



ONDOKUZ MAYIS ÜNİVERSİTESİ
MÜHENDİSLİK FAKÜLTESİ
BİLGİSAYAR MÜHENDİSLİĞİ BÖLÜMÜ

**KULLANICI VERİLERİ İLE EĞİTİLEN CHAT BOTUNA,
UYGUNSUZ İÇERİK TESPİTİ VE KULLANICI DUYGU ANALİZİNE
SAHİP MEDYA UYGULAMASI**

PROJE RAPORU

DOĞAN YILMAZ

FURKAN SEMİZ

SABRİ BAŞOĞLU

Danışman

DR. ÖĞRETİM ÜYESİ İSMAİL İŞERİ

HAZİRAN, 2021

TEŞEKKÜR

Engin bilgi ve deneyimleriyle projemizde bize yol gösteren Dr. Öğretim Üyesi İsmail İşeri'ye içtenlikle teşekkürlerimizi sunarız.

Ö Z E T

Bu çalışmada günümüz sosyal medya platformlarının temel konseptleri incelenmiş olup bu özelliklere yapay zeka güvenlik modelleri entegre edilerek yeni bir sosyal medya uygulaması geliştirilmeye çalışılmıştır. Bu süreçte, uygunsuz medya dosyalarının sınıflandırılması için test edilmiş olan model %98.2 başarı oranı göstermiş olup mobil cihazlar üzerinde derin öğrenme uyumlu grafik kartlarına sahip sunuculara ihtiyaç duymadan sınıflandırma yapabilmektedir. Geliştirilmiş olan bir diğer model olan Türkçe Duygu Analizi ve Argo Sınıflandırıcısı argo kelime sınıflandırmada %20 test verisi için MobileBert sınıflandırma algoritması ile %99.1 başarı oranı, Average Word Vector algoritması ile %92.6 başarı oranı elde etmiştir. Aynı veri seti üzerinden geliştirilmiş olan duygu analizi modeli ise MobileBert sınıflandırma algoritması ile %94.4 başarı oranı, Average Word Vector algoritması ile %71.3 başarı oranı elde etmiştir. Bu iki modülün ürettiği sonuçlarla birlikte kullanıcının aldığı şikayetler ve cezalara göre ceza kararı vermesi için geliştirilen Yönetim Karar Modülü ise 7 farklı algoritma ve 5.000 test verisi ile test edilerek Random Forest algoritması ve Karar Ağacı algoritması ile %100 başarı sunarken diğer algoritmalarda yanlış pozitif sonuçlar oluştuğu ve kullanıcıların haksız cezalar almasına neden olabileceği için kullanıma uygun bulunmamıştır. Ayrıca geliştirilen bu uygulamada doğal dil işleme yöntemleri ile konu tespiti yapabilen bir chat botu bulunmaktadır.

Anahtar Kelimeler: chat bot, duygu analizi, uygunsuz resim sınıflandırma, sosyal medya

İÇİNDEKİLER

I GİRİŞ

1	AMAÇ	2
2	LİTERATÜR ÖZETİ	3
2.1	Sosyal Medya İle İlgili Psikolojik Ve Sosyolojik Araştırmalar	3
2.1.1	Fomo Etkisi	3
2.1.2	Akış Teorisi	4
2.2	Görüntü İşleme Ve Sınıflandırma Üzerine Araştırmalar	5
2.3	Duygu Analizi Üzerine Araştırmalar	6
2.4	Chat Botları	7

II MATERYAL

3	KULLANILAN TEKNOLOJİLER	10
3.1	Kullanılan Diller	10
3.1.1	Dart	10
3.1.2	Python	11
3.2	Platformlar	11
3.2.1	Flutter	11
3.3	Firebase Çözümleri	12
3.3.1	Firebase Firestore	12
3.3.2	Firebase Bulut Mesajlaşma Servisi	12

3.3.3	Firestore Depolama	13
3.3.4	Firestore Fonksiyonlar	13
3.3.5	Firestore Kimlik Doğrulama Servisi	14
3.3.6	Firestore Makine Öğrenmesi	14
3.4	Yapay Zeka Araçları	15
3.4.1	Keras	15
3.4.2	TensorFlow	15
3.4.3	Google A.I Platform	15
3.5	Algoritmalar	16
3.5.1	Evrimsel Sinir Ağları	16
3.5.2	Karar Ağaçları	16
3.5.3	Rassal Ormanlar	17
3.6	Dil İşleme Araçları	17
3.6.1	Dialogflow	17
 III YAPAY ZEKA		
4	CEZA KARAR MODÜLÜ	20
4.1	Hedef	20
4.2	Hazırlık	21
4.3	Eğitim	22
4.3.1	Orange Üzerinden Eğitim	22
4.3.2	Python Üzerinden Eğitim	24
4.4	Sonuç	26
5	DUYGU ANALIZİ MODÜLÜ	27
5.1	Hedef	27

5.2	Hazırlık	27
5.3	Eğitim	28
5.4	Sonuç	29
6	UYGUNSUZ MEDYA SINIFLANDIRMA MODÜLÜ	30
6.1	Hedef	30
6.2	Hazırlık	30
6.3	Test	31
6.4	Sonuç	31
 IV UYGULAMA		
7	GENEL MİMARİ TASARIM VE YÖNTEM	33
7.1	Genel Kullanım	33
8	ARAYÜZLER	34
8.1	Giriş Ekranı	34
8.2	Kayıt Ekranı	34
8.3	Üye Giriş Ekranı	34
8.4	Profil Oluşturma Ekranı	35
8.5	Ana Sayfa	36
8.6	Mesajlar	37
8.7	Gruplar	38
8.8	Profil	39
8.9	Gönderi Ekranı	40
8.10	Gönderi Paylaşımı	41
9	YAPAY ZEKA ÖZELLİKLERİNİN ENTEGRASYONU	49
9.1	Mesaj Yakalama	49

9.2	Paylaşım Yakalama	49
9.3	Şikayet Sınıflandırma	50
9.4	Chat Bot	51
10	GELİŞTİRİLMİŞ YAN ÜRÜNLER	56
10.1	Fire_autho Kütüphanesi	56
10.2	BERT Modellerinin Flutter Platformu Üzerinde Çalıştırılması	56
V	SONUÇ	
11	SONUÇ	58
	KAYNAKÇA	59

ŞEKİLLER LİSTESİ

1	Firestore Fonksiyonlar Servisinin Birlikte Çalışabildiği Sistemler [1]	14
2	Evrişimsel Sinir Ağları İşleyiş Şeması	16
3	Karar Ağacı Örneği [2]	18
4	Normal Dağılıma Göre Rastgele Sayı Üretici	21
5	Orange Üzerinde Tasarlanmış Eğitim Şeması	22
6	Orange Test Sonuçları	23
7	Random Forest Karışıklık Matrisi Sonucu	24
8	Naiv Bayes Karışıklık Matrisi Sonucu	24
9	Mozaik Grafiği	25
10	Python Eğitim Sonuçları	25
11	Duygu Analizi AWV Sonucu	28
12	Duygu Analizi BERT Sonucu	28
13	Argo Analizi AWV Sonucu	29
14	Argo Analizi BERT Sonucu	29
15	BERT Model Yapısı	29
16	Giriş Ekranı	35
17	Kayıt Ekranı	36
18	Profil Oluşturma Ekranı	37
19	Ana Sayfa	38

20	Mesajlar Ekranı	39
21	Gruplar Ekranı	40
22	Grup Sohbet Ekranı	41
23	Profil Bilgi Ekranı	42
24	Profil Galerisi Ekranı	43
25	Profil Gönderi Ekranı	44
26	Gönderi Ekranı	45
27	Gönderi Menüsü	46
28	Galeri Gönderi Ekranı	47
29	Durum Gönderi Ekranı	48
30	Uygunsuz Mesaj Örneği	50
31	Uygunsuz Paylaşım Denemesi Sonucu	51
32	Uygun İçerik Şikayet Ekranı	52
33	Uygunsuz İçerik Şikayet Ekranı	53
34	Bot Eğitim Ekranı	54
35	Bot Sohbet Ekranı	55

BÖLÜM: I

GİRİŞ

AMAÇ

Projemizin amacı sosyal medya kullanımı hakkında yapılan arařtırmaların ışığında güncel sosyal medya uygulamalarının özelliklerini içinde barındıran yapay zeka özelliklerine sahip bir sosyal medya uygulaması geliřtirmektir. Bu özellikler metin içeriğinden duygu analizi, görüntü işleme ile kullanıcı görsellerinin taranması ve uygunsuz içeriklerin yakalanması, kullanıcı verileri ile eğitilebilen bir chat bot tasarlanmasıdır. Bu özellikler sayesinde siber zorbalığın önüne geçilmesi, daha temiz ve güvenli bir sosyal mecra yaratılması amaçlanmıştır.

LİTERATÜR ÖZETİ

Yapılan pazar araştırmasında başarılı sosyal medyaların yazılım ve psikoloji gibi birden çok önemli temeli olduğu gözlemlenmiştir. Bu sebeple sistem tasarımı ve kullanıcı deneyimi planmasından öncelikle bu psikolojik temeller hakkında literatür taraması yapılmıştır.

2.1 *Sosyal Medya İle İlgili Psikolojik Ve Sosyolojik Araştırmalar*

2.1.1 *Fomo Etkisi*

Fomo kavramı gelişmeleri kaçırma korkusu anlamına gelen Fear Of Missing Out teriminin kısaltılmasıdır. İlk olarak 1996 yılında Dr. Dan Herman tarafından ortaya atılmıştır [3]. Kaçırma korkusu, başkalarının daha iyi hayatlar yaşadığı ya da sizden daha çok eğlendiği veya daha iyi şeyler deneyimlediği hissidir. Derin bir kıskançlık duygusu içerir ve kişinin benlik saygısını etkilemektedir. Genellikle Instagram, SnapChat ve Facebook gibi sosyal medya uygulamaları bu hedonistik korkuyu şiddetlendirmektedir [4, 5]. Sosyal medya uygulamaları bu korkuyu kullanıcıları elde tutmak ve kullanım sürelerini artırmak için kullanır.

Pensilvanya Üniversitesi'nde yapılan araştırmada bir haftalık temel izleme süreci sonrasında 143 lisans öğrencisi Facebook, Instagram ve Snapchat kullanımını platform başına günde 10 dakika ile sınırlandırmak ya da sosyal medya uygulamalarını üç hafta boyunca normalde kullandıkları gibi kullanmak üzere rastgele atandı. Sınırlı sosyal medya kullanımına sahip grubunun üç haftalık süreçte kontrol grubuna göre daha az yalnızlık hissettiği ve depresyonlarında düşüş olduğu gözlemlendi. Her iki grupta da anksiyete ve fomo önemli düşüşler gösterdi. Araştırmanın sonucunda sosyal medya kullanımının günde yaklaşık 30 dakika ile sınırlandırılmasının yaşam kalitelerinde önemli bir iyileşmeye yol açabileceği belirtilmiştir [6].

2.1.2 Akış Teorisi

Akış(Flow) teorisi kişinin tamamen bir noktada toplanmış motivasyon halini ifade eder. Bu hal kişinin kendini o işe vermesi olarak da tanımlanabilir. Akış teorisi bu hali etkileyen faktörleri açıklamaya çalışan bir teoridir. Akış tanımı Mihaly Csikszentmihályi tarafından akışa kapılmaya benzetilerek ortaya atılan bir tanımdır [7].

Akış Unsurları

- Kişinin o an tam bir konsantrasyon durumuna sahip olması.
- Kişinin öz farkındalığını(self-consciousness) kaybetmiş olması.
- Kişinin iş üzerinde bir kontrole sahip olduğunu hissetmesi.
- Kişinin zaman algısının değişmesi.
- Kişinin işi yaparak ödülleniş hissetmesi ve yararlı bir deneyim kazandığını düşünmesi.

Akış hali eğitim hayatında öğrencilerin ders esnasında erişimleri istenen bir tam katılım halidir. Spor dallarında ve oyunlarda da bu hali gözlemleriz. Oyun oynamamızın temel sebebi bu tam katılım ve odak halidir [8].

Akış teorisinin sosyal medya uygulamaları üzerine uygulanabilirliği ile ilgili yapılan bir çalışmada 804 ergen Facebook kullanıcısı deney grubu olarak alındı. Bu araştırma sonucunda Facebook üzerinde akışı sağlayan unsurlar becerilerini gösterme, uygulama ile olan etkileşim, sosyal etkileşim, konsantrasyon, keyif alma ve gülünç içerikler olarak tespit edildi [9].

2.2 Görüntü İşleme Ve Sınıflandırma Üzerine Araştırmalar

Geçtiğimiz yıllarda yapay zeka ve makine öğrenme araştırmaları giderek hız kazanmıştır. Bu atılım ile birlikte görüntü işleme ve sınıflandırma araçları da giderek gelişmektedir. Bu gelişmeler sayesinde kullanıcıların paylaşmakta olduğu görsellerin içerikleri sınıflandırılabilir hale gelmiştir. Towards Data Science'ın yaptığı bir çalışmada 564 adet eğitim verisi, 1.056 adet test verisi ile beş farklı algoritma kıyaslandı. 421 uygunsuz resim içerisinde 399 adedini doğru tahmin ederek 92.80% doğruluk oranına ulaşmıştır. Bu çalışmada kullanılan method ten rengi piksellerini ayırıştırarak çalışmaktadır [10].

Chittagong Üniversitesinde yapılan başka bir çalışmada ise Destek Vektör Makinesi(SVM) algoritması kullanılarak uygunsuz içerikler sınıflandırılmaya çalışılmıştır. Bu araştırmanın kritik noktası yüz bölgesini algılayarak ten rengine sahip alanların yüz alanına oranlanarak sınıflandırılmanın yapılmasıdır.%7 yanlış pozitif sınıflandırma ile %94.89 doğruluk oranına sahip bir model elde edilmiştir [11].

Bu konu üzerinde yapılan birçok çalışmanın yanında bulut çözümleri de mevcuttur. Bu çözümler ölçeklendirme ve hız için optimize edilmiş, yüksek doğruluk oranına sahip

hizmetlerdir. Bu hizmetlerin başında Google Vision gelmektedir [12]. Bu hizmet ile HTTP istekleri ile görsellerin sınıflandırmasını yapılabilmektedir. Güvenli Arama Algılama, bir görsel içindeki yetişkin içeriği veya şiddet barındıran uygunsuz içerikleri tespit eder. Bu özellikte beş kategori mevcuttur (yetişkinler, sahtekarlık, tıbbi, şiddet ve müstehcenlik) ve her birinin belirli bir görüntüde bulunma olasılığını sayı olarak döndürür [13].

2.3 Duygu Analizi Üzerine Araştırmalar

Duygu analizi bir metin içerisindeki hakim duyguyu tespit etmek için belirli matematiksel yöntemler kullanılarak yapılan analiz sürecidir. Duygu analizi son on beş yılda, sosyal medya uygulamalarının etkisiyle oluşan duygu içerikli veri yığınlarının oluşması ile popüler hale gelmiştir. Duygu analizleri özellikle sosyal ve psikolojik alanlarda çıkarım yapmakta ve geniş kitlelerin fikirlerini analiz etmekte çok faydalı bir araçtır.

Duygu analizi sürecinde öncelik olarak hedeflenen veri kümesinin normalize edilmesi gerekir. Bu sayede metinler sınıflandırılabilir veriye dönüşür. Bu normalizasyon işlemi her dil için farklı şekilde yapılır. Eklerin ayrılması, kelimelerin köklerinin bulunması, gereksiz karakterlerin temizlenmesi gibi işlemler doğal dil işleme araçları sayesinde yapılır. Türkçe dili için bu araçlar kümesini sağlayan Zemberek adında bir kütüphane mevcuttur [14].

Twitter uygulaması üzerinde yapılan bir çalışmada Stanford Twitter Sentiment Dataset kullanılarak Destek Vektör Makinesi sınıflandırıcısının duygu analizinde en iyi performansı gösterdiği, MNB sınıflandırıcısı bunu takip etmekte olduğu ve diğer sınıflandırıcılara göre özellikle daha büyük veri kümelerinde çok hızlı gerçekleştirildiği görülmüştür [15].

Bir diğ er arařtırmada ise T rk e tweetlerin Destek Vekt r Makinesi ve Rastgele Orman (Random Forest) algoritmaları kullanılarak iřlenmesi sonucu %55 oranında dođru duygu tahmini ger ekleřtirilmiřtir. Bu arařtırmada toplam 867 veri kullanılmıřtır [16].

2.4 Chat Botları

Chat botları dođal dilleri kullanarak kullanıcılarla etkileřime ge meyi hedefleyen bilgisayar programlarıdır. Chat botları uzun yıllardır hayatımıza girmeye  alıřan bir ara tır. Ge tiđimiz yıllarda yapay zeka ara larının ve dođal dil iřleme tekniklerinin hızla pop lerleřmesiyle tekrar g ndeme gelmiřtir. G ndelik hayatta karřımıza  ıkan chat botları tam olarak istenen sonu ları  retememektedir. Bu durum chat botlarının iře yarayıp yaramadıđını sorgulatır durumdadır. Bu konuyu ele alan bir arařtırmada A.L.I.C.E. adlı alt yapı kullanılarak eđitim, bilgi edinme, iř, e-ticaret ve eđlence gibi pratik alanlarda bařarılı kabul edilen birkaç chatbot sistemi arařtırılmıřtır [17]. Bu arařtırmanın sonucunda chat botlarının eđitim, e-ticaret ve pratik alanlarda yarar sađladıđı fakat insan-insan etkileřimine yaklařmaktan  ok uzak olduđu belirtilmiřtir [18].

Eski chat botları ile g n m zdeki chat botlarının farkı, kullanılan y ntemlerdir. G n m zde yapay sinir ađları ve derin  đrenme modelleri metin ve dialog  retimi i in daha  ok tercih edilmektedir. Buna en iyi  rnek 11 Haziran 2020 tarihinde OpenAI řirketi tarafından yayınlanan Generative Pre-trained Transformer 3 kısaca GPT-3, insanların yazdıđı metinlere benzer i erik  retmek i in derin  đrenmeyi kullanan dil modelidir. Bu modelin  rettiđi  ıktılarının yapay zeka tarihinde bir d n m noktası olduđu dile getirilmektedir. Metin olarak tarifi yapılan uygulamaları yaratmak ve istenen řairin dilinden konuřmalar yapmak gibi bir  ok inanılmaz  zelliđi i inde barındırmaktadır. Bu  rneklerin canlı  rneklerinden biri ise

Reddit forumunda gerçekleşmiştir. Gpt-3 haftalarca bu forum üzerinde içerikler paylaşmış fakat kimse tarafından bir bot olduğu fark edilememiştir [19].

BÖLÜM: II

MATERYAL

KULLANILAN TEKNOLOJILER

3.1 *Kullanılan Diller*

3.1.1 *Dart*

Dart, Google şirketi tarafından geliştirilmekte olan yüksek seviyeli bir programlama dilidir. Açık kaynaklı ve nesneye yönelik bir dildir. C# ve Java dillerine benzerliği ile bilinmektedir. Dart dili, geliştiricilerinin büyük bir bölümü Web alanında uzmanlaşmış kişiler olduğundan, yoğun olarak kullanmakta oldukları JavaScript dilinin karmaşıklığını ve eksik yönlerini gidererek reaktif yapıya uygun bir alternatif yaratmak için geliştirilmiştir. Dart dili IOS, Android, Web, Linux, Windows ve MacOS platformlarına uygulama geliştirme imkanı sunmaktadır.

Bu özellikleri ile Dart dili; IOS, Android ve Web platformlarına uygulamanın dağıtımını sağlamak için kullanılacaktır.

3.1.2 *Python*

Python; nesne yönelimli, modüler ve yüksek seviyeli bir yorumlama dilidir. Yazım kolaylığı ve yalınlığı ile öğrenilmesi en kolay dillerden biridir. Windows, Linux, Mac, Unix, Symbian ortamlarında çalışabilmektedir. Sistem programlama, ağ programlama, arayüz geliştirme, web ve mobil uygulama geliştirme gibi birçok alanda kullanılabilmektedir. Kullanım alanlarından en yeni ve güçlü olduğu alanlar makine öğrenmesi ve derin öğrenme alanlarıdır.

Yapay zeka modüllerinin geliştirilmesinde Keras, TensorFlow, Scikit kütüphaneleri için Python dili kullanılacaktır.

3.2 *Platformlar*

3.2.1 *Flutter*

Flutter, Google şirketi tarafından geliştirilen mobil uygulama geliştirme platformudur. Android, IOS, Web, Desktop platformlarına uygulama geliştirme imkanı sunmaktadır. Google kendi işletim sistemi Fuchsia için de bu geliştirme platformunu kullanmaktadır. Dart dilini temel alır. Hızlı, kolay ve maliyetsiz geliştirme imkanı sunar. Düşük bütçeli start-up firmaları uygulamalarını birçok platforma aynı anda sunmak için Flutter platformunu kullanmayı tercih etmektedir.

Firestore çözümleri ile uyumlu olarak geliştirilmesinden ve kolay şekilde IOS, Android ve Web platformlarına uygulama geliştirilebilme imkanı sunmasından dolayı tercih edilmiştir.

3.3 *Firestore Çözümleri*

3.3.1 *Firestore*

Firestore verileri depolamak ve senkronize etmek için esnek, ölçeklenebilir bulut veritabanı sistemidir. NoSQL tabanlıdır. Android, IOS, Web ve Nodejs platformlarında kullanılabilen bu hizmet diğer Firebase hizmetleri ile etkileşimde bulunabilmektedir. Uygulamalar için çevrimdışı desteği sunabilmektedir. Gerçek zamanlı iletişim için optimize edilmiştir.

Firestore Fonksiyonlar, Kimlik Doğrulama, Depolama servisleri ile entegreli çalışabilmesinden, uygun fiyatlandırmasından ve otomatik ölçeklendirilebilme özelliğinden dolayı tercih edilmiştir. Firestore ile geliştirilmiş birçok chat uygulaması açık kaynak olarak mevcuttur.

3.3.2 *Firestore Bulut Mesajlaşma Servisi*

Firestore Bulut Mesajlaşma Servisi(Cloud Messaging Service) sunucu ile cihazlar arasında Android, IOS ve Web üzerinden ücretsiz olarak mesaj ve bildirim gönderme imkanı sunan bildirim servisi. Bu haberleşme, güvenilirlik ve güç açısından verimli bir bağlantı sağlamaktadır. Bildirimler gerçek zamanlı veya kullanıcıların yerel saat dilimlerine göre iletebilme özelliğine sahiptir. Bildirimler Firestore için Google Analizler(Google Analytics) ile entegredir. Bu sayede etkileşim ve dönüşümü izlemeye olanak sağlamaktadır. Fonksiyonlar sistemi ile entegreli olarak çalışabilmektedir.

Firestore Bulut Mesajlaşma Servisi, Firestore ve Fonksiyonlar servisleri ile entegreli çalışabilmesinden ve kolay ölçeklendirilebilmesinden dolayı tercih edilmiştir.

3.3.3 *Firestore Depolama*

Firestore Depolama(Storage) hizmeti basit ve uygun maliyetli bir dosya depolama hizmetidir. Ağ kalitesinden bağımsız şekilde Firestore uygulamalar için dosya yükleme ve indirme işlemlerini gerçekleştirir. Google geliştiricilerinin gücü ile güvenlik özellikleri artırılmıştır. Bu yüklemeler ve indirmeler durdurulabilir ve kaldığı yerden devam ettirilebilir. Kimlik doğrulama için Firestore Kimlik Doğrulama Servisi ile entegre edilebilir. Uygulamaların viral olma durumu için eksabayt ölçeği ile oluşturulmuştur. Küçük bir modelden büyük bir yapıya geçiş için uygun ve zahmetsizdir.

Kimlik Doğrulama ve Fonksiyonlar servisi ile entegreli çalışabilmesinden ve otomatik olarak ölçeklendirilebilmesinden dolayı tercih edilmiştir.

3.3.4 *Firestore Fonksiyonlar*

Firestore Fonksiyonlar(Cloud Functions), bulut sunucu hizmetidir. Nodejs ile çalışmaktadır. İstenilen iş parçacıklarını sunucu üzerinde çalıştırarak bildirim gönderme, veritabanı temizliği, API çağrıları gibi işlemleri gerçekleştirebilmektedir. Normal bir sunucudan çok daha hızlı ve güvenlidir. Güvenliği sağlanması gereken onay işlemleri bu sunucu üzerinde yapılabilir. Firestore veritabanından tetiklemeler olarak modül çalıştırabilir. Tüm Firestore çözümleri ile iletişim kurabilme özelliğine sahiptir. Şekil 1'de bu iletişim gösterilmiştir.



Şekil 1: Firebase Fonksiyonlar Servisinin Birlikte Çalışabildiği Sistemler [1]

3.3.5 *Firestore Kimlik Doğrulama Servisi*

Firestore Kimlik Doğrulama Servisi(Authentication) bir bulut üyelik sistemidir. Kullanıcıların bilgilerini güvenli şekilde tutmak ve birden çok platformda bu bilgileri ulaştırmak için tasarlanmıştır. Google, Twitter, Facebook gibi platformların üyelikleri ile üye olma özelliğini de içinde barındırır. SMS onayı, mail onayı ve şifre değişimi gibi hizmetleri de sağlamaktadır. OAuth 2.0 ve OpenID Connect gibi endüstri standartlarından yararlanmaktadır.

3.3.6 *Firestore Makine Öğrenmesi*

Firestore Makine Öğrenmesi(Machine Learning), Google'ın cihaz üzerinde ve bulut üzerinde, makine öğrenmesi ve derin öğrenme özelliklerini kullanıma sunan hizmetidir. Görüntü sınıflandırma, çeviri, yüz algılama, barkod tanıma gibi özellikleri barındırır. TensorFlow Lite modellerini cihaz üzerinde çalıştırabilmektedir.

3.4 *Yapay Zeka Araçları*

3.4.1 *Keras*

Keras, Python üzerinde geliştirilmiş açık kaynaklı bir sinir ağı kütüphanesidir. Diğer kütüphanelere göre daha kolay bir kullanıma sahiptir. TensorFlow üzerine kurulmuştur. İşlemci ve ekran kartları üzerinde çalışabilmektedir. Çoğunlukla görüntü işleme, üretme ve sınıflandırma işlemleri için tercih edilmektedir.

Uygunsuz medya sınıflandırması ve chat botu için tercih edilmiştir.

3.4.2 *TensorFlow*

TensorFlow açık kaynaklı bir makine öğrenmesi kütüphanesidir. Google tarafından geliştirilmiştir. Derin öğrenme alanında en çok kullanılan kütüphanelerden biridir. R, Python, C# ve JavaScript gibi birçok dili destekleyebilmektedir.

Firestore Makine Öğrenmesi servisi ile entegreli olarak çalışabilmektedir. Bu özellik cihazlar üzerinde yapay zeka tahminleri üretmeye yardımcı olur.

3.4.3 *Google A.I Platform*

Google A.I Platform, makine öğrenimi modellerini oluşturmak, dağıtmak ve yönetmek için geliştirilmiş bir platformdur. Üzerine hazır modeller entegre edilerek dağıtıma sunulabilmektedir. Veri kümelerinin hazırlanması, derlenmesi, doğrulanması ve dağıtımını sağlayabilmektedir.

karar alma yöntemini ortaya çıkarmayı amaçlar. Görselleştirilmesi ve analizi kolaydır. Sayısal ve kategorik verileri işleyebilmektedir. Aşırı eğitime müsait olduğu için budama işlemine ihtiyaç duyabilmektedir. Şekil 3'te futbol oynama kararını almak için hava durumu verileri ile oluşturulmuş karar ağacı verilmiştir [20].

Kolay görselleştirilip anlaşılabilmesinden ve kolay yönlendirilebilmesinden ötürü ceza sistemi modülünde kullanılmak için seçilmiştir.

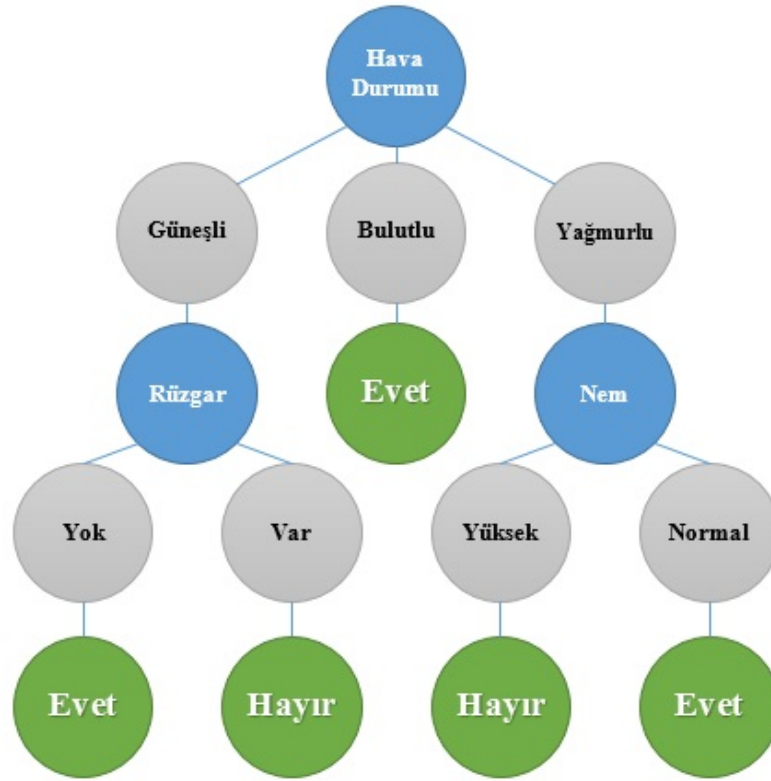
3.5.3 *Rassal Ormanlar*

Rassal Ormanlar ya da Rassal Karar Ormanları, eğitim anında çok sayıda Karar Ağacı oluşturarak çalışan bir sınıflandırma algoritmasıdır. Hiper parametre kestirimine ihtiyaç olmadan yüksek başarıda sonuçlar verebildiği için tercih edilmekte olan bir algoritmadır. Overfitting problemi de karar ağaçlarına göre kısmen daha düşüktür.

3.6 *Dil İşleme Araçları*

3.6.1 *Dialogflow*

Dialogflow bir doğal dil işleme ve sınıflandırma servsidir. Bu servis bir çok dilde verilen metinleri CNN ve LSTM modelleri kullanarak konulandırmaktadır. Bu konu başlıkları üzerinden ön tanımlı cevapları sunabilmektedir. Aynı zamanda kullanıcıların dönütlerini sınıflandırarak model eğitimini optimize etmeye olanak sağlar. Chat bot geliştirme aşamasında kullanılmıştır [21].



Şekil 3: Karar Ağacı Örneği [2]

BÖLÜM: III

YAPAY ZEKA

CEZA KARAR MODÜLÜ

4.1 *Hedef*

Ceza karar modülü, şikayet edilmiş olan bir paylaşımının argo içeriği, görsel uygunluk analizi, kullanıcının önceki şikayetleri, kullanıcının aldığı ceza verilerini dikkate alarak ceza kararı vermesi beklenen bir modeldir. Bu modülün temel beklentisi yanlış olarak kimseye ceza vermemesi yani yanlış pozitif sonuçlar üretmemesidir. Modelin temel olarak 3 sınıfı mevcuttur. Bu sınıflar; ceza verilmeyecek, ceza verilecek ve denetime gönderilecek şeklindedir.

Bu model için sınıflandırma açısından en büyük kriter görsel olarak uygunsuz içerikleri kesin olarak elemek, metin içeriği olarak belirsiz içerikleri denetime göndermektir. Görsel içerik algılama modülleri çok büyük kesinlikte çalışarak uygunsuz içerikleri yakalayabilirken metin içeriği analiz sonuçları yanlış sonuçlar doğurabilmektedir. Bu modelin asıl amacı kullanıcıları engellemekten çok, yöneticiler tarafından denetlenmeye ihtiyacı olan sorunlu hesapların filtrelenmesidir. Bu sebeple kullanıcının karma puanı ve önceden aldığı şikayetler belirleyici bir faktördür.

4.2 Hazırlık

Ceza karar modülüne uygun bir veri seti bulanamayacağı için bir veri üreticiye ihtiyaç duyulmaktadır. Bu veri üreticisi etiketlenmeye hazır şikayet verileri üretmektedir. Bu üretim sürecinde normalden uzak çöp veriler üretilmemesi için normal dağılıma uygun rastgele sayılar üreten bir fonksiyon geliştirilmiştir. Bu fonksiyon daha çok ortalama veriler üretmeye meyillidir. Örnek verilecek olursa 50 kez ceza almış bir kullanıcı hakkında verilecek karar çok kesin olarak cezadır. Bu tip uç verilerin etiketlenmesi için zaman kaybetmemek amacıyla bu yöntem geliştirilmiştir. Şekil 4'te bu metot gösterilmiştir.

