SES ANALIZIYLE DUYGU TANIMA VE DINAMIK TEPKI ÜRETIMI: AKILLI ASISTAN ÇALIŞMASI

HASAN HÜSEYIN ALAV
SAMSUN ÜNIVERSITESI
MÜHENDISLIK VE DOĞA BILIMLERI FAKÜLTESI
SAMSUN 2024

SAMSUN ÜNİVERSİTESİ

MÜHENDİSLİK VE DOĞA BILIMLERI FAKÜLTESİ

SES ANALIZIYLE DUYGU TANIMA VE DINAMIK TEPKI ÜRETIMI: AKILLI ASISTAN ÇALIŞMASI

Hasan Hüseyin Alav

BİTİRME PROJESİ

Yazılım Mühendisliği Bölümü

Danışman: Doç. Dr. Zafer Cömert

Haziran 2024

BITIRME PROJESI ONAYI

Hasan Hüseyin Alav'nın

"Ses Analiziyle Duygu Tanıma Ve Dinamik Tepki Üretimi: Akıllı Asistan Çalışması" başlıklı bitirme projesi .../.../2024 tarihinde aşağıdaki jüri tarafından değerlendirilerek "Samsun Üniversitesi Mühendislik Fakültesi Eğitim-Öğretim ve Sınav Yönetmeliği'nin ilgili maddeleri uyarınca,

Yazılım Mühendisliği Bölümünde Bitirme Projesi olarak kabul edilmiştir.

Jüri Üyeleri	**Unvanı Adı Soyadı**	**İmza**
Üye (Danışman)	:.....	.....
Üye	:.....	.....
Üye	:.....	.....
Üye	:.....	.....
Üye	:.....	.....

.....

(Unvan, Adı ve Soyadı)

Bölüm Başkanı

…/…/20....

ETİK İLKE VE KURALLARA UYGUNLUK BEYANNAMESİ

Bu bitirme projesinin bana ait, özgün bir çalışma olduğunu; çalışmamın hazırlık, veri toplama, analiz ve bilgilerin sunumu olmak üzere tüm aşamalarında bilimsel etik ilke ve kurallara uygun davrandığımı; bu çalışma kapsamında elde edilen tüm veri ve bilgiler için kaynak gösterdiğimi ve bu kaynaklara kaynakçada yer verdiğimi; bu çalışmanın Samsun Üniversitesi tarafından kullanılan "bilimsel intihal tespit programı'yla tarandığını ve hiçbir şekilde "intihal içermediğini" beyan ederim. Herhangi bir zamanda, çalışmamla ilgili yaptığım bu beyana aykırı bir durumun saptanması durumunda, ortaya çıkacak tüm ahlaki ve hukuki sonuçları kabul ettiğimi bildiririm.

…….

(İmza)

Hasan Hüseyin Alav

ÖNSÖZ

Bu tez, “Ses Analiziyle Duygu Tanıma ve Dinamik Tepki Üretme: Sesli Asistan” üzerine yapılan çalışmayı içermektedir. Bu projenin temel amacı, gelişen yapay zeka teknolojisinin insan hayatına daha da indirgemek ve hayatın her alanında kullanıma sunmaktır. Bu süreçte çalışmada sesli asistan geliştirilmesinin teorik temelleri ve gerekli olan diğer teknolojilere yer verilmiştir. Bu çalışma bilgi, deneyim ve hayal gücünün bir ürünü olup benzer disiplinlerdeki çalışmalara temel oluşturma niteliğindedir.

Bu tezde sunulan bilgi ve sonuçlar, danışmanımın rehberliğinde ve değerli geri bildirimleriyle şekillenmiştir. Bu çalışmada desteğini esirgemeyen danışman hocam Sayın Doç. Dr. Zafer Cömert’e sonsuz teşekkürlerimi sunarım. Ayrıca duygu analizi eğitiminde kullandığım veri setini benimle paylaşarak bu tezi bitirmemde katkı sağlayan Salih Fırat Canpolat’a teşekkürlerimi sunarım.

Haziran 2024 Hasan Hüseyin Alav

İÇİNDEKİLER

ÖZET

SES ANALİZİYLE DUYGU TANIMA VE DİNAMİK TEPKİ ÜRETME: SESLİ ASİSTAN

Hasan Hüseyin Alav

Yazılım Mühendisliği Bölümü

Samsun Üniversitesi, Mühendislik ve Doğa Bilimleri Fakültesi

Haziran 2024

Danışman: Doç. Dr. Zafer Cömert

Bu lisans bitirme tezinde, "Ses Analiziyle Duygu Tanıma ve Dinamik Tepki Üretme: Sesli Asistan" konusu ele alınmaktadır. Çalışma kapsamında Google Text-To-Speech, Speech-To-Text, Generative AI modellerinden Gemini-Pro ve Elevenlabs.io gibi bileşenler kullanılarak bir sesli asistan geliştirilmiştir. Çalışmanın merkezinde yapay zeka tabanlı duygu analizi modülü yatmaktadır. Bu şekilde sesli asistan insana daha yakın bir şekilde konuşma girdisinden duygu durumunu anlayarak yanıtlar üretmektedir. Tezin ilk bölümünde kullanılan teknolojiler anlatılmıştır. Ardından sanal asistan sistemlerine dair literatür taraması gerçekleştirilmiştir. Daha sonra sesli asistan sisteminin geliştirilmesi modüllere ayrılarak parça parça anlatılmıştır. Kullanıcının sesli girdisinin işlenmesi ve duygu analizine tabii tutulması aktarılmıştır. Anlık duygu durumu çıkarıldıktan sonra Gemini-Pro modeline gönderilen girdi ve duygu durumuna göre çıktı sağlanmıştır. Alınan çıktı kullanıcıya Elevenlabs.io platformu sayesinde sentezlenerek sesli olarak iletilmiştir. Tezin devamında ise duygu analizi modeline ait veri seti hazırlığı, modelin eğitilmesi ve test edilmesi anlatılmıştır. Son bölümde ise geliştirilen sistemin sonuçlarından ve gelecek çalışmalarından bahsedilmiştir.

Anahtar Sözcükler: Duygu analizi, Generative AI, STT, TTS, Sesli asistan, Elevenlabs.io

ABSTRACT

EMOTION RECOGNITION AND DYNAMIC RESPONSE GENERATION THROUGH VOICE ANALYSIS: VOICE ASSISTANT

Hasan Hüseyin Alav

Department of Software Enginerring

Samsun University, Faculty of Engineering and Natural Sciences

January 2024

Supervisor: Assoc. Prof. Zafer Cömert

In this undergraduate thesis, the topic "Emotion Recognition and Dynamic Response Generation through Voice Analysis: Voice Assistant" is addressed. As part of the study, a voice assistant was developed using components such as Google Text-To-Speech, Speech-To-Text, Generative AI models like Gemini-Pro, and Elevenlabs.io. At the core of the work lies an AI-based emotion analysis module. This allows the voice assistant to produce responses by understanding the emotional state from the voice input, making it more human-like. The first section of the thesis explains the technologies used. This is followed by a literature review on virtual assistant systems. The development of the voice assistant system is then described in modular parts. The processing of the user's voice input and its subjection to emotion analysis is detailed. Once the real-time emotional state is determined, the input along with the emotional state is sent to the Gemini-Pro model to generate an output. This output is then synthesized into speech and delivered to the user via the Elevenlabs.io platform. Subsequent sections discuss the preparation of the dataset for the emotion analysis model, the training and testing of the model. The final section covers the results of the developed system and future work.

Keywords: Emotion analysis, Generative AI, STT, TTS, Voice assistant, Elevenlabs.io

TABLOLAR LİSTESİ

ŞEKİLLER LİSTESİ

SİMGELER VE KISALTMALAR

YZ : Yapay Zeka

API : Application Programming Interface

TTS : Text-to-Speech

AI : Artificial İntelligence

LTM : Long-Term-Memory

STM : Short-Term-Memory

BLIP : Blip Image Captioning Model

BÖLÜM 1: GİRİŞ

1.1. Araştırma Konusu ve Amaç

Günümüzde insanlar bilgiye erişmek için metin tabanlı yapay zeka (AI) sistemlerinden yoğun bir şekilde yararlanmaktadır. Ancak, bu asistanların çoğu insanlarla yalnızca metin biçiminde iletişim kurmaktadır. Ses tabanlı yapay zeka sistemlerinin sayısı nispeten azdır. Bu yapay zeka sistemleri kullanıcılarla doğal dilde iletişim kurabilmeli, ihtiyaçlarını anlayabilmeli ve onlara yardımcı olabilmelidir.

Bununla birlikte, mevcut yapay zeka tabanlı asistanların çoğu, kullanıcılarla iletişim kurarken kullanıcıların ruh halinin ve ruh durumunun farkında olmadıkları için kullanıcılara istenen sonucu vermekte zorluk çekmektedir. Bu da yanlış anlamalara, hatalı sonuçlara ve kullanıcı memnuniyetsizliğine yol açabilmektedir.

Bu sistem, yapay zeka sistemlerinin insan hayatında ne kadar önemli hale geleceği sorusunun cevabı olacak. "Yapay zeka sistemleri insani duygulara sahip olamaz" şeklindeki klasik iddianın yıkılmasında belki de ilk adım olacaktır. Eğer YZ sistemleri kullanıcılarıyla sesle iletişim kurarsa, kullanıcının YZ sistemine tek girdi sesi olacaktır. Bu sesin frekans ve spektral aralıklarının incelenmesi, kullanıcının o anda içinde bulunduğu duygunun belirlenmesi, bu duygunun benimsenmesi ve duygu durumuna uygun tepkiler verilmesi ile YZ sistemlerinin gelecekte çok daha fazla gelişmesi muhtemeldir.

