P., Yan, M., Zhan, X., Tian, M., Si, Y., Sun, Y., Jiao, L., and Wu, X. (2021). Predicting Self-Reported Proactive Personality Classification With Weibo Text and Short Answer Text. *IEEE Access.* 9. 77203-77211.

97. Jiang, Y., Deng, S., Li, H., and Liu, Y. (2021). Predicting user personality with social interactions in Weibo. *Aslib Journal of Information Management.*

98. Yang, K., Yuan, H., and Lau, R. Y. K. (2022). PsyCredit: An interpretable deep learning-based credit assessment approach facilitated by psychometric natural language processing. *Expert Systems with Applications.* 198. 13.

99. Alsadhan, N., and Skillicorn, D. (18-21 Nov. 2017). *Estimating Personality from Social Media Posts*. Paper presented at 2017 IEEE International Conference on Data Mining Workshops (ICDMW).

100. Varshney, V., Varshney, A., Ahmad, T., and Khan, A. M. (21-22 Sept. 2017). *Recognising personality traits using social media*. Paper presented at 2017 IEEE International Conference on Power, Control, Signals and Instrumentation Engineering (ICPCSI).

101. Guan, Z., Wu, B., Wang, B., and Liu, H. *Personality2vec: Network Representation Learning for Personality*. Paper presented at 2020 IEEE Fifth International Conference on Data Science in Cyberspace (DSC).

102. Khan, A. S., Ahmad, H., Asghar, M. Z., Saddozai, F. K., Arir, A., and Khalid, H. A. (2020). Personality Classification from Online Text using Machine Learning Approach. *International Journal of Advanced Computer Science and Applications.* 11(3). 460-476.

103. Sun, X. G., Liu, B., Meng, Q., Cao, J. X., Luo, J. Z., and Yin, H. Z. (2020). Group-level personality detection based on text generated networks. *World Wide Web-Internet and Web Information Systems.* 23(3). 1887-1906.

104. Azucar, D., Marengo, D., and Settanni, M. (2018). Predicting the Big 5 personality traits from digital footprints on social media: A meta-analysis. *Personality and Individual Differences.* 124. 150-159.

105. İnternet: Webis. PAN Shared Tasks. URL: https://pan.webis.de/, Son Erişim Tarihi: 05.12.2020.

106. Francisco Rangel, Fabio Celli, Paolo Rosso, Martin Potthast, Benno Stein, and Daelemans, W. Overview of the 3rd Author Profiling Task at PAN 2015. Linda Cappellato, Nicola Ferro, Gareth Jones ve Juan, E. S.s: Linda Cappellato, Nicola Ferro, Gareth Jones ve Juan, E. S. CLEF 2015 Evaluation Labs and Workshop -- Working Notes Papers, Toulouse, France, CEUR-WS.org, 2015.

107. Xu, J., Tian, W., Lv, G., Liu, S., and Fan, Y. (2021). 2.5 D Facial Personality Prediction Based on Deep Learning. *Journal of Advanced Transportation.* 2021.

108. Dutta, K., Singh, V. K., Chakraborty, P., Sidhardhan, S. K., Krishna, B. S., and Dash, C. (2017). Analyzing Big-Five personality traits of Indian celebrities using online social media. *Psychological Studies.* 62(2). 113-124.

109. Gain, U., Koponen, M., and Hotti, V. *Behavioral Interventions from Trait Insights*. Paper presented at International Conference on Well-Being in the Information Society.

110. Asgha, B., Hartono, D. A., Halim, M. S., Wijaya, K., Ngan, E., and Surya, C. M. (2020). An Analysis of Personality Traits of Indonesian Start-Up Entrepreneurs Based on Social Media Footprint. *Indonesian Journal of Business and Entrepreneurship (IJBE).* 6(2). 127-127.

111. İnternet: Okrent, A. Mental Floss. 13 Words That Changed From Negative to Positive Meanings (or Vice Versa). URL: https://www.mentalfloss.com/article/65987/13-words-changed-negative-positive-or-vice-versa, Son Erişim Tarihi: 30.11.2020.

112. Bojanowski, P., Grave, E., Joulin, A., and Mikolov, T. (2017). Enriching word vectors with subword information. *Transactions of the Association for Computational Linguistics.* 5. 135-146.

113. Gers, F. A., Schmidhuber, J., and Cummins, F. (2000). Learning to forget: Continual prediction with LSTM. *Neural computation.* 12(10). 2451-2471.

114. Malki, Z., Atlam, E., Dagnew, G., Alzighaibi, A. R., Ghada, E., and Gad, I. (2020). Bidirectional Residual LSTM-based Human Activity Recognition. *Computer and Information Science.* 13(3). 1-40.

115. Ramachandran, P., Zoph, B., and Le, Q. V. (2017). Searching for activation functions. *arXiv preprint arXiv:1710.05941*.

116. Kingma, D. P., and Ba, J. (2014). Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980.*

117. Graves, A., and Schmidhuber, J. (2005). Framewise phoneme classification with bidirectional LSTM and other neural network architectures. *Neural networks.* 18(5-6). 602-610.

118. İnternet: Strickland, E. IEEE Spectrum. ANDREW NG: UNBIGGEN AI. URL: https://spectrum.ieee.org/andrew-ng-data-centric-ai, Son Erişim Tarihi: 10.02.2022.

119. Rammstedt, B., and John, O. P. (2007). Measuring personality in one minute or less: A 10-item short version of the Big Five Inventory in English and German. *Journal of research in Personality.* 41(1). 203-212.

120. Horzum, M. B., Tuncay, A., and Padir, M. A. (2017). Adaptation of big five personality traits scale to Turkish culture. *Sakarya University Journal of Education.* 7(2). 398-408.

121. Gosling, S. D., Rentfrow, P. J., and Swann Jr, W. B. (2003). A very brief measure of the Big-Five personality domains. *Journal of Research in personality.* 37(6). 504-528.

122. Atak, H. (2013). On-Maddeli Kişilik Ölçeği'nin Türk Kültürü'neUyarlanması.

123. Donnellan, M. B., Oswald, F. L., Baird, B. M., and Lucas, R. E. (2006). The mini-IPIP scales: tiny-yet-effective measures of the Big Five factors of personality. *Psychological assessment.* 18(2). 192.

124. Korkmaz, M., Somer, O., and Güngör, D. (2013). Ergen örneklemde beş faktör kişilik envanteri'nin cinsiyetlere göre ortalama ve kovaryans yapılarıyla ölçme eşdeğerliği. *Eğitim ve Bilim.* 38(170).

125. Soto, C. J., and John, O. P. (2017). The next Big Five Inventory (BFI-2): Developing and assessing a hierarchical model with 15 facets to enhance bandwidth, fidelity, and predictive power. *Journal of personality and social psychology.* 113(1). 117.

126. İnternet: Soto, C. Colby College - Personality Lab. The Big Five Inventory–2 (BFI-2). URL: https://www.colby.edu/psych/personality-lab/#4, Son Erişim Tarihi: 17.02.2020.

127. İnternet: The 3,320 IPIP Items in Alphabetical Order. URL: https://ipip.ori.org/AlphabeticalItemList.htm, Son Erişim Tarihi: 05.01.2020.

128. İnternet: Somer, O. Turkish Translation of 924 IPIP Items. URL: https://ipip.ori.org/Turkish924Items.htm, Son Erişim Tarihi: 05.01.2020.

129. İnternet: Soto, C. Colby College - Personality Lab. The Big Five Inventory–2 Short Form (BFI-2-S). URL: https://www.colby.edu/psych/wp-content/uploads/sites/50/2013/08/bfi2s-form.pdf, Son Erişim Tarihi: 17.02.2020.

130. İnternet: Soto, C. Colby College - Personality Lab. The Big Five Inventory–2 Extra-Short Form (BFI-2-XS). URL: https://www.colby.edu/psych/wp-content/uploads/sites/50/2013/08/bfi2xs-form.pdf, Son Erişim Tarihi: 17.02.2020.

131. İnternet: Cemalcılar, Z., Sümer, H. C., Sümer, N., & Baruh, L. Colby College - Personality Lab. Kişilik Ölçeği (Big Five Inventory-2). URL: https://www.colby.edu/psych/wp-content/uploads/sites/50/2013/08/bfi2-form-turkish.pdf, Son Erişim Tarihi: 17.02.2020.

132. İnternet: Schweter, S. Zenodo. BERTurk - BERT models for Turkish. URL: https://doi.org/10.5281/zenodo.3770924, Son Erişim Tarihi: 12.01.2022.

