Müşteri Geri Bildirimlerindeki Sübjektifliğin Azaltılmasına Yönelik Yüksek Performanslı Otomatik Çağrı Atama Modeli: Mobilya Sektöründen Kanıtlar

Araştırma Makalesi/Research Article

Sevgi PEHLİVAN[1,2], Hülya TORUN[3]*

[1]Fen Bilimleri Enstitüsü, Endüstri Mühendisliği Anabilim Dalı, Erciyes Üniversitesi, Kayseri, Türkiye
[2]Bellona Mobilya A.Ş., OSB Mahallesi, Kayseri, Türkiye
[3*] Endüstri Mühendisliği Bölümü, Mühendislik Fakültesi, Erciyes Üniversitesi, Kayseri, Türkiye

sevgisevgi137@gmail.com, hbehret@erciyes.edu.tr
(Geliş/Received:15.09.2025; Kabul/Accepted:18.12.2025)
DOI: 10.17671/gazibtd.1782837

Özet— Bu çalışma, mobilya sektöründe faaliyet gösteren çağrı merkezleri operasyonlarının verimliliğini ve objektifliğini artırmayı hedeflemektedir. Araştırmada, Türkiye'deki üç büyük mobilya firmasına ait müşteri geri bildirimleri, Metin Madenciliği ve Makine Öğrenmesi teknikleri kullanılarak kapsamlı analiz edilmiştir. Manuel atama süreçlerindeki sübjektifliği gidermek amacıyla, müşteri geri bildirimleri öncelikle Latent Dirichlet Allocation (LDA) yöntemiyle analiz edilmiş; elde edilen konu dağılımları, Terim Frekansı-Ters Doküman Frekansı (TF-IDF) özellikleri ile birleştirilerek sınıflandırma için geliştirilmiş bir veri seti oluşturulmuştur (Model-2). Uygulanan sınıflandırma algoritmaları arasında, Destek Vektör Makinesi (SVM), Model-2'de %89,1 doğruluk ile yüksek başarı sağlamıştır. Çalışmanın güncel literatür ile bağını kurmak ve performansın üst sınırını belirlemek amacıyla, Transformer tabanlı Sentence-BERT (S-BERT) modeli ile SVM birleştirilerek bir üçüncü model (Model-3) geliştirilmiştir. Model-3, %90,8'lik doğruluk oranı ile en yüksek performansı göstermiş ve çağrı merkezleri operasyonlarında en yüksek verimlilik potansiyelini sunmuştur. Bu sonuçlar, LDA temelli yöntemin klasik yöntemler arasında sübjektifliği gidermedeki başarısını kanıtlarken, bağlamsal gömülmelerin üstünlüğünü de ortaya koymaktadır. Çalışmanın temel katkısı, müşteri geri bildirimlerinde gizli kalan ilişkileri sistematik biçimde ortaya koyarak, çağrı temsilcilerinin sübjektif değerlendirmelerinden kaynaklanabilecek hataları en aza indiren ve yüksek doğruluk oranıyla çalışan otomatik bir çağrı atama mekanizması sunmasıdır.

Anahtar Kelimeler— metin madenciliği, makine öğrenmesi, konu modelleme, çağrı atama optimizasyonu, mobilya sektörü, transformer tabanlı gömülmeler.

High-Performance Automated Call Assignment Model to Reduce Subjectivity in Customer Feedback: Evidence from the Furniture Sector

Abstract— This study aims to enhance the efficiency and objectivity of call center operations in the furniture industry. Customer feedback collected from three major furniture companies in Türkiye was comprehensively analyzed using Text Mining and Machine Learning techniques. To eliminate subjectivity in manual call-type assignment, the feedback was first examined using the Latent Dirichlet Allocation (LDA) method, and the resulting topic distributions were combined with Term Frequency–Inverse Document Frequency (TF-IDF) features to construct an improved dataset for classification (Model-2). Among the utilized classification algorithms, Support Vector Machine (SVM) algorithm achieved strong performance on this model with an accuracy of 89,1%. To align the study with contemporary literature and determine the upper performance boundary, a third model (Model-3) integrating a Transformer-based Sentence-BERT (S-BERT) architecture with SVM was evaluated. Model-3 demonstrated the highest performance, reaching an accuracy of 90,8%, and thus presented the greatest potential for operational efficiency in call center environments. These results confirm the effectiveness of LDA-based method in reducing subjectivity within classical approaches, while also highlighting the superiority of contextual embeddings. The primary contribution of this study is the development of a high-accuracy, automated call assignment mechanism that systematically uncovers latent relationships within customer feedback, thereby minimizing errors arising from subjective assessments made by call center agents.

Keywords— text mining, machine learning, topic modeling, call assignment optimization, furniture industry, transformer based embeddings.

1. GİRİŞ (INTRODUCTION)

Artan küresel rekabet koşulları ve hızla değişen pazar dinamikleri, çağdaş işletmeleri temel stratejilerini müşteri odaklılık etrafında yeniden yapılandırmaya zorlamaktadır [1, 2]. Müşteri memnuniyetinin sürekli ve sistematik takibi ile müşterilerin beklentilerine hızlı ve etkili çözümler sunulması, işletmelerin pazardaki rekabet avantajını uzun vadede sürdürmeleri için belirleyici bir faktördür [3]. Özellikle, Dijital Dönüşüm ve Endüstri 4.0 paradigmalarının etkisiyle, işletmeler müşteri deneyimini iyileştirmek amacıyla geri bildirimleri proaktif ve sistematik biçimde analiz etmeye odaklanmıştır [4, 5]. Bu bağlamda, müşteri verilerinin etkin ve verimli yönetimi için veri odaklı yaklaşımlar ve ileri analitik stratejileri, operasyonel ve stratejik kararlar için temel bir ihtiyaç haline gelmiştir [6].

Müşteri geri bildirimleri, çağrı merkezleri, mobil uygulamalar, sosyal medya platformları), çevrimiçi şikâyet platformları ve e-posta gibi çok çeşitli dijital kanallar aracılığıyla toplanmaktadır [7]. Bu verilerin büyük çoğunluğu, yapısal olmayan doğal dil (metin) biçimindedir ve ürün/hizmet eksiklikleri, tedarik zinciri hataları, rakiplerle karşılaştırmalar, operasyonel riskler ve iyileştirme fırsatları gibi işletmeler için kritik öneme sahip değerli bilgiler içermektedir [8, 9]. Özellikle sosyal medyanın ve çevrimiçi platformların yaygınlaşması, bireylerin deneyimlerini anlık ve geniş kitlelerle paylaşmasını sağlayarak, işletmelerin bu kanallardaki geri bildirimleri analiz etme zorunluluğunu artırmıştır [10].

Giderek artan bu geniş ve karmaşık veri yığınını anlamlandırmak ve gizli ilişkileri keşfetmek için Büyük Veri Analitiği ve Metin Madenciliği gibi ileri analitik tekniklerden yararlanılmaktadır [11, 12]. Bu yöntemler, müşteri ihtiyaç ve beklentilerinin nicel ve objektif bir şekilde anlaşılmasını sağlayarak, geri bildirimlerin stratejik karar alma süreçlerine doğrudan katkı sağlamasına olanak tanır.

Bu analitik gereklilik, özellikle çağrı merkezi operasyonları için büyük önem taşımaktadır. Müşteri taleplerinin doğru ve hızlı bir şekilde ilgili birimlere/uzmanlara yönlendirilmesi (çağrı atama süreci), hem müşteri memnuniyetinin korunması hem de operasyonel maliyet ve verimlilik açısından kritiktir [13]. Ancak mevcut çağrı atama süreçlerinin birçoğu, sübjektif değerlendirmelere dayanmakta ve bu durum bilgi kayıplarına, yanlış yönlendirmelere ve dolayısıyla müşteri memnuniyetsizliğine yol açmaktadır [14].

Bu bağlamda, bu çalışma, mobilya sektöründe müşteri hizmetleri süreçlerinin etkinliğini artırmayı hedefleyen, metin madenciliği ve makine öğrenmesi tekniklerine dayalı objektif bir çağrı atama modelinin geliştirilmesine odaklanmaktadır. Çalışma, Türkiye'de üç büyük mobilya firmasına ait çevrimiçi şikâyet platformlarından (Şikayetvar.com) ve iç kaynaklardan (e-posta/SMS) elde edilen 9.930 müşteri geri bildirimini kapsamlı bir şekilde analiz etmektedir. Mobilya sektörü, ürün çeşitliliği, montaj karmaşıklığı ve lojistik/teslimat süreçlerinin kendine has zorlukları nedeniyle müşteri geri bildirimlerinin analizi açısından özel bir derinliği ve uygulama ihtiyacını barındırmaktadır.

Araştırmanın temel amacı, mobilya sektöründeki müşteri geri bildirimlerinde saklı bulunan gizli konu ilişkilerini sistematik biçimde ortaya çıkarmak ve bu yapıları kullanarak güvenilir, yüksek doğruluklu bir çağrı atama sınıflandırma modeli geliştirmektir. Bu kapsamda, öncelikle müşteri geri bildirimleri öncelikle Latent Dirichlet Allocation (LDA) yöntemiyle analiz edilmiş [15]; elde edilen konu dağılımları, Terim Frekansı-Ters Doküman Frekansı (TF-IDF) özellikleri ile birleştirilerek sınıflandırma için geliştirilmiş bir veri seti oluşturulmuştur (Model-2). Oluşturulan özellik kümesi, Destek Vektör Makinesi (SVM), LightGBM ve Yapay Sinir Ağları gibi çeşitli sınıflandırma algoritmaları ile test edilmiştir[16].

Çalışmanın güncel literatür ile bağını kurmak ve performansın üst sınırını belirlemek amacıyla, Transformer tabanlı Sentence-BERT (S-BERT) modeli ile SVM birleştirilerek bir üçüncü model (Model-3) geliştirilmiştir. Model-3 mimarisi, yorumların bağlamsal anlamını yüksek doğrulukla temsil eden S-BERT gömülmelerini, sınıf merkezlerine dayalı olasılıksal çağrı tipi atama yaklaşımı ile birleştirmektedir. Tüm test edilen yöntemler arasında en yüksek performansı göstererek %90,8 doğruluk oranına ulaşmıştır.

Bu bulgular, LDA tabanlı özellik çıkarımının klasik makine öğrenmesi modelleri için güçlü bir temel sunduğunu kanıtlarken, bağlamsal gömülmelerin semantik yapıyı yakalama kapasitesi sayesinde çağrı atama süreçlerinde çok daha yüksek doğruluk ve tutarlılık sağladığını açıkça ortaya koymaktadır. Önerilen modeller, çağrı merkezi operasyonlarında subjektif değerlendirmeleri en aza indirerek müşteri deneyimini ve operasyonel verimliliği maksimize etme potansiyeli taşımaktadır.

2. LİTERATÜR TARAMASI (LITERATURE REVIEW)

Bu bölümde, müşteri hizmetleri yönetim süreçleri ve yapısal olmayan müşteri geri bildirimlerinin analizi bağlamında, metin madenciliği ve makine öğrenmesi yöntemlerinin uygulandığı bilimsel çalışmalar incelenmiştir. Odak noktası; hizmet kalitesi ve müşteri memnuniyeti, dijitalleşmenin müşteri deneyimine etkileri, metin temsili ve konu modellemesi yaklaşımlarıdır.

2.1 Müşteri Hizmetleri Yönetimi ve Dijitalleşme

Müşteri hizmetleri yönetimi, günümüzde işletmeler için sadece bir destek fonksiyonu olmaktan çıkarak, uzun vadeli rekabet avantajı sağlayan stratejik bir yapı taşına dönüşmüştür [1]. Literatürde, hizmet kalitesini müşteri beklentileri ve algılanan performans arasındaki fark

üzerinden değerlendiren SERVQUAL modeli [17] gibi klasik çerçeveler, sektörler arası uygulamalarda hizmet kalitesinin temel boyutlarını (güvenilirlik, yanıt verebilirlik, empati vb.) ölçmek için geniş bir temel oluşturmuştur. Örneğin, Yücel [18], SERVQUAL'i bankacılık sektöründe uygulayarak müşteri memnuniyeti üzerinde güvenilirlik boyutunun kritik bir unsur olduğunu göstermiştir.

Dijital teknolojilerin ve Büyük Veri Analitiği yöntemlerinin hızla gelişmesiyle birlikte, müşteri hizmetleri yönetim süreçleri de önemli ölçüde dönüşüme uğramıştır [4]. Bu dönüşüm, müşterilerin davranışlarını ve ihtiyaçlarını çok kanallı bir perspektiften daha derinlemesine anlamayı mümkün kılmıştır [7]. Chen ve Xie [19], çevrimiçi tüketici incelemelerinin işletmeler için kritik bir stratejik değer kaynağı olduğunu ve bu geri bildirimlerin, özellikle pazarlama iletişimi ve hizmet iyileştirme kararlarına doğrudan entegre edilebileceğini vurgulamışlardır.

Costa ve ark. [20] ile Chernova ve Sharafutdinova [21], büyük veri analitiği uygulamalarının, küçük ve orta ölçekli işletmeler dâhil olmak üzere perakende ve hizmet sektörlerinde, müşteri deneyimini optimize etmek ve süreç verimliliğini artırmak için stratejik bir araç olduğunu rapor etmişlerdir. Varol ve İşler [22] ise dijital araçların bankacılık sektöründe müşteri algısı ve sadakat üzerindeki olumlu etkilerini araştırmıştır.

Kobayashi ve ark. [23], yapay zekâ destekli şikâyet yönetimi ile metin madenciliği tekniklerinin bir araya gelerek müşteri deneyimini iyileştirme potansiyeline sahip olduğunu göstermiştir.

2.2 Metin Madenciliği ve Klasik Sınıflandırma Uygulamaları

Müşteri geri bildirimlerinin büyük bir kısmını oluşturan yapısal olmayan metin verilerinin etkin analizi, işletmelerin iyileştirme fırsatlarını tespit etmesi açısından hayati öneme sahiptir [8]. Bu bağlamda, Metin Madenciliği, bu verilerden bilgi çıkarma, konu modelleme ve duygu analizi gibi alanlarda yoğunlaşarak stratejik bir araç olarak kullanılmaktadır [9].

Literatürde en yaygın kullanılan konu modelleme algoritmalarından biri, Blei ve ark. tarafından geliştirilen Latent Dirichlet Allocation (LDA) yöntemidir [15]. LDA, her bir metin belgesini belirli konu dağılımlarına göre dağıtmakta ve her konuyu kelime dağılımı ile tanımlayarak, metin yığınları içindeki gizli semantik yapıları ortaya çıkarmaktadır [24]. Hong ve Davison [25] ise bu yöntemi kısa metinlerden anlamlı konuları çıkarmak için kullanmış ve sosyal medya platformlarında konu modellemesi çalışmalarının yapılabilirliğine katkı sağlamışlardır.

LDA, özellikle müşteri yorumlarının sınıflandırılması ve analizinde geniş uygulama alanı bulmuştur:

- **Sektörel Uygulamalar:** Budak ve Sökmen [26], otel hizmetlerine yönelik çevrimiçi yorumları LDA ile analiz ederek ana temaların hizmet kalitesi ve çalışan davranışları olduğunu tespit etmiştir. Sutherland ve Kiatkawsinvd [27] da benzer şekilde konaklama platformu yorumlarını kullanarak, otel konumu ve temizlik gibi temel temaları belirlemiştir.

- **Duygu ve Sınıflandırma Analizleri:** Tuna ve ark. [28], otel yorumlarında duygu analizi için SVM, Naive Bayes ve Lojistik Regresyon gibi makine öğrenimi algoritmalarını karşılaştırmış ve SVM'in en yüksek doğruluğu sağladığını belirtmiştir. Bu durum, klasik sınıflandırma algoritmalarının metin verisi üzerindeki başarısını göstermektedir. Büyükeke ve ark. [29], Tripadvisor yorumlarını analiz ederek SVM ve Lojistik Regresyon algoritmalarının duygu sınıflandırmasında daha iyi sonuçlar verdiğini ortaya koymuştur.

- **Sektörel Odaklı Uygulamalar:** Göker ve Tekedere [30], metin madenciliği yöntemleri kullanarak yapısal olmayan veriyi, TF-IDF ağırlıklandırma yöntemiyle vektörel olarak temsil etmiş ve Ardışık Minimal Optimizasyon algoritmasının en yüksek başarıyı elde ettiğini (%88,73) göstermiştir. Erduran [31] ise bankalara yönelik müşteri şikayetlerini kümeleme ve birliktelik analizleri ile incelemiştir.

Literatürde yer alan bu uygulamalı çalışmaların çoğunda, Python, RapidMiner, Weka gibi yazılımlar kullanılmış olup, TF-IDF ve LDA gibi yöntemler metin temsilinde temel olarak kabul edilmiştir [32].

2.3 Güncel Doğal Dil İşleme (NLP) Yaklaşımları: Transformer Modelleri

Metin madenciliği alanında 2018 yılından itibaren yaşanan en önemli gelişme, bağlamsal bilgi içeren dil temsillerini mümkün kılan Transformer mimarisinin ve özellikle BERT (Bidirectional Encoder Representations from Transformers) modelinin ortaya çıkmasıdır [33, 34]. BERT ve ondan türetilen modeller (RoBERTa, GPT vb.), klasik TF-IDF ve LDA gibi frekans tabanlı temsillerin aksine, kelimelerin cümle içindeki anlamını (bağlamını) modelleyebilen güçlü bağlamsal gömülmeler üretmektedir [34, 35].

Transformer tabanlı modeller, metin sınıflandırma, cümle benzerliği ve duygu analizi gibi görevlerde geleneksel yöntemlere göre belirgin bir performans artışı sağlamıştır. Bununla birlikte, orijinal BERT mimarisinin yüksek hesaplama maliyeti ve çıkarım süresinin yavaş olması, büyük hacimli çağrı merkezi verilerinin gerçek zamanlı olarak işlenmesinde operasyonel kısıtlar yaratmaktadır. Bu nedenle çalışmada, BERT mimarisinin üzerine

siamese ve triplet network yapıları eklenerek cümle ve paragraf gömülmelerinin hızlı ve etkin şekilde üretilmesi amacıyla geliştirilen Sentence-BERT (S-BERT) modeline odaklanılmaktadır [36]. S-BERT, her cümleyi sabit boyutlu bir vektör uzayında temsil ederek kosinüs benzerliği gibi hızlı vektörel karşılaştırmaların yapılmasına olanak tanır. Böylece bağlamsal bilgiyi korurken, klasik yöntemlere yakın hızda çıkarım verimliliği sunmakta ve çalışmamızın operasyonel gereklilikleri için ideal bir ileri NLP yaklaşımını oluşturmaktadır.

2.4 Mobilya Sektöründeki Literatür Boşluğu

İncelenen literatür, müşteri hizmetleri yönetim sürecinin ve müşteri memnuniyetinin iyileştirilmesi adına metin madenciliği ve büyük veri analizi yaklaşımlarının stratejik önemini açıkça ortaya koymaktadır. Ancak, bu çalışmaların büyük çoğunluğu finans, konaklama (otel), perakende veya e-ticaret gibi sektörlere odaklanmıştır.

Özellikle, mobilya sektörüne özgü müşteri şikâyetlerinin; lojistik (teslimat), montaj süreçleri ve ürün çeşitliliğinin karmaşıklığından kaynaklanan yapısal zorlukları ele alarak çağrı atama süreçlerine yönelik bilimsel bir model geliştiren çalışmaların sınırlı olduğu görülmektedir. Örneğin, Sözen ve Bardak [37] çalışmalarında, e-ticaret mobilya yorumlarını web madenciliği ile görselleştirmiştir. Sınırlı sayıda sektöre özgü çalışmalar bulunsa da konu modellemesi ile yeni çağrı tipleri oluşturulmasını ve bu tiplere dayalı yüksek başarımlı makine öğrenmesi sınıflandırma modeli önerilmesini içeren kapsamlı bir çalışma literatürde yer almamaktadır.

Bu durum, mobilya sektöründeki operasyonel verimliliğin artırılmasına yönelik veri odaklı ve sektörel bağlamı güçlendirilmiş yaklaşımlara olan ihtiyacı gündeme getirmektedir. Mevcut çalışma, bu boşluğu doldurmayı ve klasik yöntemleri birleştirerek objektif bir çağrı atama modeli sunmayı hedeflemektedir.

3. YÖNTEM VE MATERYAL (METHODS AND MATERIALS)

Çalışma kapsamında, metin madenciliği ve makine öğrenmesi yöntemleri kullanılarak mobilya sektörüne ait müşteri geri bildirimleri analiz edilmiş ve elde edilen bulgular ışığında çağrı atama süreçlerini optimize etmeye yönelik bir sınıflandırma modeli önerilmiştir. Yöntem adımları aşağıdaki gibidir;

1. **Metin Madenciliği ve Veri Yapılandırma**: Veri ön işleme adımları uygulanmış, TF-IDF matrisi oluşturulmuş ve LDA yöntemiyle olasılıksal konu modellemesi yapılmıştır.

2. **Sınıflandırma ve Optimizasyon (Model-1 ve Model-2)**: Model-1'de mevcut manuel konu etiketleri kullanmış ve Model-2'de, LDA'dan elde edilen yeni sınıf değerleri TF-IDF matrisi ile birleştirilerek çeşitli sınıflandırma algoritmaları (Naive Bayes, Yapay Sinir Ağları, SVM, Karar Ağaçları, LightGBM, AdaBoost ve Rastgele Orman) uygulanmıştır. Böylece LDA'nın sübjektifliği giderme başarısı test edilmiştir.

3. **Bağlamsal Gömülme ile Performans Üst Sınırının Belirlenmesi (Model 3):** En güncel NLP yaklaşımı olan Sentence-BERT (S-BERT) modeli kullanılarak, yorumların bağlamsal anlamını içeren yoğun vektörgömülmeleri elde edilmiş ve bu gömülmeler, Model- 2'de en başarılı olan SVM algoritması ile sınıflandırılarak performansın teorik üst sınırı belirlenmiştir.

3.1. Veri Kümesi ve Deney Kurulumu (Dataset and Experimental Setup)

Çalışmanın veri kümesi, Türkiye'deki üç farklı mobilya firmasına (A, B, C olarak anonimleştirilmiştir) ait toplam 9.930 müşteri geri bildiriminden oluşmaktadır. Bu veriler, analiz edilmek üzere aşağıdaki dört temel değişkende yapılandırılmıştır:

- **Kayıt Yeri** (Geri bildirimin kaynağı)
- **Marka** (Gizlilik nedeniyle anonimleştirilmiştir)
- **Çağrı Tipi** (Müşteri temsilcisi tarafından atanan manuel etiket)
- **Yorum** (Analiz edilen metin tabanlı müşteri geri bildirimi)

3.1.1. Veri Kaynakları (Data Sources)

Veri kümesi, çevrimiçi şikâyet platformu olan Şikayetvar.com'dan toplanan veriler ile firmaların iç kaynaklarından (e-posta ve SMS yoluyla iletilen geri bildirimler) elde edilen verileri kapsamaktadır. Modelin genelleme yeteneğini anlamak amacıyla, bu kaynakların yüzdesel dağılımı şöyledir:

- Şikayetvar.com (Çevrimiçi Platform): %71,20
- Diğer Kaynaklar (E-posta/SMS): %28,80

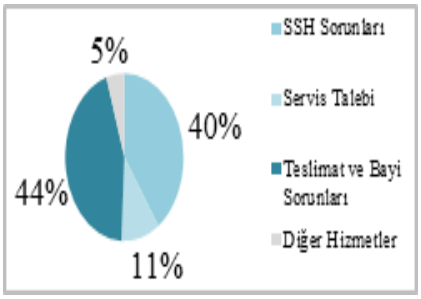
Bu dağılım, veri setinin büyük bir kısmının halka açık şikâyet platformlarından geldiğini göstermektedir. Bu, çevrimiçi platformlardan gelen kamuya açık şikayetler ile firmaların iç iletişim kanallarından gelen daha özel geri bildirimler arasındaki ton ve yapı farklılıklarını hesaba katmanın önemini göstermektedir.

Çağrı Tipi değişkeni, müşteri temsilcileri tarafından manuel olarak atanan ve mobilya sektörünün operasyonel zorluklarını yansıtan dört ana kategoriye ayrılmıştır:

- **SSH Sorunları (Satış Sonrası Hizmetler):** Garanti, tamir ve iade süreçleri gibi ürün kaynaklı sorunlar.

- **Servis Talebi (Teknik Destek/Bakım):** Montaj, kurulum veya yedek parça gereksinimi.
- **Teslimat ve Bayi Sorunları: Lojistik ve Tedarik Zinciri** süreçlerinden kaynaklanan gecikmeler, hatalı ürün gönderimi veya bayi kaynaklı aksaklıklar.
- **Diğer Hizmetler:** Fatura, bilgilendirme eksikliği veya genel müşteri hizmetleri iletişimi.

Bu kategorilerin veri setindeki yüzdesel dağılımı Şekil 1'de gösterilmektedir.



Şekil 1. Çağrı Tipi Dağılım Grafiği
(Call Type Distribution Chart)

Tablo 1, veri kümesinde yer alan temel değişkenleri ve müşteri geri bildirimlerinin genel yapısını örneklemektedir. Bu yapı, sonraki aşamalarda uygulanan ön işleme, konu modelleme ve sınıflandırma süreçlerine temel oluşturmuştur.

Tablo 1. Örnek Veri Seti
(Sample Dataset)

Kayıt Yeri	Marka	Çağrı Tipi	Yorum
Şikayetvar.com	A	Servis Talebi	Yay sesi
Diğer	B	SSH Sorunları	A'dan aldığımız koltuk takımından hiç memnun değiliz A bizi gerçekten hayal kırıklığına uğrattı #Markadan beklentileriniz nedir?# #Tamir-Düzeltme#

3.1.2. Deney Kurulumu ve Teknik Parametreler (Experimental Setup and Technical Parameters)

Analiz süreci, Python programlama dili kullanılarak Jupyter Notebook ortamında gerçekleştirilmiştir. Çalışmanın tekrarlanabilirliğini sağlamak amacıyla, deneylerde kullanılan donanım özellikleri, yazılım sürümleri, kütüphaneler ve model hiperparametreleri Tablo 2'de sunulmuştur.

Tablo 2. Parametre Tablosu
(Parameter Table)

Algoritma / Yöntem	**Hiperparametreler (Parameters)**
TF-IDF	min_df = 5 · max_df = 0.90 · ngram_range = (1,1) · use_idf = True · smooth_idf = True
LDA (Latent Dirichlet Allocation)	n_components = 4 · learning_method='batch' · max_iter = 50 · random_state = 42 · evaluate_every = None
Naive Bayes (MultinomialNB)	alpha = 1.0 (varsayılan Laplace düzeltmesi) · fit_prior = True
SVM (LinearSVC)	C = 1.0 · dual = 'auto' · max_iter = 2000
Yapay Sinir Ağı (MLPClassifier)	hidden_layer_sizes = (100, 50) · activation='relu' · solver='adam' · max_iter = 50 · early_stopping = True
Karar Ağacı (Decision Tree)	criterion='gini' · max_depth=None · min_samples_split=2 · min_samples_leaf=1 · random_state=42
Random Forest	n_estimators = 200 · criterion='gini' · max_depth=None · bootstrap=True · random_state=42
AdaBoost	n_estimators = 100 · learning_rate = 1.0 · base_estimator = DecisionTree(max_depth=1)
LightGBM	num_leaves = 31 · learning_rate = 0.1 · n_estimators = 200 · boosting_type='gbdt' · max_depth=-1
S-BERT (Embedding Model)	Model: paraphrase-multilingual-mpnet-base-v2 · Gömülme boyutu: 768

3.2. Metin Ön İşleme (Text Preprocessing)

Metin ön işleme adımı, doğal dilde elde edilen müşteri geri bildirimlerinin analiz edilebilir hale getirilmesi için Python ortamında Pandas, NLTK ve RE kütüphaneleri kullanılarak gerçekleştirilmiştir. Ön işleme sürecinde izlenen adımlar aşağıda sıralanmaktadır:

1. Gereksiz karakterlerin temizlenmesi: Noktalama işaretleri, sayılar ve HTML etiketleri metinden çıkarılarak analiz için önemsiz unsurlar temizlenmiştir.
2. Küçük harfe dönüştürme: Metindeki tüm harfler küçük harfe dönüştürülerek tutarlılık sağlanmıştır.
3. Durak kelimelerin kaldırılması: Anlam taşımayan "ve", "bir", "bu" gibi durak kelimeler metinden çıkarılmıştır.
4. Kök bulma: Kelimelerin ekleri çıkarılarak kök forma indirgenmiştir.
5. Tokenizasyon: Metin, kelime, cümle veya karakter gibi anlamlı parçalara ayrılmıştır.

Örnek bir müşteri geri bildirimi üzerinde bu adımların uygulanması Tablo 3'de gösterilmektedir.

Tablo 3. Metin Ön İşleme Örneği
(Sample of Text Preprocessing)

Ham Veri	A marka baza koku yaptı iade kabul edilmesi için servisin ürünü göndermesi gerekiyormuş yaklaşık 2 aya yakın bazasız kalacağım.
Temizlenmiş Veri	A marka baza koku yaptı iade kabul edilmesi servisin ürünü göndermesi gerekiyormuş aya yakın bazasız kalacağım
Küçük Harf Dönüşümü	a marka baza koku yaptı iade kabul edilmesi servisin ürünü göndermesi gerekiyormuş aya yakın bazasız kalacağım
Durak Kelimelerin Çıkartılması	"a marka baza koku yaptı iade kabul servis ürün gönder gerek aya yakın bazasız kalacağım
Kök Bulma	marka baza koku yap iade kabul servis ürün gönder gerek ay yakın baza kal
Tokenizasyon	"marka", "baza", "koku", "yap", "iade", "kabul", "servis", "ürün", "gönder", "gerek", "ay", "yakın", "bazasız", "kal"

3.3. Özellik Çıkarımı (Feature Extraction)

3.3.1. Metinlerin Sayısal Temsili (Numerical Representation of Texts)

Temizlenen müşteri yorumları, analiz sürecinde kullanılmak üzere iki farklı biçimde sayısal vektörlere dönüştürülmüştür.

A. Frekans Tabanlı Temsil (TF-IDF):

TF-IDF yaklaşımı, her kelimenin belge içindeki önem derecesini ölçerek metni yüksek boyutlu bir özellik vektörüne dönüştürmektedir. Bu yöntem, sık kullanılan ancak ayırt ediciliği düşük kelimelerin ağırlığını azaltırken, az görülen ancak anlam açısından kritik kelimelerin etkisini artırır. Çalışmada oluşturulan TF-IDF matrisi, konu modelleme (LDA) ve klasik makine öğrenmesi algoritmalarına girdi olarak kullanılmıştır. Bu tercih, literatürde TF-IDF'in metin sınıflandırma için yaygın ve etkili bir başlangıç temsili olarak önerilmesiyle uyumludur [24].

TF-IDF, metindeki önemli kelimeleri belirlemek için terim frekansını (TF) ve bu kelimenin kaç farklı belgede geçtiğini (IDF) dikkate alır. Bu yöntem, sık geçen genel kelimelerin önemini azaltırken, nadir fakat kritik öneme sahip kelimelere daha fazla ağırlık verir.

Terim Frekansı (TF), bir kelimenin belirli bir belgede kaç kez geçtiğini ifade eder. TF'nin matematiksel bağıntısı Eşitlik (1)'de yer almaktadır.

$$TF = \frac{Kelimenin\ Belgedeki\ Frekansı}{Belgedeki\ Toplam\ Kelime\ Sayısı} \quad (1)$$

Ters Doküman Frekansı (IDF), *k*elimenin toplam kaç belgede geçtiğini ölçer ve ters logaritmasını alır. IDF'nin matematiksel bağıntısı Eşitlik (2)'de yer almaktadır.

$$\text{IDF} = \log\left(\frac{Toplam\ Belge\ Sayısı}{Kelimenin\ Geçtiği\ Belge\ Sayısı}\right) \quad (2)$$

TF-IDF, her kelimenin TF ve IDF değerlerinin çarpımıdır. TF-IDF'nin matematiksel bağıntısı Eşitlik (3)'de yer almaktadır.

$$\text{TF-IDF} = TF * IDF \quad (3)$$

B. Bağlamsal Vektör Temsili (S-BERT Gömülmeleri):

Transformer mimarisi üzerine geliştirilen Sentence-BERT (S-BERT), metinleri bağlamsal anlamı koruyarak sabit boyutlu yoğun vektörlere dönüştüren modern bir yöntemdir. S-BERT, klasik frekans temsillerinin aksine kelime sırası, bağlam ve semantik benzerlik ilişkilerini modelleyebilir. Bu çalışmada her yorum, 768 boyutlu bir S-BERT gömülmesine dönüştürülmüş ve bu temsiller Model-3'te sınıflandırıcıya girdi olarak kullanılmıştır. S-BERT, BERT modellerine göre önemli ölçüde daha hızlı çıkarım süresi sunduğundan, operasyonel sistemlere entegrasyon için uygun bir alternatif sağlamaktadır [36].

3.4. Konu Modelleme (Topic Modeling)

Konu modelleme, büyük ölçekli metin koleksiyonlarındaki gizli temaları ortaya çıkarmayı amaçlayan bir metin madenciliği yaklaşımıdır. Çalışmada, müşteri geri bildirimlerinin olasılıksal konu dağılımlarını belirlemek amacıyla Latent Dirichlet Allocation (LDA) yöntemi uygulanmıştır. LDA, her belgenin birden fazla konu içerdiğini ve bu konuların kelime dağılımlarıyla temsil edildiğini varsayan probabilistik bir modeldir [15].

Bu çalışmada LDA, mevcut dört çağrı tipiyle ilişkili temaların ortaya çıkarılması ve özellikle manuel kategori atamalarında görülmeyen bağlamsal örüntülerin tespit edilmesi amacıyla kullanılmıştır. Bu işlem sonucunda her yoruma ait konu olasılık dağılımları elde edilmiş ve Model-1 ve Model-2 için ek özellik seti olarak kullanılmıştır.

3.4.1. LDA Algoritması (LDA Algorithm)

Latent Dirichlet Allocation (LDA), 2003 yılında Blei ve ark. [15] tarafından geliştirilen ve metinlerdeki gizli konuları keşfetmek için kullanılan bir probabilistik konu modelleme yöntemidir. LDA, yapılandırılmamış metin verilerinde gizli temaların belirlenmesi için en yaygın kullanılan modellerden biridir. Her belgeyi belirli konulara, her konuyu ise kelime dağılımlarına dayandırarak analiz eder. Geniş veri setlerinde esneklik ve ölçeklenebilirlik sunar, müşteri geri bildirim analizi ve

sosyal medya araştırmaları gibi çeşitli alanlarda kullanılır. Ancak bağlama duyarsızlık, hesaplama karmaşıklığı ve ideal konu sayısının belirlenmesindeki zorluklar gibi sınırlamaları vardır. Alternatiflerinden farklı olarak, konuların probabilistik yapısını modelleyerek belgeler arası ilişkileri daha etkili bir şekilde ortaya çıkarır. LDA'nın temel çalışma ilkeleri aşağıdaki gibidir;

1. Belge ve Kelime Dağılımı: LDA, her belgenin farklı konulardan oluştuğunu ve her konunun farklı kelimelerle temsil edildiğini varsayar. Her belge, belirli bir konu dağılımına sahiptir. Her konu, kelimelerden oluşan bir dağılıma sahiptir.

2. Probabilistik Yapı: LDA, bir belgeyi gözlemlenen kelimelerin ardışık bir dizisi olarak ele alır ve bu kelimelerin altında yatan gizli konuları modellemek için Bayes teoremini kullanır. Her kelimenin bir konudan geldiği varsayılır ve bu süreç probabilistik bir şekilde işler.

3. Varsayımlar: Konular Dirichlet dağılımı kullanılarak modellenir. Belgelerdeki kelimelerin belirli bir konudan gelme olasılığı, o konunun kelime dağılımına dayanır.

3.5. Sınıflandırma (Classification)

Bu çalışmada temel amaç, çağrı merkezi operasyonlarında manuel kategori atama sürecindeki sübjektifliği azaltarak yüksek doğruluklu bir otomatik sınıflandırıcı model geliştirmektir. Bu doğrultuda, LDA temelli klasik makine öğrenmesi modelleri (Model-1 ve Model-2) ve S-BERT temelli bağlamsal modeller (Model-3) değerlendirilmiştir.

Transformer tabanlı modeller yüksek doğruluk sunmalarına karşın gerçek zamanlı sistemlerde yüksek donanım maliyetleri nedeniyle sınırlı kullanılabilirliğe sahiptir. Bu sebeple Model-3'te, BERT'in hızlı çıkarım sağlayan bir varyantı olan S-BERT ile klasik makine öğrenmesinin yüksek performanslı temsilcisi SVM birleştirilmiştir. Böylece operasyonel maliyet açısından uygulanabilir ve literatürdeki güncel yaklaşımlarla uyumlu bir hibrit model elde edilmiştir.

Modeller, veri setinin eğitim-test bölünmüş yapısı üzerinde k-katlı çapraz doğrulama yöntemiyle test edilmiş ve performans metrikleri kullanılarak karşılaştırılmıştır.

3.5.1. Destek Vektör Makinesi (Support Vector Machine - SVM)

Destek Vektör Makinesi, yüksek boyutlu veriler üzerinde güçlü performans sergileyen doğrusal ve doğrusal olmayan sınıflandırma yöntemlerinden biridir [16]. SVM, sınıflar arasındaki ayrımı maksimize eden hiper düzlemi bulmayı amaçlar ve karar sınırı, destek vektörü adı verilen kritik veri noktaları tarafından belirlenir [38].

Metin verisi gibi yüksek boyutlu uzaylarda çekirdek fonksiyonları ile oldukça başarılı olduğu bilinmektedir. Bu nedenle hem TF-IDF ve LDA temsillerinde hem de S-BERT gömülmelerinde test edilmiştir. Literatür, SVM'in metin sınıflandırma görevlerinde geleneksel yöntemler arasında en başarılı modellerden biri olduğunu göstermektedir [26], [28], [29].

3.5.2. LightGBM

LightGBM, histogram tabanlı bölme ve yaprak odaklı büyüme stratejisi sayesinde hızlı ve düşük bellek tüketimine sahip bir gradient boosting algoritmasıdır [39]. Büyük veri kümelerinde verimli çalışması ve doğruluk performansı nedeniyle sınıflandırma görevlerinde sık tercih edilmektedir.

3.5.3. Yapay Sinir Ağları (Artificial Neural Networks)

Yapay sinir ağları, özellikle doğrusal olmayan ilişkilerin yoğun olduğu büyük veri setlerinde güçlü performans sergileyen esnek modellerdir [16]. Derin öğrenme mimarileri, çok katmanlı sinir ağları üzerinden karmaşık örüntüleri öğrenebilir. Transformer tabanlı BERT ve türevi modellerin temeli de bu çok katmanlı ağ yapısına dayanmaktadır [33]–[35]. Ancak ANN modelleri yüksek hesaplama maliyeti, uzun eğitim süresi ve aşırı öğrenme riski nedeniyle operasyonel sistemlerde daha sınırlı kullanılabilir.

3.5.4. Karar Ağaçları (Decision Trees)

Karar ağaçları, sınıflandırma ve regresyon analizlerinde kullanılan popüler bir tekniktir. Karar ağaçları, veriyi dallara ayırarak karar kuralları oluşturur ve bu kurallara dayanarak veriyi sınıflandırır veya tahmin eder [40]. Bu teknik, veriyi anlaşılır bir şekilde görselleştirir ve karmaşık karar süreçlerini basitleştirir. Kolay yorumlanabilir, görselleştirilebilir ve anlaşılırdır. Hem kategorik hem de sürekli veriyle çalışabilir. Özellikle küçük veri setlerinde iyi performans gösterir. Aşırı öğrenmeye eğilimlidir. Derin ağaçlar, veri setini gereğinden fazla öğrenerek, yeni verilerde düşük performans gösterebilir. Gelişmiş Versiyonlar; Rastgele Orman (Random Forest) ve Gradient Boosting Trees gibi yöntemler, birçok karar ağacını bir araya getirerek aşırı öğrenmeyi önleyip model performansını artırır [16].

3.5.5. AdaBoost

AdaBoost, bir dizi zayıf öğreniciyi (örneğin, karar ağaçlarını) bir araya getirerek güçlü bir model oluşturan etkili bir Boosting algoritmasıdır [41]. Her bir zayıf modelin hatalarını dikkate alarak bir sonraki modeli eğitir ve bu süreçte hataları daha iyi sınıflandırmaya çalışır. Bu, modelin doğruluğunu artıran iteratif bir süreçtir. AdaBoost, Boosting yönteminin temel bir versiyonudur ve daha gelişmiş sürümleri bulunmaktadır. Gradient Boosting ve XGBoost gibi yöntemler, AdaBoost'un prensiplerini daha karmaşık optimizasyon teknikleriyle genişleterek model performansını artırır. Bu yöntemler, daha iyi genelleme yeteneği ve daha yüksek doğruluk sunarak, AdaBoost'un sınırlamalarını aşmaya çalışır.

3.5.6. Rastgele Orman (Random Forest)

Rastgele Orman, birden fazla karar ağacını birleştirerek sınıflandırma ve regresyon problemlerinde yüksek doğruluk sağlayan güçlü bir makine öğrenimi algoritmasıdır [42]. Her ağaç, rastgele bir alt veri kümesi ve öznitelikler kullanılarak eğitilir ve sonuçlar çoğunluk oyu (sınıflandırma) veya ortalama (regresyon) ile birleştirilir. Aşırı öğrenmeye dayanıklıdır ve öznitelik önemini belirleme gibi avantajlar sunar. Ancak, hesaplama maliyeti yüksek olabilir ve yorumlanabilirliği düşüktür. Finans, sağlık, pazarlama ve e-ticaret gibi çeşitli alanlarda yaygın olarak uygulanır.

3.5.7. Naive Bayes

Naive Bayes, olasılık teorisine dayanan ve Bayes teoremini kullanan bir sınıflandırma algoritmasıdır [16]. Bu algoritma, özelliklerin birbirinden bağımsız olduğu varsayımına dayanır, bu nedenle "naive" (saf) olarak adlandırılır [16]. Hızlı ve verimlidir. Küçük veri setlerinde bile iyi performans gösterir. Özelliklerin sayısı arttıkça performansı düşmez. Çok hızlı ve verimli çalışır. Özellikle metin madenciliği, spam tespiti gibi uygulamalarda güçlüdür [15]. Özelliklerin birbirinden bağımsız olduğunu varsayar, bu genellikle gerçek dünyadaki verilerde geçerli değildir. Bu nedenle bazı veri setlerinde performansı düşük olabilir. Naive Bayes algoritmasının temelinde Bayes teoremi yer alır. Bayes teoremi, bir olayın olasılığını, diğer bir olayın gerçekleştiği bilgisine dayanarak hesaplar.

3.6. Performans Metrikleri (Performance Metrics)

Sınıflandırma modellerinin değerlendirilmesinde birden fazla performans ölçütü kullanılmıştır. Bu metrikler, modelin genel doğruluğunun yanı sıra sınıflar arasındaki dengenin korunup korunmadığını ve tahmin başarısının hangi yönünün güçlü/zayıf olduğunu analiz etmeye olanak sağlamaktadır. Çalışmada kullanılan temel metrikler: Karmaşıklık Matrisi, Doğruluk değeri, Kesinlik değeri, Hatırlama değeri, F1-Skoru, ROC eğrisi ve Kappa istatistiği'dir..

3.6.1. Karmaşıklık Matrisi (Confusion Matrix)

Karmaşıklık matrisi bir sınıflandırma modelinin performansını görsel olarak değerlendirmek için kullanılır. Bu matris, modelin doğru ve yanlış tahminlerini ayrıntılı olarak gösterir ve modelin başarısız olduğu alanları net bir şekilde görmeyi sağlar. Matris, dört farklı bileşenden oluşur. Karmaşıklık matrisi değişkenleri Tablo 4'te yer almaktadır.

Tablo 4. Karmaşıklık Matrisi
(Confusion Matrix)

	Gerçek Pozitif	Gerçek Negatif
Tahmin Pozitif	Doğru Pozitif (DP)	Yanlış Pozitif (YP)
Tahmin Negatif	Yanlış Negatif (YN)	Doğru Negatif (DN)

3.6.2. Doğruluk (Accuracy)

Modelin tüm veri setindeki doğru tahmin oranını ifade eder. Genel başarıyı gösterir. Yüksek doğruluk değeri model performansının yüksek olduğunu gösterir. Doğruluk değeri, sınıflar arasında dengesizlik olduğunda yanıltıcı olabilir. Örneğin, nadir görülen bir sınıfın tahminlerinde doğruluk yüksek olsa bile, modelin başarısı düşük olabilir. Doğruluk, ek metrikler ile daha kapsamlı bir şekilde değerlendirilmelidir. Doğruluk metriğinin denklemi Eşitlik (6)' da yer almaktadır.

$$\text{Doğruluk} = \frac{(DP+DN)}{(DP+DN+YP+YN)} \quad (6)$$

3.6.3. Kesinlik (Precision)

Pozitif olarak tahmin edilen verilerin gerçekten pozitif olma oranını ölçer. Yanlış pozitifleri azaltmada önemlidir. Yüksek kesinlik değeri model performansının yüksek olduğunu gösterir. Kesinlik metriğinin denklemi Eşitlik (7)'de yer almaktadır.

$$\text{Kesinlik} = \frac{(DP)}{(DP+YP)} \quad (7)$$

3.6.4. Hatırlama (Recall)

Pozitif sınıfa ait gerçek verilerin ne kadarının doğru tahmin edildiğini ölçer. Yanlış negatifleri azaltmada önemli bir metriğe sahiptir. Yüksek hatırlama değeri, model performansının yüksek olduğunu gösterir. Hatırlama metriğinin denklemi Eşitlik (8)'de yer almaktadır.

$$\text{Hatırlama} = \frac{(DP)}{(DP+YN} \quad (8)$$

3.6.5. F1 Skoru (F1 Score)

Kesinlik ve hatırlama arasında bir denge sağlar. Dengesiz veri setlerinde kullanışlıdır. Yüksek F1 skoru kesinlik ve hatırlama arasında iyi bir denge olduğunu gösterir. F1 skorunun denklemi Eşitlik (9)'da yer almaktadır.

$$\text{F1 Skoru} = \frac{2 * Kesinlik * Hatırlama}{(Kesimlik + Hatırlama)} \quad (9)$$

3.6.6. ROC Eğrisi (ROC Curve)

Farklı eşik değerleri için modelin hatırlama oranı ile yanlış pozitif oranını karşılaştırır. Eğrinin altında kalan alan (AUC), modelin genel performansını değerlendirir. Değerin yüksek olması, modelin farklı eşik değerleri arasında iyi bir performans sergilediğini ve pozitif sınıfları negatiflerden ayırmada başarılı olduğunu gösterir.

3.6.7. Kappa İstatistiği (Cohen's Kappa)

Rastgele tahminlere göre modelin ne kadar iyi performans gösterdiğini ölçer. Değerin yüksek olması modelin rastgele tahminlerden çok daha iyi performans gösterdiğini ifade eder. Kappa istatistiğinin denklemi Eşitlik (10)'da yer almaktadır.

$$\text{Kappa} = \frac{Gözlenen\ Doğruluk - Beklenen\ Doğruluk}{(1 - Beklenne\ Doğruluk)} \quad (10)$$

4. BULGULAR (FINDINGS)

Bu bölümde, müşteri geri bildirimleri üzerinde gerçekleştirilen metin madenciliği, konu modelleme ve sınıflandırma çalışmalarından elde edilen bulgular sunulmaktadır. Bulgular, kelime frekansları, TF-IDF'e dayalı anahtar kelime analizi, LDA ile belirlenen konu–etiket ilişkileri ve sınıflandırma modellerinin performans karşılaştırmaları çerçevesinde raporlanmıştır.

Veri setinin ön işleme sonrası analizinde toplam 7.412 benzersiz kelime tespit edilmiştir. Tablo 5'te en sık geçen kelimeler sunulmaktadır. Bu kelimeler, müşteri sorunlarının ürün kullanımı, teslimat süreçleri ve satış sonrası hizmetler etrafında yoğunlaştığını göstermektedir.

Tablo 5. Örnek Kelime Listesi
(Sample Word List)

Kelime	Toplam Kelime Sayısı	Dokümanda Bulunma Sayısı
Satın Alındı	9.063	6.390
Marka	7.197	5.593
Beklemek	5.460	5.177
Ürün	11.381	4.907
Özel	6.935	4.904
İsim	7.885	4.521
Kayıt	4.419	4.092
Hizmet	8.359	4.064
Fatura	4.243	4.027

Kelime frekansları incelendiğinde, "ürün", "satın alındı", "hizmet" ve "fatura" gibi kelimelerin yüksek tekrar oranlarına sahip olduğu görülmektedir. Bu bulgu, mobilya sektöründe geri bildirimlerin ağırlıklı olarak ürün kalitesi, satış sonrası destek ve teslimat süreçleri üzerinde yoğunlaştığını göstermektedir.

TF-IDF yöntemi ile her çağrı tipi için belirlenen anahtar kelimeler Tablo 6'da sunulmaktadır. Bu analiz, çağrı tiplerinin semantik yapıları arasındaki farklılıkları ortaya koymuştur.

Tablo 6. Örnek Anahtar Kelime Tablosu
(Sample Keyword Table)

Çağrı Tipi	Anahtar Kelimeler
Teslimat ve Bayi Sorunları	Teslimat, gecikme, şube, kargo, gönderi…
SSH Sorunları	Şikâyet, garanti, iade, tamir, destek…
Servis Talebi	Teknik destek, servis, onarım, koltuk, yatak…
Diğer Hizmetler	Hizmet, bilgilendirme, müşteri hizmetleri…

Bu sonuçlar, metin içeriklerinin çağrı tiplerine göre tutarlı ayrıştığını ve TF-IDF tabanlı kelime ağırlıklandırmanın sınıflandırma modelleri için anlamlı sinyaller ürettiğini göstermektedir.

4.1. LDA ile Konu Modelleme Sonuçları ve Etiketleme Sorunu (Results of Topic Modeling with LDA and Labeling Issue)

Veri seti LDA algoritması ile dört konu üzerinden modellenmiş ve her bir geri bildirimin bu konulara ilişkin olasılık dağılımı elde edilmiştir. Tablo 7 ve Tablo 8 ilgili sonuçları sunmaktadır.

Tablo 7. Örnek LDA Sonuç Tablosu
(Sample LDA Results Table)

Yorum Numarası	3	4
Mevcut Çağrı Tipi	Teslimat ve Bayi Sorunları	SSH Sorunları
Servis Talebi	0,08	0,03
Teslimat ve Bayi Sorunları	0,26	0,55
SSH Sorunları	0,36	0,38
Diğer Hizmetler	0,30	0,04
Yeni Çağrı Tipi	Diğer Hizmetler	Teslimat ve Bayi Sorunları

Yorum-3: "XXX mağazasından ürünleri alalı yaklaşık 2-3 yıl oldu. Koltukların yaylarında ses geliyor. Dış görünümlerinde herhangi bir olumsuzluk olmamasına rağmen rahatsız edici bir şekilde ses var. Bunun giderilmesini talep etmekteyim."

Örneğin, 3 numaralı yorum SSH Sorunları kategorisiyle %36 oranında ilişkilendirilmiş olup, aynı zamanda Teslimat ve Bayi Sorunları (%26) ve Diğer Hizmetler (%30) kategorileriyle de bağlantılıdır. Yorum metni incelendiğinde, ürünlerin kullanım sonrası yaşanan sorunlarının çözümüne yönelik talep olduğu

görülmektedir. Bu bağlamda, ilgili geri bildirimin SSH Sorunları kategorisinde değerlendirilmesi anlamlıdır.

Tablo 8. Çağrı Dağılım Olasılıkları
(Call Distribution Probabilities)

Çağrı Tipi	Servis Talebi	Teslimat ve Bayi Sorunları	SSH Sorunları	Diğer Hizmetler
Servis Talebi	0,043	0,367	0,164	0,426
Teslimat ve Bayi Sorunları	0,038	0,349	0,158	0,454
SSH Sorunları	0,032	0,361	0,156	0,450
Diğer Hizmetler	0,041	0,365	0,157	0,438

Tablo 8, tüm çağrı tiplerinin birbirleriyle olan ilişki düzeylerini göstermektedir. Bulgular, geri bildirimlerin yaklaşık %78'inin mevcut çağrı tipinden farklı bir çağrı tipine de yüksek olasılıkla atanabileceğini ortaya koymuştur. Bu durum, çağrı merkezi operasyonlarında manuel atama sürecinin doğasında olan subjektifliğin ve bunun yol açtığı çoklu konu ilişkisinin kanıtıdır ve otomatik model ihtiyacının temelini oluşturur.

4.2. Sınıflandırma Model Karşılaştırması ve Gerekçelendirme (Classification Model Comparison and Justification)

Sınıflandırma performansı, Model-1 (manuel etiketler) ve Model-2 (LDA ile güncellenmiş etiketler) üzerinden değerlendirilmiştir. Her iki model yedi makine öğrenmesi algoritmasıyla test edilmiş ve sonuçlar Tablo 9 ve 10'da sunulmuştur. Sınıflandırma verisi, %80 test ve %20 eğitim olarak bölünmüş ve modellerin sağlamlığı 10-katlı çapraz doğrulama yöntemiyle test edilmiştir.

Tablo 9. Model-1 Sınıflandırma Sonuçları
(Model-1 Classification Results)

	Doğruluk	Kappa	F1 Skor	Kesinlik	Hatırlama
Naive Bayes	0,421	-0,016	0,215	0,230	0,244
Yapay Sinir Ağları	0,381	-0,007	0,248	0,251	0,249
Destek Vektör Makinesi	0,423	-0,007	0,224	0,208	0,247
Karar Ağaçları	0,380	-0,000	0,251	0,252	0,251
LightGBM	0,382	-0,005	0,247	0,248	0,249
AdaBoost	0,431	-0,000	0,210	0,249	0,250
Rastgele Orman	0,415	-0,017	0,225	0,280	0,244

Model-1'deki algoritmaların Kappa istatistiği değerleri (örn. SVM için -0,007; Karar Ağaçları için -0,000) incelendiğinde, bu performansın rastgele tahmin seviyesinde olduğu görülmektedir. Negatif veya sıfıra yakın Kappa değerleri, mevcut çağrı tipi etiketlerinin, metin verisiyle anlamlı bir ilişki kuracak kadar güvenilir olmadığını açıkça kanıtlamaktadır. Bu durum, çağrı temsilcileri tarafından yapılan orijinal manuel etiketlemenin hatalı, tutarsız veya metin verisiyle uyumsuz olduğunu güçlü bir şekilde desteklemektedir. Bu sonuç, çalışmamızda objektif bir otomatik çağrı atama modelinin geliştirilmesi gerekliliğini bilimsel olarak doğrulamaktadır.

Tablo 10. Model-2 Sınıflandırma Sonuçları
(Model-2 Classification Results)

	Doğruluk	Kappa	F1 Skor	Kesinlik	Hatırlama
Naive Bayes	0,802	0,666	0,611	0,578	**0,878**
Yapay Sinir Ağları	0,885	0,820	0,850	**0,837**	0,867
Destek Vektör Makinesi	**0,891**	**0,828**	**0,852**	0,836	0,870
Karar Ağaçları	0,676	0,495	0,620	0,621	0,621
LightGBM	0,863	0,783	0,819	0,798	0,847
AdaBoost	0,709	0,542	0,669	0,660	0,682
Rastgele Orman	0,802	0,678	0,726	0,683	0,813

Model-2'de çağrı tipi hedef değişkeni LDA sonuçlarıyla birleştirilip konu odaklı güncellendiğinden, metin verisiyle daha uyumlu hale getirilmiş ve bu sayede sınıflandırma algoritmalarının performansı anlamlı biçimde artmıştır.

Tablo sonuçları incelendiğinde, en yüksek Doğruluk Oranı %89,1 ile Destek Vektör Makineleri (SVM) algoritmasına aittir. SVM'in Kappa değeri 0,828 olup, bu değer rastgele tahminlere kıyasla modelin tahminlerinin çok yüksek oranda güvenilir ve anlamlı olduğunu gösterir (Kappa > 0.8 genellikle "mükemmel uyum" olarak kabul edilir). Bu, Model-2'nin istatistiksel güvenilirliğini net bir şekilde ortaya koymaktadır. SVM'i sırasıyla Yapay Sinir Ağları (%88,5), LightGBM (%86,3), Naive Bayes (%80,2), Rastgele Orman (%80,2), AdaBoost (%70,9) ve Karar Ağaçları (%67,6) takip etmektedir.

Performans metrikleri ayrıntılı incelendiğinde:

- F1 Skoru: SVM, 0,852 ile tüm algoritmalar arasında en yüksek değeri elde etmiştir. Bu, kesinlik ve hatırlama dengesinin en iyi şekilde sağlandığını gösterir.
- Kesinlik (Precision): 0,836 ile SVM, yanlış pozitif oranını düşük tutarak doğru tahminlerde yüksek isabet sağlamıştır.
- Hatırlama (Recall): 0,870 değeri, modelin gerçek çağrı tiplerini yakalama oranının oldukça yüksek olduğunu gösterir.

Bu sonuçlar, SVM algoritmasının müşteri geribildirimlerini doğru çağrı tipine atamada diğer yöntemlere kıyasla daha tutarlı ve dengeli bir performans sergilediğini ortaya koymaktadır. Özellikle, yüksek Hatırlama (0,870) değeri, müşteri taleplerinin yanlış yönlendirilme riskini (YN) %87 oranında azaltma potansiyeli taşırken; yüksek Kesinlik (0,836) değeri, yanlış pozitif tahminlerden doğabilecek operasyonel maliyetleri (Yanlış birime atanan çağrılar) en aza indirir. Bu, çağrı merkezlerinde iş gücü optimizasyonunu ve müşteri memnuniyetini doğrudan artıran somut bir katkıdır.

4.3. S-BERT ile Performans Üst Sınırının Belirlenmesi (Determining the Upper Performance Bound with S-BERT)

LDA temelli Model-2, manuel etiketlerden kaynaklanan tutarsızlıkları azaltarak sınıflandırma performansını anlamlı biçimde iyileştirmiştir. Bununla birlikte, çalışmanın akademik güncelliğini sağlamak ve metinlerin bağlamsal boyutunu yakalayabilen daha gelişmiş modellerle karşılaştırma yapmak amacıyla Transformer tabanlı bir yaklaşımın test edilmesi gerekli görülmüştür. Literatürde Transformer mimarisinin, özellikle bağlamsal anlam temsilinde klasik yöntemlere kıyasla belirgin üstünlük sağladığı geniş biçimde rapor edilmektedir. Bu nedenle çalışma kapsamında Sentence-BERT (S-BERT) tabanlı bağlamsal gömülmeler kullanılarak üçüncü bir model (Model-3) geliştirilmiştir.

Güncellenmiş Model-3 mimarisinde tüm müşteri geri bildirimleri S-BERT tarafından 768 boyutlu bağlamsal vektörlere dönüştürülmüş, her çağrı tipi için sınıf merkez vektörleri hesaplanmış ve her yorum için tüm sınıflara ilişkin olasılıksal benzerlik dağılımları elde edilmiştir. En yüksek olasılıklı çağrı tipi, ilgili yorum için "Yeni Çağrı Tipi" olarak atanmış ve bu etiketler SVM sınıflandırıcısında kullanılmıştır. Bu yapı, bağlamsal gömülme ile olasılıksal sınıf tahmin mekanizmasını bir araya getirerek klasik LDA sürecinden bağımsız, modern bir Transformer tabanlı üst seviye sınıflandırma modeli ortaya koymuştur. Karşılaştırmalı sonuçlar Tablo 11'de sunulmuştur.

Tablo 11. Model-2 ve Model-3 Karşılaştırma Sonuçları
(Comparison Results of Model-2 and Model-3)

Model	Özellik Temsili	Doğruluk (Accuracy)	Kappa	F1 Skoru
Model-3	S-BERT Gömülmeleri + Olasılıksal Çağrı Tipi + SVM	%90,8	%87	%90,3
Model-2	TF-IDF + LDA + SVM	%89,1	%82,8	%85,2

Model-3 sonuçları, bağlamsal temsil gücünün klasik TF-IDF tabanlı yaklaşımların üzerine çıkabildiğini göstermektedir. Özellikle SSH gibi semantik çeşitliliği yüksek sınıflarda performansın beklendiği kadar yüksek olmaması, bağlamsal modelin sınıf dengesizliğine duyarlılığını göstermektedir. Ancak genel doğruluk ve makro F1 değerlerindeki belirgin artış, bağlamsal gömülmelerin metinsel benzerliği yakalama konusunda güçlü bir avantaj sunduğunu açıkça ortaya koymaktadır.

Bu bulgular, çalışmanın hibrit yaklaşımını destekler niteliktedir: Model-2 pratik uygulamalar için kararlı ve tutarlı sonuçlar üretirken, Model-3 çağrı atama süreçlerinde en yüksek doğruluk, en yüksek semantik tutarlılık ve en yüksek operasyonel verimliliği sunan model olarak öne çıkmaktadır. Sonuç olarak S-BERT tabanlı Model-3, mobilya sektöründe çağrı merkezleri için performans üst sınırını temsil eden ileri düzey bir çözüm ortaya koymaktadır.

5. SONUÇLAR VE ÖNERİLER (CONCLUSIONS AND RECOMMENDATIONS)

Bu çalışma, müşteri geri bildirimlerinin Metin Madenciliği ve Makine Öğrenmesi yöntemleriyle analiz edilmesi sonucunda mobilya sektöründeki müşteri hizmetleri yönetimine rehberlik edecek ve çağrı atama süreçlerini daha objektif, tutarlı ve verimli hâle getirecek bir sınıflandırma modeli önermiştir. Araştırmanın temel amacı, müşteri memnuniyetini artırmak ve çağrı yönlendirme süreçlerinde sübjektif kararların neden olabileceği hataları en aza indirmektir.

Türkiye'deki üç büyük mobilya markasına ait toplam 9.930 müşteri şikâyeti detaylı biçimde incelenmiş ve manuel etiketleme süreçlerinin metinlerle yeterli uyum göstermediği, çağrı tiplerinin çok boyutlu ve örtüşen yapılara sahip olduğu görülmüştür. Bu nedenle, LDA tabanlı konu modelleme yöntemiyle metinlerdeki temel temalar keşfedilmiş ve bu konu dağılımları TF-IDF özellikleriyle birleştirilerek daha tutarlı bir veri seti oluşturulmuştur.

5.1. Temel Bulgular ve Operasyonel Katkı

LDA ile geliştirilmiş etiketlerin kullanıldığı Model-2'de, SVM algoritması %89,1'lik Doğruluk oranına ulaşmıştır. Model-2, manuel atama sübjektifliğini giderme hedefinde büyük başarı kaydetmiştir. Çalışmanın performan süst sınırını belirlemek amacıyla, Transformer tabanlı Sentence-BERT (S-BERT) modeli ile SVM birleştirilerek Model-3 denenmiştir. Model-3, %90,8 Doğruluk oranı ile tüm modeller arasında en yüksek performansı göstermiş ve çağrı atama süreçlerinde ulaşılabilecek en yüksek verimlilik potansiyelini sunmuştur. Bu sonuçlar, modelin sektöre yönelik somut ve nicel katkısını ortaya koymaktadır.

Bu sonuçlar, modelin sektöre yönelik somut ve nicel katkısını ortaya koymaktadır:

- **İş Gücü Optimizasyonu:** Otomatik sınıflandırma sayesinde çağrı merkezlerinde doğru birime yönlendirme süresi önemli ölçüde kısalacak, personel iş yükü azalacak ve operasyonel maliyetler minimize edilecektir.

- **Memnuniyet Artışı:** Yüksek Hatırlama değeri, müşteri taleplerinin yanlış yönlendirilme riskini azaltarak müşteri memnuniyetsizliğini minimize etme potansiyeli taşımaktadır.

- **Objektiflik:** Geliştirilen Model-2 ve Model-3, çağrı temsilcilerinin sübjektif değerlendirmelerinden kaynaklanabilecek bilgi kaybının önüne geçen, yüksek doğruluk oranıyla çalışan otomatik bir çağrı atama mekanizması sunmaktadır.

5.2. Gelecek Çalışmalar İçin Öneriler

Bu çalışma, mobilya sektöründe güçlü bir temel oluşturmuştur. Model-2'nin LDA temelli maliyet etkin çözümü ve Model-3'ün S-BERT temelli sağladığı en üst düzey performans göz önüne alınarak, gelecek çalışmalar için aşağıdaki adımlar önerilmektedir:

1. **Operasyonel Uygulama ve Test:** LDA ve S-BERT ile oluşturulan olasılıkların gerçek zamanlı sistemlerde test edilmesi, operasyonel uygulanabilirliğin artırılması ve müşteri ilişkileri yönetimi süreçlerinde proaktif hizmet yaklaşımlarının geliştirilmesi açısından önemli bir adım olacaktır.

2. **Duygu Analizi ve Aciliyet Sınıflandırması:** Şikâyet metinlerindeki öfke, memnuniyetsizlik veya aciliyet göstergelerinin analize dahil edilmesi, çağrıların önceliklendirilmesi için önemli bir katkı sağlayabilir.

3. **Veri Kaynağını Zenginleştirme:** Sosyal medya paylaşımları, chatbot konuşmaları, e-posta diyalogları ve satış sonrası anketlerin eklenmesi, özellikle Model-3'ün bağlamsal performansını daha da artıracaktır.

4. **Hibrit Model Yaklaşımı:** Operasyonel sistemlerde, Model-2 birincil, hızlı ve düşük maliyetli sınıflandırıcı, Model-3 ise zorlayıcı, çelişkili veya yüksek riskli geri bildirimler için ikinci değerlendirme mekanizması olarak hibrit biçimde kullanılabilir. Bu yaklaşım performans ve maliyet dengesini optimize edecektir.

Etik Kurul Onayı (Ethical Approval)
Bu çalışma, halka açık platformlardan ve anonimleştirilmiş kurumsal verilerden yararlanmıştır. Çalışma, kişisel verilerin korunması ve anonimleştirilmesi kurallarına tam olarak uymuştur. Bu nedenle, ulusal yayın kuralları gereği ayrıca bir Etik Kurul Onayı alınması gerekmemektedir.

Çıkar Çatışması Beyanı (Conflict of Interest Statement)
Yazarlar, bu makalenin sonuçlarını etkileyebilecek herhangi bir finansal veya kişisel çıkar çatışması olmadığını beyan ederler.

Fonlama (Funding)
Bu araştırma için herhangi bir özel fon veya hibe alınmamıştır.

KAYNAKLAR (REFERENCES)

[1] M. E. Porter, **Competitive Strategy: Techniques for Analyzing Industries and Competitors (1'st ed.)**, Free Press, New York, 1998.

[2] P. Kotler, K. L. Keller, Marketing Management (15th ed.), **Pearson Education.** 2016.

[3] R. L. Oliver, "Whence consumer loyalty", *Journal of Marketing*, 63, 33–44, 1999.

[4] U. Suryani, "The Effect of Digital Transformation on Company Performance and Customer Satisfaction in the Era of Technological Disruption", *Journal of Finance and Business Digital (JFBD)*, 4(1), 33-48, 2025.

[5] J. Reis, M. Amorim, N. Melão, P. Matos, "Digital transformation: A literature review and guidelines for future research", **Advances in Intelligent Systems and Computing**, Cilt 745, Editörler: Rocha, Á., Adeli, H., Reis, L.P., Costanzo, S., In: Trends and Advances in Information Systems and Technologies, Springer, Cham, 2018.

[6] T. H. Davenport, **Big Data at Work: Dispelling the Myths, Uncovering the Opportunities**, Harvard Business Review Press, 2014.

[7] W. He, S. Zha, L. Li, "Social media competitive analysis and text mining: A case study in the pizza industry", *International Journal of Information Management*, 33(3), 464-472, 2013.

[8] R. Feldman, J. Sanger, **The Text Mining Handbook: Advanced Approaches in Analyzing Unstructured Data**, Cambridge University Press, 2007.

[9] S. Fan, R. Y. K. Lau, J. L. Zhao, "Demystifying big data analytics for business intelligence through the lens of marketing mix", Big Data Research, 2(1), 28-32, 2015.

[10] G. D. Miner, J. Elder, A. Fast, T. Hill, R. Nisbet, D. Delen, **Practical Text Mining and Statistical Analysis for Non-structured Text Data Applications**, Academic Press, 2012.

[11] K.Çalış, O. Gazdağı, O. Yıldız, "Reklam İçerikli Epostaların Metin Madenciliği Yöntemleri İle Otomatik Tespiti", *Bilişim Teknolojileri Dergisi*, 6(1), 1-7, 2013.

[12] B. Saylan, S. Çınaroğlu, "Metin Madenciliği ve Makine Öğrenmesi Teknikleri ile Sağlık Hizmetleri Pazarlamasına Yönelik Twitter Verilerinin Analizi", *Bilişim Teknolojileri Dergisi*, 17(2), 109-121, 2024.

[13] B. Karakus, G. Aydin, "Call center performance evaluation using big data analytics". **International Symposium on Networks, Computers and Communications (ISNCC)**, IEEE., 1-6, 2016.

[14] M. H. Huang, R. T. Rust, "Artificial intelligence in service", *Journal of Service Research*, 21(2), 155-172, 2018.

[15] D. M. Blei, A. Y. Ng, M.I. Jordan, "Latent Dirichlet allocation", *Journal of Machine Learning Research*, 3, 993–1022, 2003.

[16] V. Krishnaiah, G. Narsimha, N. S. Chandra, "Survey of classification techniques in data mining", *International Journal of Computer Science and Engineering*, 2(9), 65-74, 2014.

[17] L. L. Berry, "SERVQUAL: A multiple-item scale for measuring consumer perceptions of service quality", *Journal of Retailing*, 64(1), 12–40, 1988.

[18] M. Yücel, "Toplam hizmet kalitesinin SERVQUAL analizi ile ölçümü: Bankacılık sektöründe bir araştırma", *Elektronik Sosyal Bilimler Dergisi*, 12(44), 82-106, 2013.

[19] Y. Chen, J. Xie, "Online consumer review: Word-of-mouth as a new element of marketing communication mix", Management Science, 54(3), 477–491, 2008.

[20] A. C. F. Costa, N. F. Capelo, M. Espuny, A. B. T. D. Rocha, O. J. D. Oliveira, "Digitalization of customer service in small and medium-sized enterprises: drivers fort he development and improvement", *International Journal of Entrepreneurial Behavior & Research*, 30(2/3), 305-341, 2024.

[21] D.V. Chernova, N.S. Sharafutdinova, I.I. Nurtdinov, Y.S. Valeeva, L.I. Kuzmina, "The Transformation of the Customer Value of Retail Network Services Under Digitalization", Lecture Notes in Networks and Systems, Cilt 84, Editörler: S. Ashmarina, M. Vochozka, V. Mantulenko, In: Digital Age: Chances, Challenges and Future, Springer, Cham, 2020.

[22] H. C. Varol, D. B. İşler, "Dijitalleşme sürecinde müşteri deneyimi ve sadakat ilişkisi", *Uluslararası Global Turizm Araştırmaları Dergisi*, 8(1), 1–20, 2023.

[23] Y. Kobayashi, T. Uchida, T. Inoue, Y. Iwasaki, R. Ito, K. Saito, H. Akiyama, K. Tsuda, K. Yoshida, "Analysis of 50 years of health food research trends using Natural Language Processing and Generative AI", Procedia Computer Science, 246, 1875-1884, 2024.

[24] J. Han, M. Kamber, J. Pei, Data Mining: Concepts and Techniques (3rd ed.), **Morgan Kaufmann Publishers**, 2011.

[25] L. Hong, B. D. Davison, "Empirical study of topic modeling in Twitter". In Proceedings of the First Workshop on Social Media Analytics, 80–88, 2010.

[26] I. Budak, A. Sökmen, "Otel hizmetlerinin değerlendirilmesinde gizli Dirichlet Ayırımı ile analiz: Kastamonu ili örneği", *Journal of Tourism & Gastronomy Studies*, 10(4), 2942–2954, 2022.

[27] I. Sutherland, K. Kiatkawsin, "Determinants of guest experience in Airbnb: a topic modeling approach using LDA, Sustainability, 12(8), 3402, 2020.

[28] M. F. Tuna, O. Kaynar, M. Ş. Akdoğan, "Otellere ilişkin çevrimiçi geribildirimlerin makine öğrenmesi yöntemleriyle duygu analizi", *İşletme Araştırmaları Dergisi*, 13 (3), 2232-2241, 2021.

[29] A. Büyükeke, A. Sökmen, C. Gencer, "Metin Madenciliği ve Duygu Analizi Yöntemleri ile Sosyal Medya Verilerinden Rekabetçi Avantaj Elde Etme: Turizm Sektöründe Bir Araştırma", *Journal of Tourism & Gastronomy Studies*, 8(1), 322-335, 2020.

[30] H. Göker, H. Tekedere, "FATIH projesine yönelik görüşlerin metin madenciliği yöntemleri ile otomatik değerlendirilmesi". *Bilişim Teknolojileri Dergisi*, 10(3), 291-299, 2017.

[31] G. Yıldız Erduran, **Online müşteri şikayetlerinin veri madenciliği ile incelenmesi**, Doktora Tezi, Trakya Üniversitesi, 2017.

[32] C. Melek, **Metin madenciliği teknikleri ile şirketlerin vizyon ifadelerinin analizi**, Yüksek Lisans Tezi, Dokuz Eylül Üniversitesi, 2012.

[33] A.Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, I. Polosukhin "Attention is all you need", Advances in Neural Information Processing Systems, 30, 5998-6008, 2017.

[34] J. Devlin, M. W. Chang, K. Lee, K. Toutanova, "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding", arXiv preprint arXiv:1810.04805. 2018.

[35] Y. Liu, M. Ott, N. Goyal, J. Du, M. Joshi, D. Chen, , ... & V. Stoyanov, "RoBERTa: A robustly optimized bert pretraining approach", arXiv preprint arXiv:1907.11692. 2019.

[36] N. Reimers, I. Gurevych, "Sentence-BERT: Sentence Embeddings using Simese BERT-Networks", arXiv preprint arXiv:1908.10084. 2019.

[37] E. Sözen, T. Bardak, "Mobilya üretiminde farklı malzemelerin web madenciliği yöntemleri ile değerlendirilmesi", *Mobilya ve Ahşap Malzeme Araştırmaları Dergisi*, 4(2), 103-113, 2021.

[38] C. Cortes, V. Vapnik, "Support-vector networks", Machine Learning, 20(3), 273–297, 1995.

[39] G. Ke, Q. Meng, T. Finley, T, Wang, W. Chen, W. Ma, Q. Ye, T. Liu, "LightGBM: A highly efficient gradient boosting decision tree", In Advances in Neural Information Processing Systems; The MIT Press: Cambridge, MA, USA, 3146–3154, 2017.

[40] J. R. Quinlan, "Induction of decision trees", Machine Learning, 1(1), 81–106, 1986.

[41] Y. Freund, R. E. Schapire," A decision-theoretic generalization of on-line learning and boosting", *Journal of Computer and System Sciences*, 55(1), 119–139, 1997.

[42] L. Breiman, "Random forests", Machine Learning, 45(1), 5–32, 2001.