1.2. Kullanılan Teknolojiler

Projede kullanılan başlıca teknolojiler şunlardır:

1. Speech Recognition (Speech to Text)
2. Google GenAI (Generative AI)
3. OpenCV (Open Source Computer Vision)
4. Blip Image Captioning Model
5. Doğal Dil İşleme (NLP)
6. ElevenLabs (Text to Speech)
7. Flask

1.2.1. Speech Recognition (Speech To Text)

Bu kütüphane, kullanıcının söylediği kelimeleri anlık olarak metne dönüştürmek için kullanılmıştır. Google'a ait STT API'si olan Google Cloud Speech-to-Text API'yi kullanarak konuşmalar metne dönüştürülmektedir.

1.2.2. Google GenAI (Generative AI)

Google Cloud tarafından hizmete sunulmuş Gemini gibi büyük Generative AI modellerini içeren çeşitli araçlar ve altyapılar sunan platformdur. Projenin temel bileşeni olan bu modül kullanıcı YZ etkileşimini gerçekleştiren kilit rolündeki bileşendir.

1.2.3. OpenCV (Open Source Computer Vision)

Bu kütüphane, görüntü işleme ve bilgisayarla görü işleme görevleri için kullanılır. Projede ise kullanıcının sesli komutu sayesinde aktif olacak şekilde kamera girişini almaktadır.

1.2.4. Blip Image Captioning Model

Salesforce tarafından geliştirilen **BLIP Image Captioning- Large**, görselleri analiz ederek metin açıklamaları üreten bir yapay zeka modelidir. Bu model, "Görsel Dil İşleme" (VLP) olarak adlandırılan bir alanda yer alır ve metin ve görseller arasındaki bağlantıyı kurarak çalışır (Github).

1.2.5. Doğal Dil İşleme (NLP)

Projede LTM mantığını sağlayabilmek adına kullanılan teknolojidir. Metin işleme ve kelimelerin benzerliğini kıyaslamak amacıyla kullanılmıştır.

1.2.6. ElevenLabs (Text to Speech)

Yapay zeka teknolojisini kullanarak ses sentezleme, ses klonlama gibi hizmetler için kullanılan platformdur. Sesli asistanın sesini klonlamak ve sentezlemek için TTS API sağlayıcısı kullanılmıştır (Elevenlabs.io).

1.2.7. Flask

Arka uç geliştirmesinde Flask API kullanılmıştır. Flask, Python tabanlı bir web uygulama geliştirme mikro çerçevesidir. Flask, basit ve esnek bir yapıya sahiptir ve RESTful API'lerin hızlı bir şekilde oluşturulmasını sağlar. Flask'in hafif yapısı, API'lerin hızlı bir şekilde dağıtılmasına ve yönetilmesine olanak tanır. Ayrıca, Flask'in geniş bir ekosistemi vardır ve çeşitli eklentilerle geliştirme sürecini destekler (Flask Documentation,

n.d.). Bu projede Flask API, ön uç ile veri iletişimi sağlamak ve yapay zeka modeli tarafından verilen sonuçları sunmak için kullanılmıştır.

BÖLÜM 2: LITETATÜR TARAMASI

2.1. Sanal Asistan Uygulamalarına Genel Bakış

Sanal asistanlar, kullanıcıların çeşitli görevleri yerine getirmelerine yardımcı olan yazılım uygulamalarıdır. Bu asistanlar kullanıcının sözlü veya yazılı komutlarını anlar ve doğal dil işleme (NLP) ve yapay zeka (AI) teknolojilerini kullanarak bunlara yanıt verir. Sanal asistanların ilk örnekleri 1960'larda geliştirilen ELIZA gibi basit sohbet robotlarıydı (Weizenbaum, 1966). Ancak modern sanal asistanların gelişimi, özellikle 2010'lu yıllarda uzun bir yol kat etti. Apple'ın Siri'si, Google'ın Asistanı, Amazon'un Alexa'sı ve Microsoft'un Cortana'sı gibi örnekler kullanıcılarla etkileşimi daha doğal ve etkili hale getirdi (Hoy, 2018; Lopatovska & Williams, 2018).

2.1.1. Çeşitli Kullanım Alanları

Sanal asistanlar çok çeşitli uygulamalarda kullanıcıların hayatını kolaylaştırıyor. Ev otomasyonunda Amazon Alexa ve Google Home gibi cihazlar akıllı ev cihazlarını kontrol etme, müzik çalma ve çevrimiçi bilgi sağlama gibi işlevleri yerine getirmektedir (Bentley et al., 2018). Kişisel asistan hizmetlerinde Siri ve Google Asistan, takvim yönetimi, hatırlatıcılar, e-postalar ve mesaj gönderme gibi günlük görevlerde kullanıcılara yardımcı oluyor (Hoy, 2018). İş dünyasında sanal asistanlar müşteri hizmetleri, veri analizi ve toplantı organizasyonu gibi alanlarda kullanılıyor; örneğin IBM'in Watson sanal asistanı, sektörler genelinde iş süreçlerini iyileştirmek için kullanılıyor (Ferrucci et al., 2013). Sağlık sektöründe, sanal asistanlar hastaların randevu almaları, semptomları takip etmeleri ve sağlık bilgileri sağlamaları için kullanılıyor. Önde gelen bir örnek olan ADA Health, kullanıcıların sağlık durumlarını değerlendirmelerine yardımcı olmaktadır (Munsch et al., 2019).

2.2. Konuşma Tanıma Teknolojileri

Konuşma tanıma teknolojileri, insan konuşmasını dijitalleştirerek anlayabilen ve işleyebilen sistemlerdir. Bu teknolojiler, kullanıcıların doğal dilini tanımak ve komutları yürütmek veya konuşmayı metne dönüştürmek için kullanılır. Konuşma tanıma teknolojilerinin temelleri 1950'lerde Bell Laboratories tarafından geliştirilen ilk otomatik konuşma tanıma sistemleri ile atılmıştır (Davis, Biddulph, & Balashek, 1952). O zamandan bu yana ve özellikle 2000'li yıllarda, bilgisayarların artan işlem gücü ve yapay zeka

alanındaki gelişmeler sayesinde bu teknolojiler önemli ölçüde gelişmiştir (Juang & Rabiner, 2005).

2.2.1. Teknolojik Bileşenler

Konuşma tanıma teknolojileri bir dizi gelişmiş teknolojik bileşene dayanmaktadır. Akustik modelleme, sesin özelliklerini tanımlayarak farklı fonetik birimleri tanımak için ses dalgalarını dijital verilere dönüştürme işlemidir (Huang, Acero, & Hon, 2001). Dil modellemesi, konuşma sırasında ortaya çıkan kelime ve cümle yapılarını anlamak için kullanılır; dil modellemesi, belirli bir bağlamda hangi kelimelerin ortaya çıkma olasılığının en yüksek olduğunu tahmin etmek için olasılık hesaplamalarını kullanır (Jelinek, 1998). Fonetik analiz, farklı aksan ve telaffuzların belirlenmesinde kilit rol oynayan insan sesindeki fonetik farklılıkları analiz ederek doğru kelime tanımayı sağlar (Rabiner & Juang, 1993). Doğal Dil İşleme (NLP), kullanıcıların konuşmalarını anlamak ve uygun yanıtlar üretmek için doğal dil yapısını analiz ederek konuşma tanıma teknolojilerinin kullanıcılarla daha doğal ve etkili bir şekilde iletişim kurmasını sağlar (Jurafsky & Martin, 2008).

2.2.2. Uygulama Alanları

Ses tanıma teknolojileri çok çeşitli uygulamalarda kullanılmaktadır. Mobil cihazlar ve kişisel asistanlar, Siri, Google Assistant ve Amazon Alexa gibi kişisel asistanları kullanarak kullanıcı komutlarını tanımak ve yürütmek için bu teknolojileri kullanır. Müşteri hizmetlerinde, otomatik çağrı merkezi sistemleri müşteri taleplerini tanıyarak ve yanıtlayarak hizmet kalitesini ve verimliliğini artırmaktadır. Otomotiv sektöründe, araç içi bilgi-eğlence ve navigasyon sistemleri, sürücülerin araç sistemlerini sesli komutlarla kontrol etmesini sağlar. Eğitim ve sağlık hizmetlerinde, konuşma tanıma teknolojileri öğrencilerin dil öğrenmelerine yardımcı olurken, sağlık hizmetlerinde doktorların hasta bilgilerini konuşma kullanarak kaydetmelerini ve yönetmelerini sağlar.

2.3. Doğal Dil İşleme (NLP) Teknikleri

Doğal Dil İşleme (NLP), bilgisayarların insan dilini anlamasını, yorumlamasını ve yaratmasını sağlamak için tasarlanmış yapay zeka teknolojilerinin bir bileşenidir. NLP, metin ve konuşma verilerini analiz ederek bilgisayar sistemlerinin insanlarla daha doğal ve etkili bir şekilde iletişim kurmasını sağlar (Jurafsky & Martin, 2008). Bu teknolojinin önemi, dijital asistanlardan müşteri hizmetlerine, sağlık hizmetlerinden dil öğrenimine

kadar çok çeşitli alanlarda uygulanabilmesinde yatmaktadır (Chowdhury, 2003; Hirschberg & Manning, 2015).