133. Clark, K., Luong, M.-T., Le, Q. V., and Manning, C. D. (2020). Electra: Pre-training text encoders as discriminators rather than generators. *arXiv preprint arXiv:2003.10555.*

134. Grave, E., Bojanowski, P., Gupta, P., Joulin, A., and Mikolov, T. (2018). Learning word vectors for 157 languages. *arXiv preprint arXiv:1802.06893.*

135. Mikolov, T., Chen, K., Corrado, G., and Dean, J. (2013). Efficient estimation of word representations in vector space. *arXiv preprint arXiv:1301.3781.*

136. Lau, J. H., and Baldwin, T. (2016). An empirical evaluation of doc2vec with practical insights into document embedding generation. *arXiv preprint arXiv:1607.05368.*

137. Schuster, M., and Paliwal, K. K. (1997). Bidirectional recurrent neural networks. *IEEE transactions on Signal Processing.* 45(11). 2673-2681.

138. Kumar, J. A., and Abirami, S. (2021). Ensemble application of bidirectional LSTM and GRU for aspect category detection with imbalanced data. *Neural Computing and Applications.* 33(21). 14603-14621.

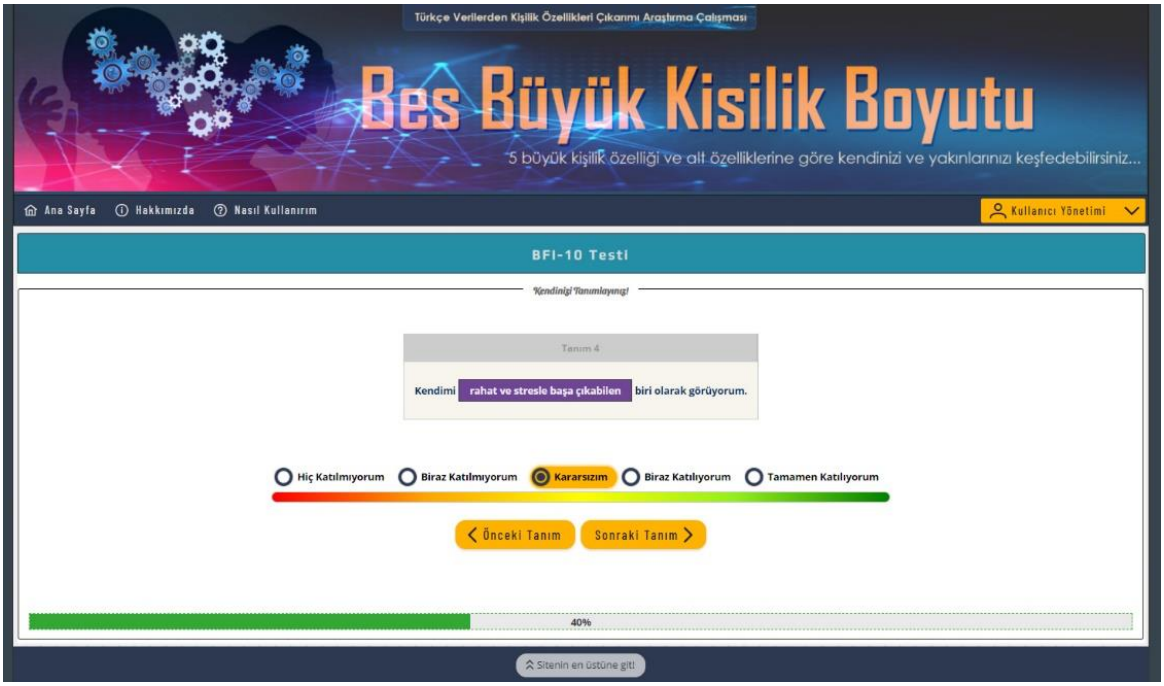
139. Kiranyaz, S., Avci, O., Abdeljaber, O., Ince, T., Gabbouj, M., and Inman, D. J. (2021). 1D convolutional neural networks and applications: A survey. *Mechanical systems and signal processing.* 151. 107398.