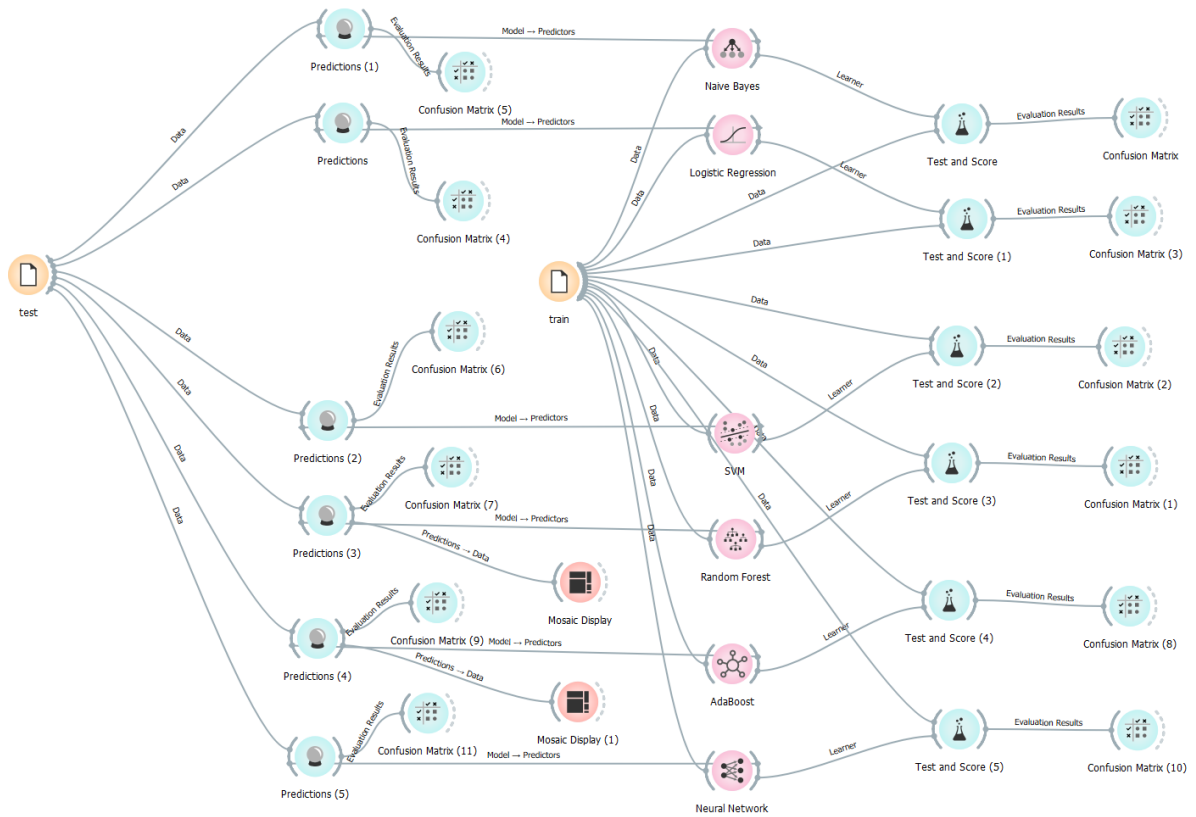
```
def yogunluklu_random(sayi=random.randint(0, 50), iter=0):  
    if sayi < 2:  
        return sayi  
    elif sayi < 7:  
        if iter > 3:  
            return sayi  
        else:  
            return yogunluklu_random(sayi=random.randint(0, 50), iter=iter + 1)  
    elif sayi < 10:  
        if iter > 5:  
            return sayi  
        else:  
            return yogunluklu_random(sayi=random.randint(0, 50), iter=iter + 1)  
    elif sayi < 30:  
        if iter > 7:  
            return sayi  
        else:  
            return yogunluklu_random(sayi=random.randint(0, 50), iter=iter + 1)  
    elif sayi <= 50:  
        if iter > 11:  
            return sayi  
        else:  
            return yogunluklu_random(sayi=random.randint(0, 50), iter=iter + 1)
```

Şekil 4: Normal Dağılıma Göre Rastgele Sayı Üretici

4.3 Eğitim

4.3.1 Orange Üzerinden Eğitim

Orange, basit kullanımı ve görselleştirme seçeneklerine sahip bir veri madenciliği ve makine öğrenmesi aracıdır [22]. Bu programın kullanılma sebebi model oluşturulmadan önce en yüksek başarı oranına sahip olacak makine öğrenmesi algoritmasını tespit etmektir. Şekil 5'te gerçekleştirilmiş olan eğitim şeması gösterilmiştir. Bu program üzerinde Naiv Bayes, Logistic Regression, SVM, Random Forest, AdaBoost ve Yapay Sinir Ağı algoritmaları denenmiştir.



Şekil 5: Orange Üzerinde Tasarlanmış Eğitim Şeması

4.3.1.1 Sonuçlar

5 k fold değeri ile yapılan test sonucu Şekil 6’da gösterilmektedir. Bu sonuçlara göre Flutter üzerinde çalıştırmaya uygun olan en başarılı model Random Forest modelidir. Şekil 7’de karşılık matrisi gösterilmiştir. Bu şemaya göre hiçbir yanlış cezalandırma gerçekleşmemiştir. Bir karşılaştırma yapılacak olursa Şekil 8’de %81 başarı oranına sahip Naiv Bayes algoritmasının karışıklık matrisi verilmiştir. Bu matrise göre 58 kullanıcı haksız olarak ceza almış, 198 kişi ceza alması gerekirken ceza almamıştır. Bu fark yanlış pozitiflik oranının ne kadar az olması gerektiğini göstermektedir.

Evaluation Results					
Model	AUC	CA	F1	Precision	Recall
SVM	0.906	0.707	0.710	0.743	0.707
Random Forest	1.000	0.997	0.997	0.997	0.997
Neural Network	0.998	0.971	0.971	0.971	0.971
Naive Bayes	0.931	0.816	0.809	0.812	0.816
Logistic Regression	0.946	0.807	0.805	0.804	0.807
AdaBoost	0.998	0.998	0.998	0.998	0.998

Şekil 6: Orange Test Sonuçları

Orange programı ile gerçekleştirilen eğitimde en başarılı bulunan Random Forest algoritmasının mozaik grafiği Şekil 9’da verilmiştir. Bu grafiğe göre daha önce alınan ceza sayısına göre ve gönderilen içeriğin analiz sonucuna göre verilen kararlar gösterilmektedir.

		Predicted			
		0	1	2	Σ
Actual	0	4962	0	0	4962
	1	0	2120	7	2127
	2	2	10	2899	2911
	Σ	4964	2130	2906	10000

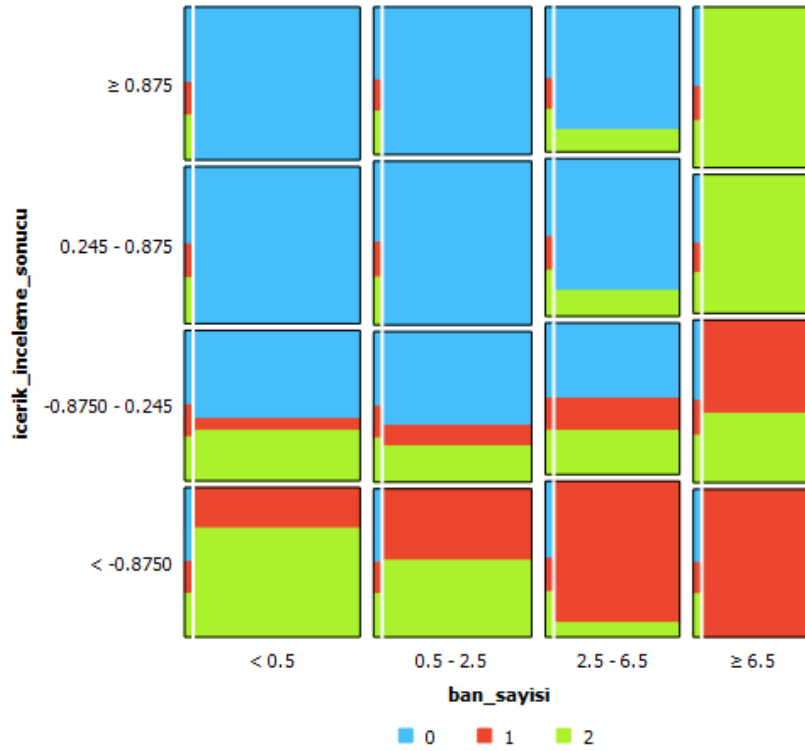
Şekil 7: Random Forest Karışıklık Matrisi Sonucu

		Predicted			
		0	1	2	Σ
Actual	0	4862	58	42	4962
	1	198	1493	436	2127
	2	605	423	1883	2911
	Σ	5665	1974	2361	10000

Şekil 8: Naiv Bayes Karışıklık Matrisi Sonucu

4.3.2 Python Üzerinden Eğitim

Dart dili üzerinde çalıştırılabilen yapay zeka algoritmaları şu an için kısıtlıdır. Bu sebeple SkLite kütüphanesi ile desteklenmekte olan algoritmalar ile Python üzerinde eğitim gerçekleştirilmiştir. Bu kütüphane Dart üzerinde çalıştırılabilen modeller üretebilmektedir [23]. Bu eğitimde Random Forest, SVC, K Neighbors, Karar Ağacı, MLP, Gaussian Neighborhood, Bernoulli Neighborhood algoritmaları test edilmiştir. Test sonuçları Şekil 10'da gösterilmiştir. Bu eğitim



Şekil 9: Mozaik Grafiği

sonucunda Orange üzerinde de elde etmiş olduğumuz Random Forest algoritması %100 başarı göstermiştir. Bununla birlikte Karar Ağacı algoritması da aynı başarıya ulaşmıştır.

```
Random Forest Score: 1.0
SVC Score: 0.5500833333333334
KNeighborsClassifier Score: 0.77225
DecisionTreeClassifier Score: 1.0
MLPClassifier Score: 0.93425
GaussianNB Score: 0.8378333333333333
BernoulliNB Score: 0.6750833333333334
```

Şekil 10: Python Eğitim Sonuçları

4.4 Sonuç

Eğitimler sonucunda verilere dayalı tutarlı kararlar alınmak isteniyorsa en doğru seçeneklerin Random Forest ve Karar Ağacı algoritmaları olduğu tespit edilmiştir. Düşük yanlış pozitif oranı hedeflenen kritik durumlar için güvenli sonuçlar elde edilmiştir. Yüksek başarı oranına sahip Random Forest modeli uygulamaya entegre edilmiştir.

DUYGU ANALIZI MODÜLÜ

5.1 *Hedef*

Duygu analizi modülünün temel hedefi uygunsuz ve argo içerikli metinleri sınıflandırmaktır. 19 bin veri içeren Türkçe veri seti, BERT ve Average Word Vector algoritmaları ile eğitilerek olumlu ve olumsuz olarak iki sınıf olarak etiketlenecektir.

5.2 *Hazırlık*

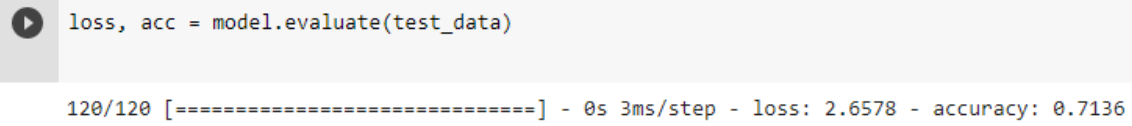
Kaynak olarak alınmış olan veri seti, metin içeriğinin olumlu ya da olumsuz olduğuna göre etiketlenmiştir. Şikayetlerde ceza unsuru olarak incelenmesi için sadece küfür ve argo içerikli metinlerin olumsuz etikelenecek şekilde yeniden sınıflandırılması gerekmektedir.

Average Word Vector uygulamasının çalıştırılabilmesi için bir kelime dağarcığına ihtiyaç vardır. Dart dili için bu kelime dağarcığını oluşturacak bir sınıf yazılmıştır. Bu dağarcıkta kelimeler numaralandırılmaktadır. Bu dağarcık kullanılarak verilen metin tokenizasyon işlemine uğrayarak sayısallaştırılır. BERT modelleri için böyle bir ön işleme ihtiyaç duyulmaz.

5.3 Eğitim

Bu modelin eğitimi Colab platformu üzerinden gerçekleştirilerek Tensorflow Lite modeline dönüştürülmüştür. Şekil 11’de %20 test verisi ile yapılan Average Word Vector algoritması sonucu verilmiştir. Bu veri setinde metnin olumlu ya da olumsuz olduğu sınıflandırılmaktadır.

Step 4. Test.



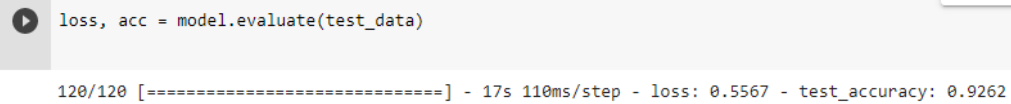
```
loss, acc = model.evaluate(test_data)
```

120/120 [=====] - 0s 3ms/step - loss: 2.6578 - accuracy: 0.7136

Şekil 11: Duygu Analizi AWV Sonucu

Şekil 12’de ise BERT modeli ile aynı verinin eğitim sonucu gösterilmiştir. BERT modelleri çok daha başarılı sonuçlar ortaya çıkartmaktadır.

Step 4. Test.



```
loss, acc = model.evaluate(test_data)
```

120/120 [=====] - 17s 110ms/step - loss: 0.5567 - test_accuracy: 0.9262

Şekil 12: Duygu Analizi BERT Sonucu

Şekil 13’te ise Average Word Vector algoritması ile sadece küfür ve argo içeriklerin sınıflandırıldığı veri seti eğitilmiştir. Bu veri setinin çok daha kolay sınıflandırılabilir olduğu görülmüştür.

Şekil 14’te aynı veri setinin BERT algoritması ile sınıflandırma sonucu verilmiştir.

```
[10] loss, acc = model.evaluate(test_data)
```

```
120/120 [=====] - 35s 267ms/step - loss: 0.3920 - test_accuracy: 0.9444
```

Şekil 13: Argo Analizi AWV Sonucu

```
[12] loss, acc = model.evaluate(test_data)
```

```
119/119 [=====] - 13s 77ms/step - loss: 0.0721 - test_accuracy: 0.9916
```

Şekil 14: Argo Analizi BERT Sonucu

5.4 Sonuç

Yapılan eğitim sonucunda BERT modellerinin %20 test verisi ile yüksek başarı sağladığı görülmüştür. Fakat veri setinde çok uzun cümleler bulunması ve kelime dağarcığında bulunmayan kelimeler ile oluşturulan kısa metinlerde yanlış tahminler yapılabildiği görülmüştür. Elde edilen modelin yapısı Şekil 15'te verilmiştir.

```
] model.summary()
```

Model: "model"

Layer (type)	Output Shape	Param #	Connected to
input_word_ids (InputLayer)	[(1, 128)]	0	
input_mask (InputLayer)	[(1, 128)]	0	
input_type_ids (InputLayer)	[(1, 128)]	0	
keras_layer (KerasLayer)	[(None, 768), (None, 109482241		input_word_ids[0][0] input_mask[0][0] input_type_ids[0][0]
dropout (Dropout)	(None, 768)	0	keras_layer[0][0]
output (Dense)	(None, 2)	1538	dropout[0][0]
Total params: 109,483,779			
Trainable params: 109,483,778			
Non-trainable params: 1			

Şekil 15: BERT Model Yapısı

UYGUNSUZ MEDYA SINIFLANDIRMA MODÜLÜ

6.1 *Hedef*

Bu modülün amacı resim dosyalarını işleyerek uygunsuz içeriğe sahip olup olmadığını tespit etmektir. Ceza kararlarında etkili olacağından, tutarlığı çok yüksek ve yanlış pozitif sonuçlar üretmeyen bir modele ihtiyaç duyulmaktadır.

6.2 *Hazırlık*

Sınıflandırmayı yapacak olan Tensorflow Lite modeli belirli sabit boyutta görselleri alabilmektedir. Bu nedenle Dart dili üzerinde verilen resimleri sabit boyuta indirgeyen ve float32 tipinde veri dizisine çeviren bir metot yazılmıştır.

6.3 Test

Colab platformu belirli miktarda depolama ve hafıza sunduğu için yüksek boyutlu NudeNet veri seti CNN algoritması ile çalıştırılamamıştır. Bu nedenle bu veri seti ile eğitilmiş olan Tensorflow Lite modeli alınarak yeterli tutarlılığı sağlayıp sağlayamadığını kontrol etmek amacıyla 100 adet yarısı uygunsuz, kalan yarısı ise yanlış sınıflandırılabilir ya da tamamen normal görseller ile test edilmiştir. Test sonucunda 3 adet %50-60 olasılıklarla uygunsuz seçilmiş görsel bulunmuştur. Başarı oranı %97 olarak belirlense de yanlış etiketlenen görsellerin olasılık değerlerinin düşük olmasından ve karar modülünün %80 kararlılıktan az olan gönderileri zararsız saymasından ötürü tamamen güvenli bir model olduğuna karar verilmiştir.

6.4 Sonuç

Kullanılmış olan sınıflandırma modeli göstermektedir ki uygunsuz içeriklerin sansürlenmesi ve tespit edilmesi için CNN modellerinin mobil cihazlarda kullanılması güvenli ve maliyetsiz bir çözümdür.

BÖLÜM: IV

UYGULAMA

GENEL MİMARİ TASARIM VE YÖNTEM

7.1 Genel Kullanım

Geliştirilmiş olan Flutter uygulamasında kullanıcılar mail, Twitter ya da Google hesapları ile üyelik oluşturmaktadır. Kullanıcılar hesap oluşturma işlemi sonrasında profilleri için biyografi, bölüm bilgisi ve profil fotoğraflarını seçerler. Bu işlemde sonra kayıt tamamlanmış olmaktadır. Kullanıcılar giriş yaptıktan sonra ana sayfaya aktarılır. Bu sayfa 5 sekmeden oluşmaktadır. Akış bölümünde kullanıcılar takip etmiş oldukları kişilerin paylaşımlarını görürler. Mesajlar bölümünde sohbet ettikleri arkadaşları ile olan mesajları ve "yeni biri" butonu bulunmaktadır. Bu buton kullanıcıların birbirleri ile sosyalleşmeleri amacıyla eklenmiştir. Paylaşım bölümünde kullanıcılar galeri paylaşımı ve durum paylaşımı yapabilmektedir. Durum paylaşımları metin, video ve resim şeklinde olabilmektedir. Paylaşım ekranında eklenen görseller uygunsuz içerik yakalayıcı tarafından taranır. Gruplar bölümünde fakülte grupları ve diğer sohbet grupları bulunur. Profil bölümünde kullanıcıların galeri ve durum paylaşımları yer alır.

ARAYÜZLER

8.1 Giriş Ekranı

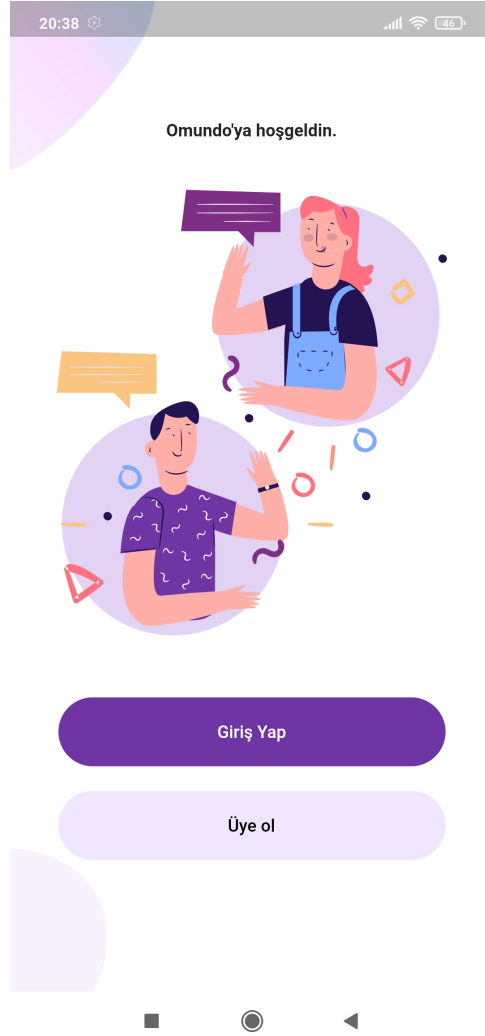
Kullanıcıların giriş metodunu seçmekte olduğu giriş sayfası Şekil 16’da verilmiştir.

8.2 Kayıt Ekranı

Kullanıcıların Twitter, Google ve mail ile kayıt olabildikleri üyelik sayfası Şekil 17’de verilmiştir.

8.3 Üye Girişi Ekranı

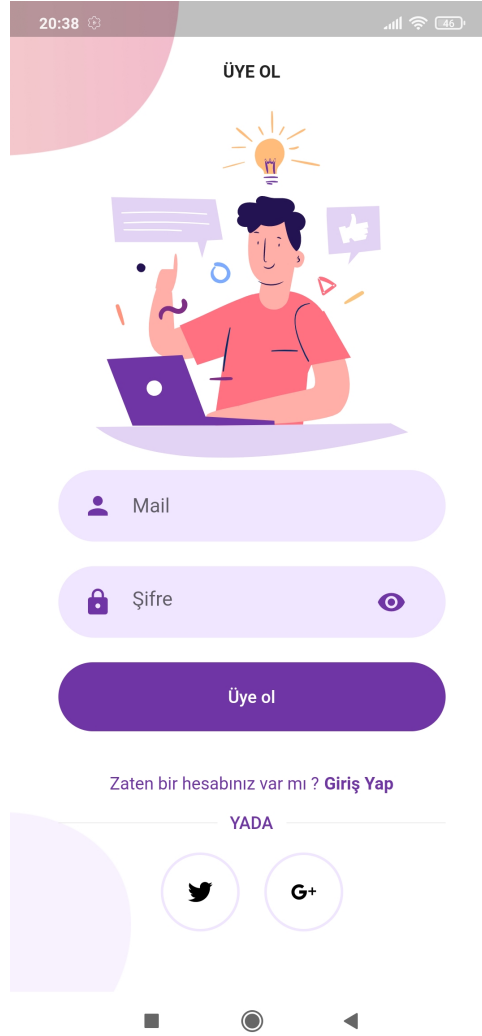
Kullanıcıların Twitter, Google ve mail ile giriş yapabildikleri üye giriş sayfası Şekil 17’de verilmiş olan arayüz ile aynıdır. Twitter ve Google giriş fonksiyonları üye kaydı oluşturma ile aynı fonksiyonelliği göstermektedir.



Şekil 16: Giriş Ekranı

8.4 Profil Oluşturma Ekranı

Profil oluşturma ekranı, hesap kaydı yapılmış kullanıcıların profillerini oluşturmak için yönlendirildiği sayfadır. Kullanıcılar bu ekranda profil fotoğraflarını, kullanıcı adlarını ve bölümlerini girerler. Şekil 18’de görünümü verilmiştir.



Şekil 17: Kayıt Ekranı

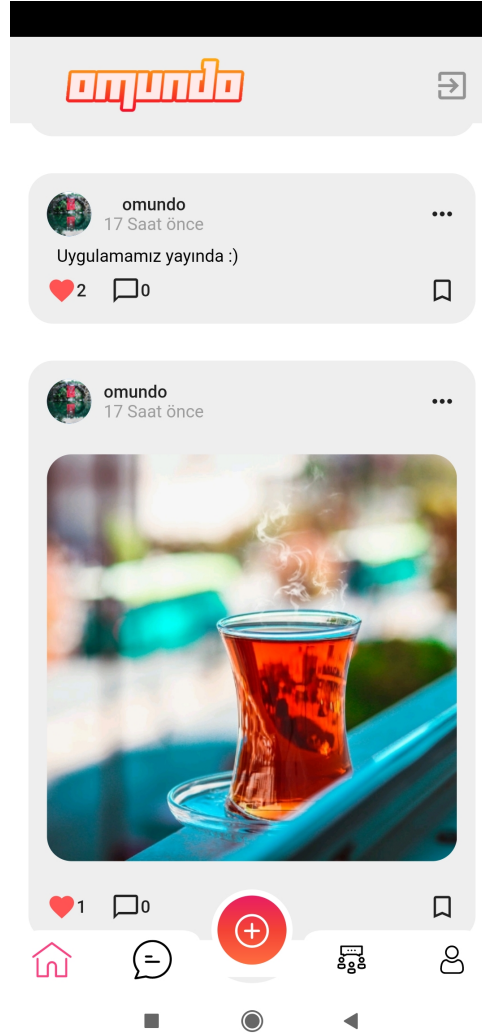
8.5 Ana Sayfa

Uygulamanın merkez bölümü olan ana sayfa; akış, mesajlaşma, gruplar ve profil sayfasını içinde barındırmaktadır. Giriş yapıldığında ilk olarak akış sayfası kullanıcıyı karşılamaktadır. Bu bölümde kullanıcılar takip ettikleri üyelerin paylaşımlarını görebilmektedir. Şekil 19'da görünümü verilmiştir.

Şekil 18: Profil Oluşturma Ekranı

8.6 Mesajlar

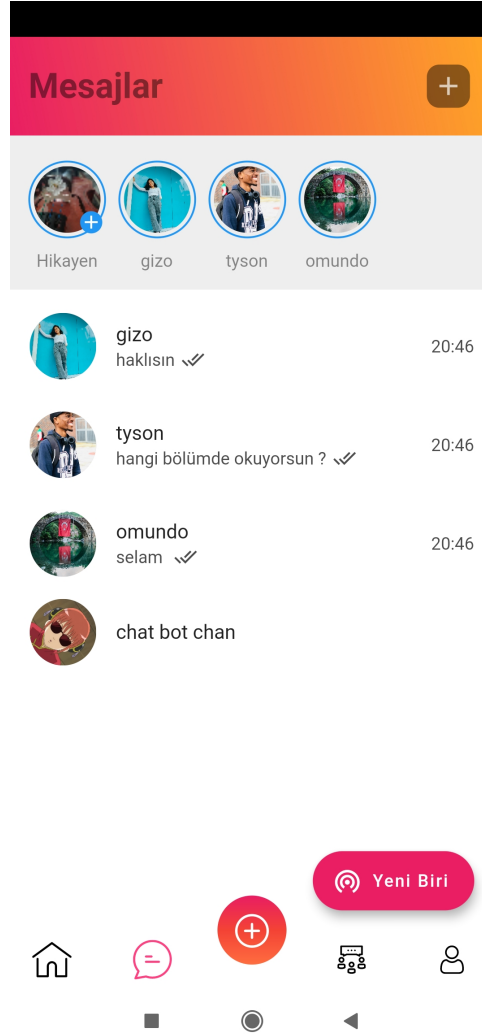
Mesajlar bölümünde kullanıcıların diyalog bilgileri bulunmaktadır. Aynı zamanda uygulama yardımcısı olan bot ile de bu ekran üzerinden iletişime geçilebilmektedir. Yeni biri butonu ile kullanıcılar rastgele olarak diğer kullanıcılar ile sosyalleşebilmektedir. Şekil 20’de mesajlar ekranı verilmiştir.



Şekil 19: Ana Sayfa

8.7 Gruplar

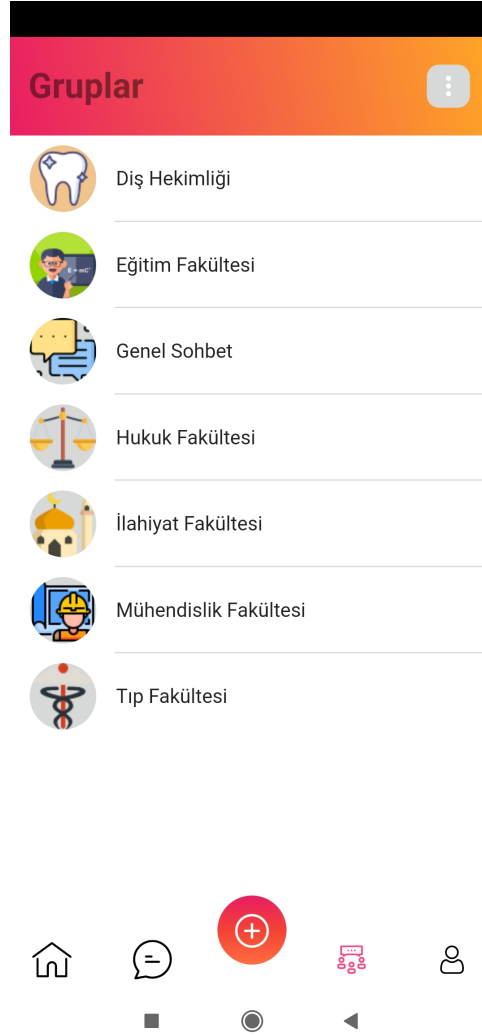
Gruplar ekranı fakülte grupları ve genel sohbet gruplarını içeren kısımdır. Kullanıcılar kendi bölümlerinden kişilerle bu bölümlerde yardımlaşabilmektedir. Şekil 21’de gruplar ekranı, Şekil 22’de ise sohbet ekran görüntüsü verilmiştir.



Şekil 20: Mesajlar Ekranı

8.8 Profil

Profil ekranı kullanıcıların kendi bilgi ve paylaşımlarını görebildikleri ekrandır. Şekil 23, Şekil 24 ve Şekil 25'te bölümleri gösterilmiştir.



Şekil 21: Gruplar Ekranı

8.9 Gönderi Ekranı

Gönderi ekranı kullanıcıların paylaşmış oldukları içeriklerin ve yorumlarının görüntülendiği sayfadır. Şekil 26'da bu ekran gösterilmiştir.



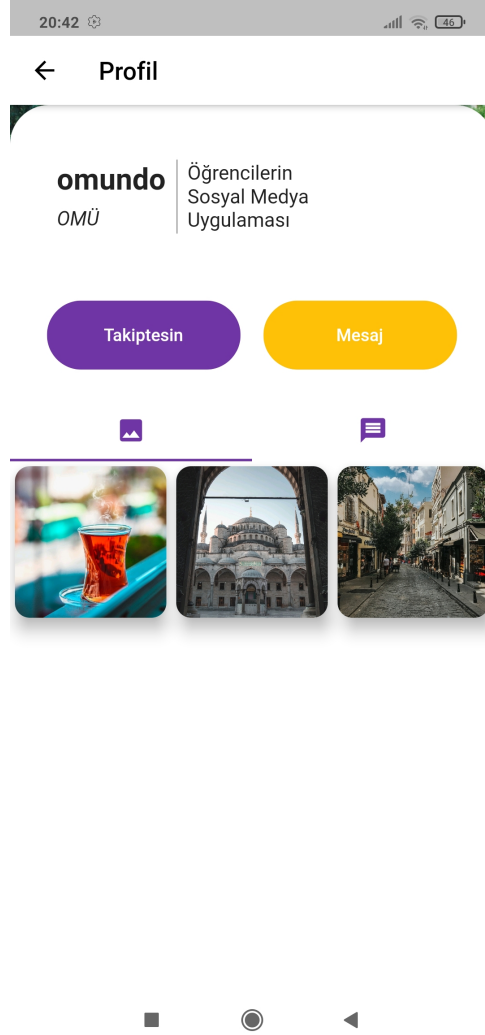
Şekil 22: Grup Sohbet Ekranı

8.10 Gönderi Paylaşımı

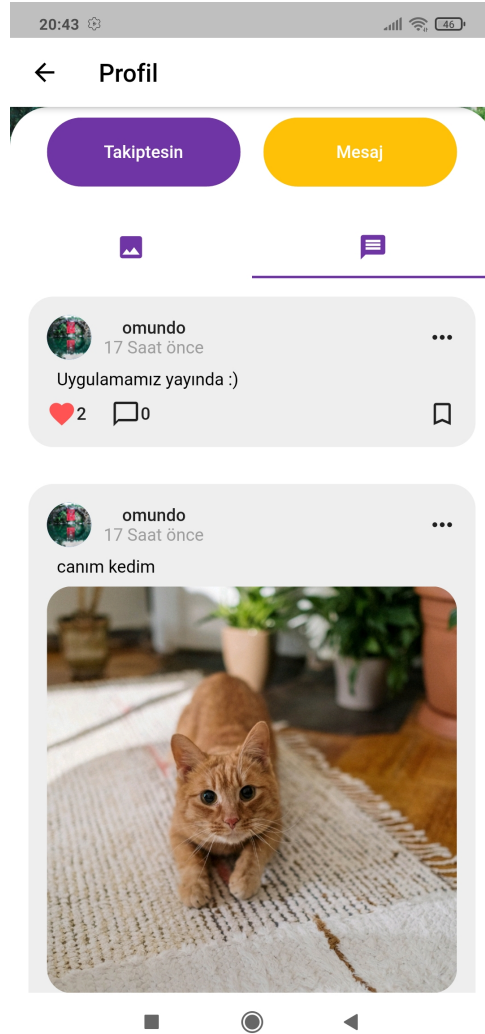
Gönderi paylaşımı ana sayfa üzerinde bulunan paylaşım butonu ile yapılmaktadır. Gönderi menüsü Şekil 27’de verilmiştir. Şekil 28 ve Şekil 29’da gönderi paylaşımı yapılan ekranlar verilmiştir. Kullanıcılar seçtikleri medyaları ve açıklama metinlerini bu ekrandan düzenlemektedir.



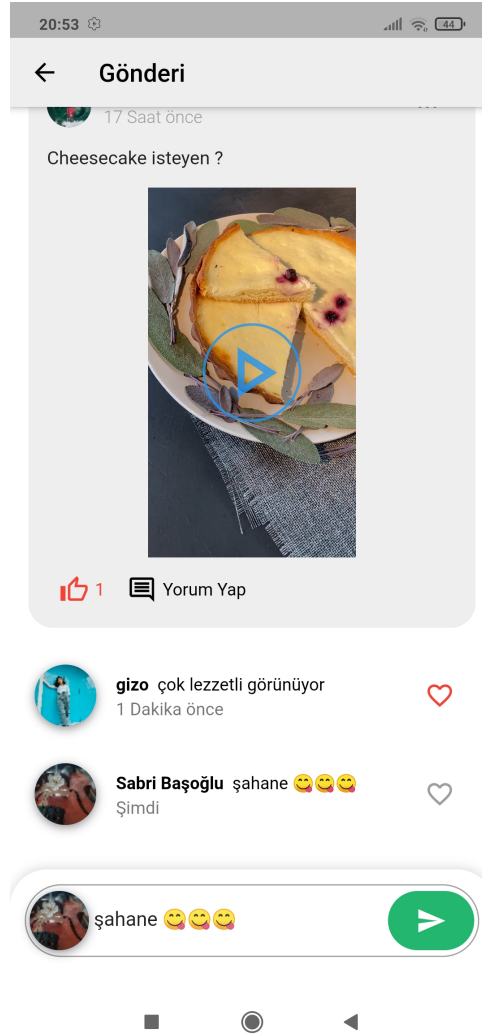
Şekil 23: Profil Bilgi Ekranı



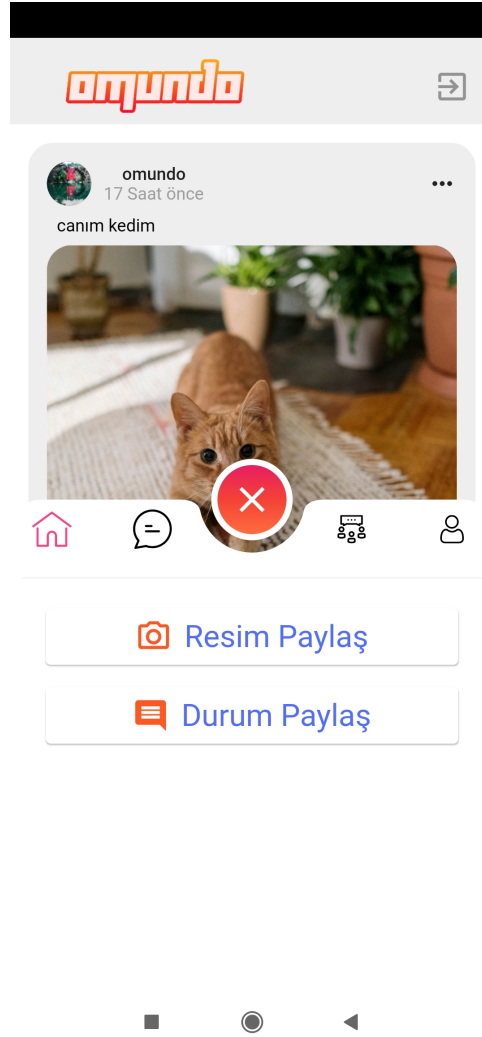
Şekil 24: Profil Galerisi Ekranı



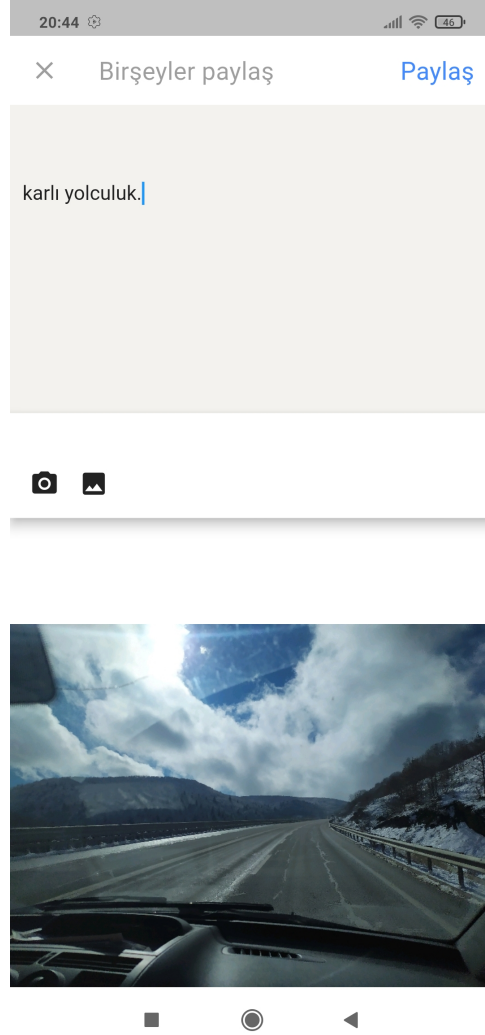
Şekil 25: Profil Gönderi Ekranı



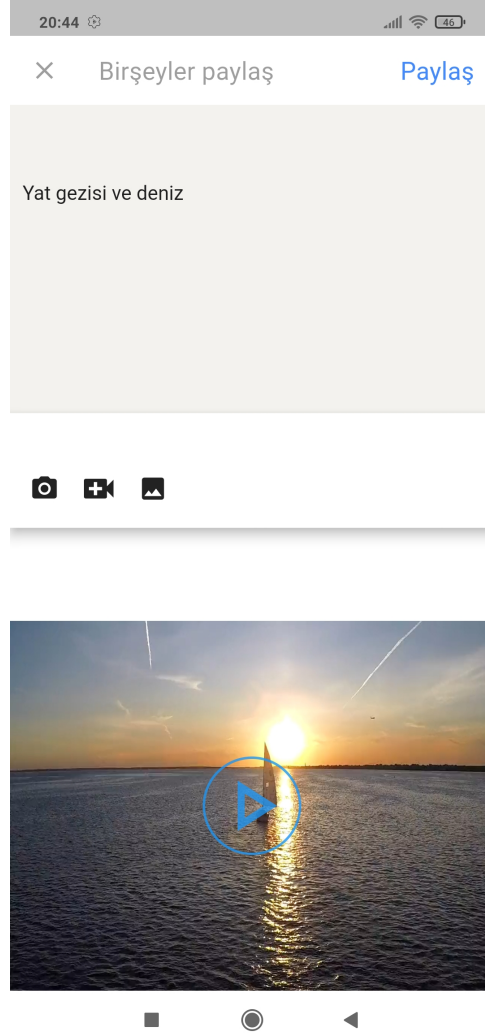
Şekil 26: Gönderi Ekranı



Şekil 27: Gönderi Menüsü



Şekil 28: Galeri Gönderi Ekranı



Şekil 29: Durum Gönderi Ekranı

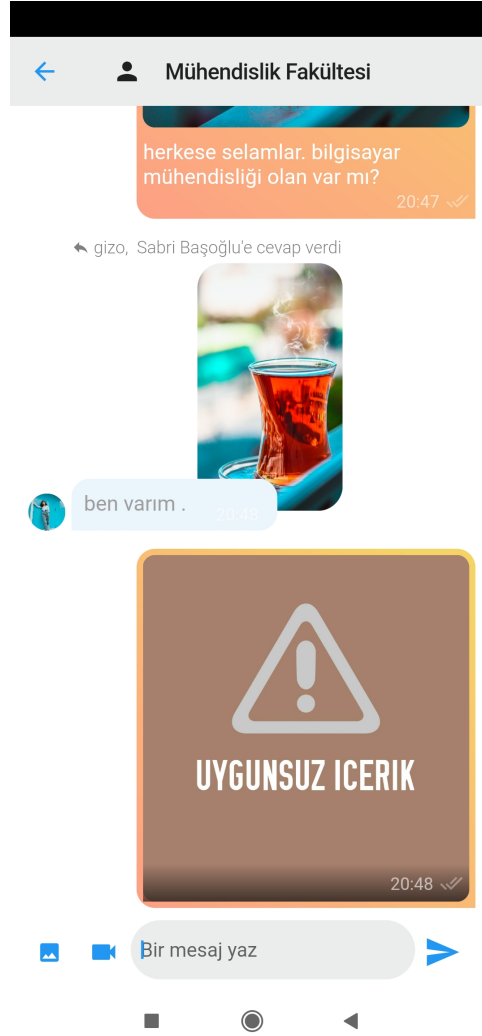
YAPAY ZEKA ÖZELLİKLERİNİN ENTEGRASYONU

9.1 Mesaj Yakalama

Kullanıcıların mesajlaşma esnasında uygunsuz görüntüler paylaşımlarını engellemek için Bölüm 6’da anlatılmış olan görüntü işleme modeli mesajlaşma sistemine entegre edilmiştir. Bu entegrasyonun örneği Şekil 30’da gösterilmiştir. Kullanıcı uygunsuz bir medya dosyası eklediğinde bu görüntü veritabanına kaydedildikten sonra uyarı görseli ile değiştirilir. Gönderilmiş içerik kullanıcının sonraki şikayetlerinde incelenmek üzere uygunsuzluk puanlandırması ile birlikte depolanır. Belirli sayıda uygunsuz içerikli gönderi paylaşmış kullanıcılar Bölüm 4’te anlatılmış olan Ceza Karar Modülü ile cezalandırılır.

9.2 Paylaşım Yakalama

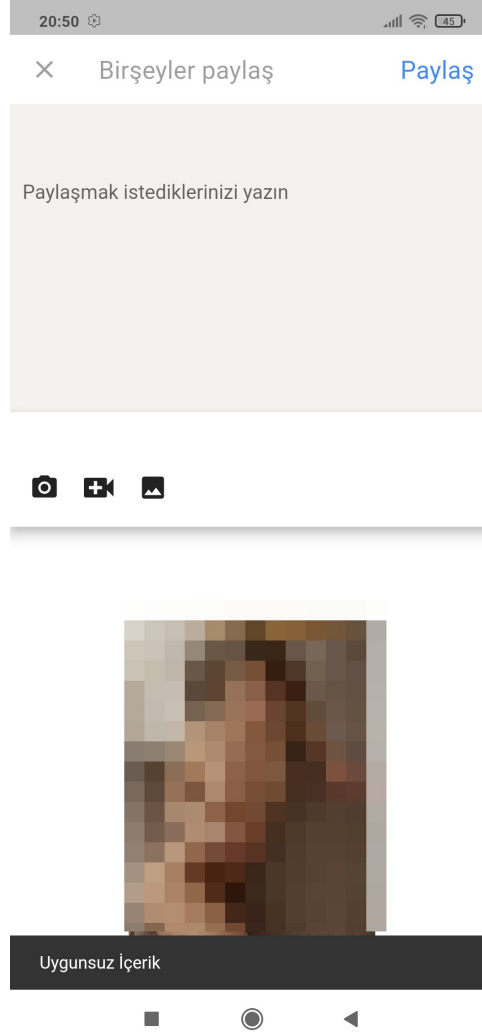
Kullanıcılar uygunsuz içerikler yüklemeye çalışırken Bölüm 6’da anlatılmış olan görüntü işleme modeli devreye girerek yükleme işlemini durdurmaktadır. Bu işlem Şekil 31’de gösterilmiştir.



Şekil 30: Uygunsuz Mesaj Örneği

9.3 Şikayet Sınıflandırma

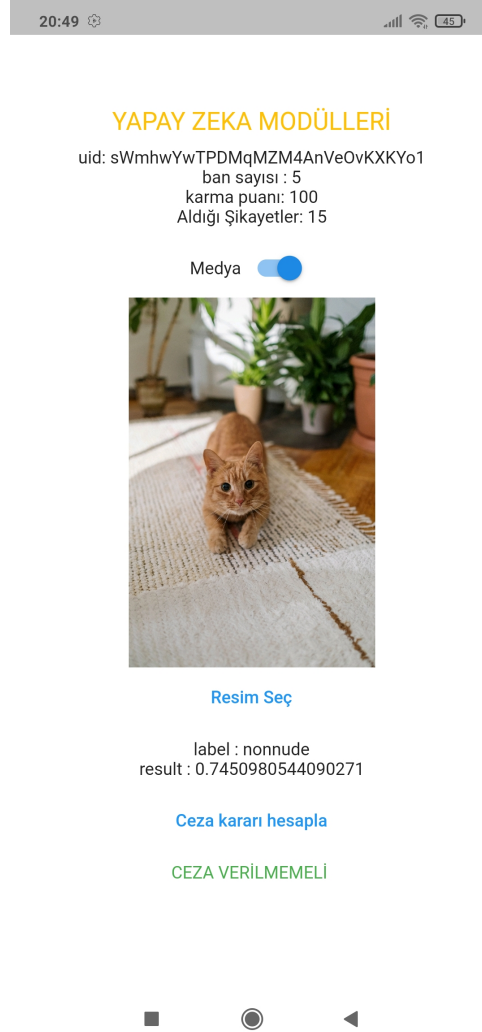
Şikayet sınıflandırma ekranı şikayet edilen uygulamanın duygu analizi ve uygunsuz medya sınıflandırma sonuçlarını gösteren ekrandır. Elde edilen metin ya da görsel sınıflandırma sonucu ceza karar modülü tarafından işlenerek ceza sonucunu gösterir. Şekil 32’de normal bir gönderinin ceza sonucu gösterilmiştir. Şekil 32’de ise uygunsuz içerik sonucu gösterilmiştir.



Şekil 31: Uygunsuz Paylaşım Denemesi Sonucu

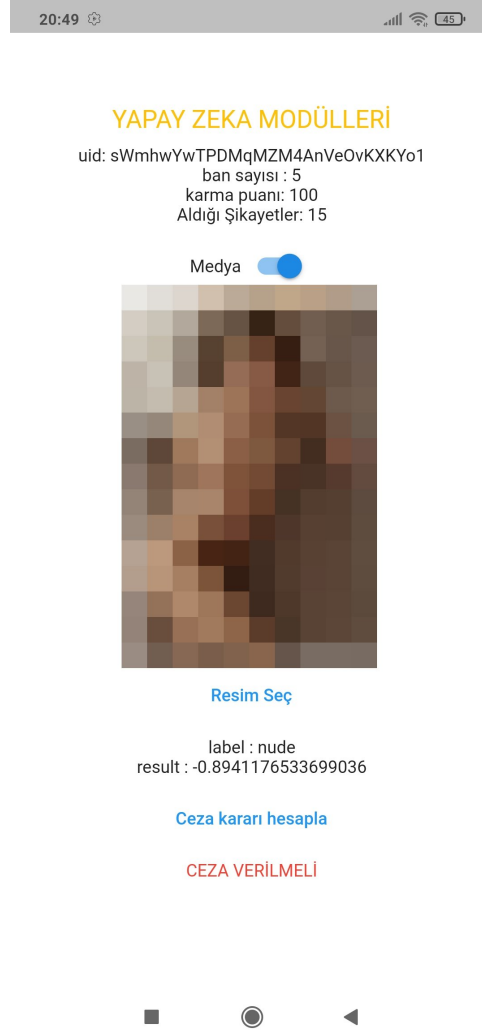
9.4 Chat Bot

Chat bot geliştirilme aşamasında Dialogflow sistemi kullanılmıştır. Bu sistem doğal dil işleme yöntemlerini kullanarak metin içeriğini konulandırmaktadır. Bu proje için toplamda yüz konu başlığı belirlenmiş olup eğitilen model ile verilen metinlerin bu yüz konu içerisinden hangisine dahil olduğunu bularak bu konuların standart cevaplarını kullanıcıya göndermek amaçlanmıştır. Aynı zamanda kullanıcılar bota komut vererek uygulama üzerinde bazı işlemleri de gerçekleştirebilmektedir. Dialogflow eğitim görüntüsü Şekil 34’de verilmiştir.



Şekil 32: Uygun İçerik Şikayet Ekranı

Bu içerikte kullanıcı mesajlarının silinmesini istediğini uzun bir metin ile belirtmiştir. Dialogflow modeli bu içeriği mesaj sil konusuna dahil olarak algılamış ve mesaj temizleme fonksiyonunu aktif etmiştir. Hazırlanan bot önceden tanımlanmış uygulama içi özellikler aktif hale getirilebilir. Şekil 35'te bot sohbet ekranı gösterilmiştir.

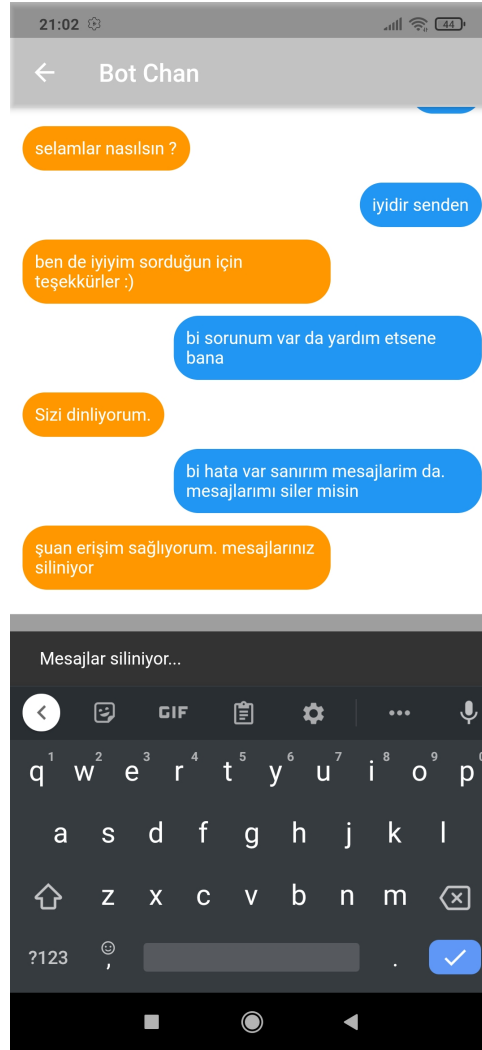


Şekil 33: Uygunsuz İçerik Şikayet Ekranı

USER SAYS	niyeymiş	🗑️
⚠️ INTENT	Click to assign	
USER SAYS	uygulamada yeniyim bana yardım et	✅
INTENT	smalltalk.agent.can_you_help	🗑️
USER SAYS	şuan işim var. sonra konuşuruz	✅
INTENT	smalltalk.user.busy	🗑️
USER SAYS	yok teşekkürler	✅
INTENT	smalltalk.confirmation.no	🗑️
USER SAYS	güle güle	✅
INTENT	smalltalk.greetings.bye	🗑️

CLOSEAPPROVE

Şekil 34: Bot Eğitim Ekranı



Şekil 35: Bot Sohbet Ekranı

GELİŞTİRİLMİŞ YAN ÜRÜNLER

10.1 *Fire_autho Kütüphanesi*

Fire_autho kütüphanesi, bu projenin üyelik işlemleri için geliştirilmiş bir Firebase Auth kimliklendirme yönetimi kütüphanesidir. Bu kütüphane anlık kullanıcı değişikliklerini takip edebilecek şekilde Provider olarak oluşturulmuştur. Mail, Twitter, Google, SMS ve anonim olarak hesap oluşturma; hesapları birbirine bağlama, onaylama, giriş ve güncelleme gibi işlemleri yapabilmektedir. Bu kütüphane Web, IOS ve Android platformlarında çalışabilmektedir. Pub.dev platformu üzerinde fire_autho adı ile dağıtımına sunulmuştur.

10.2 *BERT Modellerinin Flutter Platformu Üzerinde Çalıştırılması*

Geliştirilmiş olan SentimentManager sınıfı, metot kanalları ile Flutter platformu üzerinde BERT modelleri kullanımına olanak sağlayan ilk pakettir. Yapay zeka çalışmalarının Dart dili ve Flutter platformuna taşınmasına katkı sağlamak için Python ile geliştirilecek sunucu çözümleri tercih edilmemiştir.

BÖLÜM: V

SONUÇ

SONUÇ

Yapılan araştırma ve geliştirmeler sonucunda yapay zeka modellerinin mobil cihazlar üzerindeki işlem performansı, güvenlik ve yönetim kararları alma konusundaki başarıları bir sosyal medya uygulaması üzerinde test edilmiştir. Bu testler sonucunda önceki karar verileri ile Rassal Ağaçların ve Karar Ağaçlarının çok yüksek doğrulukta karar alabileceği bununla birlikte ceza kararları için de güvenli olduğu izlenmiştir. Elde edilen bu bilgilerin yanı sıra CNN görüntü sınıflandırıcılarının da mobil cihazlar üzerinde hızlı ve yüksek doğrulukta görüntü sınıflandırabildiği görülmüştür. Uygunsuz içerik sınıflandırma amacıyla test edilmiş modelin ceza kararı verilebilecek kadar yüksek doğruluğa sahip olduğu rahatlıkla söylenebilmektedir. BERT modelleri ile Türkçe metin sınıflandırmada elde edilen sonuç ise bu modellerin ön işleme ihtiyaç duymadan kritik olmayan kararlar için yeterli sonuçlar üretebildiğidir. Geliştirilmiş olan sosyal medya uygulaması üzerinde yapılan testler sonucu yapay zeka modellerinin yerel olarak performans sorunu yaratmadan fonksiyonel şekilde çalışabildiği görülmüştür.

KAYNAKÇA

- [1] Laptrinhx, “Cloud functions and google analytics for firebase,” <https://laptrinhx.com/cloud-functions-and-google-analytics-for-firebase-3771578338/>, 2020, [Çevrimiçi; Erişim Tarihi: 09-Ocak-2021].
- [2] E. Uzun, “Decision tree (karar ağacı),” https://erdincuzun.com/makine_ogrenmesi/decision-tree-karar-agaci-id3-algoritmasi-classification-siniflama, 2020, [Çevrimiçi; Erişim Tarihi: 09-Ocak-2021].
- [3] D. Herman, “Introducing short-term brands: A new branding tool for a new consumer reality,” *Journal of Brand Management*, vol. 7, no. 5, pp. 330—340, 2000.
- [4] H. Krasnova, T. Widjaja, H. Wenninger, and P. Buxmann, “Envy on facebook: A hidden threat to users’ life satisfaction?” *Semantic Scholar*, 2013.
- [5] K. Gemmell, “Why snapchat is the leading cause of fomo,” *The Odyssey*, 2016.
- [6] M. G. Hunt, R. Marx, C. Lipson, and J. Young, “No more fomo: Limiting social media decreases loneliness and depression,” *Journal of Social and Clinical Psychology*, vol. 37, no. 10, 2018.
- [7] M. Csíkszentmihályi, “Beyond boredom and anxiety,” *Jossey-Bass Publishers*, 1975.
- [8] C. Murphy, “Why games work and the science of learning,” *goodgamesbydesign.com*, 2011.

- [9] P. Kaur, A. Dhir, S. Chen, and R. Rajala, “Flow in context: Development and validation of the flow experience instrument for social networking,” *Computers in Human Behavior*, vol. 59, pp. 358–367.
- [10] R. Ap-apid, “An algorithm for nudity detection,” *College of Computer Studies De La Salle University*.
- [11] R. Y. Emon, “A novel nudity detection algorithm for web and mobile application development,” *Department of Computer Science and Engineering Chittagong University of Engineering and Technology*, 2020.
- [12] Google, “Automl vision documentation | cloud auto ml vision | google cloud,” <https://cloud.google.com/vision/automl/docs> [Çevrimiçi; Erişim Tarihi: 22-Aralık-2020].
- [13] Google, “Detect explicit content google vision,” <https://cloud.google.com/vision/docs/detecting-safe-search> [Çevrimiçi; Erişim Tarihi: 23-Aralık-2020].
- [14] M. Akin, “Zemberek, an open source nlp framework for turkic languages,” vol. 10, 12 2020.
- [15] B. Engüllü, “Twitter sentiment analysis,” Ph.D. dissertation, 2018.
- [16] E. Evirgen, “Sentiment analysis of turkish tweets,” Ph.D. dissertation, 2016.
- [17] E. Atwell and B. Shawar, “A.l.i.c.e,” *Artificial Intelligence Foundation*, 01 2003.
- [18] B. Shawar and E. Atwell, “Chatbots: Are they really useful?” *LDV Forum*, vol. 22, pp. 29–49, 01 2007.
- [19] W. D. Heaven, “A gpt-3 bot posted comments on reddit for a week and no one noticed,” <https://www.technologyreview.com/2020/10/08/1009845/>

a-gpt-3-bot-posted-comments-on-reddit-for-a-week-and-no-one-noticed [Çevrimiçi;

Erişim Tarihi: 24-Aralık-2020].

- [20] B. Kaminski, M. Jakubczyk, and P. Szufel, “A framework for sensitivity analysis of decision trees,” *Central European Journal of Operations Research*, vol. 26, 03 2018.
- [21] Wikipedia, “Dialogflow,” <http://en.wikipedia.org/w/index.php?title=Dialogflow&oldid=1006427880>, 2021, [Çevrimiçi; Erişim Tarihi: 13-Haziran-2021].
- [22] Wikipedia, “Orange (software),” [http://en.wikipedia.org/w/index.php?title=Orange%20\(software\)&oldid=1023238051](http://en.wikipedia.org/w/index.php?title=Orange%20(software)&oldid=1023238051), 2021, [Çevrimiçi; Erişim Tarihi 12-Haziran-2021].
- [23] axegon, “SkLite-dart,” <https://github.com/axegon/SkLite-dart>, 2021, [Çevrimiçi; Erişim Tarihi 10-Haziran-2021].