



Veri Madenciliği

Tüketicilerden Şikâyetlerinden Sorun Türü Tahmini: CFPB
Consumer Complaints Verisiyle Açıklanabilir Metin
Madenciliği Uygulaması

AVENGERS

Başar Yasin ŞAHİN	22040301029	barisyasinsahin@stu.topkapi.edu.tr
Musa ULUĞ	21040301044	musaulug @stu.topkapi.edu.tr
Salih İmran BÜKER	22040301062	salihimranbuker @stu.topkapi.edu.tr
Yusuf YENİGÜN	22040301052	yusufyenigun @stu.topkapi.edu.tr

https://github.com/YusufYenigun/YUSUF_YEN-G-N_22040301052_AVENGERS
https://github.com/barisYasi/Veri_Madencilik_proje
https://github.com/mUsaulug/Veri_Madenciliği_Final/blob/main/21040301044_Musa_Uluğ_Final.ipynb
https://github.com/mersieS/salih_imran_b-ker_veri_madenciliği_final

1) Problem Tanımı

İş / Bilimsel Soru

Consumer Financial Protection Bureau (CFPB) tarafından yayımlanan tüketici finans şikayetleri veri setinde, her şikayet kaydı için ürün türü (mortgage, kredi kartı, borç tahsili vb.), şikayet konusu (issue/sub_issue), şirketin verdiği yanıt türü, yanıtın zamanında olup olmaması, şikayet metninin uzunluğu ve yanıt süresi gibi yapısal özelliklere bakarak, tüketicinin bu şikayetin resmî olarak "dispute" (itiraz) edip etmeyeceğini önceden tahmin etmeyi hedefliyoruz. İş sorusunu şöyle özetliyoruz:

"Verilen yapısal özellikler kullanılarak, bir şikayetin consumer_disputed? = Yes olup olmayacağı önceden tahmin edilebilir mi? Bu tahmin, müşteri ilişkileri ve şikayet yönetimi süreçlerini iyileştirmek için anlamlı bir sinyal üretebilir mi?"

Görev Türü:

- Görev türü: İkili Sınıflandırma (Binary Classification)
- Etiketler: Yes ve No
- Pozitif sınıf: Yes (müşterinin firmaya resmî itiraz yaptığı durum)

Hedef Değişken(ler):

- Hedef değişken: consumer_disputed?
- Değerler: "Yes" ve "No"
- Pozitif sınıf: Yes (müşterinin firmaya resmî itiraz yaptığı durum)
- Etki alanı: Tüketici finansı, şikayet yönetimi, müşteri memnuniyeti ve risk yönetimi

Başarı Kriterleri:

Veri seti dengesiz ($\approx 80\text{ No} / 20\text{ Yes}$) olduğu için yalnızca accuracy yeterli değil. Başarı kriterlerimizi şöyle tanımlıyoruz:

- **Baseline (Dummy model - herkese "No" tahmin eden):**
 - ROC AUC $\approx 0,50$
 - Recall(Yes) = 0,00
- **Vize aşaması hedefleri:**
 - Baseline'a göre ROC AUC $> 0,60$ elde etmek
 - Pozitif sınıf için Recall(Yes) $\geq 0,60$ seviyesine çıkmak

2) Proje Yönetimi

Kilometre Taşları ve Zaman Çizelgesi

Final dönemi proje zaman çizelgesi aşağıdaki şekilde planlanmıştır:

1. Hafta (1–7 Aralık):

- Vize sonuçlarının değerlendirilmesi
- Base modellerin güçlü ve zayıf yönlerinin analiz edilmesi
- Final aşamasında kullanılacak advanced/deep learning modellerin belirlenmesi
- Ortak değerlendirme metriklerinin (Accuracy, Precision, Recall, F1, ROC AUC) netleştirilmesi
- Sorumlu: Tüm grup

2. Hafta (8–15 Aralık):

- Metin tabanlı özelliklerin (tokenization, embedding tabanlı yaklaşımlar) hazırlanması
- Deep learning modeller için veri formatlarının düzenlenmesi
- Eğitim/test ayırmalarının ve sınıf dağılımlarının kontrol edilmesi
- Sorumlu: Barış (ortak veri yapısı), tüm grup gözden geçirme

3. Hafta (16–22 Aralık)

- Her grup üyesinin kendi advanced modellerini geliştirmesi (en az 2 model)
- CNN, BiLSTM, Attention ve MLP tabanlı mimarilerin eğitimi
- İlk performans sonuçlarının alınması
- Sorumlu: Her üye kendi modelinden sorumlu

4. Hafta (23–31 Aralık)

- Hiperparametre ayarlamaları (learning rate, epoch, batch size, model derinliği vb.)
- Model çıktılarının karşılaştırılması
- Overfitting ve sınıf dengesizliği etkilerinin analizi
- Sorumlu: Her üye + grup içi karşılıklı değerlendirme

5. Hafta (1–7 Ocak)

- Nihai modellerin belirlenmesi
- Confusion matrix, ROC eğrileri ve sınıf bazlı performans metriklerinin çıkarılması
- Model performanslarının tablo haline getirilmesi
- Sorumlu: Tüm grup

6. Hafta (8–15 Ocak)

- En iyi modelin seçimi ve sonuçların yorumlanması
- Feature engineering ve modelleme sonuçlarının raporlanması
- Kaynakça ve literatür bölümünün genişletilmesi
- Sorumlu: Salih (rapor koordinasyonu), tüm grup katkı

7. Hafta (16–29 Ocak)

- Final raporunun yazımı ve biçimsel düzenlemeleri
- Tablolar, şekiller ve grafiklerin son kontrolleri
- Dil ve tutarlılık kontrolleri
- Sorumlu: Tüm grup

30 Ocak – Final Teslim

- Nihai raporun ALMS sistemine yüklenmesi
- Bireysel Jupyter Notebook dosyalarının teslimi

Roller ve Sorumluluklar (Final Aşaması)

Final aşamasında roller vize dönemindeki görev dağılımı korunarak, ileri seviye modelleme ve analiz çalışmalarına genişletilmiştir:

- **Barış Yasin Şahin**

- CNN tabanlı modellerin geliştirilmesi
- Ortak veri yapısının korunması
- Model sonuçlarının karşılaştırılmasına destek

- **Musa Uluğ**
 - TextCNN ve Attention BiLSTM modellerinin geliştirilmesi
 - Metin tabanlı derin öğrenme yaklaşımlarının analizi
- **Salih İmran Büker**
 - Attention BiLSTM ve MLP tabanlı modeller
 - Model performanslarının raporlanması
 - Final raporunun koordinasyonu ve bütünlüğü
- **Yusuf Yenigün**
 - MLP tabanlı derin öğrenme modelleri
 - Model çıktılarının sınıf bazlı değerlendirilmesi

3) İlgili Çalışmalar (Literatür İncelemesi)

Tüketicilerin şikayet verilerinin analizi, hem akademik literatürde hem de endüstriyel uygulamalarda müşteri memnuniyeti, risk yönetimi ve süreç iyileştirme açısından önemli bir araştırma alanıdır. Consumer Financial Protection Bureau (CFPB) tarafından yayımlanan tüketici finans şikayetleri veri seti, büyük hacmi ve gerçek dünya problemlerini yansıtması nedeniyle bu alanda sıkılıkla kullanılmaktadır.

CFPB ve Benzeri Veri Setleri Üzerine Çalışmalar

Literatürde CFPB veri seti çoğunlukla iki ana perspektiften ele alınmaktadır. İlk yaklaşım, şikayet metinlerinin doğal dil işleme (NLP) yöntemleriyle sınıflandırılmasıdır. Bu çalışmalarında TF-IDF, word embeddings ve derin öğrenme tabanlı modeller (CNN, LSTM, BiLSTM) kullanılarak şikayetlerin konu, ürün veya sonuç bazlı tahmini hedeflenmektedir. Kaggle platformunda yer alan birçok çalışma, Random Forest, Gradient Boosting ve Logistic Regression gibi geleneksel modellerle birlikte metin tabanlı özelliklerin performansa etkisini incelemiştir.

İkinci yaklaşım ise yapısal (structured) özellikler üzerinden tahmin yapılmasıdır. Bu çalışmalarında ürün türü, yanıt süresi, yanıtın zamanında olup olmaması ve şirket tepkisi gibi değişkenler kullanılarak müşteri davranışlarının (şikayetin itiraza dönüşmesi gibi) öngörülmesi amaçlanmaktadır. Bu yaklaşım, özellikle açıklanabilirlik (explainability) ve iş süreçlerine doğrudan katkı sağlama açısından öne çıkmaktadır.

Sınıf Dengesizliği Problemi

CFPB veri seti gibi gerçek dünya veri setlerinde sınıf dengesizliği yaygın bir problemdir. Literatürde bu problem; fraud detection, churn prediction ve customer complaint analysis gibi alanlarda sıkılıkla ele alınmıştır. Araştırmalar, yalnızca accuracy metriğine odaklanması yaniltıcı sonuçlar doğurabileceğini, özellikle azınlık sınıfı için Recall, F1-score ve ROC AUC gibi metriklerin değerlendirilmesi gerektiğini vurgulamaktadır. Class weight kullanımı, eşik (threshold) ayarlamaları ve yeniden örneklemme (SMOTE, undersampling) yöntemleri bu problemin çözümünde yaygın olarak önerilmektedir.

Derin Öğrenme ve Metin Sınıflandırma Yaklaşımı

Son yıllarda metin sınıflandırma problemlerinde derin öğrenme tabanlı modellerin başarısı dikkat çekmektedir. Kim (2014) tarafından önerilen CNN tabanlı metin sınıflandırma yaklaşımı, kısa ve uzun metinlerde etkili özellik çıkarımı sağlamaktadır. LSTM ve BiLSTM modelleri ise metin içerisindeki sıralı bağımlılıkları yakalamada güçlündür. Attention mekanizmaları, modelin metnin hangi bölümlerine odaklandığını belirleyerek hem performansı artırmakta hem de yorumlanabilirliği güçlendirmektedir.

Bu çalışmalar, büyük ve dengesiz veri setlerinde derin öğrenme modellerinin uygun hiperparametre ayarlamalarıyla geleneksel modellere kıyasla daha dengeli performans sunduğunu göstermektedir.

Bu Projenin Literatüre Katkısı

Bu proje, mevcut çalışmalarlardan aşağıdaki yönleriyle ayrılmaktadır:

1. **Hibrit yaklaşım:** Yapısal özellikler ve metin tabanlı derin öğrenme modellerinin birlikte değerlendirilmesi.
2. **Sistematik karşılaştırma:** Her grup üyesinin farklı advanced model mimarileri kullanarak sonuçları ortak metriklerle kıyaslaması.
3. **Sınıf dengesizliği odağı:** Model seçiminde özellikle pozitif sınıf (consumer_disputed? = Yes) performansının önceliklendirilmesi.
4. **Uygulanabilirlik:** Elde edilen sonuçların şikayet yönetimi ve müşteri ilişkileri süreçlerine doğrudan katkı sağlayabilecek şekilde yorumlanması.

Bu yönleriyle çalışma, CFPB veri seti üzerinde gerçekleştirilen mevcut literatürü tamamlayıcı ve genişletici bir nitelik taşımaktadır.

4) Veri Tanımı ve Yönetimi

Veri Seti

- Adı: US Consumer Finance Complaints
- Kaynak: CFPB'nin kamu açık şikayet veri tabanı
- Kaggle üzerinden erişilen türetilmiş veri seti
(kaggle.com/datasets/kaggle/us-consumer-finance-complaints)
- Kullanım hakkı: Eğitim/araştırma amaçlı açık veri olarak paylaşılmaktadır. Nihai lisans detayları, Kaggle sayfası ve CFPB kullanım şartları üzerinden kontrol edilecektir.

Veri Şeması

- Satır sayısı: 555.957
- Sütun sayısı: 18
- Sütun türleri:
 - Çoğu sütun object (kategorik/metinsel)
 - complaint_id: int64
 - date_received ve date_sent_to_company: tarih tipine dönüştürülmektedir

Önemli Sütunlar

Sütun Türü	Sütun Adları
Ürün ve konu	product, sub_product, issue, sub_issue
Müşteri açıklaması	consumer_complaint_narrative
Firma yanıtı	company_response_to_consumer, company_public_response, timely_response
Süreç bilgileri	date_received, date_sent_to_company, submitted_via
Hedef değişken	consumer_disputed?

Boyut ve Sınıf Dengesi

- Toplam Gözlem: 555.957
- Hedef dağılımı:
 - No: 443.823 (%79,83)
 - Yes: 112.134 (%20,17)

Veri seti belirgin şekilde sınıf dengesiz; bu durum modelleme ve değerlendirme aşamalarında özel olarak ele alınacaktır

Veri Erişim Planı

- Veri seti, Kaggle üzerinden CSV dosyası olarak indirilip her grup üyesinin yerel ortamında `consumer_complaints.csv` adıyla saklanmaktadır.
- İsteğe bağlı olarak GitHub reposunda küçük bir örnek subset (sample.csv) tutulacaktır.
- Veri güncellenmeyecektir; proje boyunca aynı snapshot kullanılacaktır.

Etik, Gizlilik, Önyargı

- Veri, CFPB tarafından önceden anonimleştirilmiş ve kamuya açık hale getirilmiş durumdadır.
- Yine de state ve zipcode gibi sütunlar dolaylı olarak sosyo-ekonomik ve bölgesel farkları yansıtabilir; bu nedenle model sonuçlarının belirli alt gruplar üzerinde sistematik önyargı üretip üretmediği final aşamasında değerlendirilecektir.
- Proje, yalnızca eğitim/araştırma amaçlıdır; ticari veya gerçek operasyonel karar alma sürecinde kullanılmayacaktır.

5) Keşifsel Veri Analizi (Exploratory Data Analysis)

Veri Kalitesi Kontrolleri

Genel inceleme:

- `df.info()` ile her sütunun tipleri ve non-null sayıları incelenmiştir.
- Veri setinin boyutu: 555.957 satır × 18 sütun

Eksik değer analizi:

- `consumer_complaint_narrative`, `tags`, `company_public_response`, `consumer_consent_provided` gibi sütunlarda çok yüksek oranda eksik değer bulunmaktadır.
- `sub_product` ve `sub_issue` sütunlarında sırasıyla yaklaşık %28 ve %62 civarında eksik değer vardır.
- `state` ve `zipcode` sütunlarında ise eksik oranı %1'in altındadır.

Yinelenen kayıtlar ve aşırı uç değerler:

- Vize aşamasında özellikle `response_days` için aşırı uçlar incelenmiş, negatif değerler 0'a kliplenmiştir.
- Tam duplicate kayıt kontrolü final aşamasında yapılacaktır.

Dağılımlar ve Denge

Hedef değişken dağılımı:

- Bar grafikleri ile sınıf dengesizliği görselleştirilmiştir (%80 No / %20 Yes)

Product sütununda en çok şikayet gelen ürünler:

1. Mortgage
2. Debt collection
3. Credit reporting
4. Credit card
5. Bank account or service

Response_days değişkeni:

- Histogram çizilerek yanıt süresinin çoğunlukla 0–4 gün aralığında kümelendiği gözlemlenmiştir.
- Az sayıda da olsa 100+ gün süren yanıtlar bulunmaktadır.

Özellik-Hedef İlişkileri (Plan)

Final aşamasında ise şunlar analiz edilmiştir:

- `product × consumer_disputed?` oranları
- `company_response_to_consumer` kategorilerine göre dispute oranları
- `response_days` ve `complaint_length_words` için kutu grafikleri (dispute olup olmamasına göre)

Görselleştirme Planı

Vize aşamasında kullanılanlar:

- Hedef dağılımı için bar grafikleri
- Ürün, issue ve yanıt türleri için bar grafikleri
- response_days ve complaint_length_words için histogram / boxplot

Final aşamasında kullanılanlar:

- Karışıklık matrisleri
- ROC eğrileri
- Feature importance grafikleri

6) Veri Hazırlama Planı

Temizleme

1. Tarih sütunları:

- `date_received`, `date_sent_to_company` sütunlarının datetime tipine dönüştürülmesi
- Parse edilemeyen değerlerin NaT yapılması

2. Response_days türetilirken:

- Negatif değerlerin 0'a çekilmesi
- Uç değerlerin (ör. 365+ gün) gerekirse winsorization veya log-transform ile dengelenmesi (final aşamasında düşünülebilir)

3. Duplicate kayıtlar:

- Gerekirse duplicate kayıtların tespit edilip kaldırılması

İmputasyon Stratejisi

Sayısal sütunlar (complaint_length_words, response_days):

- Eksik değerleri median ile doldurma

Kategorik sütunlar (product, issue, company_response_to_consumer, timely_response, genişletilmiş set için sub_product, sub_issue, submitted_via, state):

- Eksik değerleri en sık görülen kategori ile doldurma

Önemli not: İmputasyon işlemleri, sizıntıyı önlemek için mutlaka Pipeline içinde ve yalnızca train verisi üzerinden fit edilecektir

Dönüşümler

Kategorik değişkenler:

- OneHotEncoder(handle_unknown="ignore") ile One-Hot Encoding

Sayısal değişkenler:

- Vize aşamasında ölçeklendirme yapılmamıştır
- Final aşamasında Logistic Regression / LinearSVC için StandardScaler eklenmesi değerlendirilecektir

Pipeline yapısı:

- Tüm dönüşümler ColumnTransformer içinde tanımlanmış
- Her model için tekrar kullanılabilen preprocessor objesi ile uygulanmaktadır

Özellik Mühendisliği

Şu anda yapılan özellikler:

Şu anda yapılan / planlanan özellik mühendisliği adımları:

- response_days: date_sent_to_company - date_received farkından türetilen yanıt süresi.
- complaint_length_words ve complaint_length_chars: Şikayet metninden türetilen basit metin uzunluğu ölçütleri.
- Final aşamasında potansiyel ek özellikler:
 - Yoğun saat/gün bilgisi (ör. hafta sonu şikayetleri için farklı davranış)
 - Ürün/kategori kombinasyonlarının frekansına dayalı etkileşim özelliklerı.

Final aşamasında potansiyel ek özellikler:

- Yoğun saat/gün bilgisi (ör. hafta sonu şikayetleri için farklı davranış)
- Ürün/kategori kombinasyonlarının frekansına dayalı etkileşim özelliklerı

Özellik Seçimi ve Boyut İndirgeme

Her grup üyesi farklı feature selection yaklaşımı denemektedir:

Filter yöntemleri:

- Barış: SelectKBest(f_classif, k=100)
- Musa: SelectKBest(chi², k=50)
- Salih: SelectKBest(mutual_info_classif, k=50 ve k=200)

• Embedded yöntemler:

Yusuf: SelectFromModel(DecisionTreeClassifier, threshold="median")

Not: Boyut indirgeme için PCA benzeri yöntemler şu aşamada zorunlu görülmemektedir; ancak One-Hot sonrası feature uzayı fazla büyürse final aşamasında değerlendirilebilir.

7) Modelleme Planı

Baseline Modeller

Dumb Baseline (Referans Model)

Model: DummyClassifier(strategy="most_frequent")

Açıklama: Tüm gözlemlere "No" tahmini yaparak, neden yalnızca accuracy'ye bakmanın sorunlu olduğunu göstermek için referans modeli

Beklenen performans:

- Accuracy $\approx 0,80$
- ROC AUC = 0,50
- Recall(Yes) = 0,0

Basit Baseline

Model: Sınıf dengesizliği dikkate alınmadan eğitilen Logistic Regression

Açıklama: class_weight parametresi kullanmayan temel model

Beklenen performans:

- Accuracy $\approx 0,80$
- Pozitif sınıfı neredeyse hiç yakalamama

Amaç: Bu iki baseline, "class_weight kullanan" modellerle karşılaştırma için temel alınacaktır.

Model Ailesi	Örnekler	Kullanım Gerekçesi	Sorumlu Kişi
Doğrusal Modeller	LogisticRegression, LinearSVC	<ul style="list-style-type: none"> Yorumlanabilirlik (özellik katsayıları üzerinden) Yüksek boyutlu sparse veri ile uyumlu olmaları Hızlı eğitim ve tahmin 	<ul style="list-style-type: none"> Başar: Logistic Regression (class_weight="balanced") + SelectKBest(f_classif) Salih: Logistic Regression (C=0.5, balanced) ve LinearSVC (balanced) + mutual_info_classif
Ağaç Tabanlı Modeller	DecisionTreeClassifier	<ul style="list-style-type: none"> Non-lineer ilişkileri yakalayabilme Feature importance üzerinden açıklanabilirlik Kategorik değişkenlerle doğal uyum 	<ul style="list-style-type: none"> Başar: Temel feature set üzerinde Decision Tree (balanced) Yusuf: Geniş feature set + Decision Tree (balanced) + SelectFromModel
Olasılıksal Modeller	ComplementNB	<ul style="list-style-type: none"> Çok kategorili, One-Hot'lanmış özelliklerde hızlı baseline Şikâyet metnine dair türetilmiş özelliklerle uyumlu Sınıf dengesizliğinde iyi performans 	<ul style="list-style-type: none"> Musa: Complement Naive Bayes (tam feature set) + χ^2 tabanlı SelectKBest
Mesafe Tabanlı Modeller	KNeighborsClassifier	<ul style="list-style-type: none"> Benzer şikayet örneklerine göre tahmin yapan sezgisel yöntem Non-parametrik yaklaşım Lokal yapıları yakalama 	<ul style="list-style-type: none"> Musa: KNN (k=5, weights="distance") *Not: Hesaplama maliyeti nedeniyle eğitimde 50K, teste 20K örneklik stratified alt kümeler kullanılıyor*

Vize aşamasında geliştirilen bu baseline ve klasik makine öğrenmesi modelleri, final aşamasında kullanılan ileri seviye derin öğrenme modellerinin performansını değerlendirmek ve karşılaştırmak amacıyla referans (benchmark) olarak korunmuştur. Final aşamasında bu modeller üzerinde ek bir genişletme yapılmamış; elde edilen sonuçlar, ileri seviye modellerin sağladığı katkının daha net biçimde gözlemlenebilmesi için kullanılmıştır.

Hiper-Parametre Ayarlama

Vize Aşaması

Hiper-parametreler ağırlıklı olarak mantıklı başlangıç değerleriyle elle seçilmiştir:

- Decision Tree: max_depth, min_samples_leaf, min_samples_split
- Logistic Regression: C, max_iter, class_weight
- KNN: n_neighbors, weights
- LinearSVC: C, class_weight

Final Aşaması

Final aşamasında, seçilen modeller üzerinde hiper-parametre ayarlama çalışmaları gerçekleştirılmıştır. Bu kapsamda, model performansını artırmak ve daha dengeli sonuçlar elde etmek amacıyla hiper-parametre aralıkları sistematik biçimde taramıştır.

- Seçilen modeller için Grid Search ve Random Search yaklaşımları kullanılmıştır.
- Değerlendirme sürecinde Stratified k-fold çapraz doğrulama tercih edilerek, sınıf dağılımının her katmanda korunması sağlanmıştır.
- Modeller, çapraz doğrulama sonucunda elde edilen ortalama ve standart sapma değerleri üzerinden karşılaştırılmıştır.

Sınıf Dengesizliği Stratejileri

Şu Anda Kullanılan Stratejiler

Class weight:

- Logistic Regression, Decision Tree, LinearSVC için `class_weight="balanced"` kullanımı
- Azınlık sınıfının (Yes) öğrenme sürecinde daha fazla ağırlık almasını sağlama

Eşik ayarı (Threshold Tuning):

- ROC eğrisi ve Precision–Recall eğrisi üzerinden pozitif sınıf için farklı karar eşiklerinin test edilmesi

Yeniden örneklem:

- Gerekirse SMOTE ve/veya random undersampling ile eğitim verisinde sınıfları kısmen dengeleme

8) İleri Seviye (Advanced / Deep Learning) Modelleme Yaklaşımları

Final aşamasında, vize döneminde geliştirilen yapısal özelliklere dayalı baseline ve klasik makine öğrenmesi modellerinin ötesine geçilerek, tüketici şikayet metinlerinden doğrudan öğrenme yapabilen ileri seviye derin öğrenme modelleri geliştirilmiştir. Bu yaklaşımın temel amacı, şikayet metinleri içerisinde yer alan anlamsal ve bağlamsal bilgiyi daha etkin biçimde kullanarak, özellikle azınlık sınıf olan *consumer_disputed?* = Yes örnekleri için daha dengeli ve güçlü performans elde etmektir.

Bu kapsamda, her grup üyesi en az iki farklı derin öğrenme mimarisi üzerinde çalışmış ve elde edilen modeller, ortak değerlendirme metrikleri kullanılarak karşılaştırılmıştır.

Metin Verisinin Hazırlanması ve Temsili

İleri seviye modellerde temel girdi olarak, veri setinde yer alan *consumer_complaint_narrative* alanı kullanılmıştır. Bu alan, tüketicilerin şikayetlerini serbest metin biçiminde ifade ettikleri bölümü içermektedir.

Modelleme öncesinde metin verisi aşağıdaki adımlardan geçirilmiştir:

- Metinlerin küçük harfe dönüştürülmesi
- Noktalama işaretleri ve özel karakterlerin temizlenmesi
- Tokenization işlemi
- Maksimum metin uzunluğuna göre padding ve truncation uygulanması
- Token'ların sayısal vektör temsillerine dönüştürülmesi (embedding)

Bu işlemler sonucunda elde edilen sabit boyutlu metin temsilleri, tüm derin öğrenme modelleri için ortak girdi yapısı olarak kullanılmıştır.

Convolutional Neural Network (CNN) Tabanlı Modeller

CNN tabanlı modeller, metin sınıflandırma problemlerinde yerel örüntüleri (n-gram benzeri yapılar) yakalama konusunda etkili olmaları nedeniyle tercih edilmiştir. Bu projede CNN mimarileri, şikayet metinlerinde belirli kelime gruplarının ve ifadelerin *dispute* davranışıyla olan ilişkisini öğrenmek amacıyla kullanılmıştır.

Genel CNN mimarisi şu bileşenlerden oluşmaktadır:

- Embedding katmanı
- Bir veya birden fazla 1D Convolution katmanı
- ReLU aktivasyon fonksiyonu
- Global Max Pooling katmanı
- Fully Connected (Dense) katman
- Sigmoid aktivasyon fonksiyonu (ikili sınıflandırma)

CNN tabanlı modeller, özellikle kısa ve orta uzunluktaki şikayet metinlerinde hızlı öğrenme ve stabil performans sunmuş, baseline modellere kıyasla daha yüksek Recall ve F1-score değerleri elde edilmesine katkı sağlamıştır.

LSTM ve Attention Mekanizmalı BiLSTM Modelleri

Metin içerisindeki sıralı ve bağılamsal bağımlılıkları daha etkin biçimde modellemek amacıyla Uzun Kısa Süreli Bellek (LSTM) tabanlı mimariler kullanılmıştır. Bu çalışmada, hem tek yönlü LSTM hem de çift yönlü (BiLSTM) yapılar değerlendirilmiştir.

BiLSTM modellerine ek olarak Attention mekanizması kullanılarak, modelin şikayet metninin tamamı yerine sınıflandırma açısından daha kritik bölgülerine odaklanması sağlanmıştır. Bu sayede modelin hem performansı artırılmış hem de karar verme süreci daha yorumlanabilir hale getirilmiştir.

Attention BiLSTM mimarisi genel olarak aşağıdaki bileşenlerden oluşmaktadır:

- Embedding katmanı
- BiLSTM katmanı
- Attention katmanı
- Fully Connected katman
- Sigmoid çıkış katmanı

Bu modeller, özellikle uzun şikayet metinlerinde CNN tabanlı modellere kıyasla daha dengeli bir Precision–Recall ilişkisi sunmuş ve pozitif sınıf (Yes) için daha tutarlı sonuçlar üretmiştir.

Çok Katmanlı Algılayıcı (MLP) Modelleri

Bazı grup üyeleri tarafından, embedding vektörlerinin doğrudan giriş olarak kullanıldığı Çok Katmanlı Algılayıcı (Multi-Layer Perceptron – MLP) modelleri geliştirilmiştir. Bu modeller, doğrusal olmayan karar sınırlarını öğrenebilme yetenekleri sayesinde, derin öğrenme yaklaşımı için karşılaştırma niteliğinde bir referans oluşturmuştur.

MLP modelleri genel olarak:

- Birden fazla Dense katman
- ReLU aktivasyon fonksiyonları
- Dropout ile regularization
- Sigmoid çıkış katmanı

şeklinde yapılandırılmıştır. MLP modelleri, eğitim süresi açısından avantaj sağlamakla birlikte, sıralı bağımlılıkları doğrudan modelleyememeleri nedeniyle LSTM tabanlı modellere kıyasla sınırlı performans göstermiştir.

9) Model Performans Analizi

Bu bölümde, vize aşamasında geliştirilen baseline ve klasik makine öğrenmesi modelleri ile final aşamasında geliştirilen ileri seviye (deep learning) modellerin performansları karşılaştırımalı olarak değerlendirilmiştir. Değerlendirme sürecinde, veri setinin belirgin sınıf dengesizliği dikkate alınarak yalnızca doğruluk (accuracy) metriğine odaklanılmamış; özellikle pozitif sınıf (*consumer_disputed? = Yes*) için anlamlı sonuçlar üreten metrikler önceliklendirilmiştir.

Kullanılan Değerlendirme Metrikleri

Model performanslarının değerlendirilmesinde aşağıdaki metrikler kullanılmıştır:

- **Accuracy:** Modelin tüm sınıflar üzerindeki genel doğruluk oranını göstermektedir. Ancak sınıf dengesizliği nedeniyle tek başına yeterli bir ölçüt değildir.
- **Precision (Yes):** Yes sınıfı için yapılan tahminlerin ne kadarının doğru olduğunu göstermektedir.
- **Recall (Yes):** Gerçek Yes örneklerinin ne kadarının doğru şekilde tespit edildiğini ölçmektedir. Bu çalışma kapsamında kritik bir metriktir.
- **F1-score (Yes):** Precision ve Recall metriklerinin harmonik ortalamasıdır ve dengesiz veri setlerinde daha dengeli bir performans ölçütü sunmaktadır.
- **ROC AUC:** Modelin pozitif ve negatif sınıfları ayırt etme kabiliyetini, farklı karar eşikleri boyunca değerlendirmektedir.

Bu metrikler, özellikle azınlık sınıf olan Yes için model başarımını daha doğru biçimde yansımaktadır.

Genel Model Performanslarının Karşılaştırılması

Tabloda, tüm baseline, klasik makine öğrenmesi ve ileri seviye derin öğrenme modellerinin test seti üzerindeki genel performans sonuçları sunulmuştur.

Baseline olarak kullanılan **DummyClassifier**, yüksek accuracy değerine rağmen pozitif sınıf tamamen göz ardı etmiş; Recall(Yes) ve F1(Yes) değerlerinin sıfır olması, bu modelin pratikte kullanılabilir olmadığını açıkça göstermiştir.

Klasik makine öğrenmesi modelleri arasında:

- **Logistic Regression (balanced)** ve **Decision Tree (balanced)** modelleri, pozitif sınıf için görece yüksek Recall değerleri üretmiş; ancak düşük Precision nedeniyle F1-score değerleri sınırlı kalmıştır.
- Bu modeller, azınlık sınıfı yakalama konusunda baseline modellere kıyasla anlamlı bir iyileşme sağlamıştır.

İleri seviye derin öğrenme modellerinde ise:

- **CNN, TextCNN ve Attention BiLSTM** tabanlı modeller, ROC AUC değerleri açısından klasik modellere yakın veya daha iyi sonuçlar üretmiştir.
- Buna karşın bazı modellerin yüksek accuracy değerlerine rağmen düşük Recall(Yes) üretmesi, sınıf dengesizliğinin derin öğrenme modelleri üzerinde de belirgin bir etkisi olduğunu göstermektedir.
- Özellikle Attention mekanizması içeren BiLSTM modelleri, ROC AUC ve F1-score metrikleri açısından daha dengeli bir performans sergilemiştir.

Model	Kişi	Accuracy	Precision (Yes)	Recall (Yes)	F1 (Yes)	ROC AUC
DummyClassifier (most_frequent)	Barış	0.7983	0.0000	0.0000	0.0000	0.5000
Logistic Regression (balanced)	Barış	0.5140	0.2480	0.6934	0.3653	0.6092
Decision Tree (balanced)	Barış	0.4703	0.2414	0.7587	0.3662	0.5992
MLP	Barış	0.7983	0.0000	0.0000	0.0000	0.6105
CNN1D (1 epoch log)	Barış	0.7983	—	—	0.0000	0.6050
TextCNN (Model-1)	Musa	0.7688	0.5781	0.0317	0.0602	0.6086
AttnBiLSTM (Model-2)	Musa	0.7670	0.5556	0.0043	0.0085	0.6084
TextCNN (DL-1)	Salih	0.7668	0.6000	0.0103	0.0202	0.6055
AttnBiLSTM (DL-2)	Salih	0.5330	0.2430	0.6216	0.3494	0.5939
LogReg (baseline)	Yusuf	0.5330	0.2430	0.6216	0.3494	0.5939
Torch MLP (Weighted BCE)	Yusuf	0.5577	0.2419	0.5591	0.3377	0.5841
Torch MLP2 (Focal Loss)	Yusuf	0.7983	0.0000	0.0000	0.0000	0.5926

Sınıf Bazlı Performans Analizi

Tablolarda, modellerin **No** ve **Yes** sınıfları için ayrı ayrı Precision, Recall ve F1-score değerleri sunulmuştur. Bu analiz, modellerin hangi sınıflarda güçlü veya zayıf kaldığını daha net biçimde ortaya koymaktadır.