2.3.1. Teknik Bileşenler

Doğal dil işleme (NLP) bir dizi bileşen ve tekniğe dayanır. Metin ön işleme, metin verilerinin analiz için hazırlanmasını içerir. Bu aşamada veri temizleme, boş kelimelerin çıkarılması, kesme ve budama gibi işlemler gerçekleştirilir (Manning, Raghavan, & Schütze, 2008). Dilbilimsel modelleme, doğal dilin yapısını ve gramerini anlamak için dilbilimsel modeller kullanır ve bu modeller belirli bir dildeki kelimelerin ve cümlelerin yapılarını istatistiksel olarak temsil eder (Jurafsky & Martin, 2008). Metni kelimeler veya kelime öbekleri gibi daha küçük anlamlı birimlere ayırmak, metin analizindeki temel adımlardan biridir (Manning et al., 2008).

Adlandırılmış varlık tanıma, metinde bahsedilen kişiler, yerler, kuruluşlar vb. gibi belirli varlıkları tanımlar ve metnin anlamını ve bağlamını anlamada önemli bir rol oynar (Nadeau & Sekine, 2007). Duygu analizi, metinlerin duygusal içeriğini tanımlar ve bunları olumlu, olumsuz veya tarafsız olarak kategorize eder. Genellikle müşteri geri bildirimlerini ve sosyal medyayı analiz etmek için kullanılır (Liu, 2012). Makine çevirisi, bir dilden diğerine otomatik olarak çeviri yapma sürecidir. Google Translate gibi uygulamalar hızlı ve doğru çeviriler sağlamak için bu teknolojiyi kullanır (Bahdanau, Cho, & Bengio, 2015). Metin özeti ise uzun metinleri daha kısa bir biçimde özetlemek, büyük miktarda veriyi hızlı bir şekilde anlamak için kullanışlıdır (Nenkova & McKeown, 2012).

2.4. Generative AI Modelleri ve Uygulamaları

Generative AI modelleri, belirli bir veri kümesinden öğrenerek yeni ve özgün içerik üretebilen yapay zeka sistemleridir. Bu modeller, görüntü, metin, ses ve diğer veri türlerini kullanarak yaratıcı ve yenilikçi sonuçlar elde edebilir. Generative modellerin temelini, veri örüntülerini anlamak ve bu örüntülere dayanarak yeni örnekler üretmek oluşturur (Goodfellow et al., 2014). Bu alandaki en bilinen modellerden biri, Derin Öğrenme tekniklerine dayanan Generative Adversarial Networks (GANs) ve Otomatik Kodlayıcılar (Autoencoders) gibi yapılardır (Kingma & Welling, 2013; Radford, Metz, & Chintala, 2015).

2.4.1. Teknolojik Bileşenler

Generative AI modelleri, çeşitli ileri teknolojik bileşenlerden oluşur. Generative Adversarial Networks (GANs), iki neural ağın (üretici ve ayrımcı) rekabetçi bir şekilde eğitildiği bir yapıya sahiptir. Üretici ağ, gerçekçi veri örnekleri oluşturmaya çalışırken, ayrımcı ağ ise bu örneklerin gerçek mi yoksa sahte mi olduğunu belirlemeye çalışır (Goodfellow et al., 2014). Varyasyonel Otomatik Kodlayıcılar (VAEs), verilerin gizli temsillerini öğrenerek yeni veri örnekleri oluşturabilen bir tür derin öğrenme modelidir. Bu modeller, verilerin olasılıksal dağılımlarını öğrenir ve yeni örnekler üretir (Kingma & Welling, 2013). Dönüştürücü modeller (Transformers), özellikle metin ve dil işlemede yaygın olarak kullanılan generative modellerdir. Bu modeller, büyük dil modelleri (örneğin, GPT-3) oluşturarak doğal dilde anlamlı ve tutarlı metinler üretebilir (Vaswani et al., 2017; Brown et al., 2020).

2.4.2. Uygulama Alanları

Generative AI modelleri, birçok farklı alanda yenilikçi uygulamalara sahiptir. Görsel sanatlar ve tasarım alanında, generative modeller sanat eserleri, grafik tasarımlar ve 3D modeller oluşturmak için kullanılır. Örneğin, DeepArt ve DALL-E gibi araçlar, kullanıcıların metin açıklamalarına dayanarak yeni görseller üretebilir (Elgammal et al., 2017; Ramesh et al., 2021). Metin üretimi ve dil işleme alanında, büyük dil modelleri doğal ve tutarlı metinler oluşturarak içerik üretimi, sohbet botları ve dil çevirisi gibi alanlarda kullanılır. GPT-3, haber makaleleri, hikayeler ve hatta kod yazma gibi çeşitli metin tabanlı görevlerde etkili bir şekilde kullanılabilir (Brown et al., 2020). Müzik ve ses üretiminde, generative modeller yeni müzik parçaları ve ses efektleri oluşturmak için kullanılır. OpenAI'nin Jukedeck ve MuseNet gibi projeleri, farklı müzik stillerinde ve enstrümanlarda yeni parçalar üretebilir (Sturm et al., 2016; Payne, 2019). Oyun ve simülasyon alanında ise generative modeller rastgele haritalar, karakterler ve senaryolar oluşturmak için kullanılır, bu da oyun geliştiricilerinin daha çeşitli ve dinamik içerikler sunmasını sağlar (Summerville et al., 2018).

2.5. Duygu Analizi Yöntemleri ve Teknikleri

Duygu analizi, metin, görüntü veya ses gibi veri türlerinde ifade edilen duygusal içeriği tanımlama ve sınıflandırmaya yönelik bir yapay zeka teknolojisidir. Bu teknoloji, belirli bir içeriğin olumlu, olumsuz veya nötr duygular içerip içermediğini belirlemek için kullanılır. Duygu analizi, duygusal içeriğin tespit edilmesi ve anlaşılması amacıyla doğal

dil işleme (NLP), makine öğrenimi ve derin öğrenme gibi teknikleri içerir (Liu, 2012; Cambria, 2016).

2.5.1. Teknolojik Bileşenler

Duygu analizi, bir dizi teknik bileşen ve yönteme dayanır. Doğal Dil İşleme (NLP) teknikleri, duygu analizinde metin verilerinin kullanılmasıyla önem kazanır. Bu teknikler, metnin anlamını çıkarmak ve duygusal içeriği anlamlandırmak için kullanılır (Manning, Raghavan, & Schütze, 2008). Makine öğrenimi modelleri ise duygu analizi için eğitilmiş veri kümelerini kullanarak metin verilerini etiketlenmiş duygusal kategorilere (olumlu, olumsuz, nötr) sınıflandırmak için kullanılır (Liu, 2012). Ayrıca, duygu analizi için kullanılan duygu sözlükleri ve kaynaklar, kelimelerin duygusal yoğunluğunu ve anlamlarını belirleyerek metindeki duygusal ifadelerin tanımlanmasına yardımcı olur (Taboada et al., 2011).

Bu bileşenler, duygu analizinin temelini oluşturarak, metinlerin duygusal içeriğini belirlemede ve analiz etmede kritik rol oynar. NLP teknikleri, metinleri anlamlandırarak duygusal bağlamı çıkarırken, makine öğrenimi modelleri bu bilgiyi kullanarak metinleri doğru kategorilere ayırır. Duygu sözlükleri ve kaynaklar ise, kelimelerin duygusal yoğunluklarını belirleyerek daha hassas ve doğru analizler yapılmasını sağlar.

2.5.2. Uygulama Alanları

Duygu analizi, birçok farklı alanda kullanılmaktadır. Sosyal medya ve pazar araştırmalarında, sosyal medya platformlarındaki kullanıcı yorumları ve paylaşımların duygusal içeriği analiz edilerek markaların ve ürünlerin popülerliği ve algısı anlaşılır (Ghiassi, Skinner, & Zimbra, 2013). Müşteri hizmetleri ve geri bildirim yönetiminde ise müşteri geri bildirimleri ve şikayetlerin duygusal tonu analiz edilerek şirketlerin müşteri memnuniyetini değerlendirmesi ve hizmetlerini iyileştirmesi sağlanır (Yelpaze & Alpaydın, 2021). Sağlık sektöründe, hastaların sağlık durumları ve duygusal durumları hakkında yapılan veri analizleri, sağlık profesyonellerinin daha iyi bir tedavi ve destek sağlamalarına olanak tanır (Denecke & Deng, 2015). Siyasi analiz ve kamuoyu araştırmalarında da, siyasi konular ve kampanyalarla ilgili duygusal tepkilerin analizi, seçmen davranışlarını anlamak ve politik stratejiler oluşturmak için kullanılır (Stieglitz & Dang-Xuan, 2013).

BÖLÜM 3: SISTEM TASARIMI VE GELIŞTIRME

3.1. Sistem Mimarisine Genel Bakış

Genel İş Akışı

Sistem mimarisinin genel iş akışı şu şekildedir:

- Kullanıcı, mikrofon aracılığıyla sesli bir komut verir.
- Google Cloud Speech-to-Text hizmeti, sesli komutu metne dönüştürür.
- Ses, duygu analizi için analiz edilir.
- Google Generative AI, duygu analizine göre uygun metin yanıtını üretir.
- ElevenLabs veya Google TTS hizmeti, metni sesli yanıt haline getirir.
- Sesli yanıt kullanıcıya iletilir.

Bu mimari, kullanıcıdan alınan sesli girdilerin etkili bir şekilde işlenmesini, anlamlı ve duygusal olarak uygun yanıtların oluşturulmasını ve bu yanıtların kullanıcıya iletilmesini sağlar. Sistem bileşenlerinin birbirleriyle uyumlu çalışması, sesli asistanın işlevselliğini ve kullanıcı memnuniyetini artırır.

3.1.1. Kullanıcı Girdisi

Sistem, kullanıcıdan alınan sesli komutlarla etkileşime girer. Kullanıcı, mikrofon aracılığıyla sesli bir komut verir ve bu sesli girdi, Google Cloud Speech-to-Text (STT) hizmeti kullanılarak metne dönüştürülür. Bu adım, sesli komutun doğru ve hızlı bir şekilde metin formatına çevrilmesi için kritik öneme sahiptir.

3.1.2. Duygu Analizi

Kullanıcının sesli komutu duygu analizine tabi tutulur. Bu adımda, kullanıcının verdiği komutun duygusal tonu belirlenir. Duygu analizi, komutun içerdiği kelime ve cümle yapılarından yola çıkarak komutun “sinirli”, “üzgün”, “mutlu” veya “sakin” bir duygu taşıyıp taşımadığını saptar. Bu bilgi, yanıtın duygusal tonunun belirlenmesinde kullanılır.

3.1.3. Metin Üretimi

Duygu analizinin sonucuna göre, kullanıcının komutuna uygun bir metin yanıtı oluşturulur. Bu işlem, Google Generative AI gibi doğal dil işleme modelleri kullanılarak gerçekleştirilir. Üretilen metin yanıtı, kullanıcının sorusuna anlamlı ve bağlama uygun bir şekilde cevap verir.

3.1.4. Sesli Yanıt Üretimi

Oluşturulan metin yanıtı, sesli bir yanıt haline getirilir. Bu adımda, ElevenLabs ve Google Text-to-Speech (TTS) gibi hizmetler kullanılarak metin, kullanıcıya iletilecek ses formatına dönüştürülür. Yanıtın duygusal tonu, duygu analizinin sonuçlarına göre ayarlanır. Örneğin, üzgün bir komut için daha sakin bir ses tonu, sakin bir komut için ise doğal bir ses tonu kullanılır.

3.1.5. Yanıtın Kullanıcıya İletilmesi

Son olarak, üretilen sesli yanıt kullanıcıya geri iletilir. Bu süreç, kullanıcının komutuna verilen yanıtın tamamlanmasını sağlar ve kullanıcı deneyimini zenginleştirir. Sistem, kullanıcı ile etkili bir iletişim kurarak sesli asistan işlevini yerine getirir.

3.2. Konuşma Tanıma

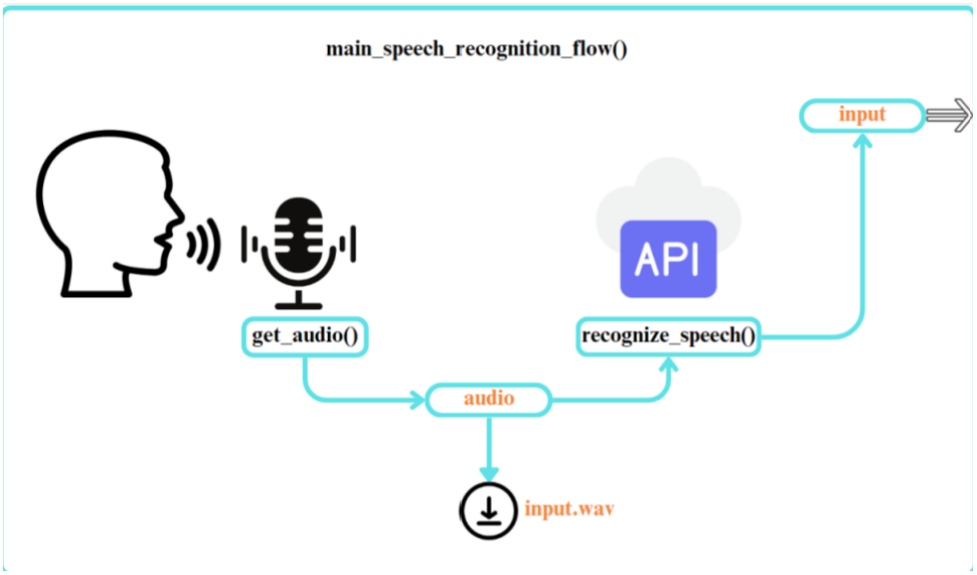
Sanal asistan uygulamasının temel bir bileşeni olan konuşma tanıma modülü, kullanıcının sesli girdisini metne dönüştürmekten sorumludur. Doğal ve etkileşimli bir kullanıcı deneyimi sunmak için bu modül kritik öneme sahiptir. Projede, **SpeechRecognition** modülü ile birlikte cihaz mikrofonuna erişim sağlanmış ve mikrofon aracılığıyla gelen veriler Google Cloud Speech-to-Text API sağlayıcısına gönderilerek kullanıcının sesli girdisi metne dönüştürülmüştür.

3.2.1. Mikrofon Erişimi

Sistemin kullanıcıdan sesli komut alabilmesi için mikrofon erişimi gereklidir. **SpeechRecognition** modülü, mikrofonu kullanarak sesli girdiyi alır ve bu veriyi işlemeye hazır hale getirir. Bu süreçte, mikrofon başlatılarak kullanıcıdan gelen sesli girdi dinlenir ve kaydedilir.

3.2.2. Konuşmayı Metne Dönüştürme

Kullanıcıdan alınan sesli girdi, Google Cloud Speech-to-Text API kullanılarak metne dönüştürülür. Bu işlem, kullanıcının konuşma komutlarını anlamak ve işlem yapmak için gereklidir. Google Cloud Speech-to-Text API, sesli girdiyi analiz ederek metin çıktısı üretir. Bu adımda, API'ye sesli girdi gönderilir ve yanıt olarak metin alınır.



Şekil 1 Mikrofon Erişimi ve Metine Dönüştürme

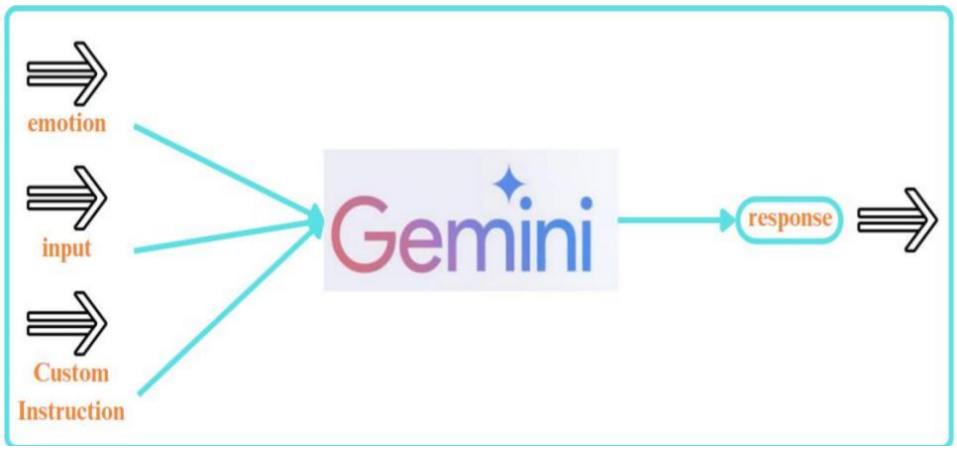
3.2.3. Hata Yönetimi ve Geri Bildirim

Sistem, konuşma tanıma sırasında oluşabilecek hataları yönetmek ve kullanıcıya geri bildirim sağlamak için tasarlanmıştır. Hata yönetimi, kullanıcı deneyimini iyileştirmek ve sistemin güvenilirliğini artırmak için kritik öneme sahiptir. Örneğin, sesli girdi anlaşılamadığında veya API hizmetine erişilemediğinde kullanıcıya durumu bildiren geri bildirimler verilir. Bu geri bildirim mekanizması, kullanıcıların sistemle etkili bir şekilde etkileşimde bulunmasını sağlar ve olası sorunlar hakkında bilgilendirir.

3.3. Generative AI Modeli Entegrasyonu

Generative AI modelleri; metin üretme, dilleri çevirme ve soruları yanıtlama gibi çeşitli görevler için kullanılabilen güçlü araçlardır. Bu projede, kullanıcı sorgularına doğal ve anlamlı yanıtlar üretmek için bir Generative AI modeli sanal asistan uygulamasına entegre edilmiştir.

Projede generative AI modellerinden olan ve Türkçe dilini destekleyen ve kullanıcı sorgularına ilişkin görevleri yerine getirebilecek kadar güçlü ve tutarlı olan GEMİNİ-PRO modeli seçilmiştir.



Şekil 2 Gemini-Pro Entegresi

Gemini-Pro, Python API aracılığıyla sanal asistan uygulamasına entegre edilmiştir. Bu API, modelin kullanıcının sorgusunu almasını, işlemesini ve bir yanıt üretmesini sağlar. API, generate_content adında bir fonksiyon sunar ve bu fonksiyon, sorgunun metnini ve modelin üretmesi gereken yanıtın türünü parametre olarak alır.

