Ekonomi, İşletme ve Bilişim Dergisi

Journal of Economics, Business and Information

ISSN: 3062-2425 Cilt/Vol. 1, Sayı/No. 1, 44-64.

Perakende Sektöründe Müşteri Duygularını Anlamak için Çok Modlu Duygu Analizi

Özlem Hakdağlı[1], Erdem Hakdağlı[2]

Makale Hakkında

Gönderim Tarihi
16 Kasım 2024
Kabul Tarihi
23 Aralık 2024
Makale Türü
Araştırma Makalesi

Özet

Bu çalışma, müşteri deneyimleri sırasında ortaya çıkan duygusal tepkileri daha kapsamlı ve doğru bir şekilde anlamak için çok modlu bir duygu analizi yöntemi geliştirmeyi amaçlamaktadır. Müşteri duygularının ölçülmesi, yalnızca müşteri memnuniyetini anlamakla kalmayıp, kişiselleştirilmiş hizmetlerin sunulması, pazarlama stratejilerinin optimize edilmesi ve müşteri sadakatinin artırılması açısından kritik öneme sahiptir. Bu doğrultuda, ses, görüntü ve metin verilerinden elde edilen duygusal ipuçları, derin öğrenme tabanlı modeller kullanılarak ayrı ayrı analiz edilmiş ve sonuçlar katsayı tabanlı bir yöntemle birleştirilmiştir. Ses verilerinin analizi için Xception modeli, yüz ifadelerindeki mikro düzeydeki farklılıkların tespiti için Xception, VGG16 ve VGG19 modelleri, metin verilerindeki bağlamsal ilişkilerin değerlendirilmesi için ise BERT ve ALBERT modelleri kullanılmıştır. Önerilen yöntemin performansı, RAVDESS, FER2013, LFW, TESS ve Beyazperde gibi açık kaynak veri setleri üzerinde test edilmiş; her bir modun duygu analizi sonuçları normalize edilerek -1 ile 1 arasında birleştirilmiştir. Deneysel bulgular, yöntemin görüntü verilerinde %98,25 F1 skoru, metin verilerinde %94,30 F1 skoru ve ses verilerinde %90,71 F1 skoru ile etkili bir performans sergilediğini ortaya koymaktadır. Geliştirilen yöntem, açık kaynak veri setlerinde başarıyla test edilmiş olup, henüz gerçek müşteri verileri üzerinde uygulanmamıştır. Bu yaklaşım, farklı veri türlerini bütüncül bir şekilde entegre ederek tek modlu analizlerin sınırlamalarını aşmakta ve müşteri duygularının kapsamlı bir şekilde anlaşılmasına olanak sağlamaktadır. Çalışma, müşteri deneyimlerinin daha derinlemesine analiz edilmesi ve kişiselleştirilmiş hizmetlerin geliştirilmesi için yenilikçi bir çözüm sunmaktadır.

Anahtar Kelimeler: çok modlu duygu analizi, sınıflandırma

Multimodal Sentiment Analysis for Understanding Customer Emotions in the Retail Sector

Abstract

Article Info

Received
November 16, 2024
Accepted
December 23, 2024
Article Type
Research Article

This study aims to develop a multimodal sentiment analysis method to comprehensively and accurately understand emotional responses during customer experiences. Measuring customer emotions is critical not only for understanding customer satisfaction but also for delivering personalized services, optimizing marketing strategies, and enhancing customer loyalty. In this context, emotional cues derived from audio, visual, and textual data were independently analyzed using deep learning-based models, and the outputs from these modalities were integrated through a weight-based method. The Xception model was utilized for analyzing audio data, while Xception, VGG16, and VGG19 models were employed to capture micro-level differences in facial expressions. For textual data, BERT and ALBERT models were used to evaluate contextual relationships. The proposed method was tested on open-source datasets such as RAVDESS, FER2013, LFW, TESS, and Beyazperde, and the sentiment analysis results from each modality were normalized between -1 and 1 for integration. Experimental findings demonstrated that the method achieved an F1 score of 98.25% for visual data, 94.30% for textual data, and 90.71% for audio data, showcasing its high performance. While the developed method has been successfully tested on open-source datasets, it has not yet been applied to real customer data. This approach surpasses the limitations of unimodal analyses by integrating diverse data types and provides an innovative solution for a comprehensive understanding of customer emotions. The study also paves the way for deeper analyses of customer experiences and the development of personalized solutions.

Keywords: multimodal sentiment analysis, classification

1 Teracity Yazılım Teknolojileri A.Ş., ozlem.hakdagli@teracity.com.tr, https://orcid.org/0000-0002-3637-4309
2 Teracity Yazılım Teknolojileri A.Ş., erdem.hakdagli@teracity.com.tr, https://orcid.org/0009-0004-2993-8921

Atıf: Hakdağlı, Ö., ve Hakdağlı, E. (2024). Perakende Sektöründe Müşteri Duygularını Anlamak için Çok Modlu Duygu Analizi. *Ekonomi, İşletme ve Bilişim Dergisi, 1*(1), 44-64.

1. Giriş

Müşteri duygularının doğru bir şekilde ölçülmesi, günümüzün müşteri odaklı iş dünyasında hayati bir öneme sahiptir. Müşterilerin yalnızca ürün veya hizmetlerden duydukları memnuniyet değil, deneyim süreçlerinde yaşadıkları duygusal tepkiler de markalar için kritik içgörüler sunmaktadır. Duygular, müşteri sadakatini artırmak, kişiselleştirilmiş hizmetler sunmak ve pazarlama stratejilerini optimize etmek için temel bir yapı taşıdır. Araştırmalar, olumlu duyguların müşteri memnuniyetini ve marka bağlılığını güçlendirdiğini; olumsuz duyguların ise müşteri kaybına yol açabileceğini ortaya koymaktadır. Bu bağlamda, müşteri duygularını ölçmek, işletmelerin bireysel müşteri ihtiyaçlarını anlamalarına ve rekabet avantajı elde etmelerine olanak sağlamaktadır.

Duygu analizi, insan davranışlarını anlamak ve bu davranışlardan anlamlı çıkarımlar yapmak amacıyla bilgi işlem alanında hızla gelişen bir araştırma alanı olarak öne çıkmaktadır. İnsan-bilgisayar etkileşimi, sosyal medya analitiği, sağlık, güvenlik ve müşteri deneyimi gibi pek çok alanda, duyguların doğru tespiti, daha etkili iletişim stratejileri geliştirme ve kişiselleştirilmiş hizmetler sunma imkânı sağlamaktadır. Duygular, bireylerin günlük yaşamda kararlarını ve sosyal ilişkilerini etkilediği gibi dijital ortamda insancıl etkileşimlerin de temelini oluşturmaktadır. Bu bağlamda, duygu analizi, insan odaklı teknolojik uygulamaların gelişimine katkı sağlayarak duygusal zeka ve empati temelli çözümler sunma potansiyeline sahiptir.

Geleneksel duygu analizi yöntemleri çoğunlukla tek bir veri türüne, yani yalnızca metin veya sese dayanmaktadır. Bu tek modlu yaklaşımlar, duygusal durumların çok boyutlu yapısını yeterince yansıtamamakta ve dolayısıyla sınırlı kalmaktadır. Örneğin, yalnızca metin analizine dayanan bir duygu analizi, yüz ifadeleri veya ses tonlamalarındaki önemli duygusal ipuçlarını içermemektedir. Bu tür sınırlamalar, analizlerin doğruluğu ve güvenilirliği üzerinde olumsuz etkiler yaratabilmektedir. Tek modlu analizlerde karşılaşılan bu sınırlamaların üstesinden gelmek amacıyla son yıllarda çok modlu duygu analizi yöntemlerine yönelik ilgi artmıştır.

Çok modlu duygu analizi, duygusal durumları daha kapsamlı ve doğru bir biçimde anlamak amacıyla birden fazla veri türünü aynı anda analiz etmeyi hedeflemektedir. Görüntü, ses ve metin gibi farklı veri türlerinin bir arada değerlendirilmesi, duyguların çeşitli yönlerden ele alınmasını mümkün kılmaktadır. Her bir veri türü, duyguları kendi özgün yapısıyla anlamlandırarak analiz sürecine benzersiz katkılar sağlamaktadır. Örneğin, yüz ifadeleri görsel ipuçları sunarken, ses analizinde tonlama, vurgu ve ritim gibi özellikler duygusal yoğunluk ve duygu türü hakkında bilgi sağlamaktadır. Metin analizinde ise kelime seçimi, cümle yapısı ve bağlam değerlendirilerek duygusal eğilimler belirlenmektedir. Bu çeşitlilik, duygu analizinin doğruluğunu artırmakla kalmayıp, duygusal durumların çok boyutlu bir yaklaşımla değerlendirilmesine de olanak tanımaktadır.

Bu çalışmada, müşteri duygularını doğru ve kapsamlı bir şekilde ölçmek amacıyla çok modlu bir duygu analizi yöntemi geliştirilmiştir. Görüntü, ses ve metin verilerinden elde edilen bilgiler, belirli ağırlık katsayıları ile birleştirilerek tek bir duygu skoru oluşturulmuştur. Her veri türüne özel analiz modelleri geliştirilmiş ve bu modellerin çıktıları entegre edilerek analiz doğruluğunun artırılması amaçlanmıştır. Bu yöntem, farklı veri türlerinden elde edilen duygusal katkıları optimize ederek, müşteri duygularının daha hassas ve dengeli bir şekilde değerlendirilmesini sağlamaktadır. Literatürdeki tek modlu veya sınırlı çok modlu analiz yaklaşımlarına kıyasla, bu çalışma, duygu analizi alanına yenilikçi bir katkı sunmaktadır. Ayrıca, her bir veri türünün belirli katsayılarla değerlendirilmesi, duygusal bilgilerin bağlama uygun şekilde yansıtılmasını sağlayarak çalışmanın özgün yönlerinden birini oluşturmaktadır. Müşteri değerlendirmelerinde duygu analizi, müşteri deneyimini ve memnuniyetini doğru bir şekilde ölçmeye olanak tanıyarak kişiselleştirilmiş hizmetlerin sağlanmasını desteklemektedir. Çalışmamız, bu doğrultuda müşteri deneyimlerinin duygusal boyutlarını derinlemesine ele alarak müşteri duygularının kapsamlı bir şekilde analiz edilmesine katkı sunmaktadır.

2. Literatür İncelemesi

Duygu analizi, bireylerin duygusal durumlarını sınıflandırma ve analiz etme amacıyla bilgi işlem dünyasında giderek önem kazanan disiplinler arası bir araştırma alanıdır. İnsan-bilgisayar etkileşimi, sosyal medya analitiği, müşteri hizmetleri, sağlık ve eğitim gibi alanlarda duyguların doğru bir şekilde tespit edilmesi, daha kişiselleştirilmiş ve etkili çözümler sunma potansiyeli taşımaktadır. Bu alandaki ilk çalışmalar, çoğunlukla metin verilerine odaklanarak temel sınıflandırma yöntemleriyle duygu çıkarımı sağlamaya çalışmıştır. Ancak teknolojideki gelişmelerle birlikte, duygu analizine yönelik çalışmalar artık tek modaliteye dayalı yaklaşımlardan çok modlu yaklaşımlara evrilmiş ve metin, görüntü ve ses gibi farklı veri türlerini bir arada analiz eden yöntemler ön plana çıkmıştır.

Metin tabanlı duygu analizi, bu alanın ilk çalışmaları arasında yer almakta olup, Pang ve Lee (2008) ve Liu (2015) gibi araştırmacıların öncü çalışmaları sayesinde kelime duyarlılığı ve duygu sözlükleri gibi tekniklerle duygusal çıkarımlar yapılmaya başlanmıştır. Bu yöntemler, metin içerisindeki duygu yüklü anahtar kelimeler ve cümle yapılarını analiz ederek duygu sınıflandırması yapmayı amaçlamaktadır. Günümüzde ise doğal dil işleme alanındaki gelişmeler, duygu analizinde daha karmaşık ve bağlamsal anlam çıkarma yeteneğine sahip modellerin geliştirilmesini sağlamıştır. Çift Yönlü Kodlayıcı Dönüştürücü Gösterimleri (Bidirectional Encoder Representations from Transformers, BERT) ve Hafif BERT (A Lite BERT, ALBERT) modelleriyle yapılan duygu analizi çalışmaları, bu modellerin bağlamsal ifadeleri doğru bir şekilde yakalayabilme yeteneği sayesinde yüksek doğruluk oranlarına ulaşmıştır. Devlin vd. (2019) BERT modeli, kelimelerin çift yönlü bağlamsal temsillerini kullanarak sosyal medya yorumlarında ve inceleme metinlerinde duygu analizinde geleneksel yöntemlere kıyasla %10-15 oranında daha yüksek doğruluk sağlamıştır. Xu vd. (2020), BERT tabanlı modellerin sosyal medya yorumlarındaki duygu analizinde %90'ın üzerinde doğruluğa ulaştığını rapor etmiştir. ALBERT modeli ise daha hafif ve verimli yapısı ile öne çıkarak, büyük veri kümeleri üzerinde düşük bellek kullanımı ile benzer doğruluğa ulaşmış ve Lan vd. (2020) çalışmasında, BERT ile kıyaslanabilir doğruluk sağladığı görülmüştür. Sun vd. (2021) ALBERT tabanlı duygu analiz çalışması ise, modelin %88 gibi yüksek bir doğruluk oranına ulaştığını ve hafif altyapılarda bile etkili sonuçlar sunduğunu göstermiştir. Bu sonuçlar, BERT ve ALBERT'in modern duygu analizi görevlerinde bağlamsal doğruluk ve esneklik açısından etkili araçlar olduğunu ortaya koymaktadır.

Görüntü tabanlı duygu analizinde, yüz ifadelerinin analiz edilmesi önemli bir rol oynamaktadır. Bu alandaki ilk çalışmalar, Ekman ve Friesen (1978) tarafından ortaya atılan yüz hareket birimi teorisine dayanmaktadır. Bu teoriye göre, belirli yüz kaslarının hareketleri çeşitli duygularla ilişkilendirilebilir. Teknolojinin gelişmesiyle birlikte, yüz ifadelerinden duygusal çıkarımlar yapmak için derin öğrenme tabanlı evrişimli sinir ağları yaygın olarak kullanılmaya başlanmıştır. Simonyan ve Zisserman (2015) tarafından geliştirilen Çok Derin Evrişimsel Ağlar (Very Deep Convolutional Networks, VGG) olan VGG16 ve VGG19 modelleri, yüz ifadelerindeki ince detayları yüksek doğrulukla ayırt etme kapasitesine sahiptir. Chollet (2017) tarafından geliştirilen Aykırı İstisna Modeli (Extreme Inception, Xception), ayrık derinlikli evrişim (depthwise separable convolution) teknolojisi ile yüz ifadelerindeki mikro detayları algılayarak daha yüksek doğruluk sağlamaktadır. Görüntü tabanlı duygu analizinin avantajı, yüz kas hareketleri gibi görsel ipuçlarını kullanarak duygusal durumları anlamaya olanak tanımasıdır. Ancak bu tür analizler, ses tonlaması gibi işitsel modalitelerin sağladığı zenginliği içermemektedir.

Ses tabanlı duygu analizi, konuşma sırasında kullanılan akustik özellikler üzerinden duygusal durumları çıkarma amacı taşımaktadır. Bu alanın ilk çalışmaları, ses tonlaması, vurgu, ritim ve hız gibi akustik özelliklerin analiz edilmesiyle duygusal sınıflandırma yapmaya odaklanmıştır. Bu çalışmalara örnek olarak Busso vd. (2008) tarafından geliştirilen Etkileşimli Duygusal İkili Hareket Yakalama (Interactive Emotional Dyadic Motion Capture, IEMOCAP) veri seti, ses tabanlı duygu analizi için temel bir kaynak

oluşturmuş ve bu alandaki birçok çalışmada referans veri olarak kullanılmıştır. Derin öğrenme modellerinin gelişimiyle birlikte, tekrarlayan sinir ağları (Recurrent Neural Networks, RNN) ve uzun kısa vadeli bellek ağları (Long Short-Term Memory, LSTM) gibi zaman serisi verilerini işleme kapasitesine sahip modeller, konuşma tabanlı duygu analizi alanında daha etkili hale gelmiştir. Özellikle Lotfian ve Busso (2019) tarafından yapılan bir çalışmada, ses tonlamaları ve konuşma sırasındaki duygusal geçişleri tespit etmek için LSTM tabanlı modeller kullanılmış ve duygu değişimlerinin doğru bir şekilde sınıflandırılması sağlanmıştır. Bu tür modeller, ses tabanlı duygu analizi çalışmalarında bağlam bağımlı bilgiye erişim sağlayarak konuşma akışındaki duygusal ipuçlarını yakalamakta başarılı bulunmuştur. Demirtaş ve Hakdağlı (2022) tarafından geliştirilen Dönüştürücü-CNN modeli, Türkçe konuşma verisinden duygu tanıma üzerine odaklanarak, farklı veri artırma teknikleri ve özniteliklerin etkisini incelemiştir. Bu model, çeşitli veri setlerinin birleştirilmesi ve akustik özniteliklerin Log-Mel spektrogramları ile analiz edilmesi sayesinde duygusal tanımlamada yüksek doğruluk sağlamıştır. Bu tür modeller, ses tabanlı duygu analizi çalışmalarında bağlam bağımlı bilgiye erişim sağlayarak konuşma akışındaki duygusal ipuçlarını yakalamakta başarılı bulunmuştur.

Son yıllarda, duygu analizi çalışmalarında çok modlu yaklaşımlar ön plana çıkmaktadır. Çok modlu duygu analizi, metin, görüntü ve ses gibi farklı veri türlerini birleştirerek daha geniş ve kapsamlı bir duygu analizi yapılmasını sağlar. Bu alanda yapılan çalışmalardan biri olan Poria vd. (2017), metin, ses ve görüntü verilerinin entegrasyonunun, duygu analizinde daha yüksek doğruluk ve bağlamsal zenginlik sunduğunu göstermiştir. Zadeh vd. (2018) ise, çok modlu duygusal davranış analiz modeli ile her bir modlu elde edilen bilgilerin birleştirilmesi sayesinde daha başarılı duygu sınıflandırmaları elde etmişlerdir. Bu çalışmalar incelendiğinde çok modlu analizlerin tek modlu analizlerin sınırlamalarını aşarak duygusal çıkarımlarda doğruluğu artırdığını göstermektedir.

Literatürde çok modlu duygu analizine artan ilgi, müşteri duyguları, sağlık ve sosyal medya gibi alanlarda daha derinlemesine ve anlamlı çıkarımlar yapma ihtiyacından kaynaklanmaktadır. Farklı modların entegrasyonu sayesinde elde edilen zengin duygusal bilgiler, duygu analizinde modelin genelleme kapasitesini artırarak farklı bağlamlarda daha etkin bir şekilde kullanılabilmesini sağlamaktadır. Özellikle perakende sektöründe, müşteri duygularını anlamaya yönelik analizler, müşterilerin duygusal tepkilerini daha iyi kavrayarak kişiselleştirilmiş hizmetlerin sunulmasına ve müşteri deneyimlerinin geliştirilmesine önemli katkılar sunmaktadır. Çok modlu duygu analizi, her bir modun güçlü yanlarından yararlanarak daha kapsamlı ve bütüncül bir duygu profili oluşturmakta; bu yönüyle tek modlu analiz yöntemlerine kıyasla önemli bir üstünlük sağlamaktadır. Bu yaklaşım, duygusal durumların doğru ve çok boyutlu bir şekilde değerlendirilmesini mümkün kılarken, analizlerin doğruluk ve güvenilirliğini artırmaktadır.

3. Veri Setleri ve Yöntemler

Bu bölümde, çalışmada kullanılan veri setleri ve uygulanan yöntemler detaylandırılmaktadır. Analiz sürecinde kullanılan verilerin özellikleri ile duygu analizi için seçilen modeller açıklanmıştır.

3.1. Veri Setleri

Çalışmada, duygu analizi amacıyla farklı kaynaklardan çeşitli veri setleri araştırılmış ve kapsamlı bir değerlendirme sürecinden geçirilmiştir. Araştırılan veri setleri arasından uygun bulunan ve onaylanan metin, ses ve görüntü verileri çalışma kapsamında kullanılmıştır. Her bir veri türü, duygusal durumları anlamak ve daha kapsamlı bir analiz sağlamak amacıyla seçilmiş ve kullanılmıştır. Çalışma kapsamında "şaşkın" ifadesi hem negatif hem de pozitif olabileceği için şaşkınlık içeren veriler çalışma dışında tutulmuştur. Bu bölümde seçilen veri setleri ses, görüntü ve metin olmak üzere üç ana kategori altında ayrıntılı bir şekilde ele alınmıştır**.**

3.1.1. Ses Verileri

Duygu analizi amacıyla geniş bir yelpazeye sahip, çeşitli duygusal ses veri tabanları kullanılmıştır. Özellikle Türkçe dilindeki duygusal konuşma veri tabanlarının sınırlılığı ve mevcut veri setlerine erişim kısıtlamaları, özgün bir Türkçe veri seti oluşturulmasında zorluk yaratmaktadır. Bu nedenle bulunan tüm uygun veri tabanları çalışma kapsamında kullanılmıştır. Çalışmada kullanılan veri tabanlarının temel özellikleri aşağıda verilmiştir.

- RAVDESS (Ryerson Audio-Visual Database of Emotional Speech and Song): Livingstone ve Russo (2018) tarafından oluşturulan veri seti İngilizce dilinde, 12 kadın ve 12 erkek aktör tarafından seslendirilen 1.440 kayıttan oluşmaktadır. Stüdyo ortamında kaydedilen bu veri tabanında 8 farklı duygusal durum simüle edilmiştir. Sakin ve nötr duygular birleştirilerek duygu sınıfı sayısı 7'ye indirilmiştir. RAVDESS veri setinin kaydedilme süreci 2018 yılında tamamlanmıştır. Ancak, verilerin tam olarak hangi tarihlerde toplandığı belirtilmemiştir.
- TESS (Toronto Emotional Speech Set): Dupuis ve Pichora-Fuller (2010) tarafından oluşturulan veri seti 26-64 yaş aralığındaki iki kadın konuşmacı tarafından İngilizce olarak seslendirilen 2.800 kayıttan oluşmaktadır. Yedi duygu sınıfı içeren bu veri setinde, şaşırma duygusu pozitif bir ifade olarak sınıflandırılmıştır. TESS veri seti 2010 yılında yayımlanmış olup, kayıtların hangi tarihlerde toplandığına dair spesifik bir bilgi verilmemiştir.
- CREMA-D (Crowd-sourced Emotional Multimodal Actors Dataset): Cao vd. (2014) tarafından oluşturulan veri seti farklı demografik özelliklere sahip 91 aktör tarafından kaydedilen toplam 7.442 kayıt içermektedir. İngilizce olarak hazırlanan veri tabanı, 6 duygu sınıfı ve geniş bir yaş ile etnik çeşitlilik sunmaktadır. Verilerin toplanma tarihine dair açık bir bilgi bulunmamakla birlikte, veri seti 2014 yılında yayımlanmıştır.
- SAVEE (Surrey Audio-Visual Expressed Emotion): Jackson ve Haq (2014) tarafından oluşturulan veri seti İngilizce dilinde, 4 erkek konuşmacı tarafından profesyonel stüdyo ortamında seslendirilmiş toplam 480 kayıttan oluşmaktadır. Veri setinde 7 farklı duygusal durum simüle edilmiştir. Verilerin toplanma sürecine ilişkin spesifik bir bilgi bulunmamakla birlikte, veri seti 2011 yılında yayımlanmıştır.
- Emo-DB (Berlin Database of Emotional Speech): Burkhardt vd. (2005) tarafından oluşturulan veri seti Almanca dilinde, 5 kadın ve 5 erkek konuşmacı tarafından seslendirilmiş 535 kayıttan oluşmaktadır. Yedi duygu sınıfına sahip olan bu veri seti, eğitim ve test amaçlı olarak yaygın bir şekilde kullanılmaktadır. Verilerin 2005 yılında yayımlandığı bilinmekle birlikte, kayıtların toplandığı tarihlere dair net bilgi bulunmamaktadır.
- Türkçe Duygu Veri Seti: Demirtaş ve Hakdağlı (2022) tarafından oluşturulan Türkçe dilinde, toplamda 209 kadın ve 227 erkek konuşmacı tarafından seslendirilen 1.227 kayıttan oluşmaktadır. Dört duygu sınıfı içeren veri seti, dizi, film ve kişisel videolardan derlenen doğal duygu ifadelerini kapsamaktadır. Veri seti, 2022 yılında açık veri kaynaklarından derlenen verilerle oluşturulmuştur.