Analiz sonucunda:

- Çoğu modelin **No** sınıfında yüksek Precision ve Recall değerleri elde ettiği,
- Buna karşın **Yes** sınıfında performansın daha sınırlı kaldığı,
- Class weight ve özel kayıp fonksiyonları (örneğin Focal Loss) kullanılan modellerin, **Yes** sınıfını yakalama konusunda göreli olarak daha dengeli sonuçlar ürettiği

gözlemlenmiştir.

Bu bulgular, problem bağlamında sınıf dengesizliğinin model başarımı üzerindeki belirleyici etkisini doğrulamaktadır.

Barış – Baseline/Klasik

Model	Sınıf	Precision	Recall	F1	Support
DummyClassifier	0	0.7983	1.0000	0.8878	88765
DummyClassifier	1	0.0000	0.0000	0.0000	22427
LogReg (balanced)	0	0.8581	0.4687	0.6063	88765
LogReg (balanced)	1	0.2480	0.6934	0.3653	22427
Decision Tree (balanced)	0	0.8670	0.3975	0.5451	88765
Decision Tree (balanced)	1	0.2414	0.7587	0.3662	22427

Musa – Deep Learning

Model	Sınıf	Precision	Recall	F1	Support
TextCNN (Model-1)	0	0.7713	0.9930	0.8682	3834
TextCNN (Model-1)	1	0.5781	0.0317	0.0602	1166
BiLSTM-Attn (Model-2)	0	0.7674	0.9990	0.8680	3834
BiLSTM-Attn (Model-2)	1	0.5556	0.0043	0.0085	1166

Salih – Deep Learning

Model	Sınıf	Precision	Recall	F1	Support
TextCNN (DL-1)	0	0.7675	0.9979	0.8677	3830
TextCNN (DL-1)	1	0.6000	0.0103	0.0202	1170
BiLSTM-Attn (DL-2)	0	0.7678	0.9963	0.8673	3830
BiLSTM-Attn (DL-2)	1	0.5333	0.0137	0.0267	1170

Yusuf – Torch Modeller

Model	Sınıf	Precision	Recall	F1	Support
LogReg (baseline)	0	0.8423	0.5107	0.6358	23949
LogReg (baseline)	1	0.2430	0.6216	0.3494	6051
Torch MLP (Weighted BCE)	0	0.8334	0.5574	0.6680	23949
Torch MLP (Weighted BCE)	1	0.2419	0.5591	0.3377	6051
Torch MLP2 (Focal Loss)	0	0.7983	1.0000	0.8878	23949
Torch MLP2 (Focal Loss)	1	0.0000	0.0000	0.0000	6051

ROC Eğrileri ve Ayırt Edicilik Analizi

Seçilen modeller için çizilen ROC eğrileri, farklı karar eşikleri altında modellerin pozitif ve negatif sınıfları ayırt etme kabiliyetini göstermektedir. ROC AUC değerleri incelemişinde:

- Baseline modellerin ayırt edicilik gücünün sınırlı olduğu,
- Klasik makine öğrenmesi modellerinin ROC AUC ≈ 0.60 seviyelerinde performans gösterdiği,
- İleri seviye derin öğrenme modellerinin ise benzer veya daha yüksek ROC AUC değerlerine ulaşabildiği

görmüştür.

Bu sonuçlar, derin öğrenme modellerinin özellikle metin tabanlı bilgiyi kullanarak daha güçlü ayırt edicilik sağladığını göstermektedir.

10) En İyi Modelin Seçimi ve Sonuçların Yorumlanması

Bu çalışmada geliştirilen modellerin karşılaştırılması sürecinde temel amaç, yalnızca yüksek doğruluk (accuracy) elde etmek değil; aynı zamanda veri setinin belirgin sınıf dengesizliği altında azınlık sınıf olan *consumer_disputed? = Yes* örneklerini anlamlı ve tutarlı biçimde tespit edebilen bir model belirlemektir. Bu nedenle model seçimi sürecinde değerlendirme kriterleri, klasik doğruluk odaklı yaklaşımların ötesine taşınmıştır.

Model Seçim Kriterleri

Model performansları değerlendirilirken aşağıdaki kriterler esas alınmıştır:

- Recall (Yes): Gerçek itiraz edilen şikayetlerin ne kadarının doğru şekilde tespit edilebildiğini göstermesi nedeniyle kritik öneme sahiptir.
- Precision (Yes): İtiraz tahmini yapılan örneklerin ne kadarının gerçekten itiraz olduğu bilgisini sunarak, yanlış alarm oranını kontrol etmeye yardımcı olur.
- F1-score (Yes): Precision ve Recall arasında denge sağlayarak, dengesiz veri setlerinde daha güvenilir bir performans ölçütü sunar.
- ROC AUC: Modelin farklı karar eşikleri boyunca pozitif ve negatif sınıfları ayırt edebilme yeteneğini ölçerek, genel ayırt edicilik gücünü ortaya koyar.

Bu metrikler birlikte ele alınarak, yalnızca tek bir ölçüte bağlı kalmadan bütüncül bir değerlendirme yapılmıştır.

Baseline ve Klasik Modellerin Değerlendirilmesi

Baseline olarak kullanılan DummyClassifier, yüksek accuracy değerine rağmen pozitif sınıfı tamamen göz ardı etmiş; Recall(Yes) ve F1-score değerlerinin sıfır olması, bu yaklaşımın problem bağlamında anlamsız olduğunu açıkça ortaya koymuştur. Bu durum, yalnızca accuracy metriğine dayalı değerlendirmelerin dengesiz veri setlerinde ne derece yaniltıcı olabileceğini göstermektedir.

Class weight kullanılarak eğitilen Logistic Regression ve Decision Tree modelleri, pozitif sınıfı yakalama konusunda baseline modellere kıyasla belirgin bir iyileşme sağlamıştır. Özellikle Recall(Yes) değerlerinde gözlemlenen artış, bu modellerin azınlık sınıfı öğrenme sürecine daha fazla dahil edebildiğini göstermektedir. Ancak bu kazanım, düşük Precision değerleri ile birlikte değerlendirmiştir; modelin çok sayıda yanlış pozitif tahmin ürettiği gözlemlenmiştir. Bu durum, iş açısından gereksiz alarm üretme riskini beraberinde getirmektedir.

Dolayısıyla klasik modeller, azınlık sınıfı yakalama konusunda önemli bir referans noktası oluştursa da, tek başına nihai çözüm olarak yeterli görülmemiştir.

İleri Seviye Derin Öğrenme Modellerinin Değerlendirilmesi

Geliştirilen CNN, TextCNN, BiLSTM ve Attention BiLSTM tabanlı derin öğrenme modelleri, şikayet metinlerinden doğrudan anlamsal ve bağlamsal bilgi öğrenebilme avantajına sahiptir. Bu modeller, yapısal özelliklere dayalı yaklaşımların ötesine geçerek, metnin içeriğindeki gizli örüntülerini yakalayabilmektedir.

Performans sonuçları incelendiğinde, bazı derin öğrenme modellerinin yüksek accuracy değerlerine rağmen düşük Recall(Yes) ürettiği görülmüştür. Bu durum, sınıf dengesizliğinin derin öğrenme modelleri üzerinde de önemli bir zorluk oluşturduğunu göstermektedir. Ancak ROC AUC değerleri dikkate alındığında, bu modellerin genel ayırt edicilik kabiliyetinin klasik modellere kıyasla daha güçlü olduğu anlaşılmaktadır.

Özellikle Attention mekanizması içeren BiLSTM modelleri, metin içerisindeki sınıflandırma açısından kritik böülümlere odaklanabilme yetenekleri sayesinde daha dengeli Precision–Recall profilleri sergilemiştir. Attention yapısı, modelin yalnızca kelime sıralamasını değil, hangi kelime ve ifadelerin karar sürecinde daha etkili olduğunu öğrenmesini sağlamıştır. Bu durum, modelin hem performans hem de yorumlanabilirlik açısından avantaj kazanmasına katkıda bulunmuştur.

En Uygun Modelin Belirlenmesi

Tüm değerlendirmeler birlikte ele alındığında, **Attention BiLSTM tabanlı derin öğrenme modeli**, problem bağlamında en uygun model olarak belirlenmiştir. Bu model:

- Pozitif sınıf için klasik modellere kıyasla daha dengeli Recall–Precision ilişkisi sunmuş,
- ROC AUC metriği açısından daha güçlü bir ayırt edicilik sergilemiş,
- Şikâyet metinlerinden bağlamaşal bilgi öğrenerek daha anlamlı karar sınırları oluşturabilmiştir.

İş problemi açısından değerlendirildiğinde, bu modelin sunduğu performans, müşteri şikayetlerinin erken aşamada riskli (itiraza dönüşme potansiyeli yüksek) olarak işaretlenmesini mümkün kılmaktadır. Bu sayede firmalar, yüksek riskli şikayetlere öncelik vererek müşteri memnuniyetini artırabilecek ve olası itiraz süreçlerini daha etkin biçimde yönetebilecektir.

Ensemble Yaklaşımların Değerlendirilmesi

Bu çalışmada, single modellerin performanslarını artırma potansiyeli bulunan ensemble yaklaşımlar da değerlendirilmiştir. Ensemble yöntemler, farklı model mimarilerinin güçlü yönlerini birleştirerek daha kararlı ve genellenebilir tahminler üretmeyi amaçlamaktadır. Özellikle sınıf dengesizliği bulunan problemler için, farklı karar sınırlarına sahip modellerin birlikte kullanılması performans açısından avantaj sağlayabilmektedir.

Değerlendirilen Ensemble Stratejileri

Çalışma kapsamında aşağıdaki ensemble stratejileri teorik ve deneyel olarak ele alınmıştır:

- Soft Voting:
Logistic Regression, Decision Tree ve LinearSVC gibi klasik makine öğrenmesi modellerinin olasılık çıktılarının ortalaması alınarak nihai tahmin üretilmesi.
- Hard Voting:
Birden fazla modelin sınıf tahminleri üzerinden çoğunluk oylaması yapılması.
- Hibrit Ensemble (ML + Deep Learning):
Klasik makine öğrenmesi modelleri ile metin tabanlı derin öğrenme modellerinin (örneğin TextCNN ve Attention BiLSTM) birlikte değerlendirilmesi.

Bu yaklaşımlar, modellerin birbirinden farklı hata profillerine sahip olması durumunda daha dengeli sonuçlar üretme potansiyeline sahiptir.

Ensemble Yaklaşımların Sınıf Dengesizliği Bağlamında Analizi

Ensemble yöntemler, teorik olarak azınlık sınıfının daha iyi yakalanmasına katkı sağlayabilmektedir. Ancak bu çalışmada yapılan değerlendirmelerde, ensemble yapılarının sınıf dengesizliği problemi karşısında tekil en iyi derin öğrenme modeline kıyasla sınırlı bir ek katkı sunduğu gözlemlenmiştir.

Bunun temel nedenleri şu şekilde özetlenebilir:

- Derin öğrenme modellerinin çoğunluk sınıf tahminine yönelik eğiliminin ensemble içinde de baskın kalması
- Olasılık çıktılarının (probability calibration) farklı model aileleri arasında yeterince uyumlu olmaması
- Ensemble yapılarının, pozitif sınıf (Yes) için Recall artışı sağlarken Precision değerini daha da düşürebilmesi

Bu durum, ensemble yöntemlerin her problem için otomatik olarak daha iyi sonuçlar üretmediğini; problem bağlamı, veri yapısı ve sınıf dağılımına bağlı olarak dikkatli biçimde değerlendirilmesi gerektiğini göstermektedir.