3.4. Hafıza Yönetimi

Sanal asistan uygulamaları, kullanıcılarla etkileşim kurarken ve görevleri yerine getirirken çeşitli bilgileri depolamak ve yönetmek zorundadır. Bu bilgiler, kullanıcı profilleri, geçmiş sorgular, bağlamsal bilgiler ve dünya hakkındaki genel bilgilere kadar uzanabilir. Hafıza yönetimi, bu bilgilerin etkili ve verimli bir şekilde depolanmasını ve erişilebilmesini sağlayan önemli bir bileşendir.

Ancak generative AI API sağlayıcılarında herhangi bir hafıza yönetimi bulunmamaktadır. Gönderilen her bir sorgu sanki sıfırdan gönderiliyormuş gibi çalışmaktadır. Bu görevi yerine getirebilmek adına kullanıcının sanal asistanla gerçekleştirdiği tüm konuşmaları ve yanıtları proje dizininde bulunan DataSet.json adlı dosyada belirlenen formata uygun şekilde depolanmaktadır. Gerekli olan tüm hafıza yönetim işlemleri bu dosya aracılığı ile sağlanmaktadır. Ayrıca kullanıcı isterse bu dizini gireceği sesli komut ile sıfırlayabilmekte ve temiz bir başlangıç yapabilmektedir.

3.4.1. STM

Kısa süreli hafıza (STM), kullanıcıyla yapılan anlık etkileşimle ilgili bilgileri depolamak için kullanılan geçici bir hafıza sistemidir. Kısa süreli hafıza görevini basit bir mantıkla kullanıcı ile AI arasında geçen son beş sorgu ve yanıt DataSet.json dosyasından alınarak anlık girilen sorgu ile birlikte “Geçmiş Konuşmalar + Sorgu” şeklinde

gönderilmekte ve bu sayede AI modeli güncel sorguyu son beş sorgu ve yanıt ile birlikte değerlendirebilmektedir.

3.4.2. LTM

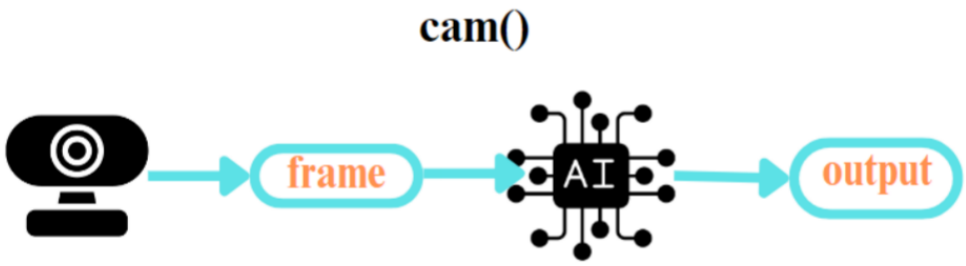
LTM, kullanıcı profilleri, geçmiş sorgular, kelime dağarcığı ve genel bilgiler gibi daha kalıcı bilgileri depolamak için kullanılır. Projede ise STM algoritması aksine DataSet.json dosyasının tamamını NLP ile birlikte geliştirilen bir algoritma sayesinde LTM mantığı işlenmiştir.

Geliştirilen algoritma, kullanıcının sorgusuna benzer geçmiş sorguları tarayıp en fazla üç tane olmak üzere seçmektedir. Algoritmada her bir kelimeye puanlama sistemi eklenerek güncel sorguya en yakın kelimeler puanlanmaktadır. Daha sonra puanlanan bu kelimeler birbirleri ile karşılaştırılarak en benzer üç sorgu API sağlayıcısına gönderilecek olan sorguya dahil edilmektedir. Bu sayede generative AI, kullanıcının sorgusunu daha iyi anlayabilmekte ve daha alakalı cevaplar üretilebilmektedir.

3.5. Kamera ile Nesne Tespit Modülü

Kamera ile nesne tespit modülü, sanal asistan uygulamasına işlevsellik ve erişilebilirlik katan değerli bir ekleme sağlar. Bu modül, görsel etkileşim ve bilgi edinme imkanı sunarak kullanıcı deneyimini önemli ölçüde geliştirme potansiyeline sahiptir. Ayrıca bu modül, sesli asistan uygulamasında kullanıcının çevresindeki nesneleri tanımlamak ve bu nesneler hakkında bilgi vermek amacıyla kullanılacaktır.

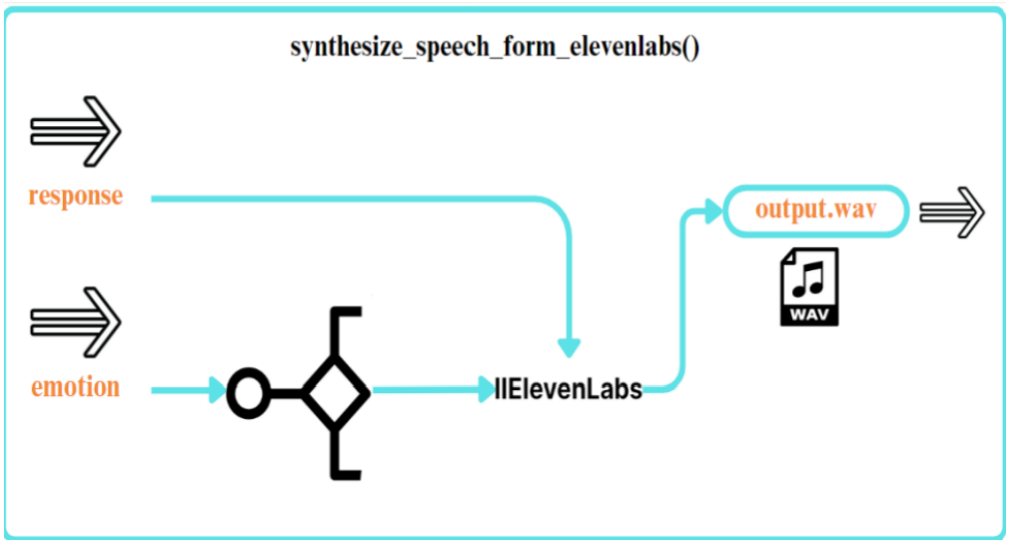
Projede bu kısım kameradan alınan anlık bir frame görüntüsünü BLIP modeli ile işleyerek metin üretimi sağlanmaktadır. Alınan metin kamera modeline özel bir Custom_Instruction ile birlikte AI modeline gönderilmektedir. AI modeli ise anlık kamera frame görüntüsünü işleyerek bir çıkarım oluşturmaktadır.



Şekil 3 Kamera Modülü Şematik Anlatım

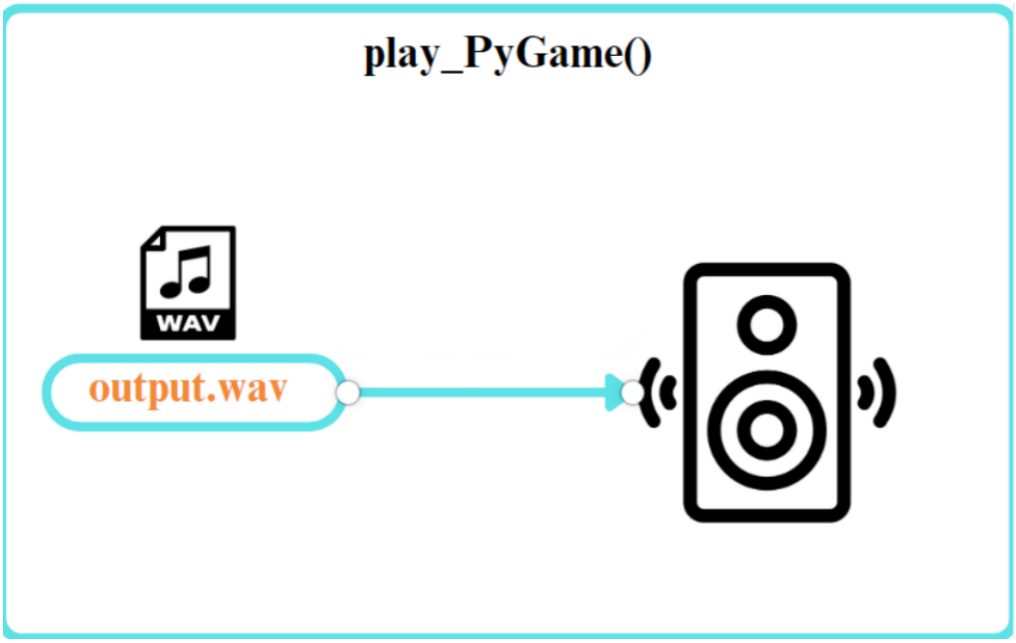
3.6. Metinden Ses Dönüştürme

Sanal asistan uygulamasının olmazsa olmazlarında olan bu modül Elevenlabs.io platformu sayesinde geliştirilmiştir. Elevenlabs.io ile ses klonlama işlemi daha önceden yapılmıştır. Bu sayede yapay robot sesinden arındırılmış gerçek bir insan (Scarlett Johansson) sesi kullanılmıştır.



Şekil 4 Ses Sentezleme

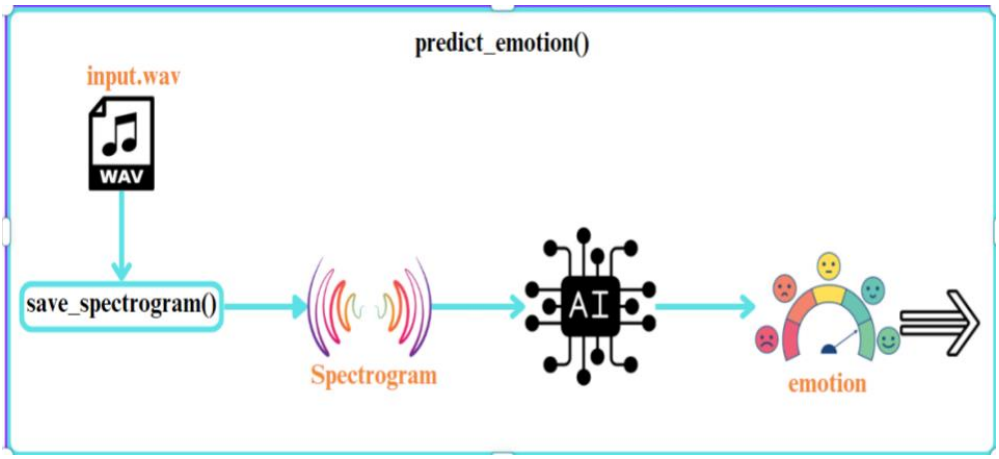
Elevenlabs API sağlayıcısı sayesinde sentezlenen sesli çıktılar daha sonra kullanıcıya iletilir.



Şekil 5 Sesin Kullanıcıya İletilmesi

3.7. Duygu Analizi

Projenin duygu analizi modülü, kullanıcının sesli girdilerini analiz ederek duygu durumlarını belirlemektedir. Bu modül, ses dosyalarını görüntüye dönüştüren spektrogramlar kullanarak çalışan bir Evrişimsel Sinir Ağı (CNN) modeline dayanmaktadır. Model, farklı duyguları doğru bir şekilde sınıflandırmak için eğitilmiştir ve kullanıcının duygu durumuna uygun tepkiler verebilmektedir. Bu sayede, sesli asistan daha doğal ve insancıl etkileşimler sağlayarak kullanıcı deneyimini geliştirmektedir.



Şekil 6 Duygu Analizi

3.8. Hatırlatıcı Modülü

Bu modül, kullanıcıların önemli olayları ve yapılacak işleri hatırlamalarına yardımcı olmak için ve işlevsellik katmak amacıyla tasarlanmıştır.

Sanal asistana yapılan özel bir komut sayesinde bu modül aktif edilerek hatırlatıcı tarih- saati ve hatırlatıcı bilgisi istenmektedir. Kullanıcının girdiği bu bilgiler proje dizininde yer alan reminder.json dosyasına belirlenen formatta kaydedilerek tutulmaktadır. Yine kullanıcıdan alınacak özel bir komut ile kaydedilen hatırlatıcılar sesli olarak kullanıcıya bildirilmektedir.

3.9. Kullanıcı Arayüzü Tasarımı

Projenin arayüz tasarımı, kullanıcı deneyimini en üst düzeye çıkarmak amacıyla sade ve kullanıcı dostu bir şekilde geliştirilmiştir. Arayüz, kullanıcıların kolaylıkla etkileşime girebileceği şekilde düzenlenmiş, modern ve minimal bir tasarım benimsenmiştir. Ana ekran, kullanıcıların sesli komutlarını girebilecekleri bir mikrofon simgesi ve çeşitli ayarlara erişim sağlayan ikonlarla donatılmıştır. Kullanıcı girdileri, duygu analizleri ve yanıtların görsel olarak takip edilebilmesi için şeffaf bilgi kutuları kullanılmıştır. Arayüz, duyarlılık ayarları, özel komutlar ve hatırlatıcı yönetimi gibi özellikleri de içermektedir. Tüm bu bileşenler, kullanıcıların sesli asistan ile etkileşimini doğal ve sezgisel hale getirmektedir.

Şekil 7 Kullanıcı Arayüzü

BÖLÜM 4: DUYGU ANALİZİ EĞİTİM VE DEĞERLENDİRME

4.1. Veri Seti

Projede duygu analizi yapmak için kullanılan veri seti, "Turkish Emotion Voice Database" (TurEV-DB) olarak adlandırılmaktadır. Bu veri seti, Türkçe konuşma verilerinden oluşan ve farklı duygusal durumları içeren kapsamlı bir veri tabanıdır.

4.1.1. Veri Setinin Özellikleri

- **Duygular:** Veri seti, dört temel duyguyu (kızgın, üzgün, sakin ve mutlu) içeren ses kayıtlarını barındırmaktadır. Bu kayıtlar, farklı konuşmacılardan elde edilmiş ve her duygu için birçok örnek içermektedir.
- **Ses Kayıtları:** Her duygu durumu için toplanan ses kayıtları, yüksek kaliteli mikrofonlar kullanılarak kaydedilmiştir. Bu kayıtlar, veri setinin doğruluğunu ve güvenilirliğini artırmak amacıyla dikkatle seçilmiştir.
- **Spektrogramlar:** Ses dosyaları, makine öğrenmesi modellerinin daha iyi analiz edebilmesi için spektrogramlara dönüştürülmüştür. Bu spektrogramlar, ses sinyallerinin frekans bileşenlerini zaman ekseninde görselleştirmektedir.

4.1.2. Veri Setinin Kullanımı

TurEV-DB, duygu analizi modellerinin eğitimi ve test edilmesi için kullanılmıştır. Ses kayıtları spektrogramlara dönüştürülerek görüntü tabanlı analiz yapılmış ve bu görüntüler Evrişimsel Sinir Ağları (CNN) kullanılarak işlenmiştir. Model, her bir duygu durumu için yüksek doğruluk oranları elde etmiştir ve çeşitli metriklerle performansı değerlendirilmiştir.

Bu veri seti, Türkçe konuşma verileri üzerinde duygu analizi çalışmaları yapmak isteyen araştırmacılar ve geliştiriciler için değerli bir kaynak sağlamaktadır.

4.2. Model Eğitimi

4.2.1. Verilerin Hazırlanması

Ses verilerini işlemek ve modele uygun hale getirmek için ses dosyaları spektrogram görüntülerine dönüştürülmüştür. Spektrogramlar, zaman ve frekans bilgilerini bir arada sunarak ses sinyallerinin görsel temsilini sağlar. Bu, makine öğrenmesi modellerinin ses sinyallerini analiz etmesini kolaylaştırır. "*save_spectrogram"* fonksiyonu, ses dosyalarını yükleyip frekans bilgilerini çıkararak spektrogram görüntülerini oluşturur ve bu görüntüleri PNG formatında kaydeder. PNG formatı, verinin görsel olarak sıkıştırılmasını ve daha sonra kolayca yüklenmesini sağlar.

Her bir duygu için "*process_audio_files"* fonksiyonu, ilgili klasördeki tüm ses dosyalarını işleyerek spektrogramlarını kaydeder. Bu işlem, her duygu için ayrı bir klasörde saklanır.

4.2.2. Verilerin Yüklenmesi ve Ön İşleme

Spektrogram görüntüleri yüklendikten sonra, her görüntü gri tonlamaya çevrilmiş ve 128x128 piksel boyutunda yeniden boyutlandırılmıştır. Bu işlem, verinin işlenmesini ve modele beslenmesini kolaylaştırır. "*load_data_and_labels"* fonksiyonu, tüm spektrogram görüntülerini yükler ve ilgili duygu etiketleriyle birlikte döner. Etiketler sayısal değerlere dönüştürülerek one-hot encoding yöntemiyle işlenmiştir.

4.2.3. Model Mimarisi

Model, üç katmanlı bir Evrişimsel Sinir Ağı (CNN) olarak tasarlanmıştır. CNN'ler, görüntü verilerini işlemek için çok uygundur çünkü yerel bağlantıları ve paylaşılmış ağırlıkları kullanarak öğrenme yeteneğine sahiptirler. Üç katmanlı bir CNN kullanmanın nedeni, modelin yeterince karmaşık olmasını sağlarken aşırı uyumdan (overfitting) kaçınmaktır. İlk katman, temel kenar ve köşe gibi düşük seviye özellikleri öğrenirken, orta katman daha karmaşık şekilleri ve desenleri tanır. Son katman ise yüksek seviye özellikleri ve belirgin duygu işaretlerini tanır.

Modelde her bir evrişim katmanı (Conv2D) ardından bir normalizasyon (BatchNormalization), havuzlama (MaxPooling2D) ve dropout katmanı bulunmaktadır. Bu yapı, modelin genelleme yeteneğini artırır ve aşırı uyumdan korunmasını sağlar. Softmax çıkış katmanı, her sınıfa bir olasılık değeri atar ve toplamı 1 olan olasılıklar döndürür. Bu, sınıflandırma problemleri için idealdir çünkü hangi sınıfın en yüksek olasılıkla doğru olduğunu belirlemeye yardımcı olur.

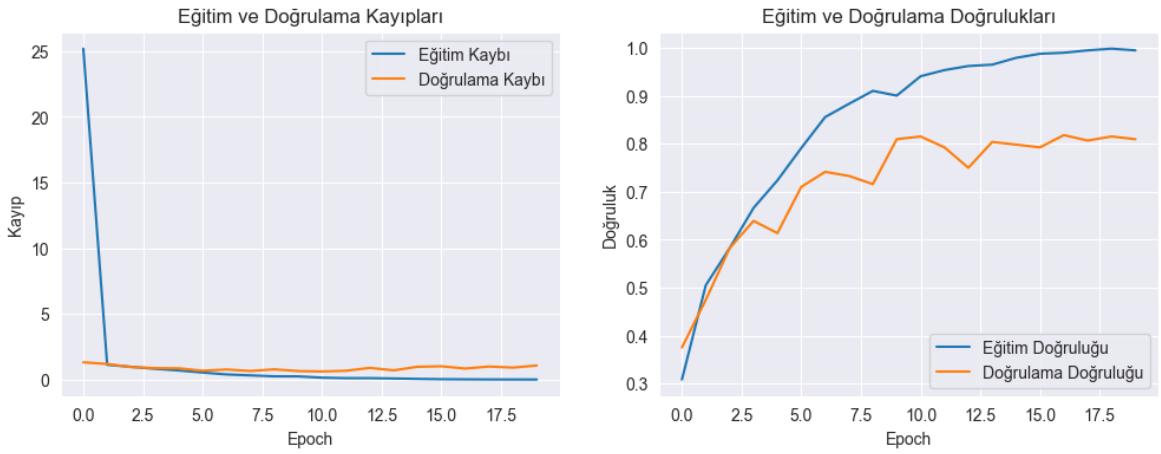
4.2.4. Modelin Derlenmesi ve Eğitimi

Model, Adam optimizasyon algoritması kullanılarak derlenmiştir. Kayıp fonksiyonu olarak kategorik çapraz entropi (categorical crossentropy) kullanılmıştır. Bu fonksiyon, modelin farklı sınıflar arasındaki farkı öğrenmesini sağlar. Model, 30 epoch boyunca eğitim verileri üzerinde eğitilmiştir.

4.3. Performans Değerlendirme ve Analizi

4.3.1. Eğitim ve Doğrulama Kayıpları ve Doğrulukları

Eğitim ve doğrulama kayıplarını ve doğruluklarını gösteren grafikler, modelin eğitim süreci boyunca nasıl performans gösterdiğini özetlemektedir. Aşağıdaki grafikler, eğitim ve doğrulama kayıplarının ve doğruluklarının epoch'lar boyunca nasıl değiştiğini göstermektedir:

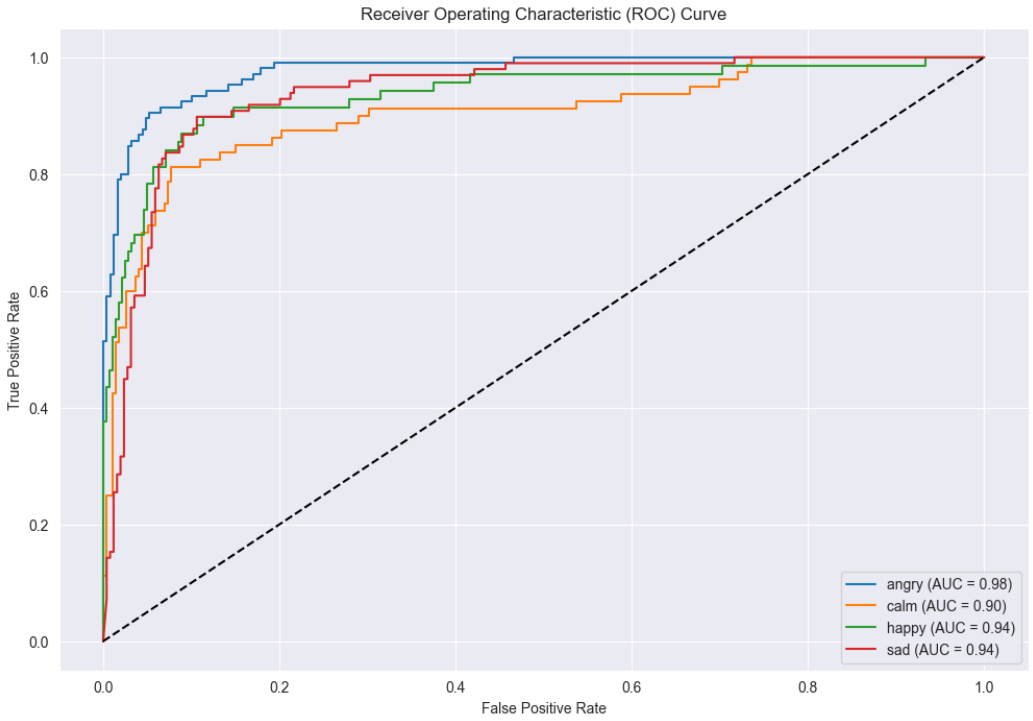


Şekil 8. Eğitim Ve Doğrulama Grafiği

Eğitim kaybı ve doğrulama kaybı, modelin ilk birkaç epoch'da hızla azaldığını ve daha sonra stabil hale geldiğini göstermektedir. Eğitim doğruluğu ve doğrulama doğruluğu ise benzer şekilde hızla artarak yüksek seviyelere ulaşmıştır. Bu durum, modelin hem eğitim verileri üzerinde iyi performans gösterdiğini hem de genelleme yeteneğinin güçlü olduğunu göstermektedir.

4.3.2. ROC Eğrisi ve AUC Değerleri

Modelin farklı duygu sınıflarındaki performansını değerlendirmek için ROC (Receiver Operating Characteristic) eğrileri ve AUC (Area Under Curve) değerleri kullanılmıştır. Aşağıdaki grafik, her bir duygu sınıfı için ROC eğrilerini göstermektedir:



Şekil 9. ROC Eğrisi

ROC eğrileri, modelin her bir duygu sınıfı için yüksek performans gösterdiğini ortaya koymaktadır. Özellikle "angry" sınıfı için AUC değeri 0.98, "happy" ve "sad" sınıfları için 0.94 ve "calm" sınıfı için 0.90 olarak hesaplanmıştır. Bu yüksek AUC değerleri, modelin her bir duygu sınıfında oldukça başarılı olduğunu göstermektedir.

4.3.3. Sınıflandırma Raporu

Modelin performansı ayrıca precision, recall ve f1-score gibi metriklerle de değerlendirilmiştir. Aşağıdaki sınıflandırma raporu, modelin her bir duygu sınıfı için bu metrikler üzerindeki performansını göstermektedir:

	precision	recall	f1-score	support
Angry	0.98	0.84	0.90	97
Calm	0.79	0.78	0.79	88
Happy	0.88	0.89	0.89	74
Sad	0.79	0.90	0.84	100
Accuracy			0.85	359
Macro avg	0.86	0.85	0.85	359
Weighted avg	0.86	0.85	0.85	359

Tablo 1 Sınıflandırma Raporu

Bu rapor, modelin genel doğruluk oranının %85 olduğunu ve her bir sınıf için f1-score değerlerinin oldukça yüksek olduğunu göstermektedir. Özellikle "angry" ve "happy" sınıfları için f1-score değerleri sırasıyla 0.90 ve 0.89 ile oldukça yüksektir. "calm" ve "sad" sınıfları için de f1-score değerleri sırasıyla 0.79 ve 0.84'tür.

Genel olarak, sesli asistan projesinde kullanılan duygu analizi modeli, yüksek doğruluk oranları ve AUC değerleri ile başarılı bir performans sergilemiştir. Eğitim ve doğrulama süreçlerindeki düşük kayıp değerleri ve yüksek doğruluk oranları, modelin hem eğitim hem de doğrulama verileri üzerinde iyi performans gösterdiğini ve genelleme yeteneğinin güçlü olduğunu göstermektedir. ROC eğrileri ve sınıflandırma raporu da modelin her bir duygu sınıfında yüksek performans sergilediğini doğrulamaktadır.

BÖLÜM 5: SONUÇ VE TARTIŞMA

5.1. Projenin Önemi ve Katkıları

Sanal asistanlar, günlük hayatı kolaylaştırır ve erişilebilirliği artırır. Günlük yaşamı kolaylaştırır; hava durumu bilgisi alma, hatırlatıcılar gibi günlük görevleri hızlı ve kolay bir şekilde halletmelerine imkan tanır. Erişilebilirliği artırır; görme engelli veya işitme engelli gibi özel gereksinimli bireyler için bilgiye erişimi ve iletişimi kolaylaştırır.

5.2. Sınırlamalar ve Gelecekteki Çalışmalar

Projenin sınırlamaları ve gelecekteki çalışmaları göz önünde bulundurularak, modelin performansı ve uygulanabilirliği değerlendirilmiştir. TurEV-DB veri seti, Türkçe konuşan sınırlı sayıda bireyden toplandığı için modelin farklı dillerde ve kültürlerdeki

duygu ifadelerini doğru bir şekilde tanıması sınırlı olabilir. Ayrıca, veri seti yalnızca dört temel duygu kategorisini (kızgın, üzgün, sakin ve mutlu) içerdiğinden, daha geniş bir duygu yelpazesini tanıma kapasitesi sınırlıdır. Modelin kontrollü ortamda kaydedilen verilerle eğitilmiş olması, gerçek dünya senaryolarında arka plan gürültüsü ve farklı konuşma hızları gibi faktörlerin performansını olumsuz etkileyebileceği anlamına gelmektedir.

Gelecekteki çalışmalar, daha geniş ve çeşitli veri setleri toplayarak modelin duygu tanıma kapasitesini artırmayı hedefleyebilir. Farklı yaş gruplarından, cinsiyetlerden ve kültürel arka planlardan bireylerin ses verilerini içeren veri setleri, modelin genelleme yeteneğini güçlendirebilir. Ayrıca, şaşkınlık, korku ve tiksinti gibi yeni duygu kategorilerinin modele eklenmesi, duygu analizinin kapsamını genişletebilir. Gerçek zamanlı duygu analizi yapabilme kapasitesini artırmak için optimizasyon çalışmaları yapılabilir, bu da müşteri hizmetleri ve insan-robot etkileşimleri gibi uygulama alanlarında büyük fayda sağlayabilir. Gelecekteki çalışmalar, ses verisiyle birlikte yüz ifadeleri ve vücut dili gibi diğer modaliteleri de kullanarak daha bütüncül bir duygu analizi yapmayı hedefleyebilir. Bu tür çapraz modalite analizleri, daha doğru ve güvenilir sonuçlar elde edilmesini sağlayabilir. Proje, duygu analizi alanında önemli bir adım olsa da, bu sınırlamaların farkında olmak ve gelecekteki çalışmalarla bu alanı geliştirmek önemlidir. Genişletilmiş veri setleri, yeni duygu kategorileri ve gerçek zamanlı uygulamalar, bu alandaki ilerlemeyi hızlandıracaktır.

Bölüm 6: Kaynaklar

Salesforce AI Research (2024). BLIP: Bootstrapping Language-Image Pre-training for Unified Vision-Language Understanding. Retrieved from https://github.com/salesforce/BLIP.

Ren, S., Liu, Z., Shu, R., Zhou, J., & Gu, J. (2021). Unified Vision-Language Pre-training with Dual-Encoder. Retrieved from https://arxiv.org/abs/2104.06997.

ElevenLabs. (2024). Text to Speech API. Retrieved from https://elevenlabs.io.

Gartner. ElevenLabs TTS API hizmetlerine yönelik kullanıcı incelemeleri.

Grinberg, M. (2018). *Flask Web Development: Developing Web Applications with Python*. O'Reilly Media.

Flask Documentation. (n.d.). Retrieved from https://flask.palletsprojects.com.

Weizenbaum, J. (1966). ELIZA—A computer program for the study of natural language communication between man and machine. *Communications of the ACM, 9*(1), 36-45.
Hoy, M. B. (2018). Alexa, Siri, Cortana, and Google Assistant: A comparison of speech-based natural language user interfaces. *Medical Reference Services Quarterly, 37*(1), 81-88.

Lopatovska, I., & Williams, H. (2018). Personification of the Amazon Alexa: BFF or a mindless companion. In *Proceedings of the 2018 Conference on Human Information Interaction & Retrieval* (pp. 265-268).

Bentley, F. R., Cramer, H., Hamilton, W. A., & Basapur, S. (2018). Three hours a day: Understanding current smart speaker use. *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, 1-13.
Hoy, M. B. (2018). Alexa, Siri, Cortana, and Google Assistant: A comparison of speech-based natural language user interfaces. *Medical Reference Services Quarterly, 37*(1), 81-88.
Ferrucci, D., Levas, A., Bagchi, S., Gondek, D., & Mueller, E. T. (2013). Watson: Beyond Jeopardy! *Artificial Intelligence, 199*, 93-105.

Munsch, N., Martin, A., Gruarin, S., Nateqi, J., & Abdarahmane, I. (2019). Diagnostic accuracy of web-based COVID-19 symptom checkers: comparison study. *Journal of Medical Internet Research, 21*(7), e21299.

Davis, K. H., Biddulph, R., & Balashek, S. (1952). Automatic recognition of spoken digits. *Journal of the Acoustical Society of America, 24*(6), 637-642.

Juang, B. H., & Rabiner, L. R. (2005). Automatic speech recognition – A brief history of the technology development. *Elsevier Encyclopedia of Language and Linguistics*.

Huang, X., Acero, A., & Hon, H. W. (2001). *Spoken Language Processing: A Guide to Theory, Algorithm, and System Development*. Prentice Hall.
Jelinek, F. (1998). *Statistical Methods for Speech Recognition*. MIT Press.
Rabiner, L., & Juang, B. H. (1993). *Fundamentals of Speech Recognition*. Prentice Hall.

Jurafsky, D., & Martin, J. H. (2008). *Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition*. Pearson.

Jurafsky, D., & Martin, J. H. (2008). *Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition*. Pearson.
Chowdhury, G. G. (2003). Natural language processing. *Annual Review of Information Science and Technology, 37*(1), 51-89.

Hirschberg, J., & Manning, C. D. (2015). Advances in natural language processing. *Science, 349*(6245), 261-266.

Manning, C. D., Raghavan, P., & Schütze, H. (2008). *Introduction to Information Retrieval*. Cambridge University Press.
Jurafsky, D., & Martin, J. H. (2008). *Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition*. Pearson.
Nadeau, D., & Sekine, S. (2007). A survey of named entity recognition and classification. *Lingvisticae Investigationes, 30*(1), 3-26.
Liu, B. (2012). *Sentiment Analysis and Opinion Mining*. Morgan & Claypool Publishers.
Bahdanau, D., Cho, K., & Bengio, Y. (2015). Neural machine translation by jointly learning to align and translate. *arXiv preprint arXiv:1409.0473*.

Nenkova, A., & McKeown, K. (2012). A survey of text summarization techniques. *In Mining Text Data* (pp. 43-76). Springer, Boston, MA.

Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., ... & Bengio, Y. (2014). Generative adversarial nets. In *Advances in Neural Information Processing Systems* (pp. 2672-2680).
Kingma, D. P., & Welling, M. (2013). Auto-Encoding Variational Bayes. *arXiv preprint arXiv:1312.6114*.

Radford, A., Metz, L., & Chintala, S. (2015). Unsupervised representation learning with deep convolutional generative adversarial networks. *arXiv preprint arXiv:1511.06434*.

Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., ... & Bengio, Y. (2014). Generative adversarial nets. In *Advances in Neural Information Processing Systems* (pp. 2672-2680).

Kingma, D. P., & Welling, M. (2013). Auto-Encoding Variational Bayes. *arXiv preprint arXiv:1312.6114*.
Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., ... & Polosukhin, I. (2017). Attention is all you need. In *Advances in Neural Information Processing Systems* (pp. 5998-6008).

Brown, T., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., ... & Amodei, D. (2020). Language models are few-shot learners. *arXiv preprint arXiv:2005.14165*.

Elgammal, A., Liu, B., Elhoseiny, M., & Mazzone, M. (2017). CAN: Creative Adversarial Networks, Generating" Art" by Learning About Styles and Deviating from Style Norms. *arXiv preprint arXiv:1706.07068*.

Ramesh, A., Pavlov, M., Goh, G., Gray, S., Voss, C., Radford, A., ... & Sutskever, I. (2021). Zero-shot text-to-image generation. *arXiv preprint arXiv:2102.12092*.

Brown, T., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., ... & Amodei, D. (2020). Language models are few-shot learners. *arXiv preprint arXiv:2005.14165*.

Sturm, B. L., Santos, J. F., Ben-Tal, O., & Korshunova, I. (2016). Music transcription modelling and composition using deep learning. *arXiv preprint arXiv:1604.08723*.

Payne, C. (2019). MuseNet. OpenAI. Retrieved from https://openai.com/blog/musenet.

Summerville, A., Snodgrass, S., Mateas, M., & Ontañón, S. (2018). Procedural content generation via machine learning (PCGML). *IEEE Transactions on Games, 10*(3), 257-270.

Liu, B. (2012). *Sentiment Analysis and Opinion Mining*. Morgan & Claypool Publishers.

Cambria, E. (2016). Affective computing and sentiment analysis. *IEEE Intelligent Systems, 31*(2), 102-107.

Manning, C. D., Raghavan, P., & Schütze, H. (2008). *Introduction to Information Retrieval*. Cambridge University Press.

Liu, B. (2012). *Sentiment Analysis and Opinion Mining*. Morgan & Claypool Publishers.

Taboada, M., Brooke, J., Tofiloski, M., Voll, K., & Stede, M. (2011). Lexicon-based methods for sentiment analysis. *Computational Linguistics, 37*(2), 267-307.

Ghiassi, M., Skinner, J., & Zimbra, D. (2013). Twitter brand sentiment analysis: A hybrid system using n-gram analysis and dynamic artificial neural network. *Expert Systems with Applications, 40*(16), 6266-6282.

Yelpaze, E., & Alpaydın, E. (2021). Customer sentiment analysis for service quality: A systematic literature review. *International Journal of Information Management Data Insights, 1*(2), 100035.

Denecke, K., & Deng, Y. (2015). Sentiment analysis in medical settings: New opportunities and challenges. *Artificial Intelligence in Medicine, 64*(1), 17-27.

Stieglitz, S., & Dang-Xuan, L. (2013). Emotions and information diffusion in social media—Sentiment of microblogs and sharing behavior. *Journal of Management Information Systems, 29*(4), 217-248.

Turkish Emotion Voice Database (TurEV-DB), https://github.com/Xeonen/TurEV-DB

ÖZGEÇMİŞ

Adı Soyadı : Hasan Hüseyin Alav

Yabancı Dil : İngilizce

Doğum Yeri ve Yılı : Selçuklu/Konya 2002

E-Posta : hhsynalv@gmail.com

Eğitim ve Mesleki Geçmişi:

- 2020-…, Samsun Üniversitesi, Mühendislik ve Doğa Bilimleri Fakültesi, Yazılım Mühendisliği
- 2022, Staj Öğrencisi, Doç. Dr. Zafer Cömert, Samsun Üniversitesi Uzem Birimi
- 2024, Staj Öğrencisi, Dr. Hakan Can Altunay, Cybernova Siber Güvenlik ve Yazılım Teknolojileri Akademisi