

Sinir Ağı Dil Modelleri ve Evrensel Cümle Kodlayıcı Kullanarak Havayolu Müşteri Yorumlarının Duygu Analizi

Semih Osman SAKA^{1*}, Zafer CÖMERT²

¹ Uzaktan Eğitim Uygulama ve Araştırma Merkezi, Samsun Üniversitesi, Samsun, Türkiye

² Yazılım Mühendisliği, Mühendislik ve Doğa Bilimleri Fakültesi, Samsun Üniversitesi, Samsun, Türkiye

*¹ semih.saka@samsun.edu.tr, ² zcomert@samsun.edu.tr

(Geliş/Received: 29/07/2024;

Kabul/Accepted: 04/10/2024)

Öz: Duygu analizi, metin tabanlı verilerin duygusal tonlarını belirlemede kullanılan önemli bir doğal dil işleme (NLP) tekniğidir. İşletmeler, müşteri memnuniyetini artırmak ve hizmet kalitesini iyileştirmek için müşteri yorumlarından elde edilen duygusal içgörülerle stratejik kararlar alabilir. Bu çalışmada ise havayolu müşteri yorumlarından oluşan bir veri seti kullanılmıştır. Veri seti, her bir yorumun doğruluk durumu, içerik, değerlendirme puanı, öneri durumu, duygu analizini içermektedir ve toplamda 1100 örnekten oluşmaktadır. Çalışmada dört farklı model incelenmiştir. Bu modellerden CNN-LSTM, sıfırdan öğrenme stratejisiyle eğitilmiştir. Ayrıca, iki farklı sinir ağı dil modeli (Neural Network Language Model, NNLM) ve evrensel cümle kodlayıcı (Universal Sentence Encoder, USE) transfer öğrenme yaklaşımıyla eğitilmiştir. CNN-LSTM modeli %92,06 doğruluk oranı ile yüksek performans göstermiştir. *nmlm-en-dim50* modeli %90,87 doğruluk oranı elde ederken, *nmlm-en-dim128* modeli %92,46 doğruluk oranı ile öne çıkmıştır. En yüksek performansı ise %95,63 doğruluk oranı ile USE modeli göstermiştir. Bu sonuçlar, derin öğrenme ve transfer öğrenme tekniklerinin duygu analizinde etkili araçlar olduğunu göstermektedir. Çalışma, işletmelerin müşteri memnuniyetini artırmak ve hizmet kalitesini iyileştirmek için duygu analizi teknolojilerini nasıl etkin bir şekilde kullanabileceklerine dair önemli içgörüler sunmaktadır. Gelecek çalışmalarda, farklı veri setleri ve daha geniş örneklem büyüklükleri ile modellerin performanslarının daha detaylı incelenmesi önerilmektedir.

Anahtar kelimeler: Derin öğrenme, yapay öğrenme, doğal dil işleme, duygu analizi.

Sentiment Analysis of Airline Customer Reviews Using Neural Network Language Models and Universal Sentence Encoder

Abstract: Sentiment analysis represents a fundamental technique within the domain of natural language processing (NLP), employed for the purpose of discerning the emotional tenor of text-based data. Businesses may make strategic decisions by employing emotional insights derived from customer reviews to enhance customer satisfaction and improve service quality. In this study, a dataset comprising reviews from airline customers was employed. The dataset comprises the verification status, content, rating score, recommendation status, and sentiment analysis of each review, with a total of 1100 examples. The study examined four distinct models. The CNN-LSTM model was trained using a training-from-scratch strategy. Furthermore, two distinct neural network language models (Neural Network Language Model, NNLM) and the Universal Sentence Encoder (USE) were trained using a transfer learning approach. The CNN-LSTM model exhibited robust performance, achieving an accuracy rate of 92.06%. The *nmlm-en-dim50* model achieved an accuracy rate of 90.87%, while the *nmlm-en-dim128* model demonstrated a notably higher level of accuracy at 92.46%. The USE model exhibited the highest performance, with an accuracy rate of 95.63%. These findings suggest that deep learning and transfer learning techniques are effective tools for sentiment analysis. The study offers valuable insights into how businesses can utilize sentiment analysis technologies to enhance customer satisfaction and service quality. It is recommended that future studies investigate the performance of these models with different datasets and larger sample sizes.

Key words: Deep learning, artificial learning, natural language processing, sentiment analysis.

1. Giriş

Duygu analizi (sentiment analysis), metin tabanlı verilerin duygusal tonunu belirlemek için kullanılan bir doğal dil işleme (NLP) tekniğidir. Özellikle sosyal medya, müşteri geri bildirimleri, ürün incelemeleri ve haber yorumları gibi büyük miktarda metin verisinin analizinde önemli bir rol oynar. İşletmeler, müşteri memnuniyetini artırmak ve hizmet kalitesini geliştirmek için bu yorumlardan elde edilen duygusal içgörülerini kullanarak stratejik kararlar alabilirler [1]. Duygu analizi, metin verilerindeki olumsuz, olumlu veya nötr duyguları otomatik olarak

* Sorumlu yazar: semih.saka@samsun.edu.tr. Yazarların ORCID Numarası: ¹ 0000-0002-6241-5485, ² 0000-0001-5256-7648

tespit ederek işletmelere, müşteri deneyimlerini daha derinlemesine anlama ve buna yönelik iyileştirmeler yapma imkânı sunar. Bu sayede, işletmeler müşteri memnuniyetini artırmaya yönelik stratejik kararlar alabilirler [1].

Son yıllarda, derin öğrenme algoritmalarının gelişimiyle birlikte duygu analizinde önemli ilerlemeler kaydedilmiştir. Derin öğrenme, büyük veri kümelerinden karmaşık örüntüleri ve ilişkileri öğrenme kapasitesi sayesinde, duygu analizinin doğruluğunu ve etkinliğini artırmıştır. Makine öğrenmesi, derin öğrenme özellikle Convolutional Neural Networks (CNN), Long Short-Term Memory (LSTM), dikkat mekanizmaları ve hibrit yaklaşımlarla metin verilerinin duygusal analizinde yüksek performanslar elde edilmiştir. Bazı önemli çalışmalara bu bölüm altında kısaca yer verilmiştir. Uslu ve Akyol Türkçe haber metinlerini Destek Vektör Makineleri (SVM), Rastgele Orman (RF) ve Naive Bayes (NB) sınıflandırıcılarıyla analiz etmiş; bu analizde %91 doğruluk oranıyla en iyi performansı NB göstermiştir [2]. Tuzcu çevrimiçi kitap satış sitesinin kullanıcı yorumları üzerinde duygu analizi yapmıştır. Python ile MLP algoritması ve RapidMiner ile NB, SVM ve LR algoritmaları uygulanmıştır. 1400 yorumla oluşturulan veri setinde, MLP %89, NB %77,57, SVM %80,93 ve LR %84,07 doğruluk oranı elde etmiştir. Daha doğru analizler için demografik bilgiler ve daha büyük veri setlerinin kullanımı önerilmiştir [3]. Acı ve Çirak CNN ve Word2Vec kullanarak TTC-3600 veri kümesi üzerinde gerçekleştirdikleri metin sınıflandırma çalışmasında, Python ve Tensorflow ile kodladıkları modellerle gövdelenmiş veride %93,3, ham veride ise %90,1 doğruluk oranı elde etmişlerdir. CNN ve Word2Vec, klasik makine öğrenmesi yöntemlerine kıyasla daha yüksek bir performans göstermiştir [4]. Demircan ve arkadaşları hepsiburada.com'dan alınan ürün incelemeleri ve puanları kullanarak Türkçe duygu analizi modelleri geliştirmiştir. Olumlu, olumsuz ve nötr müşteri yorumları SVM, RF, karar ağacı (DT), LR ve k-en yakın komşu (kNN) algoritmalarıyla sınıflandırılmıştır. SVM ve RF en iyi performansı göstermiştir [5]. Chen ve ekibi ABD yüksek mahkeme belgelerini sınıflandırmak için makine öğrenmesi ve derin öğrenme yöntemlerini karşılaştırmıştır. 30.000 dava belgesinde, RF modeli Word2Vec, GloVe, BERT tabanlı TextCNN, BiLSTM ve BiLSTM-Attention yöntemlerine karşı üstün performans sergilemiştir. RF, doğruluk (accuracy), geri çağırma (recall), kesinlik (sensitivity) ve F1 skoru açısından daha başarılı sonuçlar elde etmiştir [6]. Barfar çevrimiçi propagandanın tespiti için dilbilimsel ve oyun teorisine dayalı bir model geliştirmiştir. 205.000 makaleden oluşan veri seti, LightGBM algoritması ile C-indeksi 0,9 ve F1 skoru 0,84 ile yüksek performans göstermiştir. Model, propagandist içerik tespitinde dilsel özelliklerin önemini vurgulamaktadır [7]. Liu ve arkadaşları kısa metinler için CRFA yöntemini önermiştir. CRFA, CNN ve Recurrent Neural Network (RNN) tabanlı yöntemlere kıyasla daha yüksek doğruluk sağlamıştır [8]. Colón-Ruiz ve Segura-Bedmar ilaç incelemelerinde duygu analizi yapmak için CNN, Bi-LSTM, CNN ve LSTM kombinasyonlarını, ayrıca BERT ve Bi-LSTM modellerini kullanmışlardır. BERT tabanlı modeller en iyi performansı göstermiş, ancak bu modellerin eğitim süreleri daha uzun olmuştur [9]. Lyu ve Liu, metin sınıflandırmada hibrit modeller geliştirmiştir. CNN ve RNN kullanarak iki Çince ve beş İngilizce veri seti üzerinde çalışmışlardır. Önerilen yöntem, FastText ve HAN yöntemlerine kıyasla daha yüksek doğruluk sağlamıştır [10]. Onan ve ekibi anahtar kelime çıkarma yöntemlerinin metin sınıflandırma ve topluluk öğrenme üzerindeki etkisini incelemiştir. Reuters-21578 veri seti kullanılarak yapılan çalışmada, Bagging topluluk öğrenimi ve rastgele orman algoritması ile en yüksek tahmin performansı %93,80 olarak elde edilmiştir [11]. Göker ve Tekdere FATİH projesi ile ilgili 444 yorumu sınıflandırmıştır. NB, kNN, Karar Ağacı (DT), Sequential Minimal Optimization (SMO) ve RBF Network algoritmaları uygulanmış ve en yüksek doğruluk oranı SMO algoritması ile elde edilmiştir [12]. Kılınc ve arkadaşları akademik makaleleri sınıflandırmıştır. 2000 makale özeti iki sınıfta toplanmış ve %96,67 doğruluk oranı ile kNN algoritması en iyi performansı göstermiştir [13]. Kowsari ve arkadaşları HDLTex yöntemiyle derin sinir ağları kullanarak 46.985 makale veri setinde belgeleri sınıflandırmıştır. Derin öğrenme, geleneksel yöntemlere göre performansı artırmıştır [14]. Alparslan ve Dursun Türkçe metinleri sınıflandırmak için CNN kullanmıştır. TTC-4900 ve MY-15130 veri setlerinde CNN tabanlı model, NB, kNN, SVM ve RF'ye göre daha başarılı olmuştur [15]. Guo ve arkadaşları, metin sınıflandırmada Etiket Karmaşası Modeli (LCM) önermiştir. Üç İngilizce ve iki Çince veri setinde LCM, mevcut modellerin performansını artırmıştır [16]. Alzoubi ve arkadaşları Türkçe müşteri taleplerini sınıflandırmıştır. 225239 müşteri talebi verisinde LSTM algoritması %84 doğruluk oranı ile en yüksek performansı sağlamıştır [17]. Kaşıkçı ve Gökçen e-ticaret sitelerini sınıflandırmıştır. kNN ve NB kullanılarak iki sınıf oluşturulmuş, NB %85,30 doğruluk oranı elde etmiştir [18]. Hajibabae ve ekibi sosyal medyada saldırgan dil tespiti için TF-IDF, Word2Vec ve FastText gömme yöntemlerini kullanarak çeşitli sınıflandırıcıları karşılaştırmıştır. TF-IDF ile AdaBoost, SVM ve MLP modelleri en yüksek F1 skorunu elde etmiştir [19]. Gürbüz ve arkadaşları oteller için çevreci yorumları sınıflandırmıştır. 211.790 yorumda karar ağaçları %94,59 başarı oranı ile en iyi performansı göstermiştir [20]. Neogi ve arkadaşları çiftçi protestoları ile ilgili tweetleri sınıflandırmıştır. Bag of Words ve TF-IDF teknikleri kullanılarak RF %96,6 doğruluk oranı ile en başarılı olmuştur [21]. Çelik ve Koç Türkçe haber sitelerinden elde edilen 12000 haber metnini sınıflandırmıştır. FastText vektör modeli ve SVM ile %95,75 başarı oranı elde edilmiştir [22]. Soni ve ekibi TextConvoNet adını verdikleri yeni bir CNN tabanlı model önermiştir. Bu model, Binary SST-2, Amazon Review, R8, Twitter User Airline Sentiment ve Covid Tweets

veri setlerinde mevcut modellere kıyasla daha üstün performans sergilemiştir [23]. Jang ve arkadaşları IMDB film yorumları üzerinde Bi-LSTM+CNN hibrit modeli önermiştir. Önerilen model, yüksek doğruluk ve F1 skoru sağlamıştır [24]. Lai ve arkadaşları Recurrent Convolutional Neural Network (RCNN) ile 20Newsgroups ve Fudan Set veri setlerinde metin sınıflandırma hata oranını %33 ve %19 düşürmüştür [25]. Gasparetto ve arkadaşları NB, SVM, FastText, BiLSTM, XML-CNN, BERT ve XML-R modelleri ile EnWiki-100 ve RCV1-57 veri setlerinde BERT ve XLNet'in en iyi performansı elde ettiğini tespit etmiştir [26]. Chen ve arkadaşları Lao dilindeki 2411 haber makalesini kNN ile sınıflandırmıştır. %71,4 doğruluk oranı ile veri normalizasyonu ve boyut azaltma işlemleri performansı artırmıştır [27]. Bahassine ve arkadaşları Arapça metinleri sınıflandırmak için ImpCHI modelini önermiştir. 5070 belge ile ImpCHI ve SVM %90,50 F1 skor değeri ile en iyi sonucu sağlamıştır [28]. Reusens ve arkadaşları beş görevi kapsayan 20 veri seti ve 11 model mimarisi ile basit modellerin bazı durumlarda karmaşık modellerden daha iyi performans gösterdiğini bulmuştur [29]. Wang ve arkadaşları N-gram ve CNN teknolojisini birleştirerek kısa metinler için daha yüksek doğruluk oranları elde etmiştir [30]. Bhattacharya ve ekibi, zararlı yorumları sınıflandırmak için LSTM ve CNN tabanlı hibrit bir model kullanmış ve %98,91 doğruluk oranıyla en yüksek performansı elde etmiştir [31]. İlkuçar ve Artun beş yıldızlı Fethiye otellerine ait müşteri yorumlarını analiz ederek, duygu durumlarını anlamak için makine öğrenmesi tekniklerini kullanmışlardır. 2022 yılında Tripadvisor ve Otelpuan'dan alınan 1160 yorumu içeren bir veri seti kullanılmıştır. Çalışmada, yorumlar üzerinde duygu analizi yapmak için ANN, LSTM ve CNN modelleri uygulanmıştır. Sonuçlar, müşterilerin genel olarak otel personeli ve otelin konumları hakkında olumlu duygulara sahip olduğunu göstermektedir. Modeller arasında ANN, %81 doğruluk oranıyla en iyi performansı sergilerken, CNN %73, LSTM ise %72 doğruluk oranı ile takip etmiştir [32].

Bu çalışma, derin öğrenme ve makine öğrenmesi tekniklerini kullanarak duygu analizi alanında önemli katkılar sağlamayı hedeflemektedir. Öncelikle, farklı derin öğrenme ve makine öğrenmesi modellerinin müşteri yorumları üzerinde nasıl performans gösterdiğini kapsamlı bir şekilde karşılaştırarak, hangi yöntemlerin hangi koşullarda daha etkili olduğunu belirlemektedir. Ayrıca, sıfırdan öğrenme ve transfer öğrenme stratejilerinin karşılaştırmalı analizi, işletmelerin veri kısıtlamaları altında en uygun çözümü seçmelerine yardımcı olma potansiyeli taşır. Bu çalışma, aynı zamanda hibrit modellerin duygu analizindeki etkinliğini ortaya koyarak, daha yüksek doğruluk oranları ve verimlilik sağlamayı amaçlamaktadır.

Makalenin geri kalan bölümleri şu şekilde düzenlenmiştir: İkinci bölümde materyal ve yöntemler, üçüncü bölümde bulgular, dördüncü bölümde tartışma, beşinci bölümde ise sonuç ve öneriler yer almaktadır.

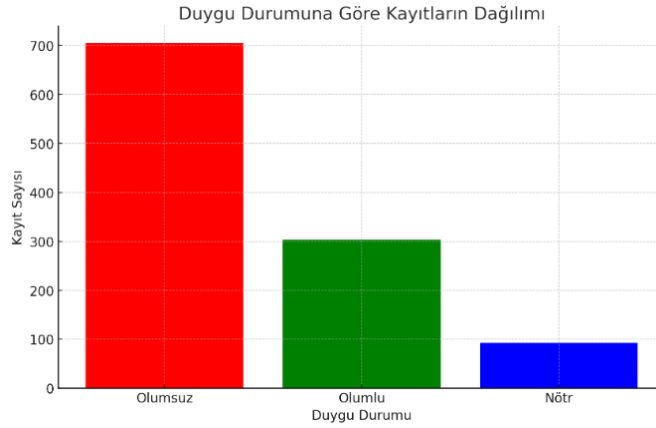
2. Materyal ve Metotlar

Bu bölüm, çalışmada kullanılan veri seti, derin öğrenme ve makine öğrenmesi modelleri ile deneysel prosedürleri detaylandırmaktadır. Çalışmanın güvenilirliğini ve geçerliliğini sağlamak amacıyla kullanılan yöntemler ve uygulamalar ayrıntılı olarak açıklanmıştır.

2.1. Veri Seti

Veri seti, hava yolu deneyimlerine ait müşteri yorumlarını içermektedir ve her bir satırda bir müşterinin doğrulama durumu, yorum içeriği, değerlendirme puanı, firmayı önerip önermediği ve duygu durumu bilgileri yer almaktadır. Veriler araştırmacıların kullanımı için açık erişimli olarak Mendeley Data platformunda yayımlanmıştır. “*Verify*” sütunu müşteri yorumunun doğruluğunu, “*Review*” sütunu müşteri yorum metnini, “*Rating*” sütunu müşterinin verdiği puanı, “*Recommended*” sütunu müşterinin havayolunu tavsiye edip etmediğini ve “*Sentiment*” sütunu yorumun duygu analizini göstermektedir. Bu veri seti, müşteri memnuniyeti ve havayolu hizmet kalitesini değerlendirmek için zengin bir bilgi kaynağı sunmaktadır [33].

Veri setinin duygu durumu dağılımı incelendiğinde, toplamda 1100 kaydın bulunduğu gözlemlenmiştir. Şekil 1'de kayıtların sınıflar arası dağılımına yer verilmiştir. Burada olumsuz ifadesi müşterinin memnun olmadığı; olumlu ise müşterinin memnun olduğu duygu durumunu ifade etmektedir. Olumsuz ya da olumlu olarak ifade edilemeyen müşteri yorumları ise nötr duygu durumu altında değerlendirilmiştir. Kayıtların 705'i olumsuz, 303'ü olumlu ve 92'si nötr duygu durumunu yansıtmaktadır. Bu dağılım, müşteri yorumlarının büyük çoğunluğunun olumsuz veya olumlu olduğunu göstermektedir. Olumsuz duygu durumu en yüksek orana sahipken, olumlu ve nötr yorumlar daha az sıklıkla görülmektedir. Bu bulgu, havayolu hizmet kalitesinin genel olarak müşteri beklentilerini karşıladığı, ancak iyileştirme alanlarının da mevcut olduğunu ortaya koymaktadır.



Şekil 1. Veri setindeki örneklerin duygu durumu dağılımı.

Deneyisel çalışmalar sırasında nötr örneklerin sayısının derin öğrenme ile anlamlı bir analiz yapmaya yetmeyecek kadar az olduğu tespit edildiğinden, bu örnekler veri setinin genel dengesini ve sonuçların güvenilirliğini olumsuz etkilememesi amacıyla analizlerden çıkarılmıştır. Sonuç olarak 705 olumsuz ve 303 olumlu örnek olmak üzere toplam 1008 örnek deneysel çalışmalar kapsamında kullanılmıştır.

Tablo 1. Veri setine ait örnek bir görünüm.

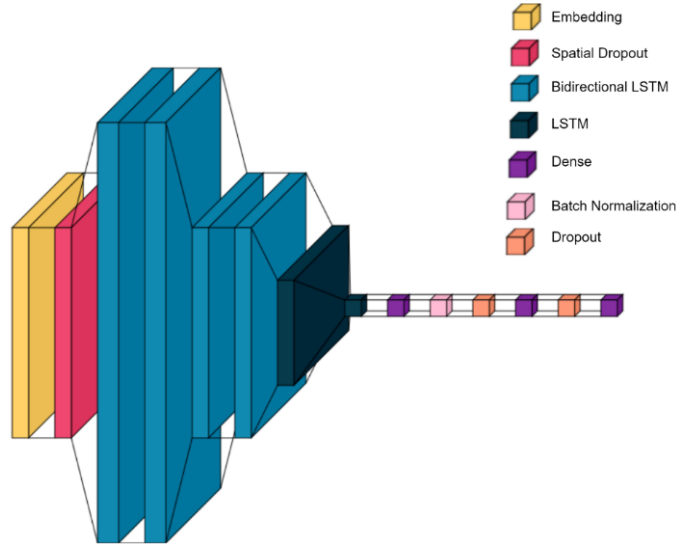
Verify	review	rating	recommended	sentiment
Yes	Tried booking 3 tickets ...	2	no	0
Yes	Flew Business Class with my wife...	1	no	0
Yes	Besides losing my baggage, I arrived ...	1	no	0
Yes	Air Canada is the only airline flying non-stop ...	9	yes	2
Yes	Bereavement Policy is a scam. First sentence ...	1	no	0

Tablo 1, çalışmada kullanılan veri setine ait örnek bir görünüm sunmaktadır. Her bir kayıt için doğrulama (verify), yorum (review), oylama (rating), öneri (recommended) ve duygu durumu (sentiment) gösterilmektedir.

2.2. CNN-LSTM Modeli

CNN, özellikle görüntü işleme ve bilgisayarla görme alanlarında yaygın olarak kullanılan bir derin öğrenme mimarisidir. CNN'ler, verilerin yerel özelliklerini yakalamak için özel olarak tasarlanmış katmanlar içerir. Temel yapı taşları arasında evrişim katmanları, aktivasyon fonksiyonları (genellikle ReLU), havuzlama katmanları ve tam bağlantılı katmanlar bulunur. Evrişim katmanları, görüntüdeki küçük bölgelere filtreler uygulayarak özellik haritaları oluşturur. Havuzlama katmanları, bu özellik haritalarının boyutunu azaltarak hesaplama yükünü ve aşırı uyumu azaltır. Genellikle maksimum havuzlama veya ortalama havuzlama kullanılır. Tam bağlantılı katmanlar, sınıflandırma veya regresyon görevleri için nihai çıktıyı üretir. Bu katmanlar, tüm özellik haritalarını düzleştirerek tek bir vektör halinde giriş alır ve son karar katmanına iletir [34-36]. CNN'ler, özellikle büyük veri kümeleri ve güçlü hesaplama kaynakları ile yüksek doğruluk oranları sağlar. Şekil 2'de önerilen özelleştirilmiş model ve bu modelin içerdiği katmanlara yer verilmiştir.

Bu modeldeki en önemli katman uzun-kısa süreli bellek (LSTM) katmanıdır. LSTM katmanı, özellikle sıralı veri ile çalışan derin öğrenme modellerinde yaygın olarak kullanılan bir tür RNN katmanıdır. LSTM'ler, uzun vadeli bağımlılıkları öğrenme ve koruma yetenekleri sayesinde zaman serileri, metin verileri ve diğer ardışık veriler üzerinde başarılı performans gösterirler [37-39].



Şekil 2. CNN-LSTM modelinin mimarisi.

LSTM'nin temel birimi olan hafıza hücresi, bilgiyi zaman içinde saklar ve manipüle eder. Hücre durumu C_t olarak adlandırılır. LSTM, bilgiyi hücre durumu boyunca kontrol etmek için üç farklı kapı kullanır: (1) Unutma Kapısı (*Forget Gate*): Hücre durumundan hangi bilgilerin atılacağını belirler. (2) Giriş Kapısı (*Input Gate*): Yeni bilgi eklemek için hücre durumunu günceller. (3) Çıkış Kapısı (*Output Gate*): Hücre durumunun hangi kısmının çıktısı olarak kullanılacağını belirler. Her kapı, sigmoid aktivasyon fonksiyonu (σ) ve bir ağırlık matrisi kullanarak hesaplanır. Unutma kapısı f_t Denklem 1'deki gibi hesaplanır:

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \quad (1)$$

Burada W_f unutma kapsının ağırlık matrisi, b_f unutma kapısının bias vektörü, h_{t-1} önceki gizli durum, x_t şu anki girdi ve σ sigmoid aktivasyon fonksiyonudur. Giriş kapısı \tilde{C}_t ise Denklem 2'deki gibi hesaplanır:

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \quad (2)$$

$$\tilde{C}_t = \tan h(W_C \cdot [h_{t-1}, x_t] + b_C)$$

Burada W_i giriş kapısının ağırlık matrisi, b_i giriş kapısının bias vektörü, W_C hücre durumu adaylarının ağırlık matrisi, b_C hücre durumu adaylarının bias vektörü, $\tan h$ hiperbolik tanjant fonksiyonudur. Hücre durumu C_t ise Denklem 3'deki gibi güncellenir:

$$C_t = f_t \cdot C_{t-1} + i_t \cdot \tilde{C}_t \quad (3)$$

Çıkış kapısının çıktısı o_t ve güncellenmiş gizli durum h_t Denklem 4'deki gibi hesaplanır:

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \quad (4)$$

$$h_t = o_t \cdot \tan h(C_t)$$

Burada W_o çıkış kapısının ağırlık matrisi, b_o çıkış kapısının bias vektörüdür.

2.3. NNLM

Neural Network Language Model (NNLM), kelimelerin bağlamını dikkate alarak dil modelleri oluşturan bir yapay sinir ağı modelidir. Kelime vektörlerini kullanarak doğal dil işleme görevlerinde anlamlı ve ilişkili temsiller sağlar [10]. Bu kapsamda iki farklı model çalışmada kullanılır:

- 1) *nmlm-en-dim50*, Google'ın TensorFlow Hub'da sağladığı modeli, İngilizce dilinde 50 boyutlu bir metin gömme (embedding) modelidir. Bu model, metin verilerini 50 boyutlu yoğun vektörlere dönüştürerek NLP görevlerinde kullanılır. Model, geniş bir metin veri kümesi üzerinde eğitilmiştir ve kelime anlamlarını temsil eden vektörler oluşturur. Bu vektörler, metin sınıflandırma, duygu analizi ve benzeri NLP görevlerinde kullanılabilir.
- 2) *nmlm-en-dim128*, Google tarafından geliştirilen ve NLP görevlerinde kullanılan bir dil modelidir. Bu model, İngilizce metinlerdeki kelimeleri ve cümleleri anlamlandırmak için NNLM mimarisini kullanır ve her kelimeyi 128 boyutlu bir vektörle temsil eder. *nmlm-en-dim128*, kelime vektörlerini oluştururken kelimelerin bağlamını dikkate alarak daha anlamlı ve ilişkili temsiller üretir. Bu model, metin sınıflandırma, duygu analizi, makine çevirisi ve diğer NLP görevlerinde etkili bir şekilde kullanılabilir. Eğitim sürecinde geniş bir İngilizce metin veri kümesi kullanılarak kelimeler arasındaki anlamsal ilişkiler öğrenilmiştir. 128 boyutlu vektörler, modelin hafızasını ve işlem süresini optimize ederek hızlı ve verimli sonuçlar elde etmeyi sağlar. *nmlm-en-dim128*, özellikle dil anlama ve metin işleme uygulamalarında yüksek performans göstermesiyle bilinir.

2.4. USE

Universal Sentence Encoder (USE), metni yüksek boyutlu vektörlere kodlayan bir modeldir. Bu vektörler, cümlelerin anlamsal bilgisini yakaladıkları için çeşitli NLP görevlerinde kullanışlıdır. USE, derin öğrenme tekniklerini kullanarak cümle benzerliği, kümeleme ve sınıflandırma gibi görevler için etkili bir şekilde kullanılabilir. Cümle gömme fonksiyonu Denklem 5'deki gibi tanımlanır:

$$E = f(S) \quad (5)$$

Burada E gömme vektörünü, S giriş cümlesini ve f model tarafından öğrenilen gömme fonksiyonunu temsil eder. Her bir cümle S bir dizi token $\{t_1, t_2, \dots, t_n\}$ olarak tokenize edilir, burada n cümledeki token sayısıdır. Tokenler, bir gömme matrisi W kullanılarak token gömme matrisine dönüştürülür:

$$T_i = W \cdot t_i \quad (6)$$

Denklem 6'da T_i , i - inci tokenin gömüsüdür. Token gömme matrisi, bağlamsal gömme matrisi üretmek için çeşitli sinir ağı katmanlarından (örn. Transformer veya Deep Averaging Network) geçirilir:

$$C_i = NN(T_i) \quad (7)$$

Denklem 7'de C_i , i - inci tokenin bağlamsal gömesidir. Token gömme matrisi, tek bir sabit boyutlu vektöre dönüştürmek için bir havuzlama mekanizması (örn. ortalama veya maksimum havuzlama) Denklem 8'deki gibi uygulanır:

$$E = \text{pooling}(\{C_1, C_2, \dots, C_n\}) \quad (8)$$

Ortalama Havuzlama, Denklem 9'da tüm token gömmelerinin ortalamasını alır:

$$E = \frac{1}{n} \sum_{i=1}^n C_i \quad (9)$$

Maksimum Havuzlama, her boyut için tüm token gömmeleri arasından maksimum değeri alır:

$$E_j = \max_{i \in \{1, \dots, m\}} C_{i,j} \quad (10)$$

Denklem 10’da E_j , gömme vektörü E_j ’nin j – inci boyutudur.

3. Bulgular

Deneyisel çalışmalar, NVIDIA P6000 24 GB ekran kartına ve Intel (R) Xeon (R) Gold 6132 3.2 GHz işlemciye sahip bir iş istasyonu üzerinde gerçekleştirilmiştir. Python 3.10.9 programlama dili kullanılmıştır.

Tablo 2. Performans metrikleri.

Metrikler	Formula	Kısa açıklama
Doğruluk	$\frac{DP + DN}{DP + YP + YN + DN}$	Modelin tüm tahminlerinin ne kadar doğru olduğunu, doğru sınıflandırılan verilerin toplam veriye oranı ile gösterir.
Hassasiyet	$\frac{DP}{DP + YN}$	Pozitif sınıfların ne kadar iyi tahmin edildiğini, doğru pozitif tahminlerin toplam pozitif tahminlere oranı ile ölçer.
Özgüllük	$\frac{DN}{DN + YP}$	Negatif sınıfların ne kadar iyi tanındığını, doğru negatif tahminlerin toplam negatiflere oranı ile ifade eder.
F1 Skor	$\frac{2DP}{2DP + YP + YN}$	Hassasiyet ve geri çağırmanın dengeli bir ölçüsü olup, doğru pozitif tahminlerin ortalama performansını değerlendirir.

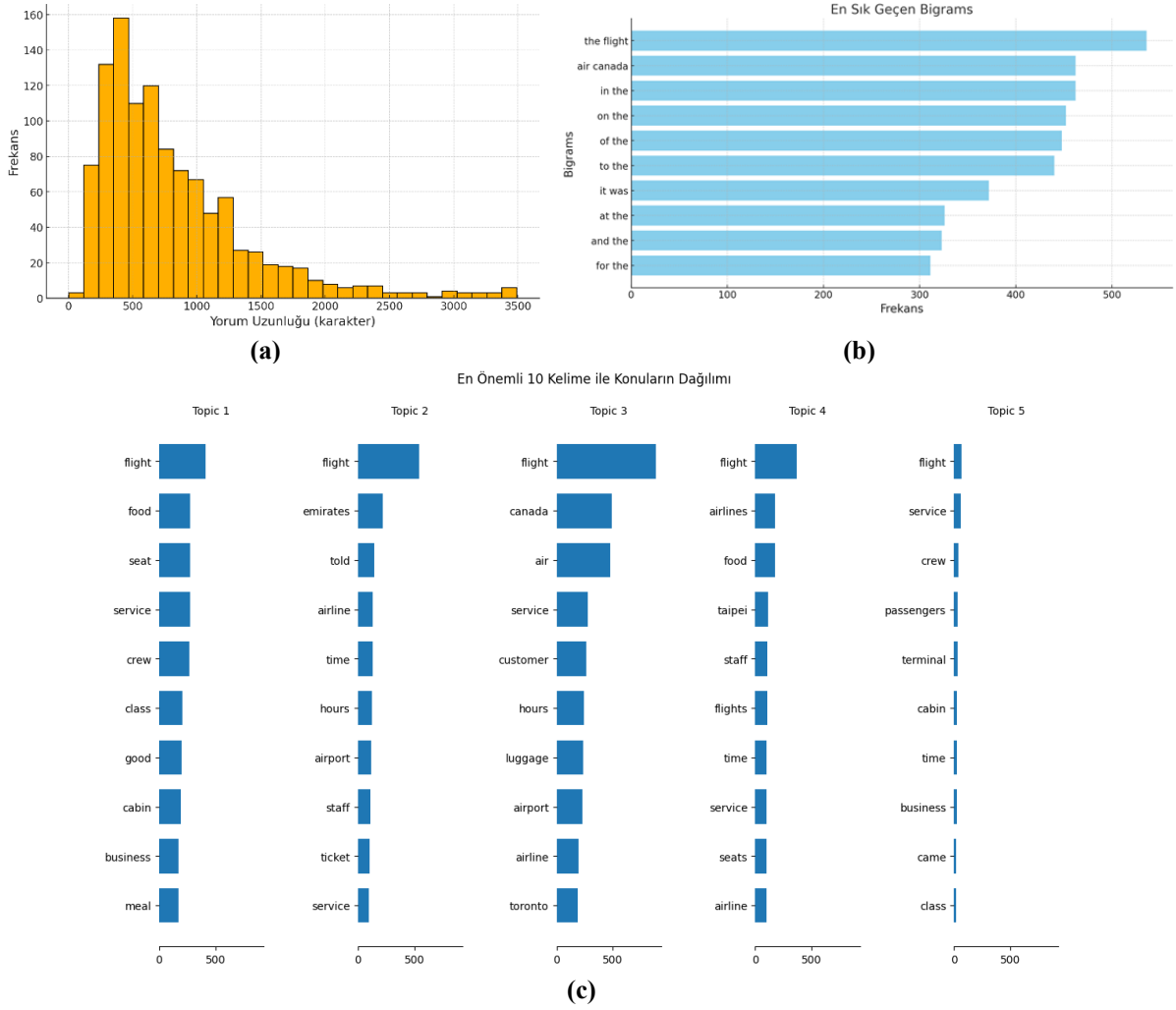
Tablo 2’de model performansını değerlendirmek üzere kullanılan performans metrikleri tanımlanmıştır. Doğru pozitif (DP), çalışmamızda olumsuz yorumların doğru bir şekilde tespit edildiği durumu temsil eder. Doğru negatif (DN), olumlu yorumların doğru olarak tanımlandığı örneklerin sayısını ifade eder. Yanlış pozitif (YP), olumlu bir yorumun yanlışlıkla olumsuz olarak sınıflandırıldığı durumu gösterir. Yanlış negatif (YN) ise, olumsuz bir yorumun yanlış bir şekilde olumlu olarak sınıflandırılması anlamına gelir. Bu terimler, modelin sınıflandırma performansını değerlendirirken doğru ve yanlış sınıflandırmaların nasıl yapıldığını açıkça gösterir.

Deneyisel çalışmaların ilk adımında veri setinin analizi gerçekleştirilmiştir. Bu çerçevede Şekil 3’te yorum uzunlukları, bigrams ve konu başlıklarına ait grafiklere yer verilmiştir.

Şekil 3(a)’da gösterilen yorum uzunluğu analizi, müşteri yorumlarının karakter sayısına göre dağılımını inceleyerek hizmet hakkında ne kadar ayrıntılı geri bildirim verildiğini anlamamıza yardımcı olur. Yorumlar 3 ile 3491 karakter arasında değişirken, ortalama uzunluk 820 karakterdir. Medyan değer 638 karakter olup, yorumların çoğunluğu 406 ile 1060 karakter arasında yoğunlaşmaktadır. Bu dağılım, müşterilerin genellikle ayrıntılı geri bildirim verdiğini göstermektedir. Uzun yorumlar, daha güçlü duygular ve detaylı deneyimler içerir, bu da müşteri memnuniyetini değerlendirmede ve hizmet kalitesini iyileştirmede önemli bilgiler sunar.

Şekil 3 (b)’de gösterilen bigram diyagramı bazı kelimelerin frekansına dayalı olarak görselleştirmek sağlamaktadır. Veri seti analizi sırasında bigramlar, yorumlardaki kelime çiftlerinin sıklığını incelemek amacıyla belirlenmiştir. Her bir yorumdaki ardışık iki kelimedenden oluşan bigramlar, veri setinde ne sıklıkla tekrarlandığına göre sıralanmıştır. Bigram frekansları, yorumlarda sıkça vurgulanan konuları ve kelime öbeklerini ortaya çıkarmada kullanılmıştır. “*The flight*” bigramı 536 kez geçerek en yaygın kullanılan öbek olmuştur, bu da uçuş deneyimlerinin sıkça tartışıldığını gösterir. “*Air Canada*” ve “*in the*” bigramları 462 kez geçmiştir, bu da havayolu şirketi hakkında yoğun geri bildirim olduğunu belirtir. Diğer yaygın bigramlar arasında “*on the*”, “*of the*”, ve “*to the*” yer almakta, her biri 440 defadan fazla geçmektedir. Bu bigramlar, müşterilerin hizmetle ilgili genel deneyimlerini ve sıkça karşılaşılan konuları anlamamıza yardımcı olmaktadır.

Şekil 3 (c)’de müşteri yorumları üzerinden *Latent Dirichlet Allocation* (LDA) yöntemi kullanılarak belirlenen beş ana konu ve bu konuların öne çıkan kelimeleri gösterilmektedir. Bu yöntem, metinlerin altında yatan temaları keşfetmek için kullanılan bir makine öğrenmesi tekniğidir. Bu çalışmada belirlenen beş ana konu, “uçuş deneyimi”, “hizmet kalitesi”, “yiyecek ve içecek servisi”, “personel”, ve “zaman yönetimi” konularıyla ilişkilidir.



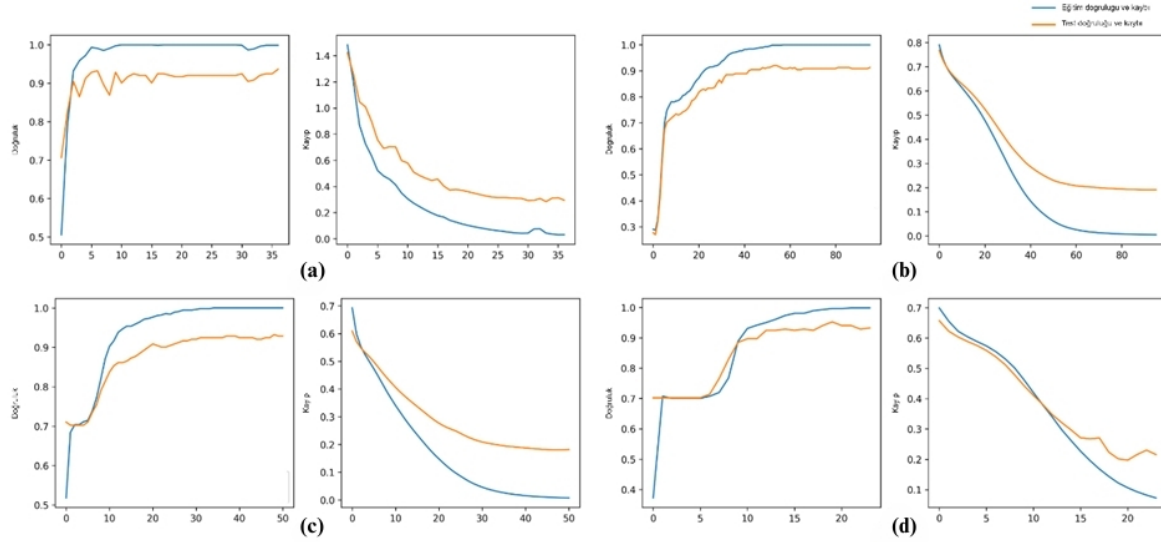
Şekil 3. Veri seti analizi. (a) Yorum uzunlukları (karakter olarak), (b) Bigrams grafiği, (c) Konu başlıklarının keşfedilmesi.

Bu başlıklar, yorumlarda en sık kullanılan kelimelere dayalı olarak çıkarılmıştır. Konu başlıklarının belirlenmesi, duygu analiziyle yakından ilişkilidir çünkü müşterilerin belirli temalar etrafında olumlu, olumsuz veya nötr duygularını ifade ettikleri konuları daha iyi anlamaya olanak sağlar. Her bir konu, ilgili kelimelerle temsil edilmektedir. Örneğin, "Topic 1" için "flight", "food", "seat", "service" ve "crew" gibi kelimeler öne çıkmaktadır. Diğer konular da benzer şekilde, "airline", "time", "hours", "staff", "terminal" gibi kelimelerle temsil edilmektedir. Bu bulgular, müşteri yorumlarının belirli ana temalar etrafında toplandığını ve bu temaların hizmet kalitesi, uçuş deneyimi, yiyecek ve içecek servisi gibi unsurları kapsadığını göstermektedir. Bu analiz, müşteri memnuniyetinin anahtar bileşenlerini anlamak için önemli içgörüler sağlamaktadır.

Deneysel çalışmanın ikinci adımında özelleştirilmiş ve önceden eğitilmiş modellerin ve sıfırdan ve transfer öğrenme yaklaşımı ile eğitilmesi ve doğrulanması sağlanmıştır. Eğitim için hiper-parametrelerin ayarlanması kapsamında; Embedding katmanı, giriş boyutunu " $len(word\ index) + 1$ " ve çıkış boyutunu 128 olarak belirlenmiştir. Giriş sekanslarının uzunluğu 100'dür. Spatial Dropout ve LSTM katmanları için Dropout oranı 0,2 olarak ayarlanmıştır. Dense katmanlarında sırasıyla 64 ve 32 nöron bulunur ve ReLU aktivasyon fonksiyonu kullanılır. Batch Normalization katmanı ağda yer alır ve Dropout oranı her iki dense katmanından sonra 0,2'dir. Çıkış katmanı Softmax aktivasyon fonksiyonunu kullanır. Kayıp fonksiyonu olarak Sparse Categorical Crossentropy ve optimizasyon algoritması olarak Adam seçilmiştir.

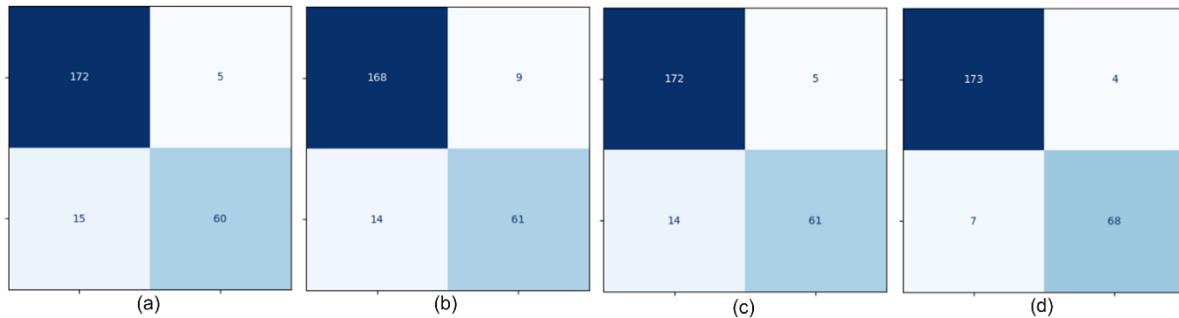
Modellerin doğrulanması için dışarda-tutma (hold-out) yöntemi tercih edilmiştir. Bu doğrultuda, veri seti %75 eğitim ve %25 test olacak şekilde rastgele örnekleme seçimiyle ikiye bölünmüştür. Eğitim kümesi, modellerin

öğrenme süreçlerinde kullanılmış, test kümesi ise performanslarının değerlendirilmesi amacıyla ayrılmıştır. Bu yöntem, modellerin yeni ve görülmemiş veriler üzerindeki başarısını objektif bir şekilde ölçmeye olanak tanımıştır. Ayrıca, tüm modellerin aynı eğitim ve test kümeleri üzerinde eğitilmesi ve doğrulanması sağlanarak, modeller arasında adil bir karşılaştırma yapılması hedeflenmiştir. Bu yaklaşım, modellerin veri seti üzerindeki genel performansını belirlemek ve olası ezberleme sorunlarını önlemek için etkili bir strateji sunmuştur. Bu çerçevede eğitim ve doğrulama eğrileri Şekil 4'te gösterildiği gibi elde edilmiştir.



Şekil 4. Eğitim ve doğrulama eğrileri (a) CNN-LSTM (b) *nlm-en-dim50* (c) *nlm-en-dim128* (d) USE.

Dört farklı modelin eğitim sonuçları incelendiğinde, CNN-LSTM modeli ile eğitilen ilk modelin eğitim ve doğrulama oranı oldukça yüksek olmasına rağmen, doğrulama kaybı değerlerinde dalgalanmalar gözlemlenmiştir, bu da modelin overfitting eğilimi gösterebileceğini düşündürmektedir. *nlm-en-dim50* modeli, daha istikrarlı doğrulama oranı ve kaybı sergilemiş, 80 epoch sonunda %91 doğrulama oranına ulaşmıştır. Bu model, doğrulama kaybının sürekli azaldığı ve eğitim kaybının düşük olduğu görülmektedir, bu da iyi bir genel performans ve düşük overfitting riskini göstermektedir. *nlm-en-dim128* modelinde, doğrulama oranı %93 ile en yüksek seviyeye ulaşmış ve doğrulama kaybı eğitimin sonunda stabilize olmuştur. Bu model, daha büyük embedding boyutu sayesinde daha zengin özellikler yakalayabilmiş ve genel performansını artırmıştır. Ancak, eğitim doğruluğu neredeyse %100'e ulaşmış ve doğrulama kaybı düşük seviyelerde kalarak ezberleme riskine işaret etmiştir. USE modelinde ise doğrulama oranı %88 seviyesinde kalmıştır. Eğitim ve doğrulama kaybı değerlerinde belirgin bir düşüş görülmüş, ancak doğrulama performansında dalgalanmalar gözlemlenmiştir. Bu model, diğer modellerle karşılaştırıldığında daha düşük doğrulama oranına sahip olmasına rağmen, belirli bir dengeye ulaşmıştır. Genel olarak, *nlm-en-dim128* modeli en yüksek doğrulama oranını elde etmiş olup, *nlm-en-dim50* modeli ise daha istikrarlı ve düşük ezberleme riski ile en dengeli performansı sergilemiştir. CNN-LSTM ve USE modelleri, bazı dalgalanmalara ve ezberleme belirtilerine rağmen, genel olarak kabul edilebilir performanslar sunmuştur.



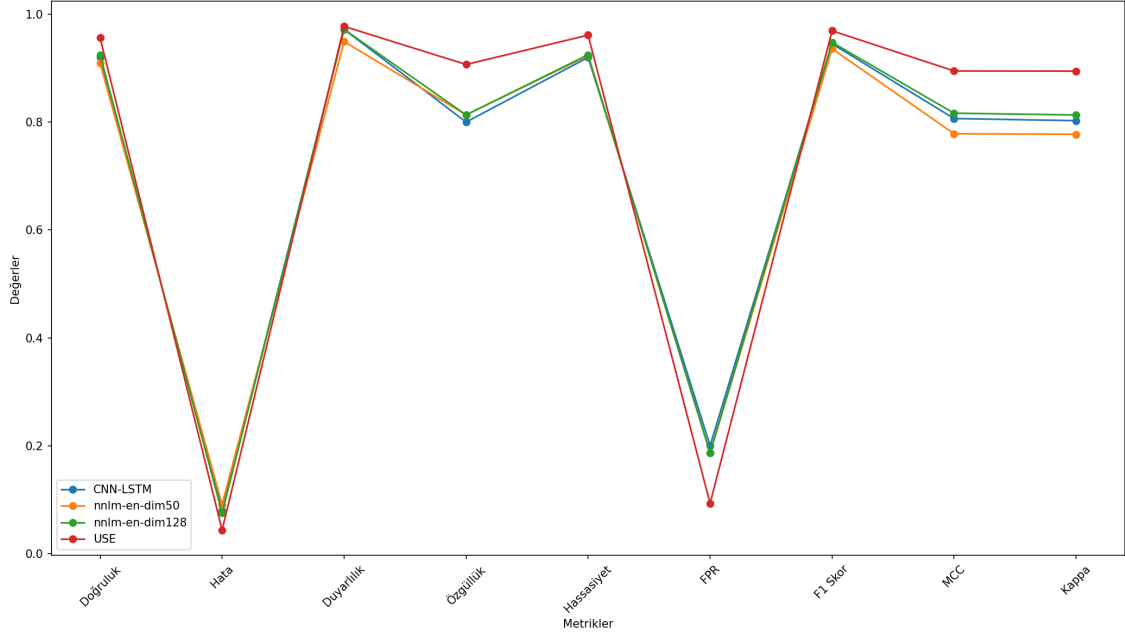
Şekil 5. Hata matrisleri, a) CNN-LSTM (b) *nlm-en-dim50* (c) *nlm-en-dim128* (d) USE.

Eğitim sonrası doğrulama testi üzerinde elde edilen hata matrislerine Şekil 5’de yer verilmiştir. Dört modelin hata matrisleri incelendiğinde, CNN-LSTM modeli (a) 172 doğru pozitif ve 60 doğru negatif tahmine ve 15 yanlış negatif ve 5 yanlış pozitif tahmini bulunmaktadır. *nmlm-en-dim50* modeli (b), 168 doğru pozitif ve 61 doğru negatif tahmin ile benzer bir performans sergilemekte, ancak 14 yanlış negatif ve 9 yanlış pozitif tahmin yapmaktadır. *nmlm-en-dim128* modeli (c), 172 doğru pozitif ve 61 doğru negatif tahminle neredeyse mükemmel bir performans göstermektedir ve sadece 14 yanlış negatif ve 5 yanlış pozitif tahmin yapmaktadır. USE modeli (d) ise 173 doğru pozitif ve 68 doğru negatif tahmin ile en iyi performansı sergilemekte, yalnızca 7 yanlış negatif ve 4 yanlış pozitif tahmin yapmaktadır. Genel olarak, *nmlm-en-dim128* ve USE modelleri daha dengeli ve yüksek performans sergileyen modeller olarak öne çıkmaktadır. CNN-LSTM modeli, yüksek doğruluk oranlarına rağmen yanlış negatif sayısının fazla olması nedeniyle performansını biraz düşürmektedir. *nmlm-en-dim50* modeli ise, diğer modellere kıyasla daha fazla yanlış pozitif tahmine sahip olmasına rağmen genel performansı oldukça iyidir. Hata matrislerine bağlı olarak elde edilen performans metrikleri Tablo 3’te sunulmuştur.

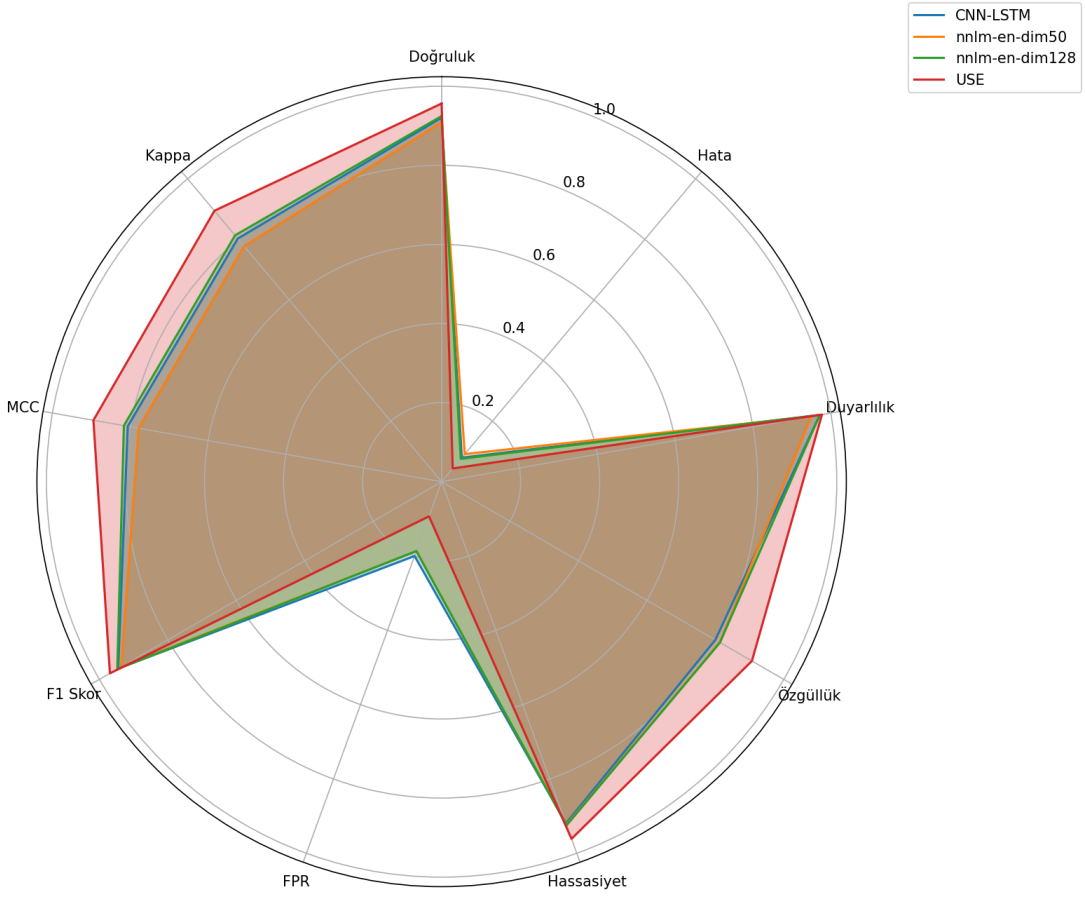
Tablo 3. Modellere ait performans sonuçları.

Metrikler	CNN-LSTM	<i>nmlm-en-dim50</i>	<i>nmlm-en-dim128</i>	USE
Doğruluk	0,9206	0,9087	0,9246	0,9563
Hata	0,0794	0,0913	0,0754	0,0437
Duyarlılık	0,9718	0,9492	0,9718	0,9774
Özgüllük	0,8000	0,8133	0,8133	0,9067
Hassasiyet	0,9198	0,9231	0,9247	0,9611
Yanlış Pozitif Oranı (FPR)	0,2000	0,1867	0,1867	0,0933
F1 Skor	0,9451	0,9359	0,9477	0,9692
MCC*	0,8065	0,7783	0,8164	0,8947
Kappa	0,8026	0,7774	0,8132	0,8944
* Matthews Korelasyon Katsayısı				

Tablo 3’teki performans metriklerine göre dört farklı modelin değerlendirilmesi yapılmıştır. CNN-LSTM modeli, %92,06 doğruluk oranı ve %7,94 hata oranı ile oldukça iyi performans göstermektedir. Bu modelin duyarlılığı %97,18, özgüllüğü ise %80,00 olup, hassasiyeti %91,98 ve F1 skoru %94,51 olarak hesaplanmıştır. MCC değeri 0,8065, Kappa değeri ise 0,8026’dır. *nmlm-en-dim50* modeli ise %90,87 doğruluk oranı ve %9,13 hata oranı ile performans göstermektedir. Duyarlılık %94,92, özgüllük %81,33, hassasiyet %92,31, F1 skoru %93,59 olup, MCC değeri 0,7783 ve Kappa değeri 0,7774’tür. *nmlm-en-dim128* modeli %92,46 doğruluk oranı ve %7,54 hata oranı ile dikkat çekmektedir. Bu modelin duyarlılığı %97,18, özgüllüğü %81,33, hassasiyeti %92,47, F1 skoru %94,77, MCC değeri 0,8164 ve Kappa değeri 0,8132’dir. Son olarak, USE modeli %95,63 doğruluk oranı ve %4,37 hata oranı ile en yüksek performansı göstermektedir. Duyarlılığı %97,74, özgüllüğü %90,67, hassasiyeti %96,11 ve F1 skoru %96,92 olan USE modeli, MCC değeri 0,8947 ve Kappa değeri 0,8944 ile diğer modellerden üstün olduğunu kanıtlamaktadır. Şekil 6(a)’da, CNN-LSTM, *nmlm-en-dim50*, *nmlm-en-dim128* ve USE modellerinin çeşitli performans metrikleri üzerinden karşılaştırılması gösterilmektedir. Doğruluk, hata, duyarlılık, özgüllük, F1 skoru, MCC ve Kappa değerleri incelenmiştir. CNN-LSTM ve USE modelleri arasında genel olarak benzer performanslar görülmekle birlikte, özellikle duyarlılık ve özgüllük metriklerinde tüm modellerin belirgin farklılıklar sergilediği gözlemlenmektedir. Yanlış pozitif oranı (FPR) ve hata oranlarında bazı düşüşler olsa da genel doğruluk ve diğer metriklerde modellerin güçlü performanslar sergilediği anlaşılmaktadır. Sonuçlar, CNN-LSTM’nin diğer modellere göre daha dengeli bir performans sunduğunu göstermektedir. Şekil 6(b)’de, CNN-LSTM, *nmlm-en-dim50*, *nmlm-en-dim128* ve USE modellerinin performans metriklerinin radar grafiği formatında karşılaştırılması sunulmuştur. Radar grafiği, her modelin farklı metriklerdeki performansını tek bir görselde toplu olarak inceleme imkânı sunar ve modellerin birbirine göre avantajlı ve dezavantajlı olduğu alanları net bir şekilde görmeyi sağlar. Bulgular, CNN-LSTM modelinin, özellikle duyarlılık ve özgüllük metriklerinde diğer modellere göre daha dengeli ve tutarlı performans sergilediğini göstermektedir. Diğer yandan, F1 skoru ve MCC gibi kritik metriklerde USE modelinin güçlü sonuçlar verdiği gözlemlenmiştir.

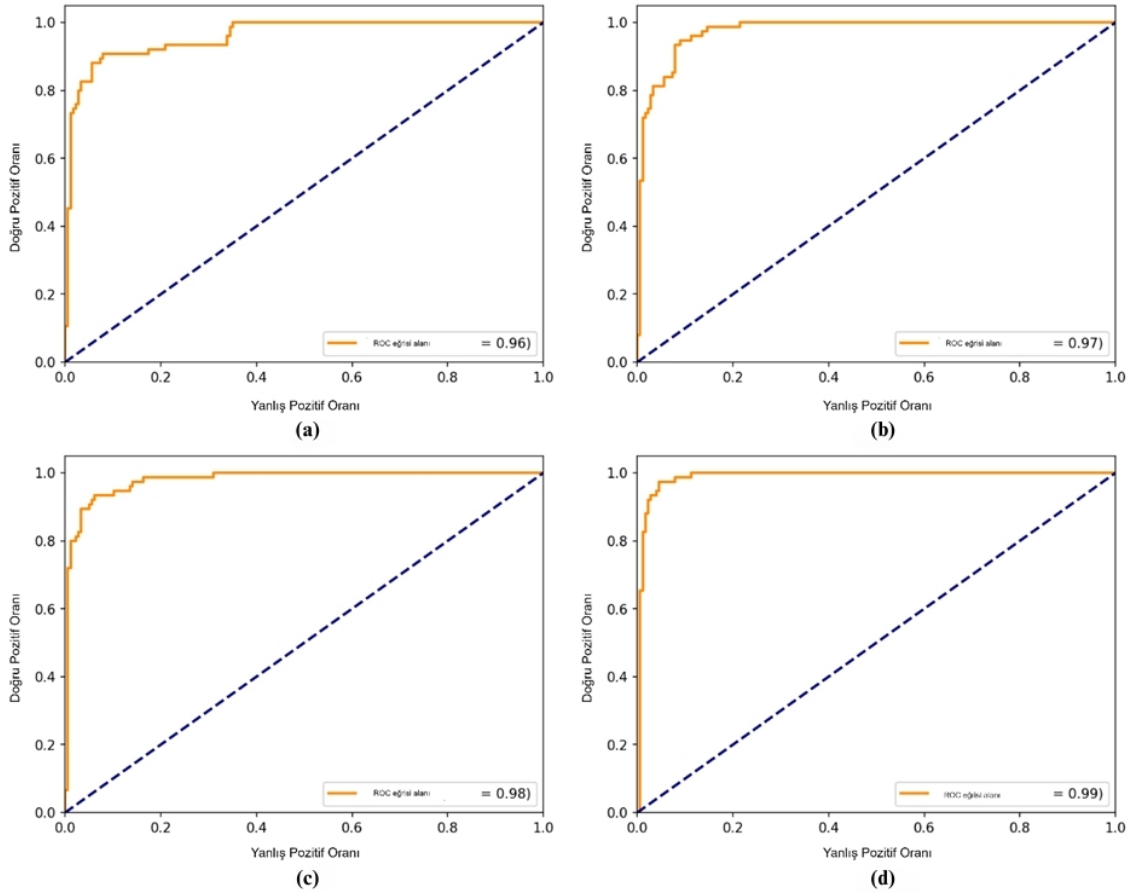


(a)



(b)

Şekil 6. (a) Modellerin performans metrikleri karşılaştırması, (b) Modellerin radar grafiğinde performans metriklerinin karşılaştırılması.



Şekil 7. ROC eğrileri (a) CNN-LSTM (b) *nlm-en-dim50* (c) *nlm-en-dim128* (d) USE.

Şekil 7’de modellere ait Alıcı İşletim Karakteristiği (Receiver Operating Characteristic, ROC) eğrileri gösterilmektedir. ROC eğrisi, modelin sınıflandırma performansını değerlendirmek için yaygın olarak kullanılan bir grafikdir ve Doğru Pozitif Oranı (True Positive Rate) ile Yanlış Pozitif Oranı (False Positive Rate) arasındaki ilişkiyi gösterir. ROC eğrisi altında kalan alan (AUC - Area Under the Curve) ne kadar büyükse, modelin sınıflandırma performansı o kadar iyidir. CNN-LSTM modelinin ROC eğrisi altında kalan alanı (AUC) 0,96 olarak hesaplanmıştır. Bu, modelin yüksek doğrulukla çalıştığını ve sınıflandırma performansının oldukça iyi olduğunu göstermektedir. *nlm-en-dim50* modelinin AUC değeri 0,97 olup, bu modelin de yüksek bir performans sergilediğini göstermektedir. *nlm-en-dim128* modelinin AUC değeri ise 0,98 olarak bulunmuş, bu da modelin sınıflandırma görevinde üstün bir performans gösterdiğini ifade etmektedir. USE modeli ise 0,99 AUC değeri ile en yüksek performansı göstermiştir, bu da modelin sınıflandırma doğruluğunun oldukça yüksek olduğunu ve çok iyi bir performans sergilediğini göstermektedir. Genel olarak, ROC eğrileri ve AUC değerleri, tüm modellerin duygu analizinde etkili olduğunu ve yüksek doğruluk oranlarıyla çalıştığını göstermektedir. Ancak, USE modeli en yüksek AUC değeri ile diğer modellerden daha üstün bir performans sergilemektedir. Bu bulgular, USE modelinin metin tabanlı duygu analizi görevlerinde en iyi performansı sağladığını ve diğer modellere göre daha güvenilir sonuçlar verdiğini göstermektedir. CNN-LSTM, *nlm-en-dim50* ve *nlm-en-dim128* modelleri de başarılı performanslar sergilemekte olup, farklı uygulama senaryolarında kullanılabilir.

Transfer öğrenme, önceden eğitilmiş bir modelin yeni bir görev için yeniden kullanılmasını sağlayan bir stratejidir. Bu yaklaşım, geniş veri kümelerinde eğitilmiş modellerin bilgilerini aktararak, özellikle sınırlı veri kümelerinde etkili sonuçlar elde eder. Sıfırdan öğrenme ise bir modelin tamamen yeni ve mevcut veri setiyle eğitildiği, tüm parametrelerin bu veri setine dayalı olarak optimize edildiği bir stratejidir. Transfer öğrenme, eğitim süresini kısaltırken genellikle daha yüksek performans sağlar; sıfırdan öğrenme ise modelin veri setine özgü ince ayar yapabilmesi açısından avantajlıdır. Bu çalışmada, CNN-LSTM modeli sıfırdan öğrenme yaklaşımı ile eğitilmiş ve %92,06 doğruluk oranına ulaşmıştır. Öte yandan, *nlm-en-dim50*, *nlm-en-dim128* ve USE modelleri

ise transfer öğrenme yaklaşımı ile eğitilmiştir ve sırasıyla %90,87, %95,63 ve %95,63 doğruluk performansı göstermiştir.

4. Tartışma

Tablo 4'te, çeşitli duygu analizi yöntemlerinin farklı veri setleri üzerindeki performansları karşılaştırılmıştır. Bu karşılaştırma, metin tabanlı duygu analizinde kullanılan yöntemlerin çeşitliliğini ve bu yöntemlerin performansını değerlendirmeye olanak tanımaktadır. Örneğin, CNN ve Word2Vec metodunun TTC-3600 veri kümesi üzerinde %93,3 doğruluk oranı elde ettiği görülmektedir. Bu yöntem, özellikle büyük veri setlerinde yüksek doğruluk oranları sağlamaktadır. Benzer şekilde, CNN ile TF-IDF, DK Filtre ve Kök Bulma metodlarının kullanıldığı TTC-4900 ve MY-15130 veri setlerinde sırasıyla %91,70 ve %94 doğruluk oranları elde edilmiştir. Bu sonuçlar, CNN tabanlı yöntemlerin metin sınıflandırma görevlerinde etkinliğini göstermektedir. LSTM yöntemi, 225.239 örnek ve 14 sınıf içeren özel bir veri seti üzerinde %84 doğruluk oranı ile daha düşük bir performans sergilemiştir. Bu durum, LSTM'nin büyük ve karmaşık veri setlerinde performansının sınırlı olabileceğini göstermektedir. Öte yandan, FastText ve SVM kombinasyonu, 12.000 örnek ve 6 sınıf içeren özel bir veri setinde %95,75 doğruluk oranı ile yüksek performans göstermektedir. Bu da FastText ve SVM yöntemlerinin metin sınıflandırmada güçlü araçlar olduğunu ortaya koymaktadır. BERT modeli, EnWiki-100 ve RCV1-57 veri setlerinde sırasıyla %75,28 ve %73,48 doğruluk performansı göstermiştir. Bu durum, BERT modelinin büyük ve çeşitli sınıflara sahip veri setlerinde bileşenlerinin optimizasyon gereksinimlerine işaret etmektedir. Bi-LSTM modeli, IMDB veri setinde %91,41 doğruluk oranı ile dikkat çekici bir performans sergilemiştir. Son olarak, TextConvoNet modeli, farklı veri setleri üzerinde geniş bir performans aralığı sergilemiştir. Örneğin, Reuters-21578 veri setinde %99,20 doğruluk oranı ile en yüksek performansı elde etmiştir. Bu durum, TextConvoNet'in genel metin sınıflandırma görevlerinde oldukça başarılı olduğunu göstermektedir.

Bu çalışma, literatürde yer alan çeşitli duygu analizi yöntemleriyle karşılaştırıldığında, sıfırdan eğitilen CNN-LSTM ve transfer öğrenme yaklaşımı ile eğitilmiş üç farklı önceden eğitilmiş modelin performansını inceleyerek önemli katkılar sunmaktadır. Literatürde CNN ve Word2Vec gibi klasik yöntemlerin yanı sıra, FastText-SVM ve BERT gibi derin öğrenme tabanlı yaklaşımlar incelenmiştir. Bu çalışmada genellikle %84 ile %99 arasında değişen doğruluk oranları elde edilmiştir. Önerilen çalışmada, transfer öğrenme yöntemiyle eğitilen USE modelinin %95,63 doğruluk oranı ile literatürdeki pek çok çalışmaya kıyasla üstün performans sergilediği gözlemlenmiştir. Ayrıca, CNN-LSTM modelinin %92,06 doğruluk oranı ile geniş veri setleri üzerinde etkili bir yöntem olduğu ortaya konulmuştur. Bu bulgular, çalışmamızın hem CNN-LSTM'nin sıfırdan öğrenme stratejisi hem de transfer öğrenme tekniklerinin, metin tabanlı duygu analizinde yüksek doğruluk oranlarına ulaşılmasında etkili yaklaşımlar olduğunu göstererek literatüre önemli katkılar sağladığını göstermektedir.

Bu bulgular, duygu analizinde kullanılan çeşitli yöntemlerin performanslarını karşılaştırarak, belirli veri setlerine ve uygulama senaryolarına göre hangi yöntemlerin daha uygun olabileceğine dair önemli bilgiler sunmaktadır. Tartışma başlıkları olarak, yöntemlerin veri setine göre değişen performansları, özel ve açık erişimli veri setlerinin etkisi, yöntemlerin karmaşıklığı ve optimizasyon gereksinimleri ile ilgili konular ele alınabilir. Bu değerlendirmeler, metin tabanlı duygu analizi çalışmalarında en uygun yöntemin seçilmesine yönelik stratejik kararlar alınmasına yardımcı olacaktır.

Tablo 4. İlgili çalışmaların karşılaştırması.

Metotlar	Veri Seti	Örnek sayısı	Sınıf sayısı	Açık Erişim	Doğruluk (%)
CNN ve Word2Vec Metodu [4]	TTC-3600 Veri Kümesi	3600	6	Evet	93,3
CNN + TF-IDF + DK Filtre + Kök Bulma [15]	TTC-4900	4900	7	Evet	91,70
	MY-15130	15170	2	Evet	94,00
LSTM [17]	Özel veri seti	225239	14	Hayır	84,00
FastText + SVM [22]	Özel veri seti	12000	6	Hayır	95,75
kNN [27]	Özel veri seti	2411	7	Hayır	71,40
BERT [26]	EnWiki-100	329600	100	Evet	75,28
	RCV1-57	758100	57		73,48
SMO [12]	Özel veri seti	444	2	Hayır	88,73
Adaboost + TF IDF [19]	Özel veri seti	24783	3	Hayır	94,00
Bi LSTM [24]	IMDB	50000	2	Evet	91,41
kNN [18]	Özel veri seti	279	2	Hayır	83,87
RCNN [25]	20News	13919	4	Evet	96,49
	Fudan	19636	20		95,20
	ACL	202979	5		49,19
	SST	11855	5		47,21
RT + Bag of Words [21]	Özel veri seti	18000	3	Hayır	96,62
TextConvoNet [23]	Binary SST-2 Dataset	8732	2	Evet	82,20
	Amazon Review for Sentiment Analysis Dataset	20000	2	Evet	90,90
	Reuters-21578	7112	8	Evet	99,20
	Twitter User Airline Sentiment Dataset	13000	3	Evet	88,40
	Covid Tweets Dataset	7500	5	Evet	83,90
MLP [3]	Özel veri seti	1400	2	Hayır	86,00
Bu çalışma, CNN-LSTM, NNLM, Transfer Learning	Havayolu incelemeleri veri kümesi	1100	3	Evet	95,63

5. Sonuçlar ve Öneriler

Bu çalışmanın ana hedefleri, havayolu müşteri yorumlarının duygu analizi için farklı yapay zekâ ve derin öğrenme modellerinin performansını değerlendirmek ve bu değerlendirmelerden elde edilen bulgularla müşteri memnuniyetini artıracak stratejik öneriler sunmaktır. Çalışmada CNN-LSTM, *nlnm-en-dim50*, *nlnm-en-dim128* ve *USE* modelleri kullanılmış ve bu modellerin metin sınıflandırma görevlerindeki başarıları detaylı bir şekilde incelenmiştir.

CNN-LSTM modeli %92,06 doğruluk oranı ile oldukça başarılı bir performans göstermiştir. Ancak, bu modelin duyarlılığı %97,18 iken özgüllüğü %80,00 olarak hesaplanmıştır. Yanlış pozitif oranı %20,00 olan modelin F1 skoru ise %94,51'dir. Bu model, yüksek doğruluk oranına sahip olmasına rağmen bazı durumlarda aşırı uyum problemi göstermiştir. *nlnm-en-dim50* modeli %90,87 doğruluk oranı ile biraz daha düşük performans sergilemiştir. Bu modelin duyarlılığı %94,92 ve özgüllüğü %81,33'tür. Yanlış pozitif oranı %18,67 olan modelin F1 skoru %93,59 olarak hesaplanmıştır. *nlnm-en-dim50* modeli, test doğruluğu ve kaybı açısından daha istikrarlı bir performans göstermiştir. *nlnm-en-dim128* modeli %92,46 doğruluk oranı ile en yüksek doğruluğa sahip modellerden biri olmuştur. Bu modelin duyarlılığı %97,18, özgüllüğü %81,33 ve yanlış pozitif oranı %18,67'dir. F1 skoru %94,77 olan *nlnm-en-dim128* modeli, büyük gömme boyutu sayesinde daha zengin özellikler yakalayarak performansını artırmıştır. *USE* modeli %95,63 doğruluk oranı ile en yüksek performansı sergilemiştir. Bu modelin duyarlılığı %97,74, özgüllüğü %90,67 ve yanlış pozitif oranı %9,33'tür. F1 skoru %96,92 olan *USE* modeli, diğer modellere göre daha düşük hata oranı ve yüksek doğruluk ile öne çıkmıştır.

Sonuç olarak, CNN-LSTM, *nmlm-en-dim50*, *nmlm-en-dim128* ve *USE* modelleri duygu analizi için etkili araçlar olarak değerlendirilmiştir. *USE* modeli, en yüksek doğruluk oranı ve en düşük hata oranı ile en iyi performansı göstermiştir. *nmlm-en-dim128* modeli de yüksek doğruluk oranı ile dikkat çekmektedir. Bu sonuçlar, işletmelerin müşteri memnuniyetini artırmak ve hizmet kalitesini iyileştirmek için duygu analizi teknolojilerini etkin bir şekilde kullanmalarına yardımcı olacak önemli içgörüler sunmaktadır.

Gelecek çalışmalar için öneriler şunlardır: farklı veri setleri ve daha geniş örneklem büyüklükleri ile modellerin performanslarının yeniden değerlendirilmesi, modellerin hiperparametre optimizasyonlarının daha detaylı incelenmesi ve farklı yapay zekâ tekniklerinin entegrasyonu ile duygu analizinde daha yüksek doğruluk oranlarının elde edilmesi mümkün olabilir. Ayrıca, müşteri yorumlarının dil ve kültürel farklılıklarını göz önünde bulunduran modellerin geliştirilmesi de müşteri memnuniyeti analizinde önemli bir adım olacaktır. Bu öneriler, duygu analizi çalışmalarının daha ileriye taşınması ve işletmelerin müşteri geri bildirimlerinden daha derinlemesine içgörüler elde etmesi için yol gösterici nitelikte olacaktır.

Kaynaklar

- [1] Cömert Ö, Yücel N. Müşteri Duyarlılığını Keşfetmek İçin Yapay Zeka Destekli Analiz ile Çevrimiçi Ürün İncelemelerinden Anlamlı Bilgiler Elde Etme. Fırat Üniversitesi Mühendislik Bilimleri Dergisi 2023; 35(2): 679-690.
- [2] Uslu O, Akyol S. Türkçe Haber Metinlerinin Makine Öğrenmesi Yöntemleri Kullanılarak Sınıflandırılması. Estudam Bilişim Dergisi 2021; 2(1): 15–20.
- [3] Tuzcu S. Çevrimiçi Kullanıcı Yorumlarının Duygu Analizi ile Sınıflandırılması. Estudam Bilişim Dergisi 2020; 1(2): 1–5.
- [4] Acı Ç, Çırak A. Türkçe Haber Metinlerinin Konvolüsyonel Sinir Ağları ve Word2Vec Kullanılarak Sınıflandırılması. Bilişim Teknolojileri Dergisi 2019; 12(3): 219-228.
- [5] Demircan M, Seller A, Abut F, Akay MF. Developing Turkish sentiment analysis models using machine learning and e-commerce data. Int J Cognit Comput Eng 2021; 2: 202-207.
- [6] Chen H, Wu L, Chen J, Lu W, Ding J. A comparative study of automated legal text classification using random forests and deep learning. Inf Process Manag 2022; 59(2): 102798.
- [7] Barfar A. A linguistic/game-theoretic approach to detection/explanation of propaganda. Expert Syst Appl 2022; 189: 116069.
- [8] Liu Y, Li P, Hu X. Combining context-relevant features with multi-stage attention network for short text classification. Comput Speech Lang 2022; 71: 101268.
- [9] Colón-Ruiz C, Segura-Bedmar I. Comparing deep learning architectures for sentiment analysis on drug reviews. J Biomed Inform 2020; 110: 103539.
- [10] Wang R, Li Z, Cao J, Chen T, Wang L. Convolutional recurrent neural networks for text classification. In: IEEE 2019 International Joint Conference on Neural Networks (IJCNN); 14-17 July 2019; Budapest, Hungary. pp. 1-6.
- [11] Onan A, Korukoğlu S, Bulut H. Ensemble of keyword extraction methods and classifiers in text classification. Expert Syst Appl 2016; 57: 232-247.
- [12] Göker H, Tekedere H. FATİH Projesine Yönelik Görüşlerin Metin Madenciliği Yöntemleri İle Otomatik Değerlendirilmesi. Bilişim Teknolojileri Dergisi, 10(3), 291-299.
- [13] Kılınç D, Borandağ E, Yücalar F, Tunalı V, Şimşek M, Özçift A. KNN Algoritması ve R Dili ile Metin Madenciliği Kullanılarak Bilimsel Makale Tasnifi. Marmara Fen Bilimleri Dergisi 2016; 28(3): 89-94.
- [14] Kowsari K, Brown DE, Heidarysafa M, Jafari Meimandi K, Gerber MS, Barnes LE. HDLTex: Hierarchical Deep Learning for Text Classification. In 2017 16th IEEE International Conference on Machine Learning and Applications; 18-21 December 2017; Cancun, Mexico. pp. 364-371.
- [15] Alparslan G, Dursun M. Konvolüsyonel Sinir Ağları Tabanlı Türkçe Metin Sınıflandırma. Bilişim Teknolojileri Dergisi 2023; 16(1): 21-31.
- [16] Guo B, Han S, Han X, Huang H, Lu T. Label Confusion Learning to Enhance Text Classification Models. In: AAAI 2021 35th AAAI Conference on Artificial Intelligence; 2021; AAAI. Vol 14B. pp. 12929-12936.
- [17] Alzoubi YI, Topcu AE, Erkaya AE. Machine Learning-Based Text Classification Comparison: Turkish Language Context. Appl Sci 2023; 13(16): 9428.
- [18] T Kaşıkçı T, Gökçen H. Metin Madenciliği İle E-Ticaret Sitelerinin Belirlenmesi. Bilişim Teknolojileri Dergisi. 2013;7(1):0.
- [19] Hajibabae P, Malekzadeh M, Ahmadi M, Heidari M, Esmaeilzadeh A, Abdolazimi R, James Jr H. Offensive Language Detection on Social Media Based on Text Classification. In: IEEE 12th Annual Computing and Communication Workshop and Conference (CCWC); 26-29 Jan 2022; Las Vegas, NV, USA. pp. 92-98 .
- [20] Gürbüz M, Sürmeli D, Taşkın K, Cebeci Hİ. Otelere için paylaşılan çevre ile alakalı yorumların metin madenciliği ile analizi: Antalya otelleri üzerine bir araştırma. Business & Management Studies: An International Journal 2024; 12(1): 218-239.
- [21] Neogi AS, Garg KA, Mishra RK, Dwivedi YK. Sentiment analysis and classification of Indian farmers' protest using twitter data. Int J Inf Manage Data Insights 2021; 1(2): 100019.
- [22] Çelik Ö, Koç BC. TF-IDF, Word2vec ve Fasttext Vektör Model Yöntemleri ile Türkçe Haber Metinlerinin

- Sınıflandırılması. Dokuz Eylül Üniversitesi Mühendislik Fakültesi Fen ve Mühendislik Dergisi 2021; 23(67): 121-127.
- [23] Soni S, Chouhan SS, Rathore SS. TextConvoNet: a convolutional neural network based architecture for text classification. *Appl Intell* 2023; 53(11): 14249-14268.
- [24] Jang B, Kim M, Harerimana G, Kang SU, Kim JW. Bi-LSTM model to increase accuracy in text classification: Combining word2vec CNN and attention mechanism. *Appl Sci* 2020; 10(17): 5841.
- [25] Lai S, Xu L, Liu K, Zhao J. Recurrent convolutional neural networks for text classification. In: Proceedings of the National Conference on Artificial Intelligence; 2015. AAAI.
- [26] Gasparetto A, Marcuzzo M, Zangari A, Albarelli A. Survey on Text Classification Algorithms: From Text to Predictions. *Information* 2022; 13(2): 1-39.
- [27] Chen Z, Zhou LJ, Li X Da, Zhang JN, Huo WJ. The Lao text classification method based on KNN. *Procedia Comput Sci* 2020; 166:523-528.
- [28] Bahassine S, Madani A, Al-Sarem M, Kissi M. Feature selection using an improved Chi-square for Arabic text classification. *J King Saud Univ Comput Inf Sci* 2020; 32(2): 225-231.
- [29] Reusens M, Stevens A, Tonglet J, De Smedt J, Verbeke W, Vanden Broucke S, Baesens B. Evaluating text classification: A benchmark study. *Expert Syst Appl* 2024; 254:124302.
- [30] Wang H, He J, Zhang X, Liu S. A Short Text Classification Method Based on N -Gram and CNN. *Chin J Electron* 2020; 29(2): 248-254.
- [31] Bhattacharya S, Shankar BG, Pitchaimanickam B, Nithya AA. Toxic Comments Classification using LSTM and CNN. In: 2024 3rd International Conference on Applied Artificial Intelligence and Computing (ICAAIC); 5-7 June 2024; Salem, India. pp.214-221.
- [32] İlkuçar M, Artun C, Misafir Yorumlarının Makine Öğrenmesi Yardımıyla Duygu Analizi: Fethiye Beş Yıldızlı Oteller Örneği. *J Bus Digit Age* 2023; 6: 33–41.
- [33] Syed AA. Airline reviews dataset for Abstractive Summarization and Sentiment Classification. *Mendeley Data* 2023; doi: 10.17632/pc6fxc95h5.1 .
- [34] Budak Ü, Cömert Z, Çıbuk M, Şengür A. DCCMED-Net: Densely connected and concatenated multi Encoder-Decoder CNNs for retinal vessel extraction from fundus images. *Med Hypotheses* 2020; 134:109426.
- [35] Togacar M, Comert Z, Ergen B, Budak U. Brain Hemorrhage Detection based on Heat Maps, Autoencoder and CNN Architecture. In: 2019 1st International Informatics and Software Engineering Conference; 6-7 November 2019; Ankara, Türkiye. pp. 1-5.
- [36] Toğaçar M, Ergen B, Cömert Z. Classification of flower species by using features extracted from the intersection of feature selection methods in convolutional neural network models. *Measurement* 2020; 158:107703.
- [37] Toğaçar M, Cömert Z, Ergen B. Intelligent skin cancer detection applying autoencoder, MobileNetV2 and spiking neural networks. *Chaos Solitons Fractals* 2021; 144:110714.
- [38] Budak Ü, Çıbuk M, Cömert Z, Şengür A. Efficient COVID-19 Segmentation from CT Slices Exploiting Semantic Segmentation with Integrated Attention Mechanism. *J Digit Imaging* 2021; 34: 263–272
- [39] Toğaçar M, Ergen B, Cömert Z. BrainMRNet: Brain tumor detection using magnetic resonance images with a novel convolutional neural network model. *Med Hypotheses* 2020; 134:109531.
- [40] AL-Smadi M, Hammad MM, Al-Zboon SA, AL-Tawalbeh S, Cambria E. Gated recurrent unit with multilingual universal sentence encoder for Arabic aspect-based sentiment analysis. *Knowl Based Syst* 2023; 261:107540.