3.1.2. Görüntü Verileri

Çalışmada, yüz ifadelerinden duygu analizi yapabilmek amacıyla çeşitli görüntü veri setleri kullanılmıştır. Bu veri setleri, yüz ifadelerindeki duygusal durumları doğru sınıflandırmak için geniş bir görsel yelpaze sunmaktadır. Her veri seti, farklı sayıda duygu sınıfı ve çeşitli yüz ifadeleri içermekte olup, modelin farklı yüz ifadelerini öğrenmesi ve genel performansını artırması için gerekli çeşitliliği sağlamaktadır. Çalışmada kullanılan görüntü veri setlerinin özellikleri aşağıda açıklanmıştır:

- FER2013 (Facial Expression Recognition 2013): Goodfellow vd. (2013) tarafından oluşturulan veri seti, 35.887 gri tonlamalı yüz ifadesi görüntüsünden oluşmaktadır ve her bir görüntü 48x48

piksel boyutundadır. Eğitim, doğrulama ve test aşamaları için sırasıyla 28.709, 3.589 ve 3.589 görüntüye sahiptir. Yedi farklı duygu sınıfı (mutlu, üzgün, şaşkın, korkmuş, öfkeli, nötr, iğrenme) içermekte olup, geniş bir yüz ifadesi yelpazesi sunmaktadır. FER2013 veri seti, 2013 yılında Kaggle platformunda düzenlenen "Challenges in Representation Learning: Facial Expression Recognition Challenge" yarışması için oluşturulmuştur. Ancak, görüntülerin hangi yıllarda çekildiğine dair spesifik bir bilgi bulunmamaktadır.

- CK+ (Extended Cohn-Kanade Dataset): Lucey vd. (2010) tarafından oluşturulan veri seti, yüz ifadelerinin analizi için kullanılan, toplamda 920 görüntüden oluşan bir veri setidir. Veri setinde, 640x490 piksel çözünürlükte, gri tonlamalı olarak sunulan görüntüler bulunmaktadır. Sekiz farklı duygu sınıfını (mutlu, üzgün, öfkeli, korkmuş, şaşkın, iğrenme, küçümseme ve nötr) içermektedir. CK+ veri setinin orijinal kayıtları 2000 yılında yayımlanmış olup, genişletilmiş sürüm olan CK+, 2010 yılında tamamlanmıştır. CK+ veri setindeki yüz ifadeleri, benzer arka planlarda kayıt altına alınmıştır ve yüksek çözünürlükte net yüz ifadeleri sağlamaktadır.
- JAFFE (Japanese Female Facial Expression Database): Lyons vd. (1998) tarafından oluşturulan veri seti, Japon kadın deneklerden elde edilen 213 görüntüden oluşmaktadır. Veri setinde yedi farklı duygu sınıfına (mutlu, üzgün, şaşkın, öfkeli, korkmuş, nötr, iğrenme) sahiptir. JAFFE veri seti, 1996 yılında planlanmış ve oluşturulmuştur. Bu veri seti, özellikle yüz kaslarındaki ifadelerin çeşitliliğini analiz etmek için önemlidir.
- LFW (Labeled Faces in the Wild): Huang vd. (2007) tarafından oluşturulan LFW veri seti, doğal koşullarda çekilmiş 5.749 farklı kişiye ait toplam 9.330 yüz görüntüsünden oluşmaktadır. Genellikle 250x250 piksel boyutlarında, renkli olarak sunulan veri seti, yüz tanıma ve doğrulama çalışmaları için geniş bir görsel çeşitlilik sunmaktadır. Verilerin toplandığı yıllara dair kesin bir bilgi bulunmamakla birlikte, bu görüntülerin internet üzerindeki halka açık kaynaklardan derlendiği belirtilmiştir. Veri seti, doğal ortamlarda çekilmiş yüz ifadelerini içererek geniş bir çeşitlilik sağlamaktadır.

3.1.3. Metin Verileri

Çalışmada, Türkçe dilinde duygu analizi gerçekleştirmek amacıyla farklı çevrimiçi kaynaklardan toplanan ve ayrıca özgün olarak oluşturulan metin veri setleri kullanılmıştır. Farklı sosyal medya platformlarından, kullanıcı yorumlarından ve çevrimiçi içeriklerden duygu etiketli veriler toplanarak kapsam genişletilmiş, veri artırma yöntemleriyle veri setimiz zenginleştirilmiştir. Bu veri tabanı, pozitif ve negatif duyguların yanı sıra nötr duygu içeriklerini de kapsayarak duygu analizinde geniş bir veri seti oluşturulmuştur. Çalışmada kullanılan ve oluşturulan veri setleri ve temel özellikleri aşağıdaki gibidir:

- Beyazperde: 53.400 film yorumu içeren bu veri seti, yorumları pozitif ve negatif olarak etiketlenmektedir. Veri setinin toplandığı yıllara dair açık bir bilgi bulunmamaktadır.
- Otelpuan: 11.600 otel yorumundan oluşur ve pozitif-negatif sınıflandırma sunmaktadır. Verilerin toplandığı spesifik tarihler hakkında bilgi bulunmamaktadır.
- Online Alışveriş Yorumları: 11.400 alışveriş yorumu içermekte olup, yorumlar olumlu, olumsuz ve tarafsız olarak etiketlenmektedir. Veri setinin toplandığı tarihler belirtilmemiştir. Bu tür veri setleri genellikle e-ticaret sitelerindeki müşteri yorumlarından derlenmektedir.
- Türkçe Twitter Veri Seti: Twitter gönderilerinden derlenmiş, pozitif ve negatif etiketli 15.000 cümle içermektedir. Verilerin hangi tarihlerde toplandığı belirtilmemiştir; ancak, Twitter gönderilerinin tarih aralıklarına dair bilgi verilmeksizin derlenmiştir.
- Turkish Product Reviews: Çevrimiçi alışveriş sitelerinden toplanan 235.165 yorumu içermektedir. Yorumlar pozitif ve negatif olarak etiketlenmektedir. Verilerin toplandığı spesifik yıllar belirtilmemiştir; ancak, veri setinin alışveriş platformlarından derlenmiş olduğu

bilinmektedir. Bu tür veri setleri, ürün değerlendirmeleri ve müşteri memnuniyeti analizleri için kullanılmaktadır.

- Türkçe Sosyal Medya Paylaşımı Veri Seti: 11.006 sosyal medya gönderisinden oluşan bu veri setinde, gönderiler pozitif ve negatif olarak sınıflandırılmaktadır. Verilerin toplandığı tarihlere dair bilgi mevcut değildir. Bu tür veri setleri, sosyal medya duygu analizi çalışmalarında kullanılmaktadır.
- TRSAv1 (Turkish Sentiment Analysis v1): Aydoğan ve Kocaman (2022) tarafından oluşturulan veri setinde, çevrimiçi alışveriş sitelerinden toplanmış toplam 150.000 cümleyi içermektedir. Bu yorumlar pozitif, negatif ve nötr olarak sınıflandırılmaktadır. Veri seti 2022 yılında yayımlanmıştır; ancak, verilerin hangi tarihlerde toplandığı belirtilmemiştir. Bu veri seti, Türkçe duygu analizi çalışmaları için önemli bir kaynaktır.
- Çalışmada Oluşturulmuş Veri Seti: Hepsiburada.com sitesinden toplanan veriler elle etiketlenmiştir. Oluşturulan veri setinde, pozitif, negatif ve nötr olmak üzere toplam 100.000 yorum bulunmaktadır. Dengeli yapılmak istenen veri seti, %35 pozitif, %45 negatif, %20 nötr şeklinde dağılım göstermektedir. 2022 yılında Hepsiburada.com'dan toplanan verilere dayanmaktadır. Bu veri seti, Türkçe e-ticaret yorumları üzerinde duygu analizi yapmak için oluşturulmuştur.

3.2. Yöntemler

Çalışmada, çok modlu duygu analizinde doğruluğu ve kapsamı artırmak amacıyla ses, görüntü ve metin verileri üzerinde derin öğrenme tabanlı çeşitli modeller kullanılmıştır. Çok modlu analiz yaklaşımı, tek bir veri türüyle sınırlı analizlerin ötesine geçerek farklı veri türlerinden gelen duygusal işaretleri bir arada değerlendirir. Bu sayede, duygusal durumların daha kapsamlı ve güvenilir bir şekilde tespit edilmesi sağlanırken, veri türleri arasındaki ilişkiler de analiz sürecinde dikkate alınır. Son yıllarda çok modlu duygu analizine yönelik araştırmaların artışı, duygu tespiti doğruluğunu ve bağlamsal zenginliği artırma potansiyelini ortaya koymaktadır.

Çalışma kapsamında her veri türünün kendine özgü özelliklerini analiz etmek için en uygun derin öğrenme modelleri seçilmiştir. Ses verilerinin analizi için Xception kullanılmıştır. Xception modeli, ayrık derinlikli evrişim (depthwise separable convolution) teknolojisiyle her bir katmanda ayrı bir evrişim işlemi gerçekleştirerek model parametrelerini azaltarak hesaplama verimliliğini artırmaktadır. Görüntü verilerinin analizi için, yüz ifadelerinden duygusal durumlarını sınıflandırmak amacıyla VGG16, VGG19 ve Xception modelleri kullanılmıştır. Metin verilerinin analizi için, dilin bağlamsal yapısını iki yönlü olarak analiz edebilen BERT ve ALBERT modelleri kullanılmıştır.

Bu yöntemler ile her veri türüne özgü en uygun sınıflandırma modelleri kullanılarak analiz gerçekleştirilmiştir. Çok modlu analiz yaklaşımı sayesinde, her bir veri türünden gelen duygusal işaretlerin entegre edilmesiyle duygusal durumların daha bütüncül, doğru ve hassas bir şekilde tespit edilmesi amaçlanmıştır.

3.2.1. Xception Modeli

Xception modeli, Szegedy vd. (2014) tarafından geliştirilen Inception modelinin genişletilmiş bir versiyonu olarak Chollet (2017) tarafından tanıtılmış ve ayrık derinlikli evrişim teknolojisi sayesinde işlem verimliliği sunmaktadır. Bu teknoloji ile her özellik haritasını bağımsız olarak işleyerek kanallar arasındaki bağlantıları optimize eden model, hem görüntü hem de ses verisinde duygu analizi için kullanılmıştır. Görüntü verisinde yüz ifadelerindeki ince kas hareketlerini; ses verisinde ise log-Mel spektrogramları kullanılarak tonlama, vurgu ve ritim gibi detayları analiz etme kapasitesiyle model, duygu tespiti sürecinde yüksek doğruluk sağlayacağı düşünüldüğünden seçilmiştir.

3.2.2. VGG16 ve VGG19 Modelleri

Görüntü verisinin analizi için VGG16 ve VGG19 modelleri tercih edilmiştir. Simonyan ve Zisserman (2015) tarafından geliştirilen bu modeller, sırasıyla 16 ve 19 katmandan oluşan derin öğrenme mimarilerine sahiptir. VGG16, sabit filtre boyutları ve ardışık katman yapısıyla temel duygusal özelliklerin ayrıştırılmasında etkilidir ve daha az katmana sahip olması sayesinde küçük veri kümelerinde daha yüksek doğruluk sağlamaktadır. VGG19 ise ek katmanları sayesinde daha derin bir özellik çıkarımı yaparak yüz kas hareketlerindeki ince detayları algılayabilmektedir. Bu derin yapısı, yüz ifadelerindeki mikro düzeydeki duygusal farklılıkları ve karmaşık duygusal durumları tespit edebilmesini sağlar. VGG16, hızlı analiz gerektiren uygulamalar için uygunken; VGG19, daha ayrıntılı analizler yapmak üzere tercih edilmiştir.

3.2.3. BERT ve ALBERT Modelleri

Metin verisinin analizi amacıyla, BERT ve ALBERT modelleri tercih edilmiştir. Devlin vd. (2019) tarafından geliştirilen BERT, metin içindeki kelimelerin bağlamını iki yönlü olarak inceleyerek duygu analizinde yüksek doğruluk sağlamaktadır. Bu modelin maskelenmiş dil modelleme yöntemi ile eksik kelimeleri tahmin etme yeteneği, metinlerdeki duygusal tonların daha doğru bir şekilde belirlenmesine olanak tanımaktadır. BERT'in optimize edilmiş versiyonu olan ALBERT ise, Lan vd. (2020) tarafından daha az parametreyle çalışmak üzere geliştirilmiştir ve büyük veri setlerinde hız avantajı sunmaktadır. Bu kapsamda, BERT daha derin analiz gerektiren metinler için ideal bir seçenekken, ALBERT geniş veri setlerinde hızlı analiz imkanı sunarak duygu analizinde önemli bir katkı sağlamaktadır.

4. Deneyler ve Bulgular

Bu bölümde, duygu analizi görevinde kullanılan derin öğrenme modellerinin performansları ve analiz sürecine dair bulgular ele alınmaktadır. Modeller, farklı modlarda (görüntü, ses, metin) toplanan veri setleri üzerinde eğitilmiş ve her mod için model performansı, çapraz doğrulama yöntemi ile doğruluk, kesinlik, duyarlılık ve F1 skoru gibi metrikler kullanılarak değerlendirilmiştir. Çalışmada kullanılan çok modlu analiz yöntemi, tek modlu yaklaşımlara kıyasla daha yüksek doğruluk ve güvenilirlik sunmuş olup, modlar arası katsayılarla ağırlıklandırma yöntemiyle genel duygu skorları hesaplanmıştır. Elde edilen bulgular, duygu analizi sürecinin başarısını ve her bir modelin göreli etkinliğini ortaya koyarak çalışmanın metodolojik yaklaşımının doğruluğunu desteklemektedir. Modeller, AMD Ryzen 7 3700X 8-Core işlemci, NVIDIA RTX 3080 Ti (12 GB) ekran kartı donanımlı bir sistemde eğitilmiştir.

4.1. Ses Duygu Analizi

4.1.1. Verinin Hazırlanması

Ses duygu analizi için kullanılan veri setleri, her bir ses kaydının görselleştirilmesi ve analiz edilebilmesi amacıyla ön işleme tabi tutulmuştur. Çalışmada, ses sinyallerini analiz edebilmek için log-Mel spektrogramları kullanılmaktadır. İlk adımda, ses sinyalleri hızlı Fourier dönüşümü (FFT) yardımıyla frekans bileşenlerine ayrılmış ve ardından Mel ölçeğinde filtrelenmiştir. Bu işlem sayesinde, her ses kaydı hem frekans hem de zaman bilgisini içeren log-Mel spektrogramlara dönüştürülerek görsel bir forma getirilmiştir. Bu görsel format, ses verisini görüntü verisi gibi analiz edilebilir hale getirmektedir.

Ön işlem adımlarında, log-Mel spektrogramlar 96x96 piksel boyutlarına yeniden boyutlandırılmıştır. Veriyi artırmak ve modelin genellenebilirliğini artırmak amacıyla veri genişletme teknikleri uygulanmıştır. Bu teknikler arasında döndürme, kaydırma gibi işlemler yer almakta olup, modelin farklı açılardan duygu varyasyonlarını öğrenmesine katkı sağlamaktadır.

4.1.2. Model Eğitimi

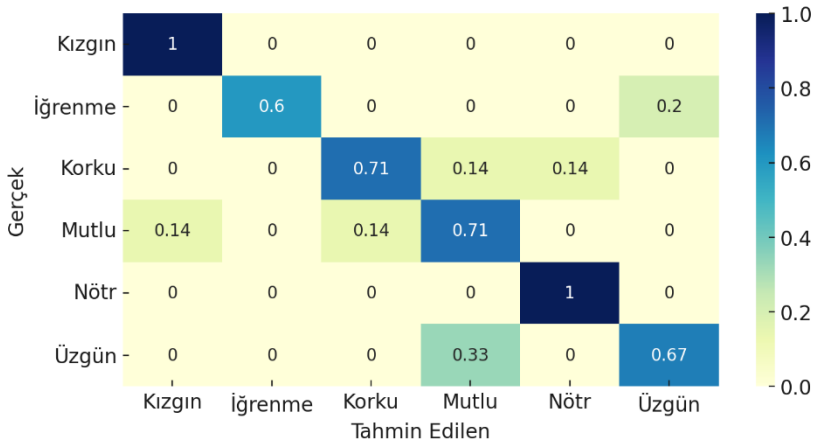
Ses duygu analizi için Xception modeli, ilk olarak Deng vd. (2009)'ın ImageNet veri setinde eğitilmiş ağırlıklarla başlatılmıştır. Bu önceden eğitilmiş ağırlıklar, modelin temel görsel özellikleri tanımasına olanak tanımakta olup, duygu analizinde log-Mel spektrogramlarından elde edilen görsel temsiller aracılığıyla ses verisindeki duygusal sinyalleri tanımlayabilmesi için ileri seviyede eğitime olanak sağlamaktadır. Bu süreçte, ilk eğitim aşamasında his kategorileri üzerinde RAVDESS, TESS, CREMA-D ve SAVEE veri setleri kullanılmıştır. İkinci eğitim aşamasında ise, duygu sınıflandırma performansını optimize etmek amacıyla Emo-DB ve Türkçe veri seti eğitim ve test olarak ayrılmış ve bu veri setleriyle modelin son eğitimi gerçekleştirilmiştir. Ses verileri 96x96 boyutunda log-Mel spektrogramlarına dönüştürülerek Xception modeline giriş olarak verilmiştir. Model, ayrık derinlikli evrişim katmanları ile frekans ve zaman bilgilerini etkili bir biçimde analiz ederek, duygusal durumları sınıflandırmayı hedeflemektedir. Eğitim sürecinde iki farklı sınıflandırma yöntemi kullanılmıştır. İlk yöntemde Xception modelinin yoğun katmanlarında ReLU aktivasyon fonksiyonları kullanılmış ve model çıkışında softmax aktivasyonu uygulanarak duygu sınıflandırması gerçekleştirilmiştir. İkinci yöntemde ise, yoğun katmanın çıkışları Cortes ve Vapnik (1995) tarafından geliştirilen Destek Vektör Makine(Support Vector Machine, SVM) sınıflandırıcısına gönderilerek farklı bir sınıflandırma yapılmıştır. Her iki yöntemde de, 0.001 öğrenme oranına sahip Adam optimizasyon algoritması kullanılarak ve erken durdurma yöntemiyle modelin aşırı öğrenmesi engellenmiştir.

Model performansını güvenilir bir şekilde değerlendirmek amacıyla 10 katlı çapraz doğrulama yöntemi uygulanmıştır. Bu yöntemle modelin genelleme kabiliyeti test edilmiş ve her bir doğrulama aşamasında doğruluk, duyarlılık, kesinlik ve F1 skoru gibi performans metrikleri hesaplanarak değerlendirilmiştir.

Bu iki aşamalı ön işleme ve model eğitimi süreci, ses verilerinin daha etkin analiz edilmesini sağlamış ve duygu sınıflandırma performansını iyileştirmiştir.

4.1.3. Model Performansı

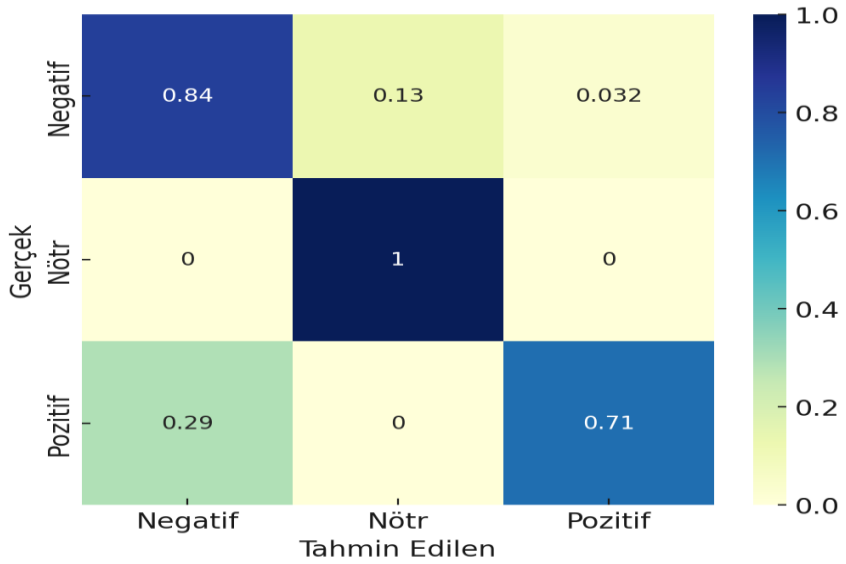
Bu çalışmada, Xception modeli kullanılarak duygu tanıma görevi gerçekleştirilmiş ve sınıflandırıcı olarak Softmax ve SVM kullanılmıştır. Aşağıda sunulan sonuçlar, Xception modelinin Softmax ve SVM sınıflandırıcıları ile elde ettiği performansı detaylı bir şekilde anlatılmıştır.



Şekil 1: Softmax sınıflandırıcı ilk eğitim karmaşıklık matrisi

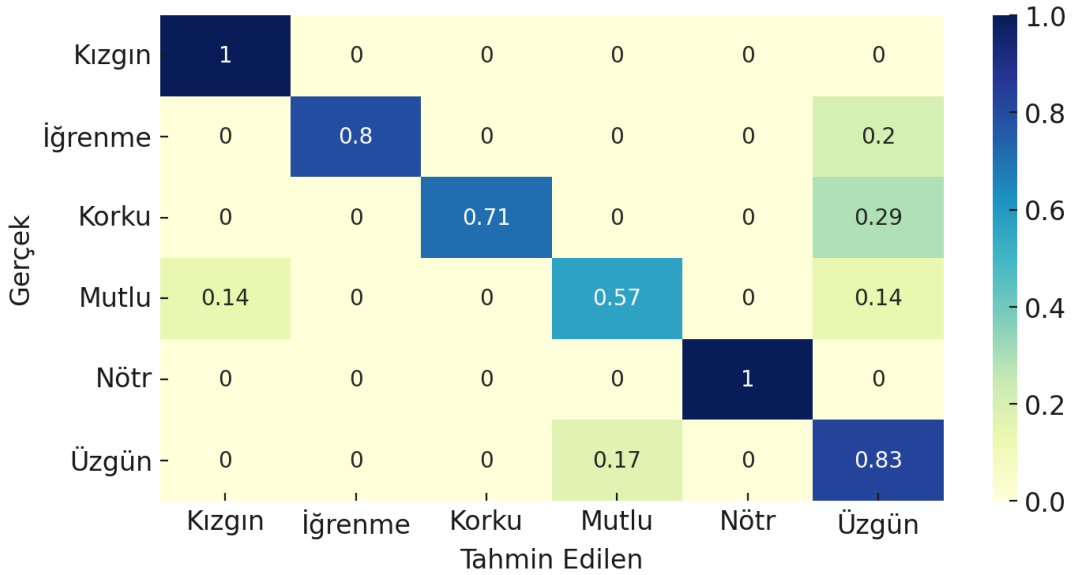
Şekil 1, Softmax sınıflandırıcısı kullanılarak modelin ilk eğitimde elde ettiği duygu sınıflarını tahmin etme performansını göstermektedir. Matriste, her bir duygu sınıfı için (Kızgın, İğrenme, Korku, Mutlu, Nötr, Üzgün) modelin doğru ve yanlış sınıflandırmaları görselleştirilmiştir. Özellikle "Kızgın" ve "Nötr" sınıflarının yüksek doğruluk oranlarına sahip olduğu, buna karşın "İğrenme" ve "Mutlu"

sınıflarında kısmi karışıklıkların olduğu görülmektedir. "Üzgün" sınıfında ise %67'lik bir doğruluk oranı gözlemlenmektedir. Bu matristeki sonuçlar, bazı duygu sınıflarının spektral özelliklerinin diğerlerine göre daha belirgin veya model tarafından daha iyi öğrenilmiş olduğunu göstermektedir.



Şekil 2: Softmax sınıflandırıcı ikinci eğitim karmaşıklık matrisi

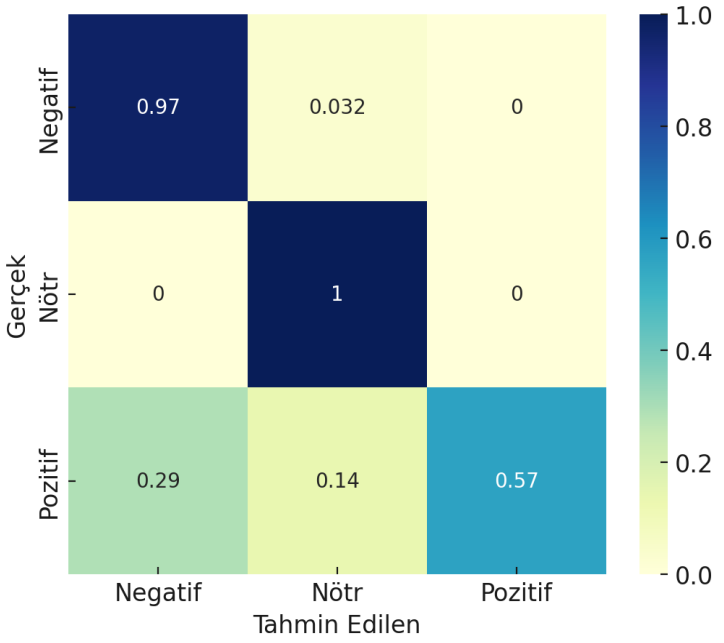
Şekil 2, Softmax sınıflandırıcısının ikinci eğitim aşamasında pozitif, negatif ve nötr sınıflar arasındaki tahmin performansını gösteren karmaşıklık matrisidir. Bu matriste, modelin genel duygu kategorilerindeki (Negatif, Nötr, Pozitif) doğruluğu değerlendirilmiştir. Negatif sınıfta %84, Nötr sınıfta %100 ve Pozitif sınıfta %71 oranında doğru sınıflandırma yapılmıştır. Nötr sınıfı en yüksek doğruluk oranına sahipken, Pozitif sınıfında kısmi bir karışıklık gözlenmiştir. Bu matristeki sonuçlar, modelin genel duygu kategorilerinde sınıflandırma performansını daha geniş bir perspektifte değerlendirme imkanı sunmaktadır.



Şekil 3: SVM sınıflandırıcı ilk eğitim karmaşıklık matrisi

Şekil 3, SVM sınıflandırıcısı kullanılarak modelin ilk eğitimde duygu sınıflarını (Kızgın, İğrenme, Korku, Mutlu, Nötr, Üzgün) tahmin etme performansını göstermektedir. Matriste her bir duygu sınıfı

için doğru ve yanlış sınıflandırmalar görselleştirilmiştir. Özellikle "Kızgın" ve "Nötr" sınıflarının yüksek doğruluk oranlarına sahip olduğu, "İğrenme" sınıfında %80 ve "Korku" sınıfında %71 doğruluk elde edildiği görülmektedir. "Mutlu" sınıfında ise %57 oranında doğru sınıflandırma yapılmış olup, kısmi karışıklık gözlenmektedir. "Üzgün" sınıfında %83 oranında doğruluk oranı elde edilmiştir. Bu sonuçlar, bazı duygu sınıflarının daha net öğrenildiğini ve diğerleri arasında karışıklık yaşandığını göstermektedir.



Şekil 4: SVM sınıflandırıcı ikinci eğitim karmaşıklık matrisi

Şekil 4, SVM sınıflandırıcısının ikinci eğitim aşamasında pozitif, negatif ve nötr sınıflar arasındaki tahmin performansını gösteren PN (Pozitif-Negatif) karmaşıklık matrisidir. Negatif sınıfta %97, Nötr sınıfta %100 ve Pozitif sınıfta %57 oranında doğru sınıflandırma yapılmıştır. Özellikle Nötr sınıfı, en yüksek doğruluk oranına sahip olup %100 doğru sınıflandırma ile öne çıkmaktadır. Pozitif sınıfta %57 oranında doğru sınıflandırma yapılmış, ancak negatif sınıfla %29 oranında karışıklık yaşanmıştır. Bu sonuçlar, modelin genel duygu kategorilerindeki performansını ve sınıflar arasındaki ayrım başarısını ortaya koymaktadır.

Tablo 1, modelin 10 katlı çapraz doğrulama ortalamalarına göre elde edilen performans değerlerini sunmaktadır. Bu tabloda, Softmax ve SVM sınıflandırıcıları ile eğitilen modelin doğruluk, kesinlik, duyarlılık ve F1 skoru metrikleri karşılaştırılmıştır. SVM sınıflandırıcısının, Softmax sınıflandırıcısına kıyasla tüm metriklerde daha yüksek performans gösterdiği gözlemlenmiştir. Özellikle kesinlik ve duyarlılık değerlerinin yüksek olması, SVM sınıflandırıcısının duygu sınıflarını daha doğru bir şekilde ayırt edebildiğini göstermektedir.

Tablo 1: Modelin Performans Değerleri(10 Katlı Çapraz Doğrulama Ortalamaları)

Sınıflandırcı	Doğruluk(%)	Kesinlik(%)	Duyarlılık(%)	F1 Skor(%)
Softmax	82,61	85,5	84,5	85,01
SVM	84,78	91,8	89,7	90,71

Bu sonuçlar, modelin genel performansını değerlendirme ve en etkili sınıflandırıcıyı belirleme açısından önemli bilgiler sağlamaktadır. Sonuç olarak, SVM sınıflandırıcısı, duygu tanıma görevinde Softmax'e kıyasla daha üstün performans sergileyerek öne çıkmaktadır. Bu veriler, modelin gelecekteki iyileştirmelerinde SVM sınıflandırıcısının kullanımını destekleyen bir temel oluşturmaktadır.

4.2. Görüntü Duygu Analizi

4.2.1. Verinin Hazırlanması

Duygu analizi amacıyla CK+, FER2013, JAFFE ve LFW veri setlerinden elde edilen görseller, modelin eğitim sürecine uygun hale getirilmiştir. Farklı veri setlerinin boyut ve özellik bakımından çeşitlilik göstermesi nedeniyle, tüm görseller üzerinde uyumlu bir ön işleme süreci uygulanmıştır. Bu sürecin amacı, modellerin daha yüksek doğruluk ve performans ile çalışmasını sağlamaktır. Uygulanan ön işleme adımları aşağıda belirtilmiştir:

- Yeniden Boyutlandırma: Veri setindeki her bir görselin boyutu farklı olduğundan, görüntüler Xception, VGG16 ve VGG19 modellerinin ihtiyaç duyduğu giriş boyutlarına uygun olarak 224x224 piksel boyutuna yeniden boyutlandırılmıştır. Bu işlem ile modelin her bir görseli eşit boyutta almasını ve eğitim sürecinin daha hızlı ilerlemesini sağlamaktadır.
- Normalizasyon: Görsellerin piksel değerlerinin modelin daha dengeli öğrenebilmesi için [0,1] aralığına normalize edilmiştir. Böylelikle modelin daha stabil bir şekilde eğitim yapabilmesini sağlanır. Ayrıca ağırlıkların aşırı büyümesini engeller.
- Veri Genişletme (Data Augmentation): Eğitim sürecinde veri setlerinin sınırlı sayıda örnek içermesi nedeniyle, modelin genellenebilirliğini artırmak amacıyla çeşitli veri artırma işlemleri uygulanmıştır. Bu işlemler, modelin yüz ifadelerinin farklı açılardan öğrenmesini ve aşırı öğrenme(overfitting) sorununu engellemeyi amaçlamaktadır. Veri artırma işlemleri olarak döndürme, kaydırma, yatay ve dikey çevirme işlemleri yapılmıştır. Böylece modelin daha fazla örnek üzerinde eğitim yapması sağlanmış ve modelin gerçek dünyadaki farklı yüz ifadeleri ile karşılaştığında daha başarılı olması hedeflenmiştir.

4.2.2. VGG16 Modelinin Eğitimi

Bu çalışmada, duygu analizi amacıyla öncelikle VGG16 mimarisi kullanılmıştır. Model, ImageNet veri seti üzerinde önceden eğitilmiş ağırlıklarla başlangıç yapmıştır. İlk aşamada CK+, FER2013, JAFFE veri setleri birleştirilerek toplamda yedi duygu sınıfı (mutlu, üzgün, şaşkın, korkmuş, kızgın, tiksinmiş, nötr) için modelin ince ayar (fine-tuning) işlemi gerçekleştirilmiştir. Bu süreçte, modelin ilk 13 evrişim katmanı dondurularak genel özellik çıkarımı korunmuş ve son katmanlar yeniden eğitime açılmıştır. Yeni eklenen tam bağlantılı katmanlarda ReLU aktivasyon fonksiyonu ve 0,5 seyreltme (dropout) oranı kullanılmıştır.

Eğitim sürecinde, Adam optimizasyon algoritması ve 0,0001 öğrenme oranı ile model 30 döngü(epoch) boyunca eğitilmiştir. Yığın boyutu(batch size) 32 olarak belirlenmiştir. Eğitim tamamlandıktan sonra elde edilen model ağırlıkları kaydedilmiştir.

İkinci aşamada, LFW veri seti üzerinde üç duygu sınıfı (negatif, nötr, pozitif) için ince ayar işlemi yapılmıştır. Önceki aşamada elde edilen model ağırlıkları bu aşamada başlangıç noktası olarak kullanılmıştır. Modelin ilk 13 evrişim katmanı yine dondurulmuş, son katmanlar ise eğitime açılmıştır. Öğrenme oranı 0,00001'e düşürülmüş ve eğitim 15 döngü(epoch) boyunca sürdürülmüştür.

Eğitim işlemlerinin ardından, modelin genelleme performansını değerlendirmek amacıyla 10 katlı çapraz doğrulama yöntemi uygulanmıştır. Bu yöntemle, veri seti 10 eşit parçaya ayrılmış ve her bir parça sırayla test verisi olarak kullanılırken geri kalan 9 parça eğitim için kullanılmıştır. Bu süreç, modelin farklı veri bölümleri üzerinde eğitilmesi ve test edilmesiyle, modelin çeşitli veri bölümlerine

uyum yeteneğini daha objektif bir şekilde ölçmemizi sağlamıştır. 10 katlı çapraz doğrulama sonucunda doğruluk, kesinlik, duyarlılık ve F1 skoru gibi performans metriklerinin ortalamaları hesaplanmış ve bu metrikler, modelin genelleme başarısını analiz etmek için kullanılmıştır.

4.2.3. VGG19 Modelinin Eğitimi

VGG19 mimarisiyle benzer bir yöntem izlenmiştir. ImageNet üzerinde önceden eğitilmiş model, CK+, FER2013, JAFFE veri setleri kullanılarak yedi duygu sınıfı için ince ayar işlemine tabi tutulmuştur. Modelin ilk 16 evrişim katmanı dondurulmuş ve son katmanlar yeniden eğitime açılmıştır. Yeni eklenen sınıflandırma katmanlarında yine ReLU aktivasyon fonksiyonu ve 0,5 seyreltme (dropout) oranı kullanılmıştır.

Eğitim sırasında, Adam optimizasyon algoritması ve 0,0001 öğrenme oranı ile model 25 döngü(epoch) boyunca eğitilmiştir. Yığın boyutu(batch size) 32 olarak belirlenmiştir. İlk aşamanın sonunda model ağırlıkları kaydedilmiştir.

İkinci aşamada, LFW veri seti üzerinde üç duygu sınıfı için ince ayar işlemi gerçekleştirilmiştir. Önceki modelin ağırlıkları kullanılarak, öğrenme oranı 0,00001'e düşürülmüş ve eğitim 15 döngü(epoch) boyunca sürdürülmüştür. Modelin ilk 16 katmanı dondurulmuş olarak kalmıştır. Bu aşamada da veri artırma teknikleri uygulanmıştır.

Eğitim işlemlerinin ardından, modelin genelleme performansını değerlendirmek amacıyla 10 katlı çapraz doğrulama yöntemi uygulanmıştır. Bu yöntemle, veri seti 10 eşit parçaya ayrılmış ve her bir parça sırayla test verisi olarak kullanılırken geri kalan 9 parça eğitim için kullanılmıştır. Bu süreç, modelin farklı veri bölümleri üzerinde eğitilmesi ve test edilmesiyle, modelin çeşitli veri bölümlerine uyum yeteneğini daha objektif bir şekilde ölçmemizi sağlamıştır. 10 katlı çapraz doğrulama sonucunda doğruluk, kesinlik, duyarlılık ve F1 skoru gibi performans metriklerinin ortalamaları hesaplanmış ve bu metrikler, modelin genelleme başarısını analiz etmek için kullanılmıştır.

4.2.4. Xception Modelinin Eğitimi

Daha karmaşık bir mimariye sahip olan Xception modeli için de aynı iki aşamalı ince ayar stratejisi uygulanmıştır. İlk olarak CK+, FER2013, JAFFE, veri setleri kullanılarak yedi duygu sınıfı için model eğitilmiştir. ImageNet üzerinde önceden eğitilmiş modelin ilk 36 derin ayrık evrişim katmanı dondurulmuş ve son katmanlar eğitime açılmıştır. Yeni eklenen tam bağlantılı katmanlarda ReLU aktivasyon fonksiyonu ve 0,5 seyreltme (dropout) oranı kullanılmıştır.

Eğitim sürecinde, Adam optimizasyon algoritması ve 0,0001 öğrenme oranı ile model 40 döngü(epoch) boyunca eğitilmiştir. Bellek gereksinimlerinden dolayı yığın boyutu(batch size) 16 olarak seçilmiştir. İlk aşama tamamlandıktan sonra model ağırlıkları kaydedilmiştir.

İkinci aşamada, LFW veri seti üzerinde üç duygu sınıfı için ince ayar işlemi yapılmıştır. Önceki aşamada elde edilen ağırlıklar başlangıç noktası olarak alınmıştır. Öğrenme oranı 0.00001'e düşürülerek model 15 döngü(epoch) boyunca eğitilmiştir. İlk 36 katman dondurulmuş olarak bırakılmıştır.

4.2.5. Modellerin Performansı

Bu çalışmada, VGG16, VGG19 ve Xception modellerinin duygu analizi performansları 10 katlı çapraz doğrulama yöntemi kullanılarak değerlendirilmiştir. Bu yöntem, modellerin genelleme yeteneğini ve farklı veri bölümleri üzerindeki tutarlılığını ölçmek için etkili bir yaklaşımdır. Her model için doğruluk, kesinlik, duyarlılık ve F1 skoru metriklerinin ortalama değerleri aşağıdaki tabloda sunulmaktadır.

Tablo 2: Modellerin Performans Değerleri(10 Katlı Çapraz Doğrulama Ortalamaları)

Model	Doğruluk(%)	Kesinlik(%)	Duyarlılık(%)	F1 Skor(%)
VGG16	74,00	73,50	73,67	73,84
VGG19	78,00	77,80	77,7	77,58
Xception	98,50	98,30	98,27	98,25

Tablo 2'de görüldüğü üzere, VGG16 ve VGG19 modelleri sırasıyla %73,84 ve %77,58 ortalama F1 skorlarına ulaşmıştır. VGG19 modelinin VGG16'ya kıyasla daha derin bir mimariye sahip olması, performansta yaklaşık %4'lük bir artış sağlamıştır. Bu durum, model derinliğinin duygu analizinde etkili özelliklerin öğrenilmesine katkı sağladığını göstermektedir. Ancak, her iki model de karmaşık duygu ifadelerinin sınıflandırılmasında sınırlı bir başarı göstermiştir.

Eğitim ve doğrulama süreçlerinde, her iki modelin de istikrarlı bir öğrenme eğrisi sergilediği gözlemlenmiştir. Ancak, doğrulama setindeki performansın eğitim setine göre daha düşük olması, modellerin aşırı öğrenme (overfitting) eğilimi gösterebileceğini işaret etmektedir. Bu durum, veri setlerinin sınırlı çeşitliliği ve modellerin genelleme kabiliyetlerinin kısıtlı olmasından kaynaklanabilir.

Xception modeli, %98,25 ortalama F1 skoru ile en yüksek performansı sergilemiştir. Bu sonuç, modelin derin ve ayrık evrişimli mimarisinin duygu analizinde daha etkili özellik çıkarımı yapabildiğini göstermektedir. Modelin yüksek kesinlik ve duyarlılık değerleri, hem pozitif hem de negatif sınıflandırmalarda tutarlı bir performans sergilediğini ifade etmektedir.

Eğitim sürecinde, Xception modelinin hızlı bir yakınsama gösterdiği ve hem eğitim hem de doğrulama setlerinde yüksek doğruluk değerlerine ulaştığı gözlemlenmiştir. Ayrıca eğitim ve doğrulama doğrulukları arasındaki farkın minimal olması, modelin genelleme kabiliyetinin yüksek olduğunu ve aşırı öğrenme sorununu minimize ettiğini göstermektedir.

4.3. Metin Duygu Analizi

4.3.1. Verinin Hazırlanması

Duygu analizi sürecinde metin verisinin doğru bir şekilde hazırlanması, modelin performansını doğrudan etkileyen en önemli adımlardan biridir. Bu çalışmada metinlerin analiz için işlenebilir hale getirilmesi amacıyla çeşitli ön işleme adımları uygulanmıştır. İlk olarak, veri kaynaklarından elde edilen metinlerde yer alan özel karakterler, noktalama işaretleri ve emojiler temizlenmiş; bu adım, modelin verideki gereksiz gürültüyü azaltarak anlamlı bilgilere odaklanmasını sağlamıştır. Ayrıca büyük ve küçük harf farklılıklarının etkisini en aza indirmek için tüm metinler küçük harfe dönüştürülmüş, böylece dil modeli açısından homojen bir yapı oluşturulmuştur.

Dil bilgisel temizlik adımlarında, gereksiz veya anlam taşımayan kelimelerin çıkarılması amacıyla stop-word listeleri kullanılarak anlamsız kelimeler filtrelenmiştir. Bu sayede modelin sadece duygu ifadeleri açısından anlam taşıyan kelimeleri dikkate alması sağlanmış; böylece analiz sürecinde anlam derinliği güçlendirilmiştir. Bu aşamalar, modelin doğru özellikleri daha sağlıklı öğrenmesine ve nihai duygu analiz sonuçlarının daha güvenilir hale gelmesine katkıda bulunmuştur.

Verinin etkili bir şekilde hazırlanmasının ardından, veri seti belirli oranlarda eğitim ve test gruplarına ayrılmıştır. Eğitim ve test süreçlerinin bağımsız olarak sürdürülebilmesi için veri setinin %80'i eğitim, %10'u doğrulama (validation), %10'u ise test verisi olarak ayrılmıştır. Eğitim sürecinde doğrulama verisi, modelin genelleme kapasitesini ve doğruluğunu gözlemlemekte kullanılmıştır. Erken durdurma (early stopping) yöntemi uygulanarak, doğrulama verisinde performans düşüşü başladığında eğitim sonlandırılmış, böylece modelin yalnızca eğitim verisine aşırı uyum sağlamasının önüne geçilmiştir. Bu

yaklaşım, test verisi üzerinde duygu analizi doğruluğunu ve güvenilirliğini en üst düzeye çıkarmak amacıyla düzenlenmiştir.

4.3.2. BERT Modelinin Eğitimi

Bu çalışmada, Türkçe duygu analizi amacıyla dönüştürücü(transformer) tabanlı bir model olan BERT kullanılmıştır. Model olarak, büyük ölçekli Türkçe metin korpusları üzerinde önceden eğitilmiş olan BERTurk modeli tercih edilmiştir. BERTurk, Türkçe dilinin anlamsal ve sentaktik özelliklerini etkin bir şekilde yakalayabilen bir dil modelidir.

Kendi oluşturduğumuz Türkçe duygu analizi veri seti kullanılarak modelin ince ayar (fine-tuning) işlemi gerçekleştirilmiştir. Veri seti, sosyal medya paylaşımları, haber metinleri ve kullanıcı yorumları gibi çeşitli kaynaklardan toplanan ve üç duygu sınıfına (negatif, nötr, pozitif) etiketlenmiş metinlerden oluşmaktadır. Metinler, BERTurk modelinin gerektirdiği biçimde ön işlemeye tabi tutulmuştur. Bu kapsamda, WordPiece tokenizasyon yöntemi kullanılarak metinler parçalara ayrılmış, maksimum sekans uzunluğu 128 token ile sınırlandırılmış ve [CLS] ile [SEP] gibi özel tokenler metinlere eklenmiştir.

BERT modelinin eğitimi sırasında, son sınıflandırma katmanı üç duygu sınıfını tahmin edecek şekilde yeniden yapılandırılmıştır. Hiperparametreler olarak öğrenme oranı 2e-5 (0,00002) olarak belirlenmiş, AdamW optimizasyon algoritması ve 0,01 ağırlık çürümesi (weight decay) kullanılmıştır. Yığın boyutu 16 seçilmiş ve model üç döngü (epoch) boyunca eğitilmiştir. Seyreltme (dropout) oranı 0,1 olarak uygulanmış, eğitim sırasında erken durdurma ve öğrenme oranı azaltma stratejileri kullanılmıştır. Modelin genelleme kabiliyetini değerlendirmek ve aşırı öğrenmeyi önlemek amacıyla, veri seti rastgele ve eşit büyüklükte 10 alt kümeye bölünerek 10 katlı çapraz doğrulama yöntemi uygulanmıştır.

4.3.3. ALBERT Modelinin Eğitimi

Türkçe duygu analizi çalışmamızda ayrıca, parametre verimliliği ve hızlı eğitim avantajları nedeniyle ALBERT modeli kullanılmıştır. Model olarak, büyük ölçekli Türkçe metin korpusları üzerinde önceden eğitilmiş olan ALBERT-TR modeli tercih edilmiştir. ALBERT-TR, Türkçe dilinin anlamsal ve sentaktik özelliklerini etkin bir şekilde yakalayabilen hafif bir dil modelidir. Veri ön işleme adımları BERT modelindekiyle aynıdır. Metinler, WordPiece tokenizasyon yöntemi ile parçalara ayrılmış, maksimum sekans uzunluğu 128 token olarak belirlenmiş ve özel tokenler metinlere eklenmiştir.

ALBERT modelinin eğitimi sırasında, son sınıflandırma katmanı üç duygu sınıfını tahmin edecek şekilde ayarlanmıştır. Hiperparametreler olarak öğrenme oranı 1e-5 (0,00001) olarak belirlenmiş, AdamW optimizasyon algoritması kullanılmış ve yığın boyutu 32 seçilmiştir. Model, dört döngü (epoch) boyunca eğitilmiştir. Seyreltme oranı 0,1 olarak uygulanmış ve eğitim sırasında erken durdurma ile öğrenme oranı azaltma stratejileri kullanılmıştır.

Her iki modelin eğitimi sırasında, hiperparametre optimizasyonu için grid search yöntemi kullanılarak farklı öğrenme oranları, yığın boyutları ve döngü sayıları denenmiştir. Modellerin performansını ve genelleme kabiliyetini değerlendirmek amacıyla aynı 10 katlı çapraz doğrulama yöntemi ALBERT modeli için de uygulanmıştır.

4.3.4. Modellerin Performansı

Bu çalışmada, BERT ve ALBERT modellerinin Türkçe duygu analizi performansları 10 katlı çapraz doğrulama yöntemi kullanılarak değerlendirilmiştir. Modellerin doğruluk, kesinlik, duyarlılık ve F1 skoru metriklerinin ortalama değerleri aşağıdaki tabloda sunulmaktadır:

Tablo 3: Modellerin Performans Değerleri(10 Katlı Çapraz Doğrulama Ortalamaları)

Model	Doğruluk(%)	Kesinlik(%)	Duyarlılık(%)	F1 Skor(%)
BERT	83,75	83,30	83,25	83,20
ALBERT	94,85	94,50	94,4	94,30

Tablo 3'te görüldüğü üzere BERT modeli, Türkçe duygu analizi görevinde %83,20 ortalama F1 skoru ile başarılı bir performans sergilemiştir. Doğruluk oranı %83,75 olup, modelin genel olarak doğru sınıflandırmalar yaptığını göstermektedir. Kesinlik ve duyarlılık değerlerinin birbirine yakın olması, modelin pozitif ve negatif sınıflandırmalar arasında dengeli bir performans gösterdiğini ifade etmektedir. Bununla birlikte, bazı duygu sınıflarının dilsel ve anlamsal benzerlikleri nedeniyle modelin bu sınıfları ayırt etmede zorlanabileceği gözlemlenmiştir. ALBERT modeli, BERT'e kıyasla üstün bir performans sergileyerek %94,30 ortalama F1 skoruna ulaşmıştır. Doğruluk oranı %94,85 olan model, Türkçe duygu analizinde yüksek bir başarı elde etmiştir. Kesinlik ve duyarlılık değerleri, modelin hem doğru pozitif oranının yüksek olduğunu hem de gerçek pozitif örnekleri başarılı bir şekilde tespit ettiğini göstermektedir. ALBERT'in hafif ve verimli mimarisi sayesinde, daha derin bir dil anlayışı ve genelleme kabiliyeti kazanmıştır.

4.4. Çok Modlu Duygu Analizi Skorunun Üretilmesi

Şekil 5'de sunulan çok modlu duygu analizi akış diyagramı, video kaydından elde edilen ses, görüntü ve metin verilerinin işlenerek birleştirilmiş bir duygu skoru elde edilmesini açıklamaktadır. Bu analiz süreci, her bir veri türünün (modun) duygu tespitine katkısını dikkate alacak şekilde ağırlıklandırılarak nihai bir duygu skoru üretilmesi üzerine kuruludur.

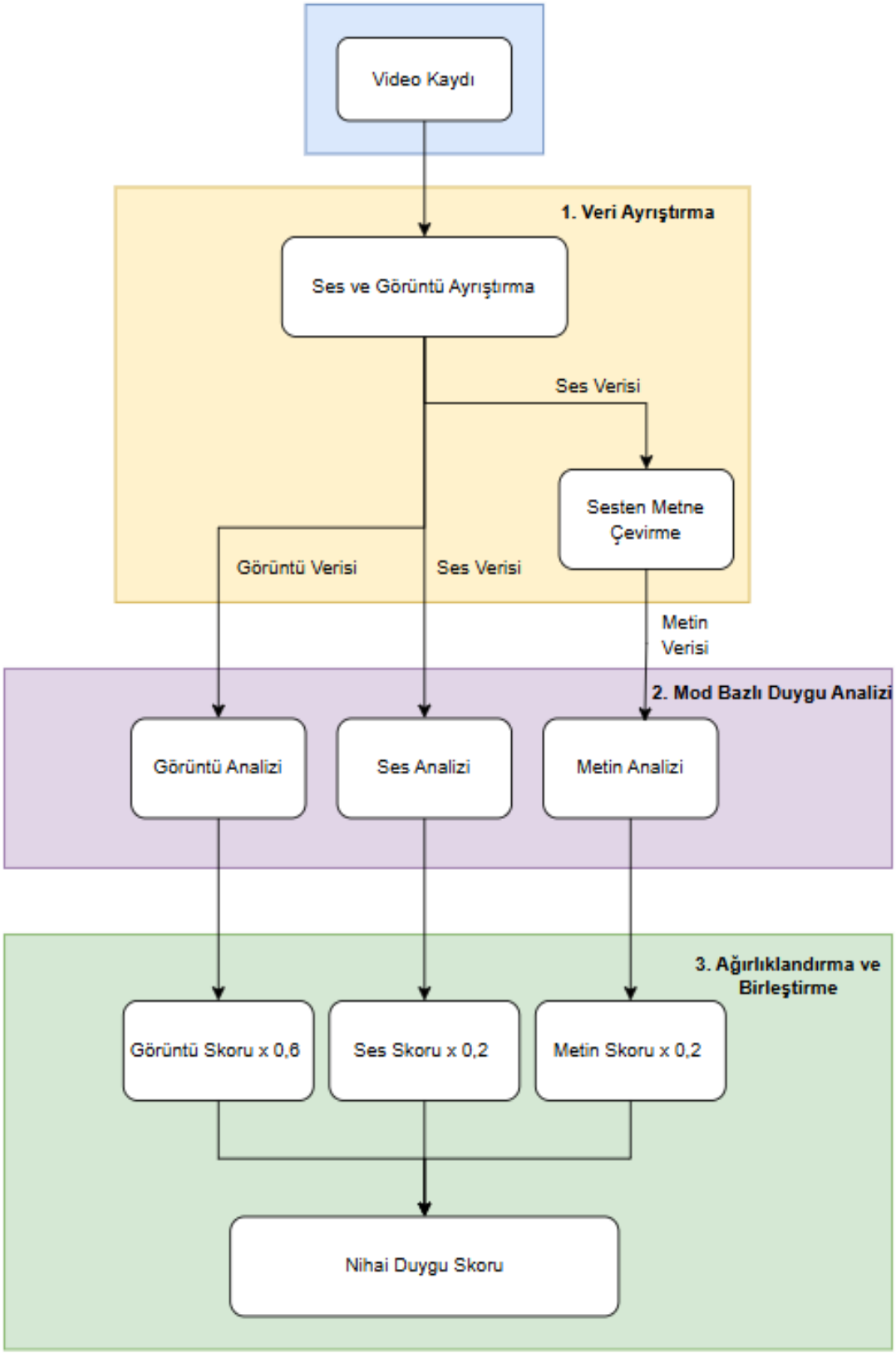
Bu işlem, üç ana aşamadan oluşmaktadır. İlk aşamada, video kaydından elde edilen görüntü ve ses verileri ayrıştırılır. Ayrıştırılan ses verisi hem doğrudan ses analizi için kullanılır hem de metin analizi için sesten metne dönüştürülür. Bu aşamadan elde edilen görüntü, ses ve metin verileri, ikinci aşamada ayrı ayrı analiz edilerek her bir veri türü için bir duygu skoru hesaplanır.

İkinci aşamada, her bir mod (görüntü, ses ve metin) üzerinden duygu analizi gerçekleştirilir. Görüntü analizi, yüz ifadelerindeki duygusal ipuçlarını inceleyerek görsel duygu skorunu üretir. Ses analizi, konuşmadaki tonlama, vurgu ve ritim gibi duygusal unsurları değerlendirir ve ses duygu skorunu elde eder. Metin analizi ise sesten metne dönüştürülmüş veriyi analiz ederek dildeki duygu ifadelerini tespit eder ve metin duygu skorunu üretir.

Üçüncü ve son aşamada ise, her bir moddan elde edilen duygu skorları, belirli katsayılarla ağırlıklandırılarak birleştirilir. Görüntü verisi, yüz ifadelerindeki ince duygusal farkları tespit etme kapasitesi nedeniyle en yüksek katkıya sahip olup 0,6 katsayısı ile ağırlıklandırılmıştır. Ses verisi, duygusal ipuçları sunmasına rağmen doğruluk oranı daha düşük olduğundan 0,2 katsayısı ile ağırlıklandırılmıştır. Benzer şekilde, metin verisi de duygu ifadelerini yansıtmakla birlikte daha sınırlı bir katkı sunduğundan 0,2 katsayısına sahiptir. Bu ağırlıklandırma işlemi ile her bir modun duygu sınıflandırmasındaki önemi dikkate alınarak -1 ile 1 arasında bir nihai duygu skoru elde edilir. Bu birleştirilmiş skor, analiz edilen kişinin duygusal durumunu daha doğru bir şekilde yansıtmayı amaçlamaktadır.

Bu çok modlu duygu analizi yöntemi, her bir veri türünün kendine özgü güçlü yönlerinden yararlanarak duygusal durumların daha bütüncül bir şekilde anlaşılmasını sağlar. Görüntü verisi, yüz ifadelerindeki duygusal değişimleri inceleyerek kişinin ruh hali hakkında ayrıntılı ipuçları sunarken, ses verisi de konuşma esnasındaki tonlama, vurgu, hız ve ritim gibi parametrelerle duygusal ton hakkında bilgi verir. Metin verisi ise konuşmanın içeriğinde yer alan kelime ve ifadeler aracılığıyla kişinin o anki hislerini

ortaya koyar. Her bir modun duygu analizi üzerindeki etkisi farklı olduğu için, görüntü analizinin daha baskın bir ağırlığa sahip olması, yüz ifadelerinin genellikle daha doğrudan ve net duygusal sinyaller taşımasıyla ilişkilidir. Ses ve metin verilerinin daha düşük ağırlıklarla birleştirilmesi ise, bu modların daha dolaylı duygusal göstergeler sunmasından kaynaklanmaktadır. Bu yapı sayesinde, analiz edilen kişinin duygusal durumunu sadece tek bir veri kaynağına dayanarak değerlendirmek yerine, tüm modları dikkate alan dengeli bir sonuç elde edilir. Böylece çok modlu duygu analizi süreçlerinde daha kapsamlı ve güvenilir bir duygu tespiti sağlanması hedeflenmiştir. Bu yaklaşım, müşteri duygularının değerlendirilmesi veya psikolojik analizler gibi alanlarda, insan duygularını daha objektif ve derinlemesine anlamak için kullanılabilir.



Şekil 5: Çok Modlu Duygu Analizi Akış Diyagramı

4.4.1. Örnek Birleştirme Süreci

Örneğin, 5 saniyelik bir videodaki duygu analizi sonuçları şu şekilde hesaplanabilir:

Videodaki yüz ifadelerinden çıkarılan duygu skoru 0.85 olarak hesaplanmıştır. Bu skor, görüntü verisi için 0.6 katsayısı ile çarpılacaktır: $0.85 \times 0.6 = 0.51$

Videodaki ses tonu ve vurgulardan çıkarılan duygu skoru -0.3 olarak hesaplanmıştır. Ses verisi için belirlenen 0.2 katsayısı ile çarpılacaktır: $-0.3 \times 0.2 = -0.06$

Videodaki konuşmadan çıkarılan duygu skoru 0.4 olarak hesaplanmıştır. Metin verisi için belirlenen 0.2 katsayısı ile çarpılacaktır: $0.4 \times 0.2 = 0.08$

Her modun ağırlıklı katkıları toplandığında, 5 saniyelik videonun nihai duygu skoru şu şekilde hesaplanır:

$0.51 + (-0.06) + 0.08 = 0.53$

Bu durumda, 5 saniyelik video için elde edilen duygu skoru 0.53'tür. Sonuca bakıldığında videodaki duygusal durumu pozitif bir şekilde yansıtmaktadır.

5. Sonuç ve Gelecek Çalışmalar

Bu çalışmada, müşteri memnuniyetini değerlendirmek amacıyla çok modlu duygu analizi yaklaşımı benimsenmiş ve ses, görüntü ile metin verileri belirli katsayılarla ağırlıklandırılarak tek bir duygu skoru elde edilmiştir. Görüntü verisinde kullanılan Xception modelinin üstün performansı, yüz ifadelerindeki mikro düzeydeki duygusal ayrıntıların etkili bir şekilde analiz edilmesini sağlamıştır. Bu bulgu, perakende sektöründe müşteri duyarlılığını ölçmek için yüz ifadelerinin güçlü bir gösterge olduğunu ortaya koymaktadır.

Ses verisinde, log-Mel spektrogramlarının kullanımı ve akustik özelliklerin detaylandırılması, ses tonlamaları ile ritimlerin model tarafından doğru algılanmasını sağlamış ve duygu tespitinde belirleyici bir katkı sunmuştur. Metin analizinde ise ALBERT modelinin yüksek performansı, bağlamsal anlamsal özelliklerin metinsel duygu tespitini iyileştirdiğini ve model doğruluğunu artırdığını göstermektedir.

Çalışmada her bir moda atanmış katsayıların belirlenmesi, sonuçları etkileyen temel bir parametre olarak öne çıkmaktadır. Görüntü verisine 0,6, ses ve metin verilerine ise 0,2 katsayılarının atanması, modların duygu tespitindeki başarılarına dayanmaktadır. Katsayıların belirlenme sürecinin daha detaylı bir şekilde açıklanması ve farklı katsayı senaryolarının analiz edilmesi, yöntemin metodolojik netliğini artırabileceği gibi sonuçların doğrulanmasına da katkı sağlayabilir. Ayrıca katsayıların farklı uygulama alanlarında nasıl uyarlanabileceğine dair yapılacak çalışmalar, yöntemin daha geniş bir uygulama yelpazesinde kullanılmasına rehberlik edebilir.

Genel olarak, bu çalışma, müşteri duygularını değerlendirme sürecinde çok modlu duygu analizinin potansiyelini ortaya koymakta ve duygu analizi alanına önemli bir katkı sunmaktadır. Gelecekte yapılacak araştırmalar, farklı veri setleri ve modlar arası katsayı optimizasyonu üzerine odaklanarak yöntemin doğruluğunu artırabilir ve genelleme kapasitesini genişletebilir. Ayrıca, maliyetleri azaltmaya yönelik model sıkıştırma tekniklerinin

incelenmesi, çok modlu duygu analizinin pratik uygulama olanaklarını artırabilir ve çeşitli sektörlerde daha yaygın bir şekilde benimsenmesini destekleyebilir.

Bu sonuçlar, çok modlu yaklaşımın müşteri duygularını değerlendirme ve duygu analizi çalışmalarında etkin bir yöntem olduğunu ortaya koymaktadır. Farklı modların güçlü yönlerini bir araya getiren bu yöntem, duygusal durumların daha kapsamlı ve güvenilir bir şekilde analiz edilmesine olanak tanımaktadır. Çalışmanın bulguları, özellikle perakende sektörü gibi müşteri etkileşiminin kritik olduğu alanlarda uygulanabilirliğe işaret etmekte ve müşteri deneyimini anlamada yenilikçi bir çözüm sunmaktadır.

Gelecekteki araştırmalarda, modlar arası katsayıların dinamik olarak belirlenmesi ve derin öğrenme modellerinin adaptif öğrenme yeteneklerinin kullanılması ile yöntemin esnekliği artırılabilir. Gerçek zamanlı duygu analizi uygulamaları için optimize edilmiş hafif modellerin geliştirilmesi, pratik uygulamalarda performansı yükseltebilir.

Kaynakça

Aydoğan, M., & Kocaman, V. (2022). TRSAv1: A new benchmark dataset for classifying user reviews on Turkish e-commerce websites. *Journal of Information Science, 49*(6),1711-1725. https://doi.org/10.1177/01655515221074328

Burkhardt, F., Paeschke, A., Rolfes, M., Sendlmeier, W. F., & Weiss, B. (2005). A database of German emotional speech. *Interspeech 2005*,1517-1520. https://doi.org/10.21437/interspeech.2005-446

Busso, C., Bulut, M., Lee, C., Kazemzadeh, A., Mower, E., Kim, S., Lee, S., Narayanan, A., & Narayanan, S. S. (2008). IEMOCAP: Interactive emotional dyadic motion capture database. *Language Resources and Evaluation*, *42*(4), 335-359. https://doi.org/10.1007/s10579-008-9076-6

Cao, H., Cooper, D. G., Keutmann, M. K., Gur, R. C., Nenkov, R., & Gur, R. E. (2014). CREMA-D: Crowd-sourced emotional multimodal actors dataset. *IEEE Transactions on Affective Computing*, *5*(4), 377-390. https://doi.org/10.1109/TAFFC.2014.2336244

Chollet, F. (2017, July 21-26). *Xception: Deep learning with depthwise separable convolutions* [Paper]. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) 1251-1258, Honolulu, HI, USA. IEEE. https://doi.org/10.1109/CVPR.2017.195

Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine Learning*, *20*(3), 273-297. https://doi.org/10.1007/BF00994018

Demirtaş, S. C., & Hakdağlı, Ö. (2022, Kasım, 24-26). *Dönüştürücü-CNN modeli ile Türkçe konuşma verisi üzerinde duygu tanıma* [Bildiri]. ELECO 2022 Elektrik - Elektronik ve Bilgisayar Mühendisliği Sempozyumu, Bursa, Türkiye. IEEE.

Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., & Fei-Fei, L. (2009). *ImageNet: A large-scale hierarchical image database* [Paper]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 248-255, Miami, FL, USA. IEEE. https://doi.org/10.1109/CVPR.2009.5206848

Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2019, June 2-7). *BERT: Pre-training of deep bidirectional transformers for language understanding* [Paper]. Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics:

Human Language Technologies, Minneapolis, MN, USA. Association for Computational Linguistics. https://doi.org/10.18653/v1/N19-1423

Dupuis, K., & Pichora-Fuller, M. K. (2010). *Toronto emotional speech set (TESS)* [Dataset] https://doi.org/10.5683/SP2/E8H2MF

Ekman, P., & Friesen, W. V. (1978*). Facial action coding system: a technique for the measurement of facial movement.* Consulting Psychologists Press. https://doi.org/10.1037/t27734-000

Goodfellow, I., Erhan, D., Carrier, P. L., Courville, A., Mirza, M., Hamner, B., Cukierski, W., Tang, Y., Thaler, D., Lee, D.-H., Zhou, Y., Ramaiah, C., Feng, F., Li, R., Wang, X., Athanasakis, D., Shawe-Taylor, J., Milakov, M., Park, J., & Bengio, Y. (2013). *Challenges in representation learning: A report on three machine learning contests*. In Proceedings of the Neural Information Processing Systems (NIPS) Workshop.

Huang, G. B., Ramesh, M., Berg, T., & Learned-Miller, E. (2007). *Labeled Faces in the Wild: A database for studying face recognition in unconstrained environments* [Paper]. Technical Report 07-49, University of Massachusetts, Amherst.

Jackson, P. J. B., & Haq, S.(2014). Surrey audio-visual expressed emotion(SAVEE) database, *University of Surrey.*

Lan, Z., Chen, M., Goodman, S., Gimpel, K., Sharma, P., & Soricut, R. (2020). *ALBERT: A lite BERT for self-supervised learning of language representations* [Paper]. International Conference on Learning Representations (ICLR), Addis Ababa, Ethiopia. https://doi.org/10.48550/arXiv.1909.11942

Livingstone, S. R., & Russo, F. A. (2018). The Ryerson audio-visual database of emotional speech and song (RAVDESS): A dynamic, multimodal set of facial and vocal expressions in North American English. *PLOS ONE, 13*(5), e0196391. https://doi.org/10.1371/journal.pone.0196391

Lotfian, R., & Busso, C. (2019, April). *Curriculum learning for speech emotion recognition from crowdsourced labels* [Paper]. *IEEE/ACM Transactions on Audio, Speech, and Language Processing, 27*(4), 815-826. https://doi.org/10.1109/TASLP.2019.2898816

Lucey, P., Cohn, J. F., Kanade, T., Saragih, J., Ambadar, Z., & Matthews, I. (2010). *The Extended Cohn-Kanade Dataset (CK+)* [Paper]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), San Francisco, CA, USA. https://doi.org/10.1109/CVPRW.2010.5543262

Lyons, M., Akamatsu, S., Kamachi, M., & Gyoba, J. (1998). *Coding facial expressions with Gabor wavelets* [Paper]. In Proceedings of the Third IEEE International Conference on Automatic Face and Gesture Recognition, 200-205, Nara, Japan. IEEE. https://doi.org/10.1109/AFGR.1998.670949

Pang, B., & Lee, L. (2008). Opinion mining and sentiment analysis. *Foundations and Trends in Information Retrieval, 2*(1-2), 1-135. https://doi.org/10.1561/1500000011

Poria, S., Cambria, E., Hazarika, D., & Mazumder, N. (2017, November 18-21). *Multi-level multiple attentions for contextual multimodal sentiment analysis* [Paper]. Proceedings of the IEEE International Conference on Data Mining (ICDM), New Orleans, LA, USA. IEEE. https://doi.org/10.1109/ICDM.2017.104

Simonyan, K., & Zisserman, A. (2015). *Very deep convolutional networks for large-scale image recognition* [Paper]. In International Conference on Learning Representations (ICLR), San Diego, CA, USA. https://doi.org/10.48550/arXiv.1409.1556

Sun, C., Huang, L., & Qiu, X. (2021, November 7-11). *Utilizing BERT and ALBERT models for sentiment analysis on social media text* [Paper]. Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing (EMNLP), Punta Cana, Dominican Republic. Association for Computational Linguistics. https://doi.org/10.18653/v1/N19-1035

Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., & Rabinovich, A. (2015). *Going deeper with convolutions* [Paper]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 1-9, Boston, MA, USA. IEEE. https://doi.org/10.1109/CVPR.2015.7298594

Xu, H., Liu, B., Shu, L., & Yu, P. S. (2020, June 2-7*). BERT post-training for review reading comprehension and aspect-based sentiment analysis* [Paper]. Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics (NAACL), Minneapolis, MN, USA. Association for Computational Linguistics. https://doi.org/10.18653/v1/N19-1242

Zadeh, A., Chen, M., Poria, S., Cambria, E., & Morency, L.-P. (2018, January-February). Multimodal sentiment intensity analysis in videos: Facial gestures and verbal messages. *IEEE Intelligent Systems*, *34*(1), 82-88. IEEE. https://doi.org/10.1109/MIS.2018.2888673