EKLER

EK-1. Veri toplama web uygulama ekran görüntüleri 1



Şekil 1.1. Veri toplama web uygulaması – ana ekran



Şekil 1.2. Veri toplama web uygulaması – test ekranı

EK-2. Veri toplama web uygulama ekran görüntüleri 2

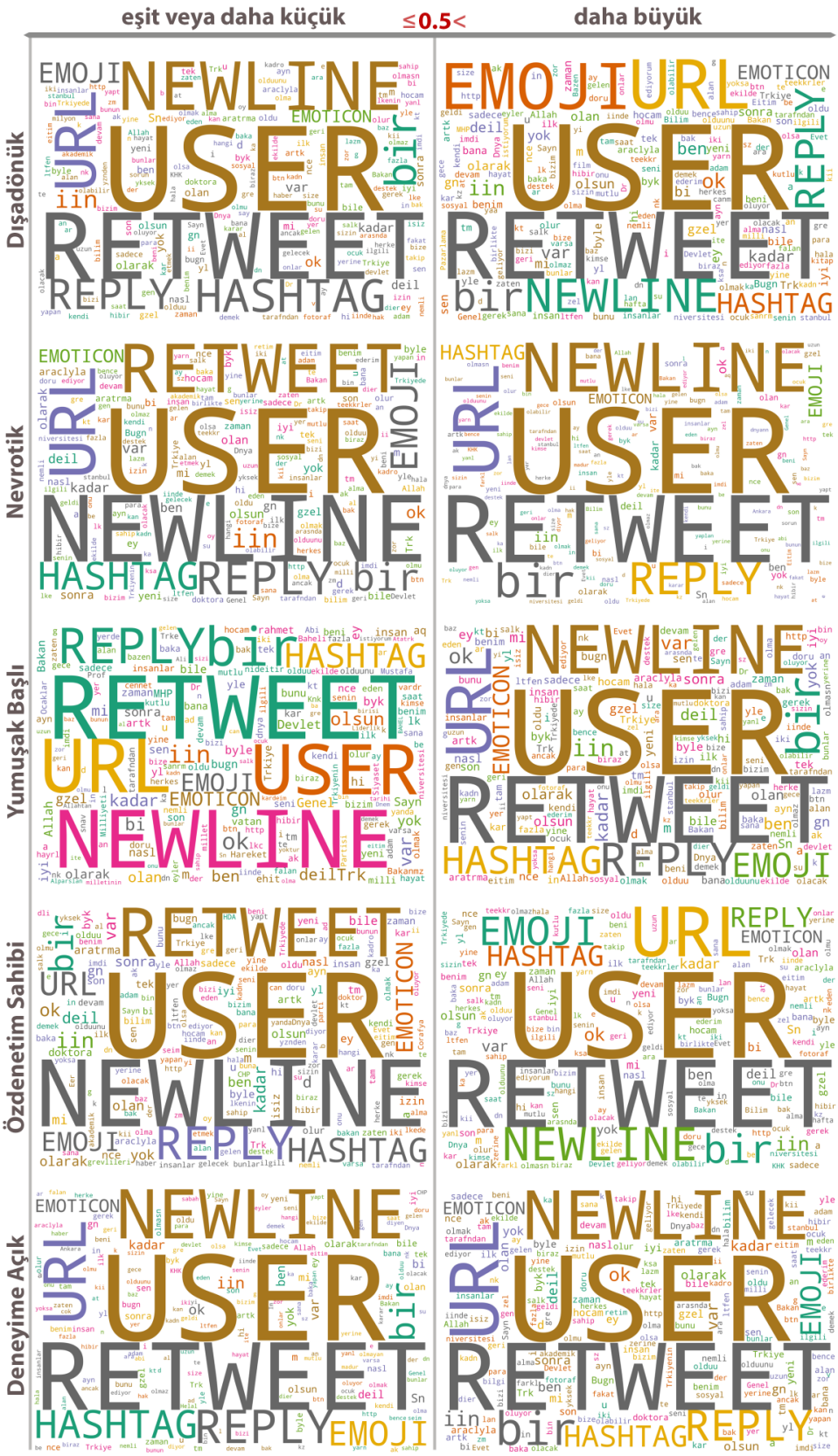


Şekil 2.1. Veri toplama web uygulaması – bfi-10 test sonuç ekranı



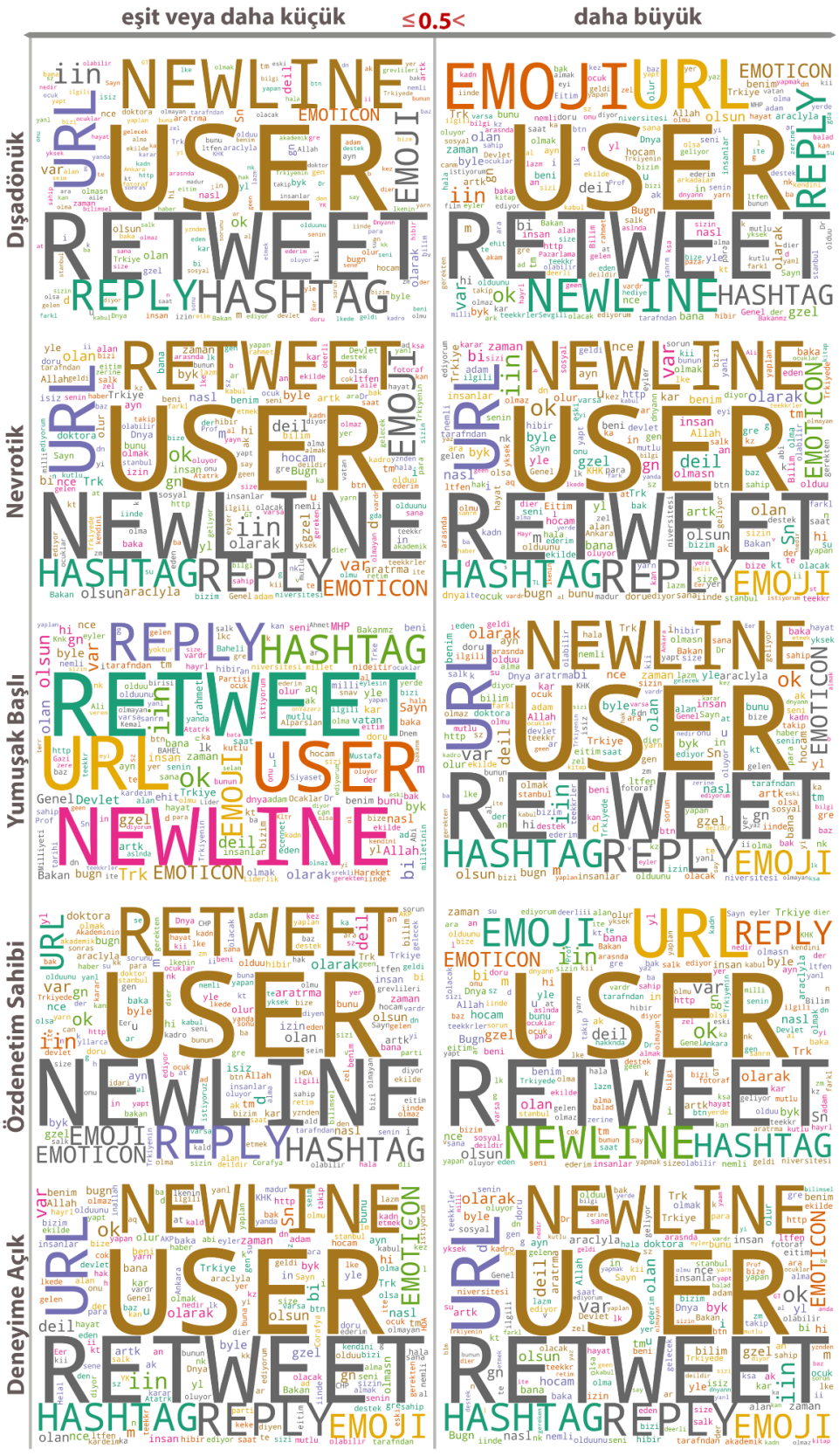
Şekil 2.2. Veri toplama web uygulaması – bfi-2 test sonuç ekranı

EK-3. Kişilik özelliklerine göre kelime/varlık yoğunluk haritaları - NLTK



Şekil 3.1. NLTK Stop Words ile oluşturulan kişilik özelliklerine göre kelime/varlık yoğunluk haritaları

EK-4. Kişilik özelliklerine göre kelime/varlık yoğunluk haritaları - Zemberek



Şekil 4.1. Zemberek Stop Words ile oluşturulan kişilik özelliklerine göre kelime/varlık yoğunluk haritaları

ÖZGEÇMİŞ

Kişisel Bilgiler

Soyadı, adı : KOŞAN, Muhammed Ali

Uyruğu : T.C.

e-mail : ceo.muhammed@gmail.com

Eğitim

Derece	Eğitim Birimi	Mezuniyet Tarihi
Doktora	Gazi Üniversitesi / Bilgisayar Bilimleri	Devam ediyor
Yüksek lisans	Gazi Üniversitesi / Bilgisayar Bilimleri	2017
Lisans	İstanbul Üniversitesi / Bilgisayar Mühendisliği	2011

İş Deneyimi

Yıl	Yer	Görev
2016-Devam ediyor	Gazi Üniversitesi	Araştırma Görevlisi
2014-Devam ediyor	Muş Alparslan Üniversitesi	Araştırma Görevlisi
2012-2013	Yonca CBS Bilişim	Proje Lideri
2011-2012	Nodeser Bilişim	Proje Lideri
2010-2011	Nodeser Bilişim	Geliştirici

Yabancı Dil

İngilizce

Yayınlar

1. Koşan, M. A., Kocatepe, A. İ., and Karacan, H. (2016, July). *A Model of Steganographic Report for Web Application Security*. Paper presented at the Global Conference on Applied Computing in Science & Engineering (ACSE), Rome.

2. Kocatepe, A. İ., Koşan, M. A., and Karacan, H. (2017, March). *An Adaptation of Matrix Encoding Technique for Data Hiding in RGB Images*. Paper presented at the International Conference on Engineering Technology and Innovation (ICETI), Sarajevo.

3. Koşan, M. A., and Karacan, H. (2017, March). *Comparative Analysis of Web Application Security Scanners*. Paper presented at the International Conference on Engineering Technology and Innovation (ICETI), Sarajevo.

4. Kocatepe, A. İ., Koşan, M. A., and Karacan, H. (2017, November). *An Altered Application of Permutative Straddling Method for Spatial Steganography*. Paper presented at the International Advanced Research and Engineering Congress (IAREC), Osmaniye.

5. Sevri, M., Koşan, M. A., and Karacan, H. (2017, November). *Applying Deep Learning to Cybersecurity Field and Its Advantages*. Paper presented at the International Advanced Researches and Engineering Congress (IAREC), Osmaniye.

6. Koşan, M. A., Kocatepe, A. İ., and Karacan, H. (2017, November). *Mobile Malware Datasets and Malware Detection*. Paper presented at the International Advanced Research and Engineering Congress (IAREC), Osmaniye.

7. Koşan, M. A., Sevri, M., and Karacan, H. (2017, November). *Spy Attacks and Protection Methods in Android Smart Phones*. Paper presented at the International Advanced Research and Engineering Congress (IAREC), Osmaniye.

8. Koşan, M. A., Yıldız, O., and Karacan, H. (2018). Kimlik Avı Web Sitelerinin Tespitinde Makine Öğrenmesi Algoritmalarının Karşılaştırmalı Analizi. *Pamukkale Üniversitesi Mühendislik Bilimleri Dergisi (PAJES)*, 24(2), 276-282. DOI: 10.5505/pajes.2017.27167

9. Koşan, M. A., Sevri, M., and Karacan, H. (2018, 27-29 June). *Comparison of Classification Algorithms used in Intrusion Detection Systems*. Paper presented at the International Conference on Applied Mathematics in Engineering (ICAME), Balıkesir.

10. Sevri, M., Koşan, M. A., and Karacan, H. (2018, 27-29 June). *Secure Multiparty Computation and Blockchain-Smart Contract Based Election System*. Paper presented at the International Conference on Applied Mathematics in Engineering (ICAME), Balıkesir.

11. Koşan, M. A., and Benzer, R. (2018, 24-26 October). *Siber Güvenlik Alanında Derin Öğrenme Yöntemlerinin Kullanımı*. Paper presented at the 5th International Management Information Systems Conference (IMISC), BTK, Ankara. DOI: 10.6084/m9.figshare.7581983.v1

12. Koşan, M. A., Coşkun, A., and Karacan, H. (2019). Yapay Zekâ Yöntemlerinde Entropi. *Journal of Information Systems and Management Research*. 1(1), 15-22. Retrieved from https://dergipark.org.tr/tr/pub/jismar/issue/51306/639389

13. Koşan, M. A., Karacan, H., and Ürgen, B. A. (2022). Predicting personality traits with semantic structures and LSTM-based neural networks. *Alexandria Engineering Journal (AEJ)*. 61(10), 8007-8025. DOI: 10.1016/j.aej.2022.01.050



GAZİLİ OLMAK AYRICALIKTIR.