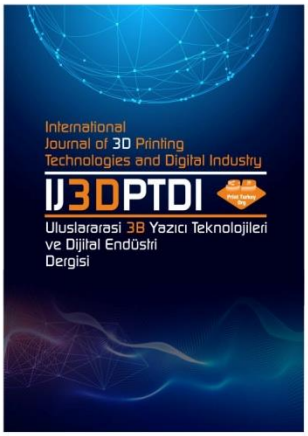



ULUSLARARASI 3B YAZICI TEKNOLOJİLERİ
VE DİJİTAL ENDÜSTRİ DERGİSİ
INTERNATIONAL JOURNAL OF 3D PRINTING
TECHNOLOGIES AND DIGITAL INDUSTRY

ISSN:2602-3350 (Online)
URL: https://dergipark.org.tr/ij3dptdi

# MÜŞTERİ GERİBİLDİRİMLERİNİN BERT TABANLI DUYGU ANALİZİNDE YILDIZ SINIFLANDIRMA TEMELLİ BİR YAKLAŞIM

## A STAR RATING-BASED APPROACH IN BERT-BASED SENTIMENT ANALYSIS OF CUSTOMER FEEDBACK

**Yazarlar (Authors):** İlkay Onay, Kıyas Kayaalp*

***Bu makaleye şu şekilde atıfta bulunabilirsiniz (To cite to this article):*** *Onay İ., Kayaalp K., "Müşteri Geribildirimlerinin Bert Tabanlı Duygu Analizinde Yıldız Sınıflandırma Temelli Bir Yaklaşım" Int. J. of 3D Printing Tech. Dig. Ind., 9(3): 556-568, (2025).*

**DOI:** 10.46519/ij3dptdi.1732179

Araştırma Makale/ Research Article

**Erişim Linki:** (To link to this article): https://dergipark.org.tr/en/pub/ij3dptdi/archive

# MÜŞTERİ GERİBİLDİRİMLERİNİN BERT TABANLI DUYGU ANALİZİNDE YILDIZ SINIFLANDIRMA TEMELLİ BİR YAKLAŞIM

İlkay Onay[a], Kıyas Kayaalp[a]*

[a]Isparta Uygulamalı Bilimler Üniversitesi, Teknoloji Fakültesi, Bilgisayar Mühendisliği Bölümü, Türkiye

* *Sorumlu Yazar:* kiyaskayaalp@isparta.edu.tr

(**Geliş/Received:** 01.07.25; **Düzeltme/Revised:** 12.11.25; **Kabul/Accepted:** 23.11.25)

## ÖZ

Bu çalışma, müşteri geri bildirimlerinin yapay zekâ tabanlı duygu analizini, geleneksel olumlu/olumsuz sınıflandırmanın ötesine taşıyarak 1'den 5'e kadar yıldız puanlarına göre çok sınıflı bir yapıda ele almaktadır. Bu amaçla, Türkçe müşteri yorumlarından oluşan yaklaşık 900.000 verilik kapsamlı bir veri seti üzerinde, "YıldızSezar" adı verilen bir dizi sınıflandırma modelleri geliştirilmiş ve optimize edilmiştir. Çalışmada, geleneksel makine öğrenmesi yöntemlerinin karmaşık dil yapıları ve ara sınıflarda (2, 3, 4 yıldız) yetersiz kaldığı görülmüş, bu nedenle Transformer tabanlı modellere odaklanılmıştır. Veri kalitesini artırmak için kapsamlı temizleme protokolleri uygulanmış ve sınıf dengesizliği sorununu gidermek amacıyla veri seti, LLaMA tabanlı bir model aracılığıyla üretilen sentetik verilerle zenginleştirilmiştir. ConvBERT mimarisi ve optimize edilmiş eğitim stratejileri kullanılarak geliştirilen nihai model YıldızSezar, test seti üzerinde 0,8996 doğruluk ve 0,7990 makro F1-skoru elde ederek, özellikle ara sınıflarda yüksek başarım göstermiştir. Bu sonuçlar, BERT tabanlı yapay zeka modellerinin, doğru veri hazırlama ve eğitim stratejileriyle, Türkçe gibi morfolojik olarak zengin dillerde dahi karmaşık duygu analizi görevlerini başarıyla yerine getirebileceğini kanıtlamaktadır.

**Anahtar Kelimeler:** Duygu Analizi, Büyük Dil Modelleri, Çok Sınıflı Sınıflandırma, Doğal Dil İşleme, Müşteri Geri Bildirimleri.

# A STAR RATİNG-BASED APPROACH İN BERT-BASED SENTİMENT ANALYSİS OF CUSTOMER FEEDBACK

## ABSTRACT

This study addresses the sentiment analysis of customer feedback using an AI-based approach, moving beyond traditional positive/negative classification to a multi-class structure based on star ratings from 1 to 5. For this purpose, a series of classification models named "YıldızSezar" were developed and optimized on a comprehensive dataset of approximately 900.000 Turkish customer reviews. The study demonstrates that traditional machine learning methods fall short in handling complex language structures and intermediate classes (2, 3, and 4 stars), thus shifting the focus to Transformer-based models. Extensive cleaning protocols were applied to enhance data quality, and the class imbalance problem was addressed by enriching the dataset with synthetic data generated via a LLaMA-based model. The final model, YıldızSezar, developed using the ConvBERT architecture and optimized training strategies, achieved 0,8996 accuracy and a macro F1-score of 0,7990 on the test set, showing high performance, especially in intermediate classes. These results prove that BERT based Artificial Intelligence models, combined with proper data preparation and training strategies, can successfully perform complex sentiment analysis tasks even for morphologically rich languages like Turkish.

**Keywords:** Sentiment Analysis, Large Language Models, Multi-Class Classification, Natural Language Processing, Customer Feedback.

## 1. GİRİŞ

Dijital çağın getirdiği en büyük dönüşümlerden biri, işletmeler ve müşteriler arasındaki etkileşimin boyut ve biçim değiştirmesidir. E-ticaret siteleri, sosyal medya platformları ve forumlar, müşterilerin ürün ve hizmetler hakkındaki deneyimlerini paylaştığı devasa bir geri bildirim havuzu oluşturmuştur. Bu geri bildirimler, işletmelerin stratejik kararlar alması, hizmet kalitesini artırması ve rekabet avantajı sağlaması için paha biçilmez bir veri kaynağıdır [1]. Ancak, her gün üretilen milyonlarca yorumun manuel olarak analiz edilmesi, hem zaman ve maliyet açısından verimsiz hem de insan hatasına açık bir süreçtir. Bu durum, bu yapılandırılmamış veriyi anlamlı içgörülere dönüştürebilecek otomatik sistemlere olan ihtiyacı doğurmuştur.

Bu ihtiyaca yanıt olarak geliştirilen duygu analizi (sentiment analysis), metinlerde ifade edilen duygusal tonu (olumlu, olumsuz, nötr) otomatik olarak belirlemeyi amaçlar. Geleneksel duygu analizi çalışmaları genellikle ikili sınıflandırmaya odaklanırken, müşteri memnuniyetinin gerçek karmaşıklığı genellikle 1'den 5'e kadar olan yıldız puanlaması gibi daha granüler bir ölçekte yatmaktadır. Özellikle Türkçe gibi sondan eklemeli, zengin morfolojiye sahip ve bağlamsal inceliklerin (ironi, kinaye, zıtlık belirten bağlaçlar) yoğun olduğu dillerde, çok sınıflı duygu analizi yapmak önemli bir zorluk teşkil etmektedir [2]. "Ürün kaliteli ama kargo çok geç geldi" gibi bir yorumu sadece olumlu veya olumsuz olarak etiketlemek, içerdiği değerli bilgiyi göz ardı etmek anlamına gelir. Bu tür yorumların 2 veya 3 yıldız gibi ara sınıflara doğru bir şekilde atanması, işletmeler için kritik öneme sahiptir.

Yapılan çalışma, bu zorlukların üstesinden gelmek için, veri zenginleştirme ve derin öğrenme adımlarını birleştiren bir yaklaşım sunmaktadır. İlk olarak, sınıf dengesizliği sorununa yönelik olarak LLaMA tabanlı bir model ile sentetik veri üretilerek veri seti güçlendirilmiştir. Ardından bu zenginleştirilmiş veri seti üzerinde, BERT tabanlı bir mimari kullanılarak "YıldızSezar" adını verdiğimiz nihai sınıflandırma modeli geliştirilmiştir. Bu sistem, Türkçe müşteri yorumlarının doğrudan 1-5 arası yıldız puanlarına göre sınıflandırılmasını hedeflemektedir. Geliştirilen model, işletmelerin müşteri memnuniyetini daha derinlemesine anlamalarına ve veriye dayalı kararlar almalarına olanak tanıyacaktır.

Çalışma konusu hakkında güncel zamanda yapılan literatür çalışmaları gözden geçirildiğinde, duygu analizi üzerine çok sayıda araştırma yapıldığı görülmektedir. Aslan [3], e-ticaret incelemelerinde özellik tabanlı duygu analizi yaparken, Tuna vd. [4] otel geri bildirimlerini makine öğrenmesi ile analiz etmiştir. Alpkoçak vd. [5], Türkçe metinler için farklı makine öğrenmesi yöntemlerini karşılaştırmış ve bu yöntemlerin belirli sınırlılıkları olduğunu ortaya koymuştur. Transformer tabanlı modellerin yükselişiyle birlikte, bu alandaki çalışmalar LLM'lere yönelmiştir. Guven [6], Türkçe ürün yorumlarında BERT, ELECTRA ve ALBERT modellerini karşılaştırırken, Acikalin vd. [7] BERT tabanlı modellerin Türkçe duygu analizindeki etkinliğini kanıtlamıştır. Hibrit yaklaşımlar da dikkat çekmektedir; Wu vd. [8] sözlük tabanlı yöntemleri sinir ağlarıyla birleştirmiş, Siddiqua vd. [9] ise kural tabanlı ve zayıf denetimli öğrenmeyi bir araya getirmiştir. Siğirci vd. [10] Google Play Store yorumları üzerinde, Atılgan ve Yoğurtcu [11] ise Twitter verileri üzerinde çalışmalar yapmıştır. Bu çalışmaların yanı sıra, model optimizasyonu için Akiba vd. [12] tarafından geliştirilen Optuna gibi araçlar ve model performansını değerlendirmek için Sathyanarayanan vd. [13] tarafından detaylandırılan metrikler de bu alandaki araştırmalara yön vermektedir. Diğer önemli çalışmalar arasında Karamollaoğlu vd. [14], Demir ve Bilgin [15], Le [16], Parkavi vd. [17], Gümüş [18], Ban vd. [19], Niu vd. [20] ve Das vd. [21] sayılabilir.

Yapılan literatür taramasında, mevcut çalışmaların çoğunlukla ikili sınıflandırmaya odaklandığı veya çok sınıflı sınıflandırmada ara sınıflarda (2, 3, 4 yıldız) düşük performans sergilediği görülmüştür. Bu çalışma, özel olarak 1-5 yıldız arası çok sınıflı sınıflandırma problemine odaklanması, sınıf dengesizliği ve veri gürültüsü sorunlarına karşı kapsamlı veri temizleme ve sentetik veri üretimi gibi sistematik çözümler sunması ve bu sayede ara sınıflarda yüksek başarım elde etmesiyle özgün bir değer taşımaktadır. Geliştirilen "YıldızSezar" model serisi, bu alandaki boşluğu doldurmayı ve pratik bir çözüm sunmayı amaçlamaktadır.

## 2. MATERYAL VE METOD

Çalışmanın dayandığı temel yöntem, Türkçe müşteri geri bildirimlerini 1'den 5'e kadar yıldız puanlarına göre sınıflandırmak için en uygun yapay zekâ modelini deneysel bir süreçle geliştirmektir. Bu süreç, geleneksel makine öğrenmesi algoritmalarıyla bir taban çizgisi oluşturulması, ardından Transformer tabanlı Büyük Dil Modelleri (LLM) ile performansın artırılması ve son olarak veri hazırlama, model mimarisi ve eğitim stratejilerinin optimize edilmesini kapsamaktadır.

Çalışmanın yöntemi 3 ana bölümden oluşmaktadır:

1. Veri Setleri ve Ön İşleme
2. Modeller ve Algoritmalar
3. Model Geliştirme ve Değerlendirme Süreci

### 2.1. Veri Setleri ve Ön İşleme

Modelin eğitimi için çeşitli kaynaklardan toplanan Türkçe müşteri yorumları birleştirilmiştir. Bu kaynaklar arasında Kaggle platformunda yer alan Hepsiburada Yorum Veri Seti [22] ve Turkish Product Review Dataset [23] ile GitHub'daki Vitamins & Supplements Reviews veri seti bulunmaktadır. Bu birleştirme sonucunda yaklaşık 900.000 yorumdan oluşan ham bir veri havuzu elde edilmiştir.

Model performansını doğrudan etkileyen veri kalitesini artırmak amacıyla kapsamlı bir ön işleme ve temizleme protokolü uygulanmıştır. İlk olarak, veri setindeki tam kopyalar (exact duplicates) ayıklanmıştır. Ardından, yalnızca birkaç kelime değişikliği içeren veya anlamsal olarak birbirinin aynısı olan yorumları tespit etmek için daha gelişmiş bir yöntem olan MinHash LSH (Locality-Sensitive Hashing) tekniği kullanılmıştır. Bu hassas analiz ile 44.741 adet anlamsal benzer (near-duplicate) yorum daha veri setinden çıkarılmıştır. Bu işlemlerden sonra daha spesifik temizliğe geçilmiştir. Bu adımlar şunları içerir:

- HTML etiketlerinin, URL ve e-posta adreslerinin kaldırılması.
- Anlamsal katkısı olmayan emojilerin temizlenmesi.
- Metinlerin küçük harfe dönüştürülmesi ve gereksiz boşlukların giderilmesi.
- Yorum başlarında bulunan "başlık:", "yorum:" gibi kalıpların silinmesi.
- Belirli bir karakter uzunluğunun (ör. 30) altındaki anlamsız veya çok kısa yorumların filtrelenmesi.

Veri seti, modelin genelleme yeteneğini güvenilir bir şekilde ölçmek amacıyla orijinal veriseti 0,70 eğitim, 0,15 doğrulama ve 0,15 test seti olarak bölünmüştür.

### 2.2. Sentetik Veri Üretimi

Veri setindeki en büyük zorluklardan biri, 5 yıldızlı yorumların diğer sınıflara göre sayıca çok fazla olmasıyla ortaya çıkan sınıf dengesizliği sorunuydu. Bu sorunu çözmek ve modelin azınlık sınıfları (özellikle 1, 2 ve 3 yıldız) daha iyi öğrenmesini sağlamak amacıyla, sentetik veri üretimi yöntemine başvurulmuştur. Bu amaçla, mehmetfevzi/Turkish-Llama-8b-DPO-v0.1 LLM'i kullanılarak her bir azınlık sınıfı için gerçekçi ve çeşitli yorumlar üretilmiştir. Üretilen sentetik veriler de aynı titiz temizleme protokolünden geçirilerek nihai eğitim setine dâhil edilmiştir.

Modelin ilgili yıldız derecelendirmesine uygun, bağlamsal ve gerçekçi yorumlar üretmesini sağlamak amacıyla aşağıdaki gibi spesifik prompt (istem) şablonları kullanılmıştır:

- **1 Yıldız:** "Bir kullanıcı olarak, bir ürünü değerlendiriyormuş gibi, ürünün kalitesinin ne kadar kötü olduğunu anlatan bir paragraf yorum yaz. Yorum 1 yıldız olarak değerlendirilsin."

- **2 Yıldız:** "Bir kullanıcı olarak, bir ürünü değerlendiriyormuş gibi, ürünün hem iyi hem de kötü yanlarından bahseden, ama genel olarak hayal kırıklığına uğratan bir paragraf yorum yaz. Yorum 2 yıldız olarak değerlendirilsin."

- **3 Yıldız:** "Bir kullanıcı olarak, bir ürünü değerlendiriyormuş gibi, ürünün ortalama olduğunu, bazı iyi yönleri ve bazı kötü yönleri olduğunu anlatan bir paragraf yorum yaz. Yorum 3 yıldız olarak değerlendirilsin."

- **4 Yıldız:** "Bir kullanıcı olarak, bir ürünü değerlendiriyormuş gibi, ürünün iyi olduğunu, ancak birkaç eksik tarafı da olduğunu anlatan bir paragraf yorum yaz. Yorum 4 yıldız olarak değerlendirilsin."

• **5 Yıldız:** "Bir kullanıcı olarak, bir ürünü değerlendiriyormuş gibi, ürünün çok iyi olduğunu anlatan bir paragraf yorum yaz. Yorum 5 yıldız olarak değerlendirilsin."

Üretilen ham sentetik verinin, LLM'lerin doğası gereği içerdiği gürültü ve tutarsızlıkları gidermek, model eğitiminde kullanılacak verinin kalitesini en üst düzeye çıkarmak için çok adımlı bir temizleme protokolü uygulanmıştır. Bu sentetik verilere uygulanan protokol şunlardır:

• LLM Artefaktlarının Temizlenmesi: Modelin ürettiği metin başlangıç/bitiş belirteçleri (<|end_of_text|>, <|start_header_id|>) gibi teknik kalıntılar kaldırılmıştır.

• "Elbette, işte bir örnek:", "Tabii ki, aşağıda bir yorum bulabilirsiniz:" gibi, prompt'a verilen yanıt niteliğindeki ifadeler metinlerden temizlenmiştir.

• Anlamsız tekrarlayan kelime veya cümle grupları içeren yorumlar tespit edilerek temizlenmiştir.

• Metinlerin başındaki ve sonundaki anlamsız boşluklar veya karakterler kırılarak, yorumun anlamlı bir kelime ile başlayıp bittiğinden emin olunmuştur.

Bu spesifik temizlik işlemlerinden sonra sentetik olmayan verisetine uygulanan temizlik adımları üzerinden de geçirilmiştir.

Temizlenmiş 903.786 adet sentetik veri, 925.114 adet orijinal eğitim verisi ile birleştirilerek 1.828.900 veriden oluşan nihai eğitim veri seti oluşturulmuştur. Sentetik veri sadece eğitim verisetine eklenmiştir. Bu sebeple sentetik veriye aşırı uyum riski olmamaktadır. Bu birleştirilmiş veri setinin sınıflara göre dağılımı, orijinal ve sentetik veri sayıları ile sentetik verinin her bir sınıftaki oransal katkısı Çizelge 1'de detaylı olarak sunulmaktadır. Çizelge 1'den de görüleceği üzere, nihai eğitim setinin 0,4942'si sentetik verilerden oluşmaktadır. Veri artırımı stratejisi, özellikle en az örneğe sahip olan 2 yıldızlı sınıfı hedeflemiş ve bu sınıftaki verilerin 0,8360'ının sentetik verilerden oluşmasını sağlayarak sınıf dengesine önemli bir katkıda bulunmuştur. Bu dengelenmiş ve zenginleştirilmiş veri seti, modelin daha sağlam ve güvenilir tahminler yapması için temel oluşturmaktadır.

**Çizelge 1.** Nihai Eğitim Veri Setinin Sınıflara Göre Dağılımı ve Sentetik Veri Oranları

| Yıldız Derecesi | Orijinal Veri Sayısı | Sentetik Veri Sayısı | Toplam Veri Sayısı | Sentetik Veri Oranı |
|---|---|---|---|---|
| **1 Yıldız** | 39.224 | 93.096 | 132.320 | 0,7036 |
| **2 Yıldız** | 22.424 | 114.321 | 136.745 | 0,8360 |
| **3 Yıldız** | 69.427 | 117.268 | 186.695 | 0,6281 |
| **4 Yıldız** | 150.658 | 126.520 | 277.178 | 0,4565 |
| **5 Yıldız** | 643.381 | 452.581 | 1.095.962 | 0,4130 |
| **Toplam** | 925.114 | 903.786 | 1.828.900 | 0,4942 |

## 2.3. Modeller ve Algoritmalar

Çalışmanın ilk aşamasında, bir temel performans (baseline) oluşturmak amacıyla Naive Bayes, Lojistik Regresyon, Rastgele Orman ve Destek Vektör Sınıflandırıcısı gibi geleneksel makine öğrenmesi algoritmaları kullanılmıştır.

Bu algoritmaların sınırlılıkları gözlemlendikten sonra, metnin bağlamsal ve anlamsal yapısını çok daha iyi kavrayabilen Büyük Dil Modelleri (LLM) üzerine odaklanılmıştır. Bu kapsamda, özellikle BERTurk projesi [24] gibi Türkçe diline özel olarak önceden eğitilmiş modellerin bulunduğu kaynaklardan yararlanılmıştır. Hugging Face platformu üzerinden, aşağıdaki BERT tabanlı modeller denenmiştir:

• berturk/distilbert-base-turkish-cased: Hızlı ve daha az kaynak gerektiren bir BERT versiyonu.

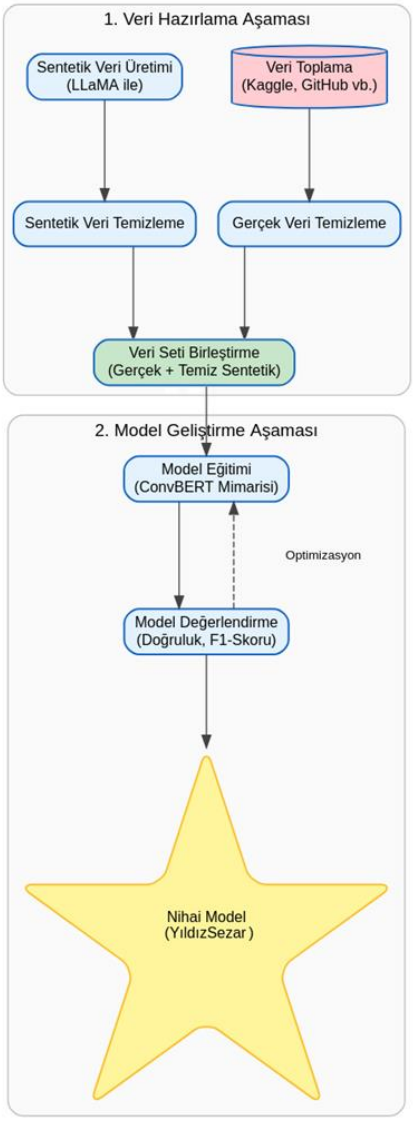
• dbmdz/electra-base-turkish-cased: Daha verimli bir eğitim süreci sunan ELECTRA mimarisi.

• dbmdz/convbert-base-turkish-mc4-uncased: Özellikle konuşma dili ve informal metinler için optimize edilmiş, çalışmanın nihai modelinde kullanılan ConvBERT mimarisi.

Model eğitimi PyTorch ve Hugging Face Transformers kütüphaneleri kullanılarak gerçekleştirilmiştir.

### 2.4. Model Geliştirme ve Değerlendirme Süreci

Model geliştirme, Şekil 1'de özetlenen iteratif bir süreçle gerçekleştirilmiştir. Her bir versiyonda ("YıldızSezar V1, V2,...") karşılaşılan bir soruna çözüm üretilmeye çalışılmıştır.




**Şekil 1.** Model Geliştirme Süreci Akış Şeması

Bu süreç aşağıdaki ana adımları içermektedir:

1. Taban Çizgisi Belirleme: Geleneksel ML algoritmalarının performansının ölçülmesi.

2. LLM'lere Geçiş: Farklı BERT tabanlı modellerin denenmesi.

3. Veri Zenginleştirme: Veri seti boyutunun artırılması.

4. Veri Kalitesini Artırma: Kapsamlı veri temizleme protokollerinin uygulanması.

5. Dengesizlikle Mücadele: Temizlenmiş sentetik verilerin veri setine entegrasyonu.

6. Eğitim Optimizasyonu: En iyi performansı veren epoch sayısının belirlenmesi ve aşırı öğrenmenin (overfitting) önlenmesi.

### 2.5. Değerlendirme Metrikleri

Geliştirilen modellerin performansını nicel olarak ölçmek ve karşılaştırmak için standart sınıflandırma metrikleri kullanılmıştır. Bu metrikler aşağıda açıklanmıştır:

• **Doğruluk (Accuracy):** Modelin doğru yaptığı tahminlerin toplam tahmin sayısına oranıdır. Genel performansı gösterse de, sınıf dağılımı dengesiz olan veri setlerinde yanıltıcı olabilir.

$$Doğruluk = (TP + TN) / (TP + TN + FP + FN) \quad (1)$$

• **Kesinlik (Precision):** Modelin pozitif olarak tahmin ettiği örneklerden kaç tanesinin gerçekten pozitif olduğunu ölçer.

$$Kesinlik = TP / (TP + FP) \quad (2)$$

• **Duyarlılık (Recall):** Gerçekte pozitif olan örneklerden kaç tanesinin model tarafından doğru bir şekilde pozitif olarak tahmin edildiğini gösterir.

$$Duyarlılık = TP / (TP + FN) \quad (3)$$

• **F1-Skoru (F1-Score):** Kesinlik ve duyarlılık metriklerinin harmonik ortalamasıdır. Özellikle sınıf dengesizliği durumunda, bu iki metriği dengeleyen daha güvenilir bir performans göstergesidir.

$$F1\text{-}Skoru = 2 * (Kesinlik * Duyarlılık) / (Kesinlik + Duyarlılık) \quad (4)$$

• **Makro F1-Skoru (Macro F1-Score):** Çok sınıflı sınıflandırma problemlerinde, her sınıf için F1-skoru ayrı ayrı hesaplanır ve bu skorların aritmetik ortalaması alınır. Bu metrik, her sınıfa eşit ağırlık vererek azınlık sınıflarındaki performansı daha adil bir şekilde yansıtır. Çalışmamızda sınıf dengesizliği

bulunduğu için makro F1-skoru, temel başarı metriklerinden biri olarak kabul edilmiştir.
(Not: TP: Gerçek Pozitif, TN: Gerçek Negatif, FP: Yanlış Pozitif, FN: Yanlış Negatif)

## 3. TARTIŞMA ve BULGULAR

Bu çalışma kapsamında yürütülen kapsamlı deneysel süreç, önemli bulgular ortaya koymuştur. Geliştirme süreci, geleneksel yöntemlerin yetersizliğinden başlayarak, adım adım optimize edilmiş bir LLM'e ulaşmıştır.

İlk olarak, geleneksel makine öğrenmesi algoritmaları ile yapılan testler, Çizelge 2'de görüldüğü gibi en iyi durumda bile doğruluk oranlarının 0,64 civarında kaldığını göstermiştir. Bu modellerin özellikle ara sınıfları (2, 3, 4 yıldız) ayırt etmede ve "ama, fakat" gibi bağlaç içeren karmaşık cümleleri anlamada başarısız olduğu gözlemlenmiştir.

**Çizelge 2.** Geleneksel Makine Öğrenmesi Algoritmalarının Doğruluk Sonuçları

| Algoritma | 5000 Örnek | 7500 Örnek | 10000 Örnek |
|---|---|---|---|
| **Naive Bayes** | 0,58 | 0,57 | 0,57 |
| **Logistic Regression** | 0,63 | 0,63 | 0,64 |
| **Random Forest** | 0,61 | 0,62 | 0,62 |
| **K-Nearest Neighbors** | 0,59 | 0,60 | 0,61 |
| **Gradient Boosting** | 0,62 | 0,63 | 0,63 |
| **Support Vector Classifier** | 0,64 | 0,64 | 0,64 |

Bu sonuçların ardından LLM'lere geçilmiş ve veri setinin yaklaşık 900.000 örneğe çıkarılması, kapsamlı veri temizliği ve sentetik veri entegrasyonu gibi adımlar sonucunda performansta kademeli olarak büyük bir artış sağlanmıştır.

Sonraki aşamada Türkçe müşteri geri bildirimlerinin çok sınıflı yıldız derecelendirmesi görevinde en yüksek performansı sergileyecek model mimarisini belirlemektir. Bu doğrultuda, önerilen nihai modelin mimari seçimini gerekçelendirmek amacıyla, farklı Transformer tabanlı Büyük Dil Modelleri (LLM) arasında sistematik bir karşılaştırma yapılmıştır. Bu analiz, model seçimimizin ampirik verilere dayandığını göstermeyi hedefler. Karşılaştırmanın adil bir zeminde yürütülmesi için, tüm Transformer tabanlı modeller (DistilBERT, ELECTRA, ConvBERT) aynı koşullar altında değerlendirilmiştir. Bu koşullar şunlardır:

1. Veri Seti: Tüm modeller, yaklaşık 700.000 yorumdan oluşan, temel düzeyde temizlenmiş aynı genişletilmiş veri seti üzerinde eğitilmiş ve test edilmiştir.

2. Değerlendirme Metrikleri**:** Modellerin performansı, genel isabeti ölçen Doğruluk (Accuracy) ve sınıf dengesizliğinin olduğu durumlarda modelin tüm sınıflardaki başarısını daha adil yansıtan Makro F1-Skoru metrikleri kullanılarak değerlendirilmiştir.

Geleneksel ML modelleri ise daha küçük bir veri setinde (10.000 örnek) test edilmiş olup, LLM'lerin sağladığı performans sıçraması için bir başlangıç referansı (baseline) olarak sunulmuştur.

Farklı yaklaşımların ve mimarilerin karşılaştırmalı performans sonuçları Çizelge 3'de özetlenmiştir.

**Çizelge 3.** Karşılaştırmalı Model Performansları

| Model Mimarisi | Yaklaşık Veri Seti Boyutu | En İyi Doğruluk | Makro F1 Skoru | Temel Gözlem |
|---|---|---|---|---|
| **DistillBERT(V5)** | 700.000 | 0,718 | 0,49 | Temel performans referansı |
| **ELECTRA(V7)** | 700.000 | 0,758 | 0,55 | Genişletilmiş veri ile performans arttı, ara sınıflar zayıf |
| **ConvBERT(V8)** | 700.000 | 0,763 | 0,585 | Hem doğruluk hem de F1-Skoru'nda en iyi temel performansı sergiledi |

DistilBERT (V5), 0,718 doğruluk ve 0,49 Makro F1-Skoru ile giriş seviyesi bir performans sergilemiştir. Düşük F1-Skoru, modelin sınıf dengesizliğinden olumsuz etkilendiğinin açık bir göstergesidir.

ELECTRA (V7), hem doğruluk oranını 0,754'e hem de Makro F1-Skoru'nu 0,55'e çıkararak DistilBERT'ten daha güçlü bir mimari olduğunu kanıtlamıştır. Bu artış, modelin ön-eğitim görevinin bu sınıflandırma problemine daha uygun olduğunu göstermektedir.

ConvBERT (V8), hem 0,763 doğruluk oranı hem de 0,585 Makro F1-Skoru ile en iyi temel performansı sergilemiştir. Özellikle F1-Skoru'ndaki artış, ConvBERT'in azınlık sınıflarını ayırt etmede diğer mimarilerden daha başarılı olduğuna işaret etmektedir. Bu üstünlüğün, ConvBERT'in konuşma dili üzerine eğitilmiş olmasından ve bu sayede müşteri yorumlarının informal dil yapısına daha iyi adapte olmasından kaynaklandığı düşünülmektedir.

Bu ampirik karşılaştırma, ConvBERT mimarisinin Türkçe müşteri geri bildirimlerini yıldız derecelendirmesine göre sınıflandırma görevinde en yüksek potansiyele sahip olduğunu göstermiştir. Bu veri odaklı bulgu, çalışmanın sonraki aşamalarında gerçekleştirilen tüm veri zenginleştirme (sentetik veri ekleme) ve eğitim optimizasyonu çabalarının neden bu mimari üzerine yoğunlaştığını gerekçelendirmektedir.

Çalışmanın kritik bulgularından biri, stratejik olarak üretilen sentetik verinin model performansı üzerindeki etkisini nicel olarak ölçmektir. Bu etkiyi izole etmek ve doğrulamak amacıyla bir ablasyon çalışması (ablation study) yürütülmüştür. Bu çalışmada, diğer tüm değişkenler (model mimarisi, hiperparametreler, veri temizleme protokolü) sabit tutularak iki temel model karşılaştırılmıştır:

1. Sentetik Veri Olmadan Eğitilen Model (YıldızSezar V15): Sadece derinlemesine temizlenmiş gerçek müşteri yorumlarından oluşan veri setiyle eğitilmiştir.

2. Sentetik Veri ile Zenginleştirilmiş Model (YıldızSezar V22): Aynı temizlenmiş gerçek verilere ek olarak, sınıf dengesizliğini gidermek amacıyla üretilen ve titizlikle temizlenen sentetik verilerle zenginleştirilmiş veri seti üzerinde eğitilmiştir.

Her iki modelin performansı da içerisinde hiç sentetik veri bulunmayan, tamamen gerçek yorumlardan oluşan aynı test seti üzerinde ve aynı şartlarda değerlendirilmiştir. Bu yaklaşım, modellerin gerçek dünya verilerine genelleme yapma yeteneğini adil bir şekilde ölçmeyi sağlamaktadır. Elde edilen karşılaştırmalı sonuçlar Çizelge 4'de özetlenmiştir.

Çizelgedeki sonuçlar, titizlikle temizlenmiş sentetik verinin entegrasyonunun model performansı üzerinde dönüştürücü bir etkiye sahip olduğunu açıkça göstermektedir. Sadece gerçek veri ile eğitilen model 0,7662 doğrulukta kalırken, sentetik veri takviyesi bu oranı yaklaşık 0,90 seviyesine taşımıştır.

**Çizelge 4.** Sentetik Veri Kullanarak ve Sentetik Veri Kullanmayarak Elde Edilen Başarı Çizelgesi

| Model Versiyonu | Veri Seti Kompozisyonu | Doğruluk | Makro F1 Skoru |
|---|---|---|---|
| **YıldızSezar V15** | Sadece Temizlenmiş Gerçek Veri | 0,7662 | 0,537 |
| **YıldızSezar V22** | Temizlenmiş Gerçek Veri + Temizlenmiş Sentetik Veri | 0,8996 | 0,799 |

En çarpıcı gelişme ise sınıf dengesizliğine karşı daha hassas olan Makro F1 skorunda yaşanmıştır. Bu metrikteki 0,537'den 0,799'a olan sıçrama, modelin başarısının sadece baskın sınıflardaki (1 ve 5 yıldız) artıştan kaynaklanmadığını; aksine, sentetik veri sayesinde sınıf dengesizliğinin aşıldığını ve daha önce düşük performansa sahip olan ara sınıfları (2, 3 ve 4 yıldız) çok daha başarılı bir şekilde öğrendiğini kesin olarak kanıtlamaktadır. Bu bulgu, karmaşık doğal dil işleme görevlerinde sınıf dengesizliğiyle mücadele etmek için yüksek kaliteli sentetik veri kullanımının ne denli etkili bir strateji olduğunu vurgulamaktadır.

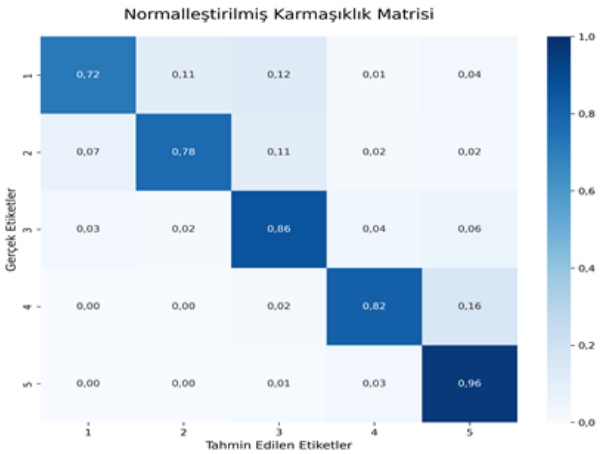
Çalışmanın nihai ürünü olan YıldızSezar V22 modeli, ConvBERT mimarisi kullanılarak, temizlenmiş ve sentetik veri ile zenginleştirilmiş veri seti üzerinde 50 epoch boyunca eğitilmiş ve en iyi performansı (en düşük doğrulama kaybı) gösteren checkpoint'i seçilmiştir. Bu modelin test seti üzerindeki detaylı sınıflandırma performansı Çizelge 5'de sunulmaktadır.

Çizelge 5'te detaylandırıldığı üzere, nihai model test seti üzerinde 0,8996 gibi yüksek bir genel doğruluğa ulaşmıştır. Modelin sınıf dengesizliğine karşı direncini ölçen en kritik metrik olan makro ortalama F1-skoru ise 0,799 (0,95 Güven Aralığı: [0,792, 0,806]) olarak hesaplanmıştır. Bu dar güven aralığı, elde edilen performansın istatistiksel olarak tutarlı olduğunu ve tek bir test bölünmesinin şans eseri bir sonucu olmadığını göstermektedir. Özellikle başlangıçtaki en büyük zorluk olan 2, 3 ve 4 yıldızlı ara sınıflarda F1-skorlarının sırasıyla 0,70, 0,79 ve 0,79 gibi yüksek değerlere ulaşması, çalışmanın temel hedefine ulaştığının en önemli kanıtıdır.

**Çizelge 5.** Nihai Model YıldızSezar Test Seti Sınıflandırma Raporu

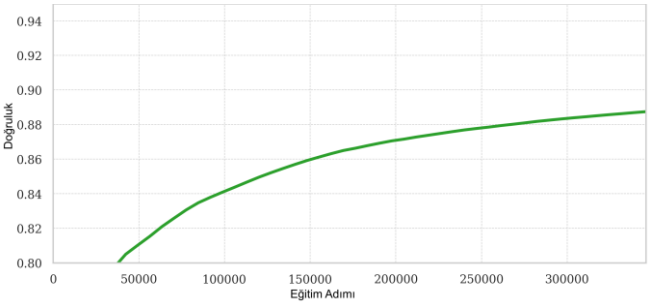
| Sınıf | Doğruluk (Precision) | Duyarlılık (Recall) | F1 Skoru | Destek |
|---|---|---|---|---|
| **1 Yıldız** | 0,77 | 0,75 | 0,76 | 935 |
| **2 Yıldız** | 0,72 | 0,68 | 0,70 | 1142 |
| **3 Yıldız** | 0,81 | 0,78 | 0,79 | 1178 |
| **4 Yıldız** | 0,82 | 0,77 | 0,79 | 1262 |
| **5 Yıldız** | 0,96 | 0,95 | 0,95 | 4530 |
| **Accuracy** | | | 0,9 | 10547 |
| **Macro Avg** | 0,81 | 0,79 | 0,8 | 10547 |
| **Weighted Avg** | 0,9 | 0,9 | 0,9 | 10547 |

Modelin sınıflandırma davranışı Şekil 2'deki karmaşıklık matrisinde görselleştirilmiştir. Matris, köşegen üzerindeki yoğunluktan da anlaşılacağı gibi, modelin çoğu örneği doğru sınıflandırdığını, ancak özellikle komşu sınıflar arasında (örneğin, 4 yıldızlı bazı yorumların 5 yıldız olarak tahmin edilmesi) hala az da olsa bir karışıklık olduğunu göstermektedir. Fakat yinede model büyük bir başarı elde etmiştir.

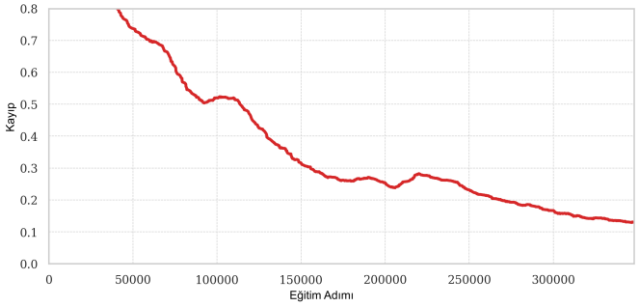



**Şekil 2.** Nihai Model YıldızSezar V22'ın Karmaşıklık Matrisi

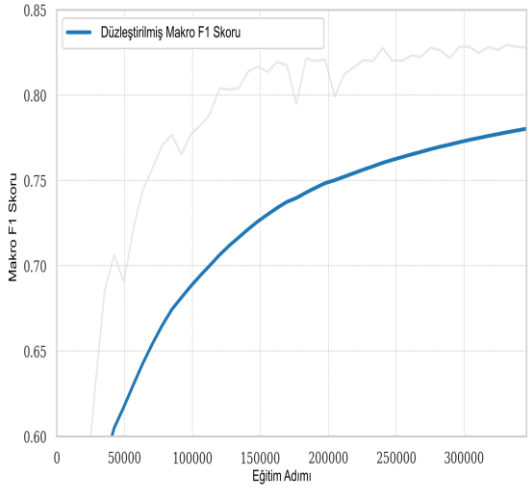
Modelin eğitim sürecinde aşırı öğrenmenin (overfitting) kontrol altına alındığı, Şekil 3, Şekil 4 ve Şekil 5'teki grafiklerden açıkça görülmektedir. Eğitim kaybı (train loss) düşmeye devam ederken, doğrulama kaybının (eval loss) belirli bir epoch'tan sonra artmaya başlaması, en iyi modelin bu dönüm noktasından önce seçilmesini sağlamıştır.




**Şekil 3.** Nihai Modelin Doğruluk Kayıp Grafiği




**Şekil 4.** Nihai Modelin Eğitim Kayıp Grafiği

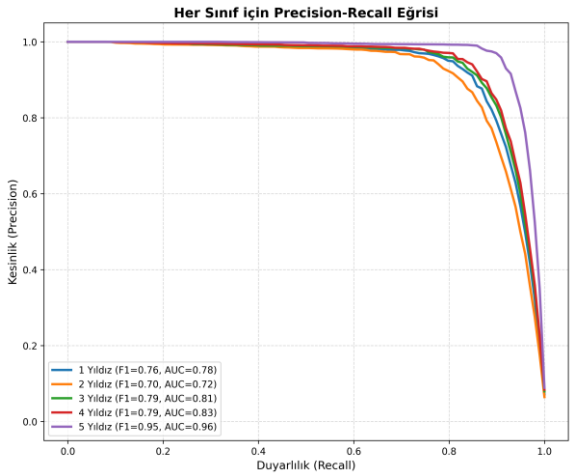



**Şekil 5.** Nihai Modelin Doğrulama Doğruluk/F1 Grafiği

Modelin sınıflandırma performansını tek bir metrikten öte, farklı güven eşiklerindeki davranışını da incelemek amacıyla, Şekil 5'te Precision-Recall (PR) eğrilerinin analizi sunulmuştur. Bu analiz, özellikle sınıf dengesizliğinin model üzerindeki etkisini değerlendirmek için kritik öneme sahiptir. Şekil 6'da görüldüğü üzere, modelin en baskın sınıf olan 5 yıldız için elde ettiği 0,96'lık AUC değeri, bu kategorideki üstün performansını teyit etmektedir. Çalışmanın odaklandığı daha zorlu ara sınıflar olan 3 ve 4 yıldız kategorilerinde sırasıyla 0,81 ve 0,82 gibi yüksek AUC değerlerine ulaşılması, modelin bu nüanslı ve genellikle birbiriyle karıştırılan yorumları da başarıyla ayırt edebildiğini göstermektedir.

En dikkat çekici bulgu ise, en zayıf F1 skoruna sahip azınlık sınıfı olan 2 yıldız kategorisinin dahi 0,72 gibi rekabetçi bir AUC değerine ulaşmasıdır.

Modelin sınıflandırma doğruluğuna ek olarak, tahminlerinin güvenilirliği de pratik uygulamalar için kritik bir önem taşımaktadır. Bu doğrultuda, modelin tahmin olasılıklarının ne kadar iyi "kalibre" edildiği, yani güven skorlarının gerçek doğruluk oranlarıyla ne kadar tutarlı olduğu incelenmiştir.



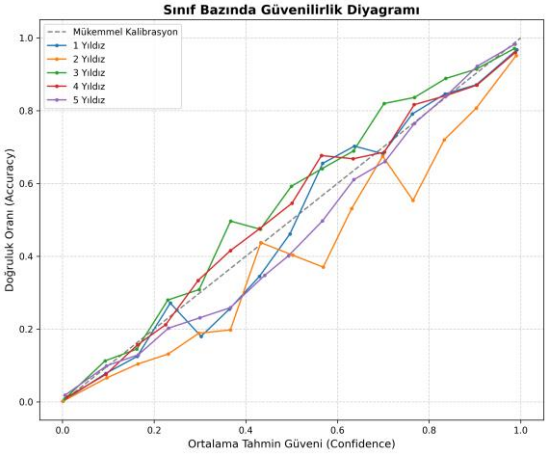

**Şekil 6.** Sınıf Bazında Precision-Recall (PR) Eğrileri ve Eğri Altında Kalan Alan (AUC) Değerleri

Analiz sonucunda, modelin tüm sınıfları kapsayan genel Brier Skoru 0,0233 gibi oldukça düşük bir değerde hesaplanmıştır (0'a yakın olması daha iyi kalibrasyonu gösterir). Benzer şekilde, modelin güven skorları ile gerçek doğruluk oranları arasındaki ortalama farkı ölçen Beklenen Kalibrasyon Hatası (ECE) ise sadece 0,0094 olarak bulunmuştur. Bu düşük hata oranları, modelin aşırı veya eksik güven sergilemeden, dengeli tahminler ürettiğine işaret etmektedir.

Bu bulgular, Şekil 7'de sunulan sınıf bazındaki güvenilirlik diyagramı ile görsel olarak da desteklenmektedir. Grafikte, her bir yıldız sınıfı için çizilen eğrilerin, "Mükemmel Kalibrasyon" olarak belirtilen referans çizgisine çok yakın seyrettiği görülmektedir. Bu durum, modelin örneğin "0,90 güvenle 5 yıldız" dediği tahminlerin, gerçekte de yaklaşık 0,90 oranında doğru olduğunu göstermektedir. Bu yüksek kalibrasyon seviyesi, YıldızSezar V22 modelinin ürettiği sonuçların sadece doğru değil, aynı zamanda güvenilir olduğunu da kanıtlamakta ve modelin kritik iş kararlarında bir destek aracı olarak kullanılabilirliğini artırmaktadır.

Bu çalışmada geliştirilen YıldızSezar modelinin akademik başarısının yanı sıra, gerçek dünya senaryolarındaki maliyeti ve uygulanabilirliği de analiz edilmiştir. Bu analiz, modelin üretim ortamlarına entegrasyonu için kritik öneme sahip olan eğitim maliyeti, depolama gereksinimi, çıkarım hızı ve kaynak tüketimi metriklerini kapsamaktadır.

YıldızSezar V22 modelinin eğitimi, büyük veri seti ve Transformer mimarisinin doğası gereği hesaplama açısından yoğun bir süreç olmuştur. Eğitim, tek bir NVIDIA GeForce RTX 3090 GPU üzerinde yaklaşık 78 saat sürmüştür. Eğitim tamamlandığında, nihai model ağırlıkları safetensors formatında disk üzerinde yalnızca 410 MB yer kaplamaktadır.




**Şekil 7**. Sınıf Bazında Güvenilirlik Diyagramı

Modelin pratik değeri en çok çıkarım performansıyla ölçülür. Bu doğrultuda, bir NVIDIA GeForce RTX 3070 GPU üzerinde yapılan testlerde aşağıdaki sonuçlar elde edilmiştir:

1.Gecikme Süresi (Latency): Tekil yorumların analizi için yapılan testlerde, model ortalama 14.42 milisaniye gibi oldukça düşük bir gecikme süresi sergilemiştir.

2.Verim (Throughput): Büyük hacimli verilerin toplu işlenmesi senaryoları için modelin verimi ölçülmüştür. 32'lik yığınlar (batches) halinde çalıştırıldığında, model saniyede yaklaşık 182 yorum işleme kapasitesine (QPS - Queries Per Second) ulaşmıştır.

3.Kaynak Tüketimi: Modelin kaynak tüketimindeki verimliliği dikkat çekicidir. Yoğun toplu işleme sırasında dahi modelin GPU belleği (VRAM) kullanımı yaklaşık 1.1 GB (1065 MiB) seviyesinde kalmıştır.

### 3.1 Kısıtlılıklar ve Gelecek Çalışmalar

Bu çalışmanın, elde edilen başarılı sonuçlara rağmen bazı kısıtlılıkları bulunmaktadır ve bu kısıtlılıklar gelecek araştırmalar için önemli fırsatlar sunmaktadır.

İlk olarak, modelin genelleme performansı, sabit bir eğitim-doğrulama-test veri seti bölünmesi kullanılarak değerlendirilmiştir. Bu yaklaşım yaygın olmakla birlikte, sonuçların veri setinin belirli bir bölünmesine duyarlı olma riskini taşır. Gelecekteki çalışmalarda, model performansının daha gürbüz ve istatistiksel olarak daha güçlü bir tahminini elde etmek için k-katlamalı çapraz doğrulama (k-fold cross-validation) gibi yöntemlerin kullanılması, bulguların geçerliliğini daha da pekiştirecektir.

İkinci olarak, bu çalışmada kullanılan veri seti ağırlıklı olarak e-ticaret platformlarındaki ürün yorumlarından oluşmaktadır. Modelin bu alandaki (in-domain) başarısı kanıtlanmış olsa da, kendine özgü dil ve jargon barındıran turizm (otel yorumları) veya gastronomi (restoran yorumları) gibi farklı alanlardaki (out-of-domain) performansı bu çalışma kapsamında test edilmemiştir. Modelin bu farklı alanlara adaptasyonu ve genelleme kabiliyetinin ölçülmesi, önemli bir gelecek araştırma yönünü temsil etmektedir.

## 4. SONUÇ

Bu çalışma, Türkçe müşteri yorumlarını 1'den 5'e kadar olan yıldız puanlarına göre sınıflandırma gibi zorlu bir problemi çözmek üzere geliştirilen "YıldızSezar" modelini ve bu modelin arkasındaki sistematik metodolojiyi sunmuştur. LLaMA tabanlı bir model ile üretilen yüksek kaliteli sentetik verilerle sınıf dengesizliği sorununun üstesinden gelinmiş; nihai model 0,90 doğruluk ve azınlık sınıflardaki başarımı yansıtan 0,799 makro F1-skoru ile alanında dikkate değer bir başarıya ulaşmıştır.

Bu çalışmanın özgün niteliği, üç temel unsurun bir araya getirilmesinden doğmaktadır: (1) Türkçe gibi sondan eklemeli, morfolojik olarak zengin ve bağlamsal inceliklerin yoğun olduğu bir dile odaklanması, (2) ikili (olumlu/olumsuz) sınıflandırmanın ötesine geçerek işletmeler için kritik değere sahip olan 1-5 arası granüler yıldız skalasını hedef alması ve (3) bu zorlukların üstesinden gelmek için büyük dil modelleriyle (LLM) sentetik veri üretme stratejisini başarılı bir şekilde uygulaması. Yapılan kapsamlı literatür taraması, bu üç unsuru birleştirerek özellikle ara sınıflarda (2, 3, 4 yıldız) yüksek başarım elde eden böylesine sistematik bir çalışmanın Türkçe için daha önce yapılmadığını

ortaya koymaktadır. Bu yönüyle çalışma, literatürde önemli bir boşluğu doldurmaktadır.
Bununla birlikte, çalışmanın bazı sınırlılıkları mevcuttur. Model, ağırlıklı olarak e-ticaret platformlarındaki ürün yorumları üzerinde eğitilmiş ve test edilmiştir. Modelin başarısı bu alanda kanıtlanmış olsa da, kendine özgü bir dil ve jargona sahip olabilen turizm (otel yorumları), gastronomi (restoran yorumları) veya mobil uygulamalar (dijital dağıtım platformu yorumları) gibi farklı hizmet sektörlerindeki performansı henüz test edilmemiştir. Ayrıca, sentetik veri üretiminde kullanılan LLM'in kapasitesi, üretilen metinlerin çeşitliliğini ve yaratıcılığını sınırlayan bir faktör olabilir.

Bu çalışma, gelecek araştırmalar için verimli bir zemin hazırlamaktadır. Geliştirilen modelin metindeki anlamsal tutarlılığı ve duygusal tonu derinlemesine anlama kabiliyeti, farklı görevler için bir temel oluşturabilir. Örneğin, model, anlamsız veya tekrar eden ifadeler içeren düşük kaliteli veya spam niteliğindeki yorumları tespit etmek için uyarlanabilir. Gelecek çalışmalarda, daha gelişmiş generative modellerle sentetik veri setinin daha da zenginleştirilmesi, verisetinin genişletilmesi ve daha gelişmiş BERT tabanlı modeller kullanılması performansın daha ileri bir seviyeye taşınmasına olanak tanıyacaktır.. Sonuç olarak, "YıldızSezar", Türkçe doğal dil işleme alanında pratik ve yüksek performanslı bir çözüm sunmakla kalmayıp, benzer zorluklara sahip diğer dillerdeki çalışmalar için de bir model teşkil etme potansiyeli taşımaktadır.

## Ek A: Şeffaflık, Tekrarlanabilirlik ve Etik Hususlar

Bu bölümde, çalışmanın şeffaflığını ve bilimsel geçerliliğini desteklemek amacıyla kullanılan veri kaynakları, gizlilik önlemleri, tekrarlanabilirlik adımları ve modele ilişkin özet bilgiler sunulmaktadır.

### 1. Veri Kaynakları ve Lisanslama

Çalışmada kullanılan veri setleri, akademik ve araştırmacı kullanımına açık olan, kamuya mal olmuş kaynaklardan derlenmiştir. Bu kaynaklar arasında Kaggle platformunda yer alan "Turkish Product Reviews" (Cebeci, 2018) ve "Turkish E-Commerce Review Dataset" (Keskin, 2020) gibi veri setleri bulunmaktadır. Tüm veriler, ilgili platformların araştırma amaçlı kullanıma izin veren lisans koşullarına uygun olarak kullanılmıştır.

### 2. Gizlilik ve KVKK Uyumu

Veri setleri, kullanıcı gizliliğini korumak amacıyla anonimleştirilmiş verilerden oluşmaktadır. Çalışmamızda uygulanan veri temizleme protokolü, Kişisel Verilerin Korunması Kanunu (KVKK) ilkeleriyle uyumlu olarak, metinler içerisinde bulunabilecek e-posta adresi, URL veya kullanıcı adı gibi kişisel olarak tanımlanabilir bilgileri (PII) otomatik olarak temizlemek üzere tasarlanmıştır. Model, yalnızca anonim yorum metinleri ve yıldız puanları üzerinde eğitilmiştir.

### 3. Tekrarlanabilirlik

Çalışmadaki sonuçların tekrarlanabilirliğini sağlamak amacıyla aşağıdaki adımlar izlenmiştir:

- **Tohum Değeri (Seed Value):** Veri bölme, model ağırlıklarının başlangıç durumu ve eğitim sürecindeki diğer tüm rastgelelik içeren adımlarda 42 olarak sabit bir tohum değeri kullanılmıştır.
- **Hiperparametreler:** Nihai modelin eğitiminde kullanılan temel hiperparametreler şunlardır:
- **Model Mimarisi:** dbmdz/convbert-base-turkish-mc4-uncased
- **Öğrenme Oranı:** 3e-5
- **Yığın Boyutu (Batch Size):** 32

**Epoch Sayısı:** 50 (Erken durdurma ile)

### 4. Model Kartı (Özet)

- **Model Adı:** YıldızSezar V22
- **Mimari:** ConvBERT (Convolutional BERT)
- **Geliştirilme Amacı:** Türkçe e-ticaret müşteri yorumlarını 1'den 5'e kadar olan yıldız puanlarına göre çok sınıflı olarak sınıflandırmak.
- **Kullanım Alanları:** Müşteri memnuniyeti analizi, ürün/hizmet kalitesi takibi, otomatik geri bildirim etiketleme.
- **Kısıtlılıklar:** Model, ağırlıklı olarak e-ticaret ürün yorumları üzerinde eğitilmiştir. Turizm, gastronomi veya sağlık gibi farklı jargona sahip alanlarda performansı düşebilir ve bu alanlarda kullanılmadan önce ek bir ince ayar (fine-tuning) gerektirebilir.

**Etik Değerlendirme:** Model, anonim verilerle eğitilmiş olup kişisel gizliliği ihlal etmez.

## KAYNAKLAR

1. Yücel, N., and Cömert, Ö., "Müşteri Duyarlılığını Keşfetmek İçin Yapay Zeka Destekli Analiz ile Çevrimiçi Ürün İncelemelerinden Anlamlı Bilgiler Elde Etme", Fırat Üniversitesi Mühendislik Bilimleri Dergisi, Cilt 35, Sayı 2, Sayfa 679-690, 2023.

2. Aytan, B., and Sakar, C. O., "Comparison of transformer-based models trained in Turkish and different languages on Turkish natural language processing problems", 30th Signal Processing and Communications Applications Conference (SIU) Sayfa 1-4, IEEE, 2022.

3. Aslan, S., "Doğal Dil İşleme Teknikleri Kullanarak E-Ticaret Kullanıcı İncelemelerinde Özellik Tabanlı Duygu Analizi", Fırat Üniversitesi Mühendislik Bilimleri Dergisi, Cilt 35, Sayı 2, Sayfa 875-882, 2023.

4. Tuna, M. F., Kaynar, O., and Akdoğan, M. Ş., "Otellere ilişkin çevrimiçi geribildirimlerin makine öğrenmesi yöntemleriyle duygu analizi", İşletme Araştırmaları Dergisi, Cilt 13, Sayı 3, Sayfa 2232-2241, 2021.

5. Alpkoçak, A., Tocoglu, M. A., Çelikten, A., ve Aygün, İ., "Türkçe metinlerde duygu analizi için farklı makine öğrenmesi yöntemlerinin karşılaştırılması", Dokuz Eylül Üniversitesi Mühendislik Fakültesi Fen ve Mühendislik Dergisi, Cilt 21, Sayı 63, Sayfa 719-725, 2019.

6. Güven, Z. A., "The effect of bert, electra and albert language models on sentiment analysis for turkish product reviews", 6th international conference on computer science and engineering (UBMK), Pages 629-632, IEEE, 2021.

7. Acikalin, U. U., Bardak, B., and Kutlu, M., "Turkish sentiment analysis using bert", 28th Signal Processing and Communications Applications Conference (SIU), Pages 1-4, IEEE, 2020.

8. Wu, X., Linghu, Y., Wang, T., and Fan, Y., "Sentiment analysis of weak-ruletext based on the combination of sentiment lexicon and neural network", IEEE 6th International Conference on Cloud Computing and Big Data Analytics (ICCCBDA), Pages 205-209, IEEE, 2021.

9. Siddiqua, U. A., Ahsan, T., and Chy, A. N., "Combining a rule-based classifier with weakly supervised learning for twitter sentiment analysis", International conference on innovations in science, engineering and technology (ICISET), Pages 1-4, IEEE, 2016.

10. Siğirci, İ. O., Özgür, H., Oluk, A., Uz, H., Çetiner, E., Oktay, H. U., and Erdemir, K., "Sentiment analysis of Turkish reviews on google play store", 5th international conference on computer science and engineering (UBMK), Pages 314-315, IEEE, 2020.

11. Atılgan, K. Ö., ve Yoğurtcu, H., "Kargo firması müşterilerinin Twitter gönderilerinin duygu analizi", Çağ Üniversitesi Sosyal Bilimler Dergisi, Cilt 18, Sayı 1, Sayfa 31-39, 2021.

12. Akiba, T., Sano, S., Yanase, T., Ohta, T., and Koyama, M., "Optuna: A next-generation hyperparameter optimization framework", 25th ACM SIGKDD international conference on knowledge discovery & data mining, Pages 2623-2631, 2019.

13. Sathyanarayanan, S., and Tantri, B. R., "Confusion matrix-based performance evaluation metrics", African Journal of Biomedical Research, Vol. 21, Issue 1, Pages, 4023-4031, 2024.

14. Karamollaoğlu, H., Doğru, İ. A., Dörterler, M., Utku, A., & Yıldız, O., "Sentiment analysis on Turkish social media shares through lexicon based approach", 3rd International Conference on Computer Science and Engineering (UBMK), Pages 45-49, IEEE, 2018.

15. Demir, E., and Bilgin, M., "Sentiment Analysis from Turkish News Texts with BERT-Based Language Models and Machine Learning Algorithms", 8th International Conference on Computer Science and Engineering (UBMK), Pages 101-104, IEEE, 2023.

16. Le, T., "An attention-based deep learning method for text sentiment analysis", International Conference on Computational Science and Computational Intelligence (CSCI), Pages 282-286, IEEE, 2020.

17. Parkavi, A., Shetty, T. B., Shibu, S. G., Ismail, R., and Muthirakalayil, R. A., "Customer Feedback Analysis Based on Emotion Detection Using Machine Learning Techniques with Privacy Preservation", International Conference on Circuit Power and Computing Technologies (ICCPCT), Pages 1617-1624, IEEE, 2023.

18. Gümüş, Ö., "Duygu analizi yönteminin pazarlama alanında kullanımına ilişkin çalışmaların değerlendirilmesi", Yüksek Lisans Tezi, Tekirdağ Namık Kemal Üniversitesi, 2022.

19. Ban, M., Zong, L., Zhou, J., and Xiao, Z., "Multimodal aspect-level sentiment analysis based on deep neural networks", 8th International

Symposium on System Security, Safety, and Reliability (ISSSR), Pages 184-188, IEEE, 2022.

20. Niu, L., Zheng, Q., and Zhang, L.," Enhance gated graph neural network with syntactic for sentiment analysis", International Conference on Advances in Electrical Engineering and Computer Applications (AEECA), Pages 1055-1060, IEEE, 2021.

21. Das, N., Gupta, S., Das, S., Yadav, S., Subramanian, T., and Sarkar, N., "A comparative study of sentiment analysis tools", International Conference on Innovative Computing, Intelligent Communication and Smart Electrical Systems (ICSES) Pages 1-7, IEEE, 2021.

22. Cebeci, I., "Turkish Product Reviews", Kaggle. https://www.kaggle.com/datasets/cebeci/turkishreviews, 2018.

23. Keskin, M., "Turkish Product Review Dataset", Kaggle. https://www.kaggle.com/datasets/mustfkeskin/turkish-product-review-dataset, 2020.

24. Schweter, S., "BERTurk: A Turkish BERT model", 3rd Workshop on E-commerce and NLP (ECNLP 3), Pages 68-73, 2020.