

BÖLÜM I

Hasta Yorumlarını Kullanarak Duygu Analizi

İpek AKPINAR¹
Ali Mete ÇIPLAK²
Ediz ENGİN³
Enes ÇAKMAK⁴
Mahmutcan SAKINCI⁵
Mustafa TETİK⁶
Nail KOCABAY⁷
Ozan DESTE⁸

1. Giriş

Dijitalleşmenin hızla arttığı günümüzde, sağlık hizmetleri yalnızca fiziksel tedavi süreçlerinden ibaret olmaktan çıkmış ve hasta memnuniyeti, hizmet kalitesi ve hasta deneyimi gibi unsurlar sağlık sisteminin değerlendirilmesinde kritik ölçütler haline gelmiştir. Özellikle sağlık kuruluşları ve çalışanları hakkında yapılan çevrimiçi yorumlar bu sistemin güçlü ve zayıf yönlerini ortaya koymak açısından büyük bir veri kaynağı sunmaktadır. Bu bağlamda hasta yorumlarının analizi yalnızca bireysel deneyimlerin değerlendirilmesiyle kalmamakta, aynı zamanda sağlık sistemlerinin iyileştirilmesi adına önemli yapısal verilere ulaşılmasını da sağlamaktadır.

¹ Marmara Üniversitesi, Teknoloji Fakültesi, Bilgisayar Mühendisliği, İstanbul/Türkiye, Orcid: 0009-0009-4474-1189, ipekakpinar@marun.edu.tr

² Marmara Üniversitesi, Teknoloji Fakültesi, Bilgisayar Mühendisliği, İstanbul/Türkiye, Orcid: 0009-0007-9324-3157, ali.mete@marun.edu.tr

³ Marmara Üniversitesi, Teknoloji Fakültesi, Bilgisayar Mühendisliği, İstanbul/Türkiye, Orcid: 0009-0006-2075-0222, edizengin@marun.edu.tr

⁴ Marmara Üniversitesi, Teknoloji Fakültesi, Bilgisayar Mühendisliği, İstanbul/Türkiye, Orcid: 0009-0001-2998-5320, enescakmak@marun.edu.tr

⁵ Marmara Üniversitesi, Teknoloji Fakültesi, Bilgisayar Mühendisliği, İstanbul/Türkiye, Orcid: 0009-0004-5289-3258, mahmutcansakinci@marun.edu.tr

⁶ Marmara Üniversitesi, Teknoloji Fakültesi, Bilgisayar Mühendisliği, İstanbul/Türkiye, Orcid: 0009-0003-8638-9730, mustafatetik@marun.edu.tr

⁷ Marmara Üniversitesi, Teknoloji Fakültesi, Bilgisayar Mühendisliği, İstanbul/Türkiye, Orcid: 0009-0001-1677-3206, nailkocabay@marun.edu.tr

⁸ Marmara Üniversitesi, Teknoloji Fakültesi, Bilgisayar Mühendisliği, İstanbul/Türkiye, Orcid: 0009-0002-1473-346X, ozandeste@marun.edu.tr

Bu çalışmada çevrimiçi platformlarda yer alan hasta yorumları incelenerek sağlık kuruluşlarının ve çalışanlarının hizmet kalitelerinin veri madenciliği teknikleri ile değerlendirilmesi amaçlanmaktadır. Özellikle tekrar eden şikayet, öneri ve olumlu geri bildirimlerin belirlenmesiyle birlikte hizmet sunucularının gelişime açık yönleri daha görünür hale getirilmektedir. Projenin temel çıktısı, her yorumu pozitif, nötr veya negatif olarak sınıflandıran bir sistem oluşturarak sağlık kurumlarının stratejik planlamalarında kullanılabilir somut sonuçlar üretmektir. Bu sistem, hasta ile sağlık kurumu arasındaki dolaylı iletişimi analiz edilebilir hale getirerek hizmet kalitesine yönelik doğrudan ve anlamlı bir geri bildirim mekanizması sunmaktadır.

Bu amaç doğrultusunda, veri madenciliği ve doğal dil işleme (NLP) tekniklerinden yararlanılmıştır. Araştırmanın ilk sürecinde K-En Yakın Komşu (KNN) algoritması, Destek Vektör Makineleri (SVM) ve çift yönlü uzun kısa süreli bellek (bi-LSTM) modelleri kullanılmış ve çeşitli sonuçlar elde edilmiştir. Son aşamasında ise bu yöntemlerin yanı sıra Evrimsel Sinir Ağı (CNN), BERT ve türevi derin öğrenme modelleri ile boosting algoritmaları (XGBoost, CatBoost gibi) entegre edilerek çok yönlü bir değerlendirme gerçekleştirilmiştir. Yöntem seçimleri, hem literatürdeki benzer çalışmalardan yararlanılarak hem de proje kapsamında test edilen model performanslarına göre belirlenmiştir.

Literatür taramaları, hasta deneyimlerine odaklanan duygu analizlerinde özellikle son yıllarda Transformer tabanlı modellerin öne çıktığını göstermektedir. BERT, XLNet ve RoBERTa gibi modellerin yorum sınıflandırmasında üstün başarı sağladığı ve geleneksel yöntemlerin ise belirli veri türlerinde hâlâ geçerliliğini koruduğu görülmüştür. Bu kapsamda proje boyunca incelenen makaleler, kullanılan veri setleri, yöntem tercihleri ve deney sonuçları bir bütün olarak değerlendirilmiş; sağlık alanında veri odaklı çözüm önerileri geliştirilmiştir.

Çalışmanın kalanı ise aşağıdaki şekilde düzenlenmiştir. 2. bölümde ilgili çalışmalar ve literatür özetlerine yer verilmiştir. 3. bölümde projede kullanılan veri setleri ve yöntemler detaylandırılmıştır. 4. bölümde deneysel çalışmaların sonuçları karşılaştırmalı biçimde

sunulmuştur. Son bölümde ise genel değerlendirme yapılmış ve gelecekte yapılabilir çalışmalara dair öneriler paylaşılmıştır.

2. İlgili Çalışmalar

2.1 Hasta Geri Bildirimleri Üzerine Otomatik Metin Analizi

Alexander ve arkadaşları(Alexander, Bahja ve Butt, 2022), NHS hasta geri bildirim verilerini analiz ederek sağlık hizmet kalitesi üzerine içgörüler sunan bir sistem geliştirmiştir. Çalışmada öncelikle metinler ön işleme tabi tutulmuş, ardından Latent Dirichlet Allocation (LDA) ile konu modelleme yapılmıştır. Ayrıca metinler Naive Bayes sınıflandırıcısı ile duygu kategorilerine ayrılmıştır.

Elde edilen sonuçlarda, olumsuz yorumların büyük çoğunluğunun randevu sistemine ve personel ilgisine yönelik olduğu gözlemlenmiştir. Sağlık sisteminde sıkça tekrar eden sorunları tespit edebilmek için makine öğrenmesi tabanlı metin analizi yöntemlerinin etkili olduğu belirtilmiştir.

Yazar	Yıl	Yöntem	Veri Seti	Metrikler	Sonuç
Alexander et al.	2022	LDA, Naive Bayes	NHS Free Text Dataset	F1	Negatif yorum oranı en yüksek; randevu/personel şikayetleri önde

2.2 Sağlık Platformlarında Duygu ve Değerlendirme Analizi

Bu çalışmada(Tang ve diğerleri, 2023), Healthgrades adlı sağlık değerlendirme platformundaki metin yorumları incelenmiştir. Yorumlara VADER duygu analiz algoritması uygulanmış ve yıldızlı değerlendirme puanları ile karşılaştırılmıştır. Ayrıca kelime frekans analizi yapılarak kullanıcıların öne çıkardığı temalar belirlenmiştir.

Çalışmanın dikkat çekici sonucu, metinlerdeki duygu skorları ile yıldız puanlamalarının yüksek korelasyon göstermesidir. Yorumların içeriği, yıldız puanlarından bağımsız olarak hizmet kalitesini daha ayrıntılı yansıtabilmektedir.

<i>Yazar</i>	<i>Yıl</i>	<i>Yöntem</i>	<i>Veri Seti</i>	<i>Metrikler</i>	<i>Sonuç</i>
Tang et al.	2023	VADER, Linear Regression	Healthgrades	VADER score, frekans	Yorum duygu skoru yıldızla ilişkili, frekans analizi temaları belirledi

2.3 BERT Tabanlı Çin Sağlık Platformu Analizi

Zhao ve ekibi (Zhao, Li, Yuan ve Deng, 2024), Çin merkezli haodf.com platformundaki hasta yorumlarını sınıflandırmak için BERT tabanlı bir sistem geliştirmiştir. BERT modeline ek olarak CNN ve Bi-GRU yapıları da kullanılarak performans karşılaştırması yapılmıştır.

Deneyler sonucunda, BERT modelinin %98,36 doğruluk oranı ile en iyi sonuç verdiği görülmüştür. Ayrıca F2 skoru da dikkate alınarak olumlu, olumsuz ve nötr sınıflandırmasında BERT'in diğer modellere üstün geldiği kanıtlanmıştır.

<i>Yazar</i>	<i>Yıl</i>	<i>Yöntem</i>	<i>Veri Seti</i>	<i>Metrikler</i>	<i>Sonuç</i>
Zhao et al.	2024	BERT, CNN, Bi-GRU	haodf.com	Accuracy, F2	BERT %98.36 doğruluk ile üstün performans sağladı

2.4 Duygu Sınıflandırmasında Derin Öğrenme Modelleri

Suresh ve arkadaşları(Sagarika Suresh Thimmanayakanapalya, Pavankumar Mulgund, ve Raj Sharman, 2022), hasta yorumlarını duygu kategorilerine ayırmak amacıyla Bi-LSTM, LSTM ve BERT modellerini karşılaştırmıştır. Harvard n2c2 veriseti ve RateMDs platformu kullanılmıştır.

Deneyler sonucunda BERT modeli %99,7 gibi oldukça yüksek bir doğruluk oranına ulaşarak en başarılı model olmuştur. Bu sonuç, transformer tabanlı modellerin geleneksel RNN tabanlı modellere göre anlamlı bir üstünlük sağladığını göstermektedir.

<i>Yazar</i>	<i>Yıl</i>	<i>Yöntem</i>	<i>Veri Seti</i>	<i>Metrikler</i>	<i>Sonuç</i>
--------------	------------	---------------	------------------	------------------	--------------

Suresh et al.	2022	BERT, Bi-LSTM, LSTM	Harvard n2c2, RateMD	Accuracy	BERT %99.7 doğruluk ile en iyi sonucu verdi
---------------	------	---------------------	----------------------	----------	---

2.5 Portekizce Hasta Yorumlarında BERTopic Kullanımı

Bu çalışmada(Osório ve Fachada, 2024), Google Review üzerinden toplanan Portekizce sağlık yorumları analiz edilmiştir. LDA, LSA ve BERTopic yöntemleri karşılaştırılarak en uyumlu konu modelleme algoritması araştırılmıştır.

Sonuçlara göre BERTopic modeli, hem konu tutarlılığı hem de duygu-konu uyumu açısından en yüksek performansı göstermiştir. Bu modelin özellikle küçük ve çok çeşitli metinlerde avantaj sağladığı belirtilmiştir.

Yazar	Yıl	Yöntem	Veri Seti	Metrikler	Sonuç
Osorio & Fachada	2024	BERTopic, LDA, LSA	Google Reviews (PT)	Accuracy, uyum skoru	BERTopic konu-uyum bakımından en güçlü sonuçları verdi

2.6 Macar Hasta Deneyimlerinin Negatifliğinin Ölçülmesi

Macaristan'dan derlenen hasta yorumları, huBERT, HIL-SBERT ve HDBSCAN yöntemleri ile analiz edilmiştir. Bu modeller kullanılarak, yorumların tematik kümeleme ve duygu sınıflandırması yapılmıştır.

Çalışma(Osváth, Yang ve Kósa, 2023), özellikle devlet hastanelerine yönelik yorumların %94,4 oranında negatif duygu içerdiğini ortaya koymuştur. Bu durum, sağlık politikaları açısından dikkate alınması gereken çarpıcı bir geribildirim sunmaktadır.

Yazar	Yıl	Yöntem	Veri Seti	Metrikler	Sonuç
Osváth et al.	2023	huBERT, HIL-SBERT, HDBSCAN	praxis.blog.hu, yorumlar	Sentiment Accuracy	%94.4 oranında hasta deneyimi negatif bulundu

2.7 Hasta Deneyimi Yorumlarında BERTopic Kullanımı

Bu çalışmada(Steele ve diğerleri, 2025), hasta deneyim anketlerindeki yorumlar analiz edilmiştir. Anketler, Kanada’da hastanelerde hizmet alan yetişkin ve çocuk hastalardan toplanmış ve BERTopic ile konu modellemesi yapılmıştır. LDA gibi klasik yöntemler yerine, önceden eğitilmiş dil modeli tabanlı BERTopic tercih edilmiştir.

Sonuçlara göre BERTopic, az ayarla bile yorumlardan anlamlı başlıklar çıkarabilmiş ve mevcut anketlerde yer almayan yeni bakım temalarını tanımlamıştır. Yorumlar genel olarak olumlu bulunmuştur. Bