



# Makine Öğrenmesi ile E-Ticaret Ürün Yorumlarının Otomatik Analizi

Müjdat Çabuk<sup>1\*</sup>, Fatih Yücalar<sup>2</sup>, Mansur Alp Toçoğlu<sup>3</sup>

<sup>1\*</sup> Manisa Celal Bayar Üniversitesi, Fen Bilimleri Enstitüsü, Yazılım Mühendisliği A.B.D, Manisa, Türkiye,  
(ORCID: 0000-0002-4863-7217), [mujdat.cabuk@yandex.com](mailto:mujdat.cabuk@yandex.com)

<sup>2</sup> Manisa Celal Bayar Üniversitesi, Hasan Ferdi Turgutlu Teknoloji Fakültesi, Yazılım Mühendisliği Bölümü, Manisa, Türkiye  
(ORCID: 0000-0002-1006-2227), [fatih.yucalar@cbu.edu.tr](mailto:fatih.yucalar@cbu.edu.tr)

<sup>3</sup> İzmir Kâtip Çelebi Üniversitesi, Mühendislik ve Mimarlık Fakültesi, Bilgisayar Mühendisliği Bölümü, İzmir, Türkiye  
(ORCID: 0000-0003-1784-9003), [mansuralp.tocoglu@ikc.edu.tr](mailto:mansuralp.tocoglu@ikc.edu.tr)

(İlk Geliş Tarihi 27 Temmuz 2023 ve Kabul Tarihi 27 Eylül 2023)

(DOI: 10.5281/zenodo.10259157)

**ATIF/REFERENCE:** Çabuk, M., Yücalar, F., & Toçoğlu, M. A. (2023). Makine Öğrenmesi ile E-Ticaret Ürün Yorumlarının Otomatik Analizi. *Avrupa Bilim ve Teknoloji Dergisi*, (52), 110-121.

## Öz

Teknolojinin gelişmesiyle birlikte e-ticaret sektörü oldukça hızlı bir büyüme sağlamıştır. Bu gelişim sayesinde ürünlere yapılan yorumlarda büyük oranlarda artış gözlemlenmektedir. Ürünlerin yorum sayıları arttıkça bunları tek tek inceleyip analiz etmek oldukça zorlaşmış ve satıcılar ürünlerine yapılan yorumların hepsini değerlendiremez duruma gelmiştir. Bu çalışma kapsamında, e-ticaret ürün yorumları üzerinden bir duygu analizi çalışması gerçekleştirilmiştir. Öncelikle Türkiye’de faaliyet gösteren e-ticaret platformlarından 15170 adet ürün yorumu toplanmıştır. Ham veri setinin etiketleme aşamasında yorumlar olumlu ve olumsuz olarak farklı zamanlarda iki kere etiketlenmiş ve iki sınıflı bir veri seti hazırlanmıştır. Bu veri seti ile Uzun-Kısa Süreli Bellek algoritması kullanılarak bir model eğitilmiş ve bu modelle yine e-ticaret platformlarından toplanan 203274 adet yorumun yer aldığı bir veri seti otomatik olarak etiketlenmiştir. Çalışmanın ikinci aşamasında, otomatik olarak etiketlenmiş veri setinin performansı dört farklı derin öğrenme algoritması (Tekrarlayan Sinir Ağları, Uzun-Kısa Süreli Bellek, Geçitli Tekrarlayan Birim, Evrişimli Sinir Ağları) kullanılarak manuel olarak etiketlenmiş veri setiyle karşılaştırılmıştır. Çalışmanın son aşamasında, yüksek sınıflandırma performansı gösteren otomatik olarak etiketlenmiş veri seti kullanılarak dört geleneksel makine öğrenmesi (Lojistik Regresyon, Destek Vektör Makineleri, Rastgele Orman, Naive Bayes) ile dört farklı derin öğrenme algoritmasının sınıflandırma performansları karşılaştırılmıştır. Yapılan deneysel çalışmalar sonucunda, derin öğrenme algoritmalarının makine öğrenmesi algoritmalarına kıyasla daha başarılı olduğu gözlemlenmiş ve geliştirilen modelle otomatik olarak etiketlenen veri setinin kullanıldığı deneylerde yüksek sonuçlar elde edilmiştir.

**Anahtar Kelimeler:** Makine Öğrenmesi, Derin Öğrenme, Metin Madenciliği, Türkçe Metinler, E-ticaret, Ürün Yorumları

## Automated Analysis of E-Commerce Product Reviews with Machine Learning

### Abstract

With the development of technology, the e-commerce sector has grown very rapidly. Thanks to this development, a large increase is observed in the comments made on the products. As the number of reviews of the products increased, it became very difficult to examine and analyze them one by one, and the sellers became unable to evaluate all the comments made on their products. Within the scope of this study, a sentiment analysis study was conducted on e-commerce product reviews. First, 15170 product reviews were collected from e-commerce platforms operating in Turkey. In the labeling phase of the raw data set, the comments were labeled twice as positive and negative at different times and a two-class data set was prepared. With this dataset, a model was trained using the

\* Sorumlu Yazar: [mujdat.cabuk@yandex.com](mailto:mujdat.cabuk@yandex.com)

Long-Short-Term Memory algorithm, and a data set containing 203274 comments collected from e-commerce platforms was automatically labeled with this model. In the second stage of the study, the performance of the automatically labeled dataset was compared with the manually labeled dataset using four different deep learning algorithms (Recurrent Neural Networks, Long-Short-Term Memory, Gated Recurrent Unit, Convolutional Neural Networks). In the last stage of the study, classification performances of four traditional machine learning (Logistic Regression, Support Vector Machines, Random Forest, Naive Bayes) and four different deep learning algorithms were compared using automatically labeled datasets with high classification performance. As a result of the experimental studies, it has been observed that deep learning algorithms are more successful than machine learning algorithms, and high results have been obtained in experiments where the data set automatically labeled with the developed model is used.

**Keywords:** Machine Learning, Deep Learning, Text Mining, Turkish Texts, E-commerce, Product Reviews

## 1. Giriş

Günümüzde teknolojik gelişmeler tüm hızıyla devam etmektedir. Tüm sektörler bu teknolojik gelişmelerden faydalandığı gibi ticaret sektörü de büyük ölçüde faydalanmaktadır. Ticaret sektörü hem perakende hem de toptan satışlar için artık dijital kanalları tercih etmektedir. Mobil cihazların ortaya çıkması ve yaygınlaşmasıyla e-ticaret sektörü adeta bir devrim yaşamıştır. İnsanlar artık kolaylıkla mobil cihazlarıyla e-ticaret platformlarına ulaşabilmekte ve saniyeler içerisinde ihtiyacı olan ürünleri alabilmektedirler. E-ticaret işletmeleri sosyal medya pazarlaması ve yapay zekâ gibi teknolojileri kullanarak müşteri deneyimini daha da geliştirmeye odaklanmaktadır. Son yıllarda yaşanan Covid-19 pandemisinin etkisiyle beraber e-ticaret sektörü hızlı bir şekilde büyümüş ve önemli bir pazar haline gelmiştir. İnsanlar online alışveriş yapmak için fiziksel mağazalara gitmek yerine interneti tercih etmektedirler ve birçok geleneksel mağazada bu trende ayak uydurarak online satışlara odaklanmaktadır.

E-ticarete ürün yorumları müşterilerin bir ürün hakkında görüşlerini paylaştığı ve diğer müşterilerin de bu görüşleri inceleyerek ürün hakkında bir fikir sahibi olabilecekleri önemli bir unsurdur. Ürün yorumları e-ticaret şirketleri için çok önemli bir geri bildirim mekanizmasıdır ve hem müşterilerin memnuniyetini arttırmaya hem de şirketlerin pazarlama stratejilerini belirlemeye yardımcı olur.

Ürün yorumları müşterilerin ürünler hakkında doğru ve açıklayıcı geri bildirimler sunmasını sağlar. Bu yorumlar müşterilerin ürünleri nasıl kullandığı, hangi özelliklerinin beğenilip beğenilmediği, ürünün kalitesi gibi unsurlar hakkında bilgi vermektedir. Bu geri bildirimler şirketlerin ürünlerini geliştirmelerine yardımcı olur ve müşteri memnuniyetini artırır. Ayrıca ürün yorumları diğer müşterilerin de ürün hakkında bilgi sahibi olmalarını sağlamaktadır. Müşterilerin gerçek deneyimlerine dayalı olarak yazdığı yorumlar diğer müşterilerin ürün hakkında karar vermelerine yardımcı olmaktadır. Bu nedenle ürün yorumları e-ticaret şirketleri için önemli bir pazarlama aracıdır.

E-ticaret sektörünün gelişmesiyle ürünlere yapılan yorumlar da muazzam bir şekilde artmıştır. Ürünlerin yorum sayıları artıkça bunları tek tek inceleyip analiz etmek oldukça zorlaşmıştır. Satıcılar ürünlerine yapılan yorumların hepsini değerlendirememekte ve yararlı olabilecek yorumları gözden kaçırabilmektedir. Ayrıca yapılan yorumların tam olarak ne hakkında olduğu da manuel olarak tespit edilemez hale gelmiştir. Müşteriler için de aynı problem geçerlidir. Müşteriler ürün hakkındaki yorumların en fazla iki üç sayfasını okuyabilmektedir. Bu problemler, e-ticaret ürün yorumlarının otomatik olarak analiz edebilme ihtiyacını ortaya çıkarmıştır.

Bu çalışmada e-ticaret ürün yorumları üzerinden bir duygu analizi çalışması gerçekleştirilmiştir. Çalışmanın ilk aşamasında Türkiye’de faaliyet gösteren e-ticaret platformlarından Türkçe yorumlar toplanarak ham bir veri seti oluşturulmuştur. Ürün yorumlarını toplamak için bir uygulama geliştirilmiş ve bu uygulama içerisinde Web Scraping yöntemi kullanılmıştır. Ürün yorumları farklı zamanlarda iki kere olumlu, olumsuz ve nötr olarak manuel sınıflandırılmıştır. Sonrasında nötr yorumlar ve kararsız kalınan yorumlar çıkarılarak iki sınıflı bir veri seti oluşturulmuştur. Çalışmanın ikinci aşamasında Uzun-Kısa Süreli Bellek Ağları (Long-Short Term Memory – LSTM) derin öğrenme yöntemi kullanılarak otomatik olarak etiketlenmiş bir veri seti oluşturulmuştur. Sonraki aşamada, oluşturulan manuel ve otomatik etiketlenmiş veri setleri üzerinde Tekrarlayan Sinir Ağları (Recurrent Neural Network – RNN), Uzun Kısa Süreli Bellek Ağları (Long-Short Term Memory – LSTM), Geçitli Tekrarlayan Birim (Gated Recurrent Unit – GRU) ve Evrişimli Sinir Ağları (Convolutional Neural Network – CNN) gibi derin öğrenme algoritmaları kullanılarak duygu analizi test sonuçları elde edilmiştir. Testler sonucunda otomatik olarak etiketlenmiş veri setiyle eğitilen makine öğrenmesi modellerinin manuel olarak etiketlenmiş veri setiyle eğitilen modellere yakın sonuçlar verdiği gözlemlenmiştir. Sonrasında otomatik olarak etiketlenmiş veri seti ile Rastgele Orman (Random Forest – RF), Lojistik Regresyon (Logistic Regression – LR), Naive Bayes (NB), Destek Vektör Makineleri (Support Vector Machine – SVM) gibi temel geleneksel makine öğrenmesi algoritmaları kullanılarak testler yapılmıştır. Bunun sonucunda derin öğrenme mimarilerinin geleneksel makine öğrenmesi algoritmalarından daha iyi sonuçlar verdiği gözlemlenmiştir.

## 2. Literatür Çalışmaları

Bu çalışma kapsamı öncesinde literatürde yer alan çalışmalar incelenmiştir. Bu aşamada öncelikle Türkçe metinler ile yapılan çalışmalar ve ağırlıklı olarak müşteri yorumlarının yer aldığı veri setleri ile yapılan çalışmalar ele alınmıştır. Ele alınan çalışmalar arasında İngilizce metinlerden oluşan veri setleri ile yapılan çalışmalar da yer almaktadır. Bu çalışmalar kullanılan teknikler için tercih edilmiştir.

Rumelli ve ark. [1], ele aldıkları çalışmalarında e-ticaret ürün yorumlarını kullanarak bir duygu analizi modeli geliştirmişlerdir. Bu çalışmada hepsiburada.com e-ticaret sitesi üzerinden 272 bin adet ürün yorumu çekilerek bir veri seti hazırlanmıştır. Bu veri setinde 1 ve 2 yıldız olan yorumlar olumsuz, 3 yıldız olanlar nötr, 4 ve 5 yıldız olanlar da olumlu olarak etiketlenmiştir. Son olarak 13 bin olumlu ve 13 bin olumsuz yorum seçilerek çalışmada kullanılacak veri seti hazırlanmıştır. Çalışmada NB, RF, SVM ve K – En

Yakın Komşu (K-Nearest Neighbors – KNN) algoritmaları kullanılarak sonuçlar alınmıştır. Çalışma sonucunda ortalama %73 doğruluk oranı elde etmişlerdir.

Topçu ve Erkaya [2], Müşteri İlişkileri Yönetimi (Customer Relations Management – CRM) yazılımlarında toplanan talep mesajlarının makine öğrenmesi yöntemleri kullanılarak kategorize edilmesi amaçlı bir çalışma yapmışlardır. Çalışmada vatandaşlar tarafından gönderilen 3 bin farklı konudan oluşan 220 bin talep mektubu kullanılmıştır. Veriler ön işlemlerden geçirildikten sonra TF-IDF algoritması kullanılarak özellik seçimi yapılmıştır. Veri seti 5 farklı algoritma kullanılarak test edilmiştir. En verimli algoritmanın %85 doğruluk oranıyla LSTM olduğu görülmüştür. Çalışma sonucunda oluşturulan model örnek olarak “1314. sokakta bozuk yollar var. Bir an önce asfalt atılmasını ve yolun onarılmasını talep etmekteyim. Gereğini arz ederim.” şeklinde gelen talebi “Asfalt Bakım Onarım” olarak kategorilendirmiştir.

Tuzcu [3], kitap yorumlarının duygu analizi yöntemleri ile sınıflandırılması çalışması yapmıştır. Popüler bir çevrimiçi kitap platformundan topladıkları okuyucu yorumlarından bir veri seti hazırlamıştır. Bu çalışmada Çok Katmanlı Algılayıcı (Multilayer Perceptron – MLP) algoritmasını uygulamıştır. Sonrasında NB, SVM ve LR gibi geleneksel makine öğrenmesi algoritmaları ile de testler yapılmış ve sonuçlar karşılaştırılmıştır. Karşılaştırma sonucunda %89 başarı oranı ile MLP algoritmasının en yüksek başarı elde ettiği görülmüştür.

Mengutaycı ve Temurtas [4], Türkçe otel yorumlarının yapay sinir ağları ile sınıflandırılması konulu bir çalışma yapmıştır. Çalışmalarında çevrimiçi otel rezervasyonu yapılan bir platformda yer alan müşteri yorumlarını incelemişlerdir. Bu yorumların 4-5 yıldız olanları olumlu, 1-2 yıldız olanları olumsuz olarak etiketlenmesini sağlayarak iki sınıflı bir çalışma gerçekleştirmişlerdir. Veri setini gerekli ön işlemlerden geçirdikten sonra TF-IDF modeli ile yorumlar bir matrise dönüştürülmüş ve her terimin sınıflandırmadaki önem derecesinin belirlenmesi sağlanmıştır. Veri seti iki farklı yöntemle eğitim ve test verilerine ayrılmış ve bu veri setlerinin başarısı test edilmiştir. Testlerini yapay sinir ağı mimarileriyle gerçekleştirmişlerdir.

Mayda ve Korkmaz [5], ortak çalışmalarında müşteri yorumlarının sınıflandırılması için sözlük tabanlı bir yaklaşım kullanmışlardır. Bu çalışma kapsamında bir sıfat sözlüğü hazırlanmıştır. Kitap satışının yapıldığı bir e-ticaret platformundan topladıkları okuyucu yorumlarından bir veri seti hazırlamışlar ve sıfat sözlüğünü kullanarak olumlu, olumsuz, nötr sınıflar için bir duygu analizi çalışması yapmışlardır. Ayrıca çalışmalarında Türkçe diline özgü kurallar geliştirerek de çalışmanın başarısını arttırmayı başarmışlardır. Çalışma sonucunda yaklaşık %62’lik doğruluk oranı elde etmişlerdir.

Gezici ve Yanıkoğlu [6], çalışmalarında makine öğrenmesi algoritmalarıyla Türkçe film incelemelerinin duygu analizini gerçekleştirmişlerdir. Bu çalışmalarında denetimli öğrenme ve sözlük tabanlı yaklaşımları birleştirerek yeni bir yöntem sunmuşlardır. Çalışmalarında “beyazperde.com” adlı film sitesinden önceden oluşturulan bir veri seti kullanılmış ve bu veri setindeki incelemelerin 4 ve 5 yıldızlı olanları olumlu, 1 ve 2 yıldız olanları olumsuz olarak etiketleyerek ikili sınıflandırma çalışması yapmışlardır. NB ve SVM algoritmalarının sonuçları karşılaştırılmıştır. Bu karşılaştırma esnasında sözlük tabanlı yaklaşım ile denetimli öğrenme yaklaşımlarının sonuçları da göz önünde bulundurulmuştur. Sonuçlarda NB algoritmasının doğruluk oranının %75 olduğu görülmüştür.

Pervan ve Keleş [7], RF algoritmasını kullanarak Türkçe yorumlarda duygu analizi çalışması yapmışlardır. Popüler bir e-ticaret platformundan elektronik kategorisi altında yer alan ürünlere yapılan müşteri yorumlarından bir veri seti hazırlamışlar ve etiketleme işlemini yıldız puanına göre otomatik yapmışlardır. RF algoritmasıyla yaklaşık %85 doğruluk oranı elde etmişlerdir.

Onan [8], bir çalışmada denetimli ve denetimsiz terim ağırlıklandırma modellerini kullanarak kapsamlı bir duygu analizi karşılaştırması gerçekleştirmiştir. Veri seti olarak Türkçe Twitter mesajları kullanılmıştır. Mesajlar pozitif ve negatif olarak iki kişi tarafından etiketlenmiş ve gerekli ön işlemlerden geçirilmiştir. Sonrasında KNN, SVM, LR ve NB algoritmalarıyla testler yapılmıştır. Bu testlere AdaBoost, Bagging ve Random SubSpace topluluk öğrenme yöntemleriyle devam edilmiştir. Testler sonucunda denetimli terim ağırlıklandırma modellerinin denetimsiz terim ağırlıklandırma modellerine göre daha başarılı olduğu gözlemlenmiştir.

Demircan ve ark. [9], ise çalışmalarında Türkçe metinler üzerinde duygu analizi yöntemlerini ele almışlardır. Veri seti olarak e-ticaret ürün yorumlarını kullanmışlar ve veri etiketlenmesi için ürün yorumlarındaki puanlardan faydalanmışlardır. SVM, RF, DT, LR ve KNN algoritmalarıyla modeller geliştirip testlerini yapmışlardır. Testler sonucunda SVM ve RF modellerinin daha iyi sonuçlar verdiğini gözlemlemişlerdir.

Toçoğlu ve ark. [10], Türkçe tweetler üzerinden duygu analizi çalışması yapmışlardır. Çalışmada sözlük tabanlı bir yaklaşım kullanarak her tweete altı duygu kategorisi için otomatik olarak açıklama eklemişlerdir. Çalışmalarında ANN, CNN, LSTM ve RNN mimarilerini inceleyerek bunlar üzerinden testleri gerçekleştirmişlerdir. Testler sonucunda sözlük tabanlı yaklaşım ile otomatik eklenen açıklamaların etkisini incelemişlerdir. Deneysel çalışmalar sonucu elde ettikleri doğruluk oranlarını karşılaştırdıklarında en yüksek puanı %74 ile CNN mimarisinin verdiğini gözlemlemişlerdir.

Çataltaş ve ark. [11], bir çalışmalarında e-ticaret olumsuz ürün yorumlarından ürünlerin kusurlu özelliklerini bulmak için bir metin analizi yöntemi önermişler ve bulunan ürün kusurlarından da yapılandırılmış bir özet oluşturmaya odaklanmışlardır. Amazon.com üzerinden ayakkabı kategorisine ait 10 bin olumsuz yorum çekerek bir veri seti oluşturmuşlardır. Ön işlemlerin ardından veri sayısı 6190’a inmiş ve çalışmayı bu verilerle gerçekleştirmişlerdir. Ürün kusurlarını bulmak için DBSCAN algoritması kullanılarak yorumlar kümelere ayrılmıştır. Bulunan her kusur daha önceden tanımlanmış Part-of-Speech (Konuşma Parçası) kalıplarını arayarak kendisinin görüş sözcüklerini bulmak için kullanılmıştır. Bulunan kusurlar ve onların görüş sözcükleri kullanılarak yapılandırılmış bir özet oluşturulmuştur.

Literatürde Türkçe metinlerle yapılan çalışmalar günümüzde gittikçe artmaktadır. İncelenen çalışmalarda çoğunlukla çevrimiçi müşteri yorumlarından oluşan veri setlerinin kullanıldığı görülmüştür. Ayrıca, ilgili çalışmalarda veri etiketleme işlemlerinin karşılaştırıldığı ve manuel olarak etiketlemenin daha doğru sonuçlar verdiği gözlemlenmiştir.

### 3. Materyal ve Metot

Bu bölümde, çalışma kapsamında oluşturulan manuel ve otomatik olarak etiketlenmiş iki veri setinin geliştirilme süreçleri, öznitelik çıkarma aşamaları ve kullanılan makine öğrenmesi algoritmaları hakkında bilgiler paylaşılmıştır.

#### 3.1. Manuel Etiketli Veri Setinin Oluşturulması

Manuel olarak etiketlenmiş veri setinin oluşturulması için ülkemizde faaliyet gösteren Hepsiburada, Trendyol, N11 gibi popüler e-ticaret platformlarından ürün yorumları toplanmıştır. Öncelikle 15.170 adet yorum toplanarak bir veri seti oluşturulmuştur. Bu veri seti için ağırlıklı olarak tıraş makinesi, çalışma koltuğu ve oyun konsolu kategorileri seçilmiştir. Toplanan veriler veri temizleme aşamasından geçirilmiştir. Bu kapsamda sadece noktalama işareti veya emoji içeren veriler çıkarılmıştır. Metinlerin arasında yer alan noktalama işaretleri ve emojiler de kaldırılarak veri setinin sadece metinlerden oluşması sağlanmıştır. Metinlerdeki yazım yanlışları da olabildiğince düzeltilmiştir. Ayrıca alakasız bir yorum içeren metinler de veri setinden çıkarılarak veri temizleme adımı tamamlanmıştır. Tüm harflerin küçük harfe çevrilmesi, sayıların kaldırılması ve StopWords olarak sınıflandırılan kelimelerin de metinlerin içerisinden çıkarılması gibi standart veri ön işleme adımları da veri seti üzerinde model eğitiminden önce uygulanmıştır.

Ham veri setinin oluşturulmasından sonra manuel olarak etiketleme işlemi gerçekleştirilmiştir. Bu etiketleme için genellikle ürün yorumlarının değerlendirme puanı kullanıldığı görülmüştür. Örneğin incelenen çalışmalarda 4 ve 5 yıldız puanına sahip yorumlar olumlu, 1 ve 2 yıldız puanına sahip yorumlar olumsuz ve 3 yıldız puanına sahip yorumlar ise nötr olarak etiketlendiği görülmüştür. E-ticaret platformlarından toplanan yorumlar tek tek okunarak olumlu, olumsuz ve nötr olarak etiketlenmiştir. Ardından farklı bir zamanda önceki etiketler gizlenerek veri seti tekrar etiketlenmiştir. Bu sayede iki farklı etiketleme sonucu ortaya çıkmıştır. Bu sonuçlar karşılaştırılarak etiketi farklı olan yorumlar veri setinden çıkarılmıştır. Örneğin ilk etiketleme işleminde olumlu olarak belirlenen bir yorum sonraki etiketleme işleminde nötr olarak etiketlendiyse, bu durum eğitilecek modeli etkileyeceği için bu yorum veri setinden çıkarılmıştır. Bu işlemin sonucunda olumlu yorumların olumsuz yorumlardan çok fazla olduğu tespit edilmiş ve e-ticaret platformlarından olumsuz yorumlar çekilmeye çalışılmıştır. Çalışma kapsamında geliştirilen uygulamanın veri çekme modülünde yer alan yıldız puanı filtresi kullanılarak kolaylıkla olumsuz yorumlar toplanmıştır. Sonrasında tekrar çekilen veriler tek tek okunarak olumlu yorumlar çıkarılmış ve sadece olumsuz yorumlar veri setine eklenmiştir. Etiketleme işleminin ardından 6818 adet olumlu, 6926 adet olumsuz ve 1393 adet nötr yorumun olduğu bir seti hazırlanmıştır. Nötr yorumların az olması ve tespit edilmesinin de zor olması nedeniyle nötr yorumlar veri setinden çıkarılmıştır. Son olarak dengeli ve iki sınıflı bir veri seti çalışmaya hazır hale getirilmiştir. Tablo 1'de yorumların etiket dağılımına yer verilmiştir.

Tablo 1. Sınıflardaki Yorum Sayısı (Table 1. Number of comments in classes)

Sınıf	Yorum Sayısı
Olumlu	6818
Olumsuz	6926
Toplam	13744

#### 3.2. Otomatik Etiketli Veri Setinin Oluşturulması

Çalışmada kullanılan bir diğer veri seti olan otomatik olarak etiketlenmiş veri setini oluşturmak için e-ticaret platformlarından 203 bin 274 adet ürün yorumu toplanmıştır. Bu veri setinde bir önceki veri setinde kullanılan ürün kategorileri dışında ayakkabı, kol saati, kulaklık, oyuncak, akıllı saat ve hoparlör gibi birçok ürün kategorisi kullanılmıştır. Toplanan verilerin otomatik etiketlenmesi için LSTM algoritmasının kullanıldığı bir model geliştirilmiştir. Bu modelin eğitilmesinde manuel olarak etiketlenen veri seti kullanılmıştır.

LSTM modelinde kullanılan parametreler Tablo 2'de yer almaktadır. Bu parametrelerden Gömme Boyutu (Embedding Dimension Size), kelime gömme yönteminde her kelimenin temsil edildiği vektörün boyutunu belirler ve örneğin 200 seçildiğinde her kelime için 200 boyutlu vektör kullanılır. Gömme Dolgu Boyutu (Embedding Padding Size) parametresi, metin verilerinin gömme işlemi öncesinde veya sonrasında yapılacak olan dolgu işlemidir ve örneğin 100 seçildiğinde metin verileri 100 kelimeye tamamlanır. Nöron Boyutu (Neuron Size) parametresi, LSTM modellerinde hücrelerin boyutunu temsil eder ve daha büyük bir nöron boyutu daha fazla bilgiyi depolayabilme ve karmaşık ilişkileri daha iyi öğrenebilme yeteneği sağlar. Bırakma Oranı (Dropout Rate) parametresi, modelin aşırı öğrenmesini azaltmak için kullanılan bir parametredir. Dönem Sayısı (Number of Epochs) parametresi, modelin veri setini kaç kez tamamladığını belirler ve yeterli değer verildiğinde daha iyi sonuçlar elde edilmesi sağlanır. Grup Boyutu (Batch Size) parametresi, verilerin kaç alt gruba ayrılacağını belirler ve bu alt gruplar sırasıyla modele sunularak eğitim süreci tamamlanır. Optimize Edici (Optimizer) parametresi, modelin ağırlıklarını güncellemek ve kayıp fonksiyonunu minimize etmek için kullanılır ve Adamax algoritması büyük veri setleri ve büyük parametre uzayları için etkili bir algoritmadır. Aktivasyon Fonksiyonu (Activation



Function) parametresi ise sinir ağlarının katmanlarında kullanılan non-lineer dönüşüm işlevini temsil eder ve her bir sinir hücresi aktivasyon fonksiyonu tarafından çıktıya dönüştürülür. Doğrultulmuş doğrusal birim (ReLU) aktivasyon fonksiyonu negatif girişleri 0'a dönüştürür, pozitif girişleri olduğu gibi bırakır. Aktivasyon fonksiyonu modelin başarısını olumlu yönde etkilemektedir. Bu parametrelerin en başarılı değerleri deneme yanılma yoluyla seçilmiştir.

Tablo 2. LSTM Modelinde Kullanılan Parametreler (Table 2. Parameters Used in LSTM Model)

Parametre Adı	Değeri
Gömme Boyutu (Embedding Dimension Size)	200
Gömme Dolgu Boyutu (Embedding Padding Size)	100
Nöron Boyutu (Neuron Size)	100
Bırakma Oranı (Dropout Rate)	0.4
Dönem Sayısı (Number of Epochs)	10
Grup Boyutu (Batch Size)	16
Optimize Edici (Optimizer)	Adamax
Aktivasyon Fonksiyonu (Activation Function)	ReLU

### 3.3. Sınıflandırma Algoritmaları

#### 3.3.1. Geleneksel Algoritmalar

Lojistik Regresyon (Logistic Regression – LR) algoritması, veri setindeki değişkenler arasındaki ilişkiyi tanımlamak için kullanılır ve sınıflandırılmış bir değer olarak çıktı verir. Bir bağımlı değişkenin iki veya daha fazla kategorik veya sayısal bir değer alabileceği veri setlerinde kullanılabilir. Bu algoritma öncelikle bir doğrusal regresyon yöntemi kullanarak bir çıktı verir ve sonra bu çıktı değerini bir olasılık değerine dönüştürür. Bu olasılık değeri, bir sınıfa ait olma olasılığını temsil eder [12].

Destek Vektör Makineleri (Support Vector Machine – SVM) algoritması sınıflandırma ve gruplandırma problemlerinde kullanılmaktadır. Veri noktalarını birbirinden ayıran bir çizgi oluşturarak sınıflandırma işlemini yapmaktadır. SVM algoritması genellikle çözümlenmesi zor olan büyük veri setlerinde etkili olduğu görülmüştür [13]. SVM algoritması, verileri birbirinden ayıran en iyi hiper düzlemi bulmak için çalışır. Hiper düzlem verilerin tam olarak ayıramadığı durumlarda da kullanılabilir. Bu durumda bir hata payı belirlenir ve hiper düzlem en düşük hata payını sağlayacak şekilde oluşturulur. Bu algoritma özellikle doğrusal olmayan veri setleri için etkilidir. Bu veri setlerinde verileri daha yüksek boyutlu bir uzaya taşır ve orada daha kolay sınıflandırılabilir hale getirir. Bu işlem çekirdek fonksiyonları kullanılarak yapılır.

Rastgele Orman (Random Forest – RF) algoritması, bir ağaç yapısı kullanarak bir sınıflandırma modeli oluşturur ve veri kümesindeki öznitelikleri test ederek sonuçlar elde eder. Bu algoritma veri setini daha küçük parçalara ayırarak her bir parçayı ayrı ayrı öğrenir ve sonrasında bu parçaları birleştirerek tahminler yapar [14]. RF algoritması yüksek boyutlu verilerin sınıflandırılmasında kullanılır ve doğrusal olmayan ilişkileri tespit edebilir. Bu algoritma her bir karar ağacının kendine özgü bir eğitim seti ve farklı özellikler ile eğitildiği bir topluluk öğrenmesi yöntemidir. Bu sayede RF algoritması öğrenme aşamasında daha esnek ve doğru sonuçlar vermektedir.

Naive Bayes (NB) algoritması, bayes teoremini kullanarak veri setindeki özellikleri ve sınıfları birbirine bağlar. Özellikle veri setindeki özelliklerin birbirinden bağımsız olduğunu varsayar ve “naive” adı da buradan gelmektedir. NB algoritması sınıflandırılacak verinin özelliklerini kullanarak her sınıfın olasılık dağılımını hesaplar. Bu dağılımlar eğitim verileri üzerinde hesaplanarak öğrenilir. Sonrasında yeni bir veri geldiğinde bu verinin her bir sınıfta olma olasılığı hesaplanır ve sonucunda en yüksek olasılığa sahip sınıf seçilir [13].

#### 3.3.2. Derin Öğrenme Algoritmaları

Tekrarlayan Sinir Ağları (Recurrent Neural Network – RNN), zaman serilerinde veya sıralı verilerde kullanılan bir derin öğrenme mimarisidir. Girdilerin sıralı olarak işlendiği bir yapıya sahiptir. Önceki adımlarda hesaplanan çıktıların ve saklanan bilgilerin bir sonraki adımda tekrar kullanılmasına olarak tanır. RNN modelleri doğal dil işleme, metin üretimi ve ses tanıma gibi uygulamalarda kullanılır [15]. RNN mimarisinde girdi verilerini işleyen gizli bir durum hücresi bulunmaktadır. Gizli durum hücresi önceki adımdaki bilgileri içerir ve sonraki adımda tahmin yapmak için kullanılır. Yeni bir veri geldiğinde her adımda bir girdi alır ve çıktı üretir. Bu çıktı bir sonraki adımda girdi olarak kullanılır ve bu şekilde süreç devam eder. RNN modelleri geriye doğru yayılma ve veri setindeki örneklerin sırasına göre eğitilir. Bu eğitim sırasında model, girdilerin ardışık adımlarında beklenen çıktılarla gerçek çıktıları karşılaştırarak hata oranını azaltmaya çalışır.

Uzun-Kısa Süreli Bellek Ağları (Long-Short Term Memory – LSTM) mimarisi, aslında RNN'nin bir sorununa çözüm olarak literatüre girmiş ve RNN'nin alt türü olarak görülmektedir. RNN mimarisindeki “uzun süreli bellek kaybı” sorununu çözmek üzere

gizli durum hücrelerine ek olarak birçok kapı mekanizması içermektedir. Bu kapılar ağına hangi bilgileri saklayıp hangilerini saklamayacağına karar vermesine yardımcı olur [16]. LSTM mimarisinde üç temel kapı vardır. Bu kapılar unutma (forgot) kapısı, giriş (input) kapısı ve çıkış (output) kapısıdır. Unutma kapısı, hafıza hücresindeki bilginin ne kadarını unutacağını belirler. Bu kapı önceki hafıza hücresindeki bilgi ve şimdiki girdiye dayanarak belirlenir. Giriş kapısı, hafıza hücresine ne kadar yeni bilgi ekleneceğine karar verir. Bu kapı şimdiki girdi ve önceki saklı durum arasındaki ilişkiye dayanarak belirlenir. Çıkış kapısı ise, hafıza hücresindeki bilgiyi kullanarak ne kadar çıktı üretileceğini belirler. Bu kapı hafıza hücresindeki bilgi ve şimdiki girdiye dayanarak belirlenir.

Geçitli Tekrarlayan Birim (Gated Recurrent Unit – GRU) mimarisi, LSTM ile benzer şekilde RNN mimarisinin alt türüdür. Özellikle doğal dil işleme gibi sıralı verilerin modellenmesi için kullanılır. GRU, LSTM ile benzer bir kapı mekanizması kullanır. Ancak, LSTM'e göre daha az parametreye sahip ve daha hızlıdır. GRU mimarisinde LSTM mimarisinden farklı olarak 2 temel kapı bulunmaktadır. Bu kapılar yenileme (reset) ve güncelleme (update) kapılarıdır. Yenileme kapısı önceki durumdaki hafıza hücresinin ne kadarının unutulacağını belirler. Bu kapı önceki durumdaki saklı durum ve şimdiki durumdaki girdi arasındaki bir ilişkiye dayanarak belirlenir. Güncelleme kapısı hafıza hücresinin ne kadarının güncelleneceğini belirler. Bu kapı önceki durumdaki saklı durum ve şimdiki durumdaki girdi arasındaki ilişkiye göre belirlenir [16].

Evrişimli Sinir Ağları (Convolutional Neural Network – CNN) mimarisi, özellikle görüntü işleme alanında kullanılan bir derin öğrenme mimarisidir. CNN mimarisi evrişim katmanlarından oluşur ve her katmanda filtreler kullanılarak verilerin özelliklerinin çıkarılmasına dayanır [15]. CNN modelleri girdi olarak alınan görüntülerin özelliklerini çıkarır ve bu işlemi yaparken filtreler kullanır. Örneğin bir görüntü için bu filtreler kenarlar, köşeler veya şekiller olabilir. Ardından bir veya birden fazla tam bağımlı katman kullanarak bu özellikleri sınıflara ayırır. Bu katmanlar son çıktıyı hesaplayarak modelin tahmin ettiği sınıfın olasılığını hesaplar. CNN mimarisinin avantajı verilerin özelliklerini otomatik öğrenebilmesidir. Bu nedenle verilerin özelliklerinin elle çıkarılmasına gerek yoktur. Ayrıca eğitim verilerindeki küçük değişikliklere dayanıklıdır ve genelleştirme yeteneği yüksektir.

### 3.4. Ölçüm Metrikleri

Duygu analizi sonuçları doğruluk (accuracy) ve F1-Puanı (F1-Score) performans ölçütleri ile değerlendirilmiştir. Doğruluk değeri, algoritmanın doğru sınıflandırdığı örnek sayısının toplam örnek sayısına oranıdır. F1-Puanı değeri, geri çağırma ve kesinlik değerlerinin harmonik ortalaması olarak görülmektedir. Geri çağırma değeri, algoritmanın pozitif örnekleri sınıflandırma başarısını ölçmektedir. Kesinlik değeri, algoritmanın pozitif olarak sınıflandırdığı örneklerin gerçekten pozitif olma olasılığını ölçmektedir. F1-Puanı değerinin yüksek olabilmesi için her iki değer de yüksek olması gerekmektedir [17].

Geleneksel makine öğrenmesi algoritmaları deneylerinde N-gram ve ağırlıklandırma metotları kullanılmıştır. N-gram metodu bir dil modelleme tekniğidir ve metindeki ardışık kelime gruplarını temsil etmektedir. Bu çalışmada üç farklı N-gram tekniği kullanılmıştır. Bunlar Unigram, Bigram ve Trigram'dır. Unigram, bir metindeki her bir kelimeyi ayrı ayrı temsil eden bir N-gram tekniğidir ve metindeki kelime dağılımını modellemek için kullanılır. Bigram, metindeki ardışık iki kelimeyi temsil eden bir N-gram tekniğidir ve bir kelimenin ardından gelen kelimeyi tahmin etmek için kullanılır. Trigram, metindeki ardışık üç kelimeyi temsil eden bir N-gram tekniğidir ve bir kelimenin iki önceki kelimesine ve bir sonraki kelimesine dayalı olarak bir kelimenin tahmin edilmesi için kullanılır. Ağırlıklandırma metodu, metinlerin temsil edilmesi ve özelliklerin çıkarılması için kullanılan yöntemdir. Bu çalışmada üç farklı ağırlıklandırma yöntemi kullanılmıştır. Bunlar Terim Varlığı (Term Presence – TP), Terim Frekansı (Term Frequency – TF) ve Terim Frekansı – Ters Doküman Frekansı (Term Frequency – Inverse Document Frequency – TF-IDF)'dir. TP, bir belgede bir terimin var olup olmadığını ifade eder ve belgede terim varsa 1, yoksa değeri 0 olur. TF, bir belgede bir terimin ne sıklıkla geçtiğini ifade eder. TF-IDF ise bir terimin belge içindeki sıklığını ve tüm belgelerdeki yaygınlığını birleştiren bir yöntemdir. TP ve TF basit bir metin temsil yöntemi sağlarken, TF-IDF daha gelişmiş bir özellik çıkarım yöntemidir ve daha ayrıntılı bir metin temsili sunar.

## 4. Bulgular

Bu bölümde ürün yorumlarından hazırlanan veri setleri ile olumlu ve olumsuz olarak iki sınıflı duygu analizi deneyleri yapılmıştır. Deneylerin ilk aşamasında manuel ve otomatik etiketlenmiş iki veri seti kullanılarak derin öğrenme algoritmaları tarafından elde edilen sonuçlar karşılaştırılmıştır. Sonraki aşamada otomatik etiketlenmiş veri seti ile geleneksel makine öğrenmesi algoritmaları ile sonuçlar alınıp derin öğrenme algoritmalarıyla alınan sonuçlar karşılaştırılmıştır.

Çeşitli sınıflandırma algoritmalarının performansını ölçmek için deneylerde Keras ve Python kullanılmıştır. 10 katlı bir çapraz doğrulama yaklaşımı ile verinin 10 eşit parçaya bölünmesi ve modelin 9 parça ile eğitilip geri kalan parça ile test edilmesi ile sonuçlar alınmıştır. Ayrıca bu işlem testte kullanılacak her parça için 10 kez tekrarlanmaktadır.

Yapılan deneylerin temel amacı manuel ve otomatik olarak etiketlenen veri setlerini kullanarak çeşitli sınıflandırma algoritmalarının performanslarını karşılaştırmaktır. Bu karşılaştırma sonucunda da otomatik etiketleme modelimizin gerçek hayatta kullanılabilirliğini göstermektedir.

Tablo 3 ve Tablo 4 manuel etiketlenmiş veri seti kullanılarak derin öğrenme algoritmaları ile elde edilen sırasıyla doğruluk ve F1-Puanı sonuçlarını göstermektedir. Doğruluk sonuçları incelendiğinde 0,96 ile en yüksek sonuca ulaşan algoritmanın GRU algoritması olduğu görülmektedir. F1-Puanı sonuçlarına bakıldığında da 0,9598 ile yine GRU algoritmasının en yüksek sonucu verdiği görülmektedir.

Tablo 3. Manuel Veri Seti Kullanılarak Derin Öğrenme Algoritmaları Tarafından Elde Edilen Doğruluk Sonuçları (Table 3. Accuracy Results Obtained by Deep Learning Algorithms Using Manual Data Set)

Gömme Boyutu	Katman Birimi	RNN	LSTM	GRU	CNN
100	100	0,9519	0,9566	0,9571	0,9574
200	100	0,9510	0,9557	0,9581	0,9549
300	100	<b>0,9525</b>	<b>0,9572</b>	0,9563	0,9562
100	200	0,9488	0,9533	0,9521	0,9544
200	200	0,9498	0,9542	0,9590	0,9573
300	200	0,9473	0,9550	0,9598	<b>0,9578</b>
100	300	0,9486	0,9556	0,9573	0,9559
200	300	0,9490	0,9564	<b>0,9600</b>	0,9559
300	300	0,9521	0,9550	0,9588	0,9574

Tablo Hata! Belgede belirtilen stilde metne rastlanmadı.. Manuel Veri Seti Kullanılarak Derin Öğrenme Algoritmaları Tarafından Elde Edilen F1-Puanı Sonuçları (Table 4. F1-Score Results Obtained by Deep Learning Algorithms Using Manual Data Set)

Gömme Boyutu	Katman Birimi	RNN	LSTM	GRU	CNN
100	100	0,9516	0,9563	0,9568	0,9573
200	100	0,9508	0,9553	0,9578	0,9547
300	100	<b>0,9524</b>	<b>0,9571</b>	0,9560	0,9559
100	200	0,9478	0,9530	0,9519	0,9541
200	200	0,9496	0,9541	0,9589	0,9571
300	200	0,9476	0,9547	0,9596	<b>0,9576</b>
100	300	0,9484	0,9554	0,9569	0,9557
200	300	0,9490	0,9562	<b>0,9598</b>	0,9557
300	300	0,9519	0,9550	0,9587	0,9573

Tablo 5 ve Tablo 6 ise otomatik etiketlenmiş veri seti kullanılarak derin öğrenme algoritmaları ile elde edilen sırasıyla doğruluk ve F1-Puanı sonuçlarını göstermektedir. Bu sonuçlarda doğruluk sonuçları incelendiğinde 0,9555 ile en yüksek sonuca ulaşan algoritmanın CNN olduğu görülmektedir. F1-Puanı sonuçlarına bakıldığında da 0,9505 ile yine CNN algoritmasının en yüksek sonucu verdiği görülmektedir.

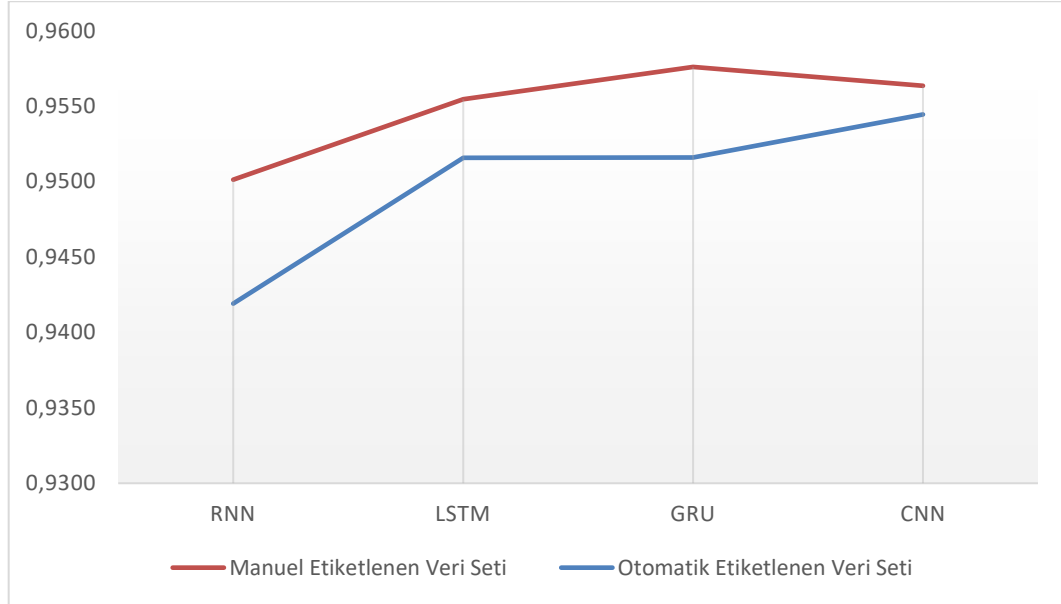
Tablo 5. Otomatik Veri Seti Kullanılarak Derin Öğrenme Algoritmaları Tarafından Elde Edilen Doğruluk Sonuçları (Table 5. Accuracy Results Obtained by Deep Learning Algorithms Using Automatic Data Set)

Gömme Boyutu	Katman Birimi	RNN	LSTM	GRU	CNN
100	100	0,9421	0,9528	0,9509	0,9539
200	100	0,9451	<b>0,9531</b>	0,9515	0,9521
300	100	<b>0,9460</b>	0,9521	0,9514	0,9538
100	200	0,9447	0,9499	0,9520	0,9552
200	200	0,9452	0,9509	0,9517	0,9550
300	200	0,9411	0,9513	0,9513	<b>0,9555</b>
100	300	0,9373	0,9513	<b>0,9525</b>	0,9538
200	300	0,9390	0,9521	0,9524	0,9554
300	300	0,9367	0,9507	0,9507	0,9552

Tablo 6. Otomatik Veri Seti Kullanılarak Derin Öğrenme Algoritmaları Tarafından Elde Edilen F1-Puanı Sonuçları (Table 6. F1-Score Results Obtained by Deep Learning Algorithms Using Automatic Data Set)

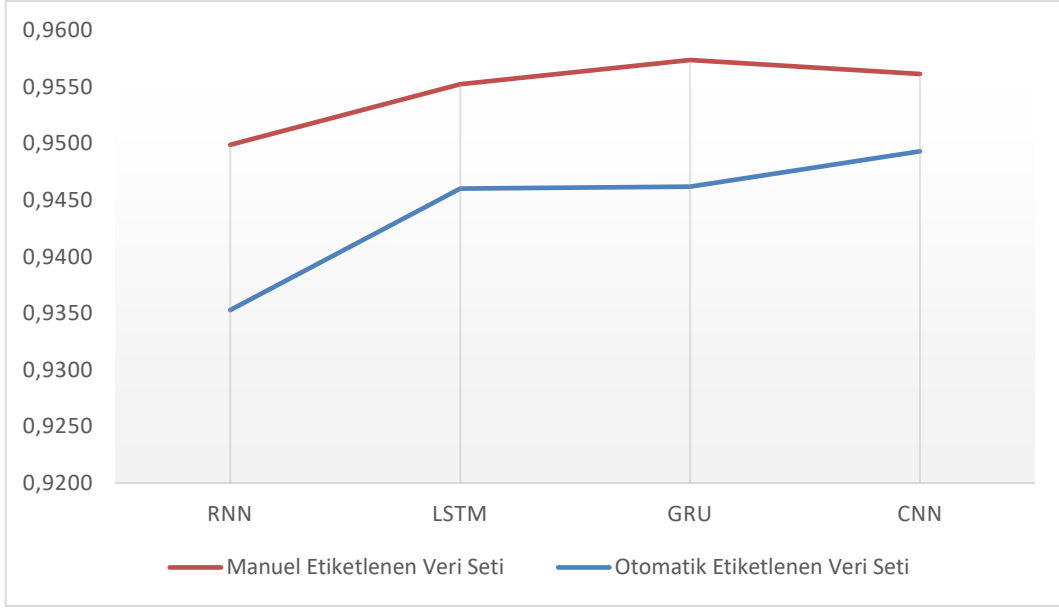
Gömme Boyutu	Katman Birimi	RNN	LSTM	GRU	CNN
100	100	0,9360	0,9471	0,9458	0,9487
200	100	0,9392	<b>0,9478</b>	0,9460	0,9462
300	100	<b>0,9397</b>	0,9469	0,9463	0,9488
100	200	0,9390	0,9438	0,9462	0,9504
200	200	0,9391	0,9453	0,9459	0,9497
300	200	0,9343	0,9458	0,9457	<b>0,9505</b>
100	300	0,9302	0,9456	0,9472	0,9487
200	300	0,9320	0,9467	<b>0,9473</b>	0,9504
300	300	0,9279	0,9453	0,9454	0,9503

Manuel ve otomatik etiketlenmiş veri setleri kullanılarak derin öğrenme algoritmaları ile elde edilen ortalama doğruluk ve F1-Puanı sonuçları Şekil 1 ve Şekil 2’de gösterilmektedir. Bu grafikler incelendiğinde CNN algoritması ile alınan sonuçların manuel ve otomatik etiketlenmiş veri setleri için birbirine yakın olduğu görülmektedir. CNN algoritmasının sonuçları daha detaylı incelendiğinde, manuel etiketlenen veri seti ile elde edilen ortalama doğruluk 0,9563 ve otomatik etiketlenen veri seti ile elde edilen ortalama doğruluk ise 0,9544 olduğu ve aradaki farkın 0,0019 olarak hesaplandığı görülmüştür. Ortalama F1-Puanı sonuçlarına bakıldığında da yine CNN algoritmasının verdiği sonuçların yüksek olduğu görülmektedir. CNN algoritmasında manuel etiketlenen veri seti ile elde edilen ortalama F1-Puanı 0,9562 ve otomatik etiketlenen veri seti ile elde edilen F1-Puanı 0,9493 olduğu görülmüştür. Bu sonuçlarda da aradaki farkın 0,0069 olarak hesaplandığı görülmüştür.



Şekil 1. Derin Öğrenme Algoritmaları Ortalama Doğruluk Sonuçları (Figure 1. Deep Learning Algorithms Average Accuracy Results)





Şekil 2. Derin Öğrenme Algoritmaları Ortalama F1-Puanı Sonuçları (Figure 2. Deep Learning Algorithms Average F1-Score Results)

Manuel ve otomatik etiketlenmiş veri setleriyle elde edilen sonuçlar karşılaştırıldığında aradaki farkların çok düşük olduğu görüldüğünden oluşturulan otomatik etiketleme modelinin kullanılabilirliği saptanmıştır. Bu sayede milyonlarca verinin olduğu büyük veri setleri bu model sayesinde dakikalar içerisinde otomatik olarak etiketlenebilecektir. Sonraki aşamada otomatik etiketlenmiş veri setinin geleneksel makine öğrenmesi ile elde edilen sonuçlar gösterilmektedir.

Tablo 7 ve Tablo 8 otomatik etiketlenmiş veri seti kullanılarak geleneksel makine öğrenmesi algoritmaları ile elde edilen sırasıyla doğruluk ve F1-Puanı sonuçlarını göstermektedir. Bu sonuçlarda doğruluk sonuçları incelendiğinde 0,9090 ile en yüksek sonuca ulaşan algoritmanın SVM olduğu görülmektedir. F1-Puanı sonuçlarına bakıldığında da 0,8986 ortalama ile yine SVM algoritmasının en yüksek sonucu verdiği görülmektedir.

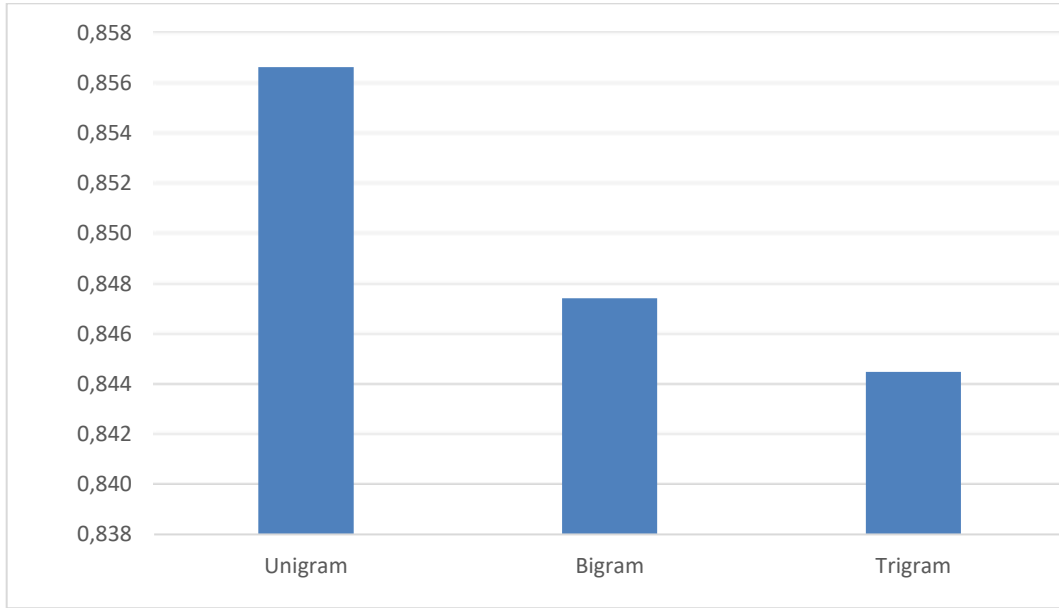
Tablo 7. Otomatik Veri Seti Kullanılarak Geleneksel Makine Öğrenmesi Algoritmaları Tarafından Elde Edilen Doğruluk Sonuçları (Table 7. Accuracy Results Obtained by Traditional Machine Learning Algorithms Using Automatic Data Set)

N-Gram ve Ağırlıklandırma Metotları	LR	SVM	RF	NB
Unigram, TF	0,9074	0,9065	0,8183	0,7983
Unigram, TP	0,8999	0,8988	0,8168	0,7963
Unigram, TF-IDF	<b>0,9081</b>	<b>0,9090</b>	<b>0,8202</b>	<b>0,7998</b>
Bigram, TF	0,8958	0,8952	0,8131	0,7873
Bigram, TP	0,8898	0,8889	0,8121	0,7863
Bigram, TF-IDF	0,8961	0,8971	0,8113	0,7960
Trigram, TF	0,8936	0,8927	0,8124	0,7841
Trigram, TP	0,8878	0,8867	0,8114	0,7832
Trigram, TF-IDF	0,8935	0,8867	0,8106	0,7910

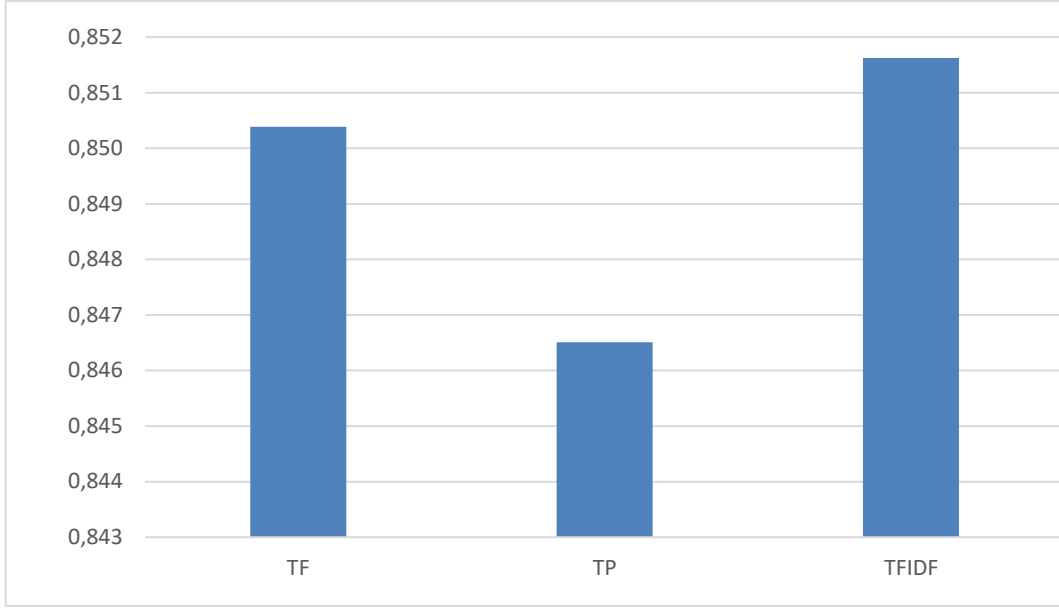
Tablo 8. Otomatik Veri Seti Kullanılarak Geleneksel Makine Öğrenmesi Algoritmaları Tarafından Elde Edilen F1-Puanı Sonuçları  
(Table 8. F1-Score Results Obtained by Traditional Machine Learning Algorithms Using Automated Dataset)

N-Gram ve Ağırlıklandırma Metotları	LR	SVM	RF	NB
Unigram, TF	<b>0,8975</b>	0,8967	<b>0,7931</b>	<b>0,7785</b>
Unigram, TP	0,8890	0,8880	0,7917	0,7756
Unigram, TF-IDF	0,8974	<b>0,8986</b>	0,7894	0,7579
Bigram, TF	0,8844	0,8838	0,7865	0,7600
Bigram, TP	0,8775	0,8766	0,7854	0,7578
Bigram, TF-IDF	0,8842	0,8856	0,7804	0,7554
Trigram, TF	0,8820	0,8811	0,7854	0,7548
Trigram, TP	0,8753	0,8742	0,7843	0,7526
Trigram, TF-IDF	0,8814	0,8742	0,7798	0,7488

Şekil 3 ve Şekil 4'te geleneksel makine öğrenmesi algoritmaları ile yapılan deneylerde kullanılan N-Gram ve ağırlıklandırma metotlarının karşılaştırılması gösterilmektedir.



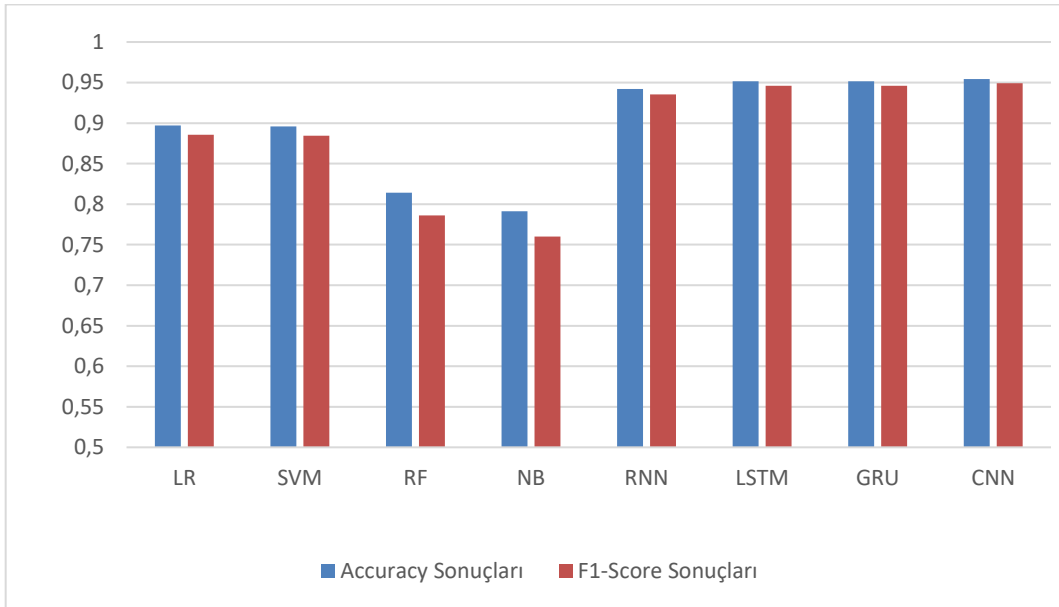
Şekil 3. Geleneksel Makine Öğrenmesi Algoritmaları N-Gram Metotları Ortalama Doğruluk Sonuçları (Figure 3. Traditional Machine Learning Algorithms N-Gram Methods Average Accuracy Results)



Şekil 4. Geleneksel Makine Öğrenmesi Algoritmaları Ağırlıklandırma Metotları Ortalama Doğruluk Sonuçları (Figure 4. Traditional Machine Learning Algorithms Weighting Methods Average Accuracy Results)

Burada görüldüğü gibi N-gram metotlarından Unigram 0,8566 ortalama ile diğer metotlara göre yüksek ortalamaya sahiptir. Ağırlıklandırma metotlarından ise TF-IDF metodu 0,8516 ortalama ile diğer metotlara göre yüksek ortalamaya sahiptir. Tersine en düşük ortalama ise 0,8445 ile Trigram ve 0,8465 ile TP yöntemlerinin olduğu görülmüştür. Yukarıda yer alan tablolara da bu metotlar özelinde bakıldığında en yüksek sonuçlar Unigram ve TF-IDF yöntemlerinin birlikte kullanıldığı deneylerde elde edildiği net bir şekilde görülmektedir.

Şekil 5'te otomatik olarak etiketlenmiş veri seti kullanılarak tüm algoritmalar ile elde edilen sonuçlar gösterilmektedir. Burada derin öğrenme algoritmalarının makine öğrenmesi algoritmalarına karşı başarısı net bir şekilde görülmektedir.



Şekil 5. Otomatik Olarak Etiketlenmiş Veri Seti Kullanılarak Elde Edilen Ortalama Sonuçlar (Figure 5. Average Results Using Automatically Tagged Data Set)

## 5. Sonuç

E-ticaret sektörünün gelişmesiyle birlikte ürünlere yapılan yorumların büyük oranlarda artmasının ardından tüm yorumların manuel analiz edilmesi hem e-ticaret satıcıları hem de müşteriler için önemli bir problem haline gelmiştir. Bu problemin çözümü için ürün yorumlarının otomatik analizini gerçekleştiren bir çalışma sunulmuştur. Sunulan bu çalışmada ürün yorumlarının otomatik sınıflandırılması sağlanarak sektöre fayda sağlanması amaçlanmıştır.

Çalışma kapsamında iki veri seti hazırlanmış ve bu veri setlerinden biri manuel biri de otomatik olarak etiketlenmiştir. Manuel olarak etiketlenen veri farklı zamanlarda iki defa etiketlenmiş ve etiketleme sonuçları karşılaştırılarak doğruluk oranı artırılmaya çalışılmıştır. Otomatik olarak etiketlenen veri seti ise bu çalışma kapsamında oluşturulan LSTM modeli ile elde edilmiştir. Bu LSTM modeli manuel olarak etiketlenen veri seti ile eğitilmiştir. Manuel ve otomatik etiketlenen veri setleriyle dört farklı derin öğrenme algoritması kullanılarak sonuçlar alınıp karşılaştırılmıştır. Elde edilen sonuçlar incelendiğinde manuel ve otomatik etiketlenmiş veri setlerinin birbirine yakın sonuçlar sağladığı gözlemlenmiştir. Çalışmanın son aşamasında otomatik etiketlenmiş veri setiyle dört farklı geleneksel makine öğrenme ve dört farklı derin öğrenme algoritmalarının sınıflandırma performansları karşılaştırılmıştır. Elde edilen sonuçlarda derin öğrenme algoritmaları geleneksel makine öğrenme algoritmalarına karşı daha başarılı olmuşlardır.

## **Kaynakça**

- [1] Rumelli, M., Akkuş, D., Kart, Ö. ve Isik, Z. (2019) Türkçe Metinlerde Makine Öğrenmesi Algoritmaları ile Duygu Analizi. 2019 Innovations in Intelligent Systems and Applications Conference (ASYU). 31 Ekim – 02 Kasım, İzmir, Türkiye. (IEEE, 1-5)
- [2] Topçu, A. E. ve Erkaya, A. E. (2020). Recent Research and Assessments for Computer Engineering. Ed.: Ahmet Ercan Topçu, Livre de Lyon, 1-14 s.
- [3] Tuzcu, S. (2020), Çevrimiçi Kullanıcı Yorumlarının Duygu Analizi ile Sınıflandırılması. ESTUDAM Bilişim Dergisi. 1(2), 1-5.
- [4] Mengutayci, Ü. ve Temurtas, H. (2021). Yapay Sinir Ağları ile Türkçe Otel Yorumlarının Sınıflandırılması. International Black Sea Coastline Countries Scientific Research Symposium – VI. 683-687
- [5] Mayda, I. ve Korkmaz, M. (2018). Sentiment Analysis with Turkish Adjective Dictionary. 2018 Innovations in Intelligent Systems and Applications Conference (ASYU). 4-6 Ekim, Adana, Türkiye. (IEEE, 1-6)
- [6] Gezici, G. ve Yanıkoğlu, B. (2018). Sentiment Analysis in Turkish. Theory and Applications of Natural Language Processing. 255–271.
- [7] Pervan, N. ve Keleş, H. Y. (2017). Sentiment Analysis Using A Random Forest Classifier On Turkish Web Comments. Communications Faculty of Sciences University of Ankara Series A2-A3 Physical Sciences and Engineering. 59(2), 69-79.
- [8] Onan, A. (2021). Ensemble of Classifiers and Term Weighting Schemes for Sentiment Analysis in Turkish. Scientific Research Communications. 1(1), 1-12.
- [9] Demircan, M., Seller, A., Abut, F. ve Akay, M. F. (2021). Developing Turkish sentiment analysis models using machine learning and e-commerce data. International Journal of Cognitive Computing in Engineering. 2, 202-207.
- [10] Toçoğlu, M. A., Öztürkmenoğlu, O. ve Alpoçak, A. (2019). Emotion Analysis from Turkish Tweets using Deep Neural Networks. IEEE Access. 7, 183061-183069.
- [11] Çataltaş, M., Doğramacı, S., Yumuşak, S. ve Öztoprak, K. (2020). Extraction of Product Defects and Opinions from Customer Reviews by Using Text Clustering and Sentiment Analysis. 2020 IEEE International Conference on Big Data (Big Data), Atlanta, ABD, 4529-4534.
- [12] Maalouf, M. (2011). Logistic regression in data analysis: an overview. International Journal of Data Analysis Techniques and Strategies. 3(3), 281-299.
- [13] Mahesh, B. (2020). Machine Learning Algorithms - A Review. International Journal of Science and Research (IJSR). 9(1), 381-386.
- [14] Drăguț, L., Belgiu, M. (2016). Random Forest in remote sensing: A review of applications and future directions. ISPRS Journal of Photogrammetry and Remote Sensing. 114, 24-31.
- [15] LeCun, Y., Bengio, Y., Hinton, G. (2015). Deep learning. Nature. 521, 436–444.
- [16] Yang, S., Yu X., Zhou, Y. (2020). LSTM and GRU Neural Network Performance Comparison Study: Taking Yelp Review Dataset as an Example. 2020 International Workshop on Electronic Communication and Artificial Intelligence (IWECAI). Shanghai, China, 98-101.
- [17] Bozuyula, M., Özçift, A. (2022). Developing a fake news identification model with advanced deep language transformers for Turkish COVID-19 misinformation data. 30, 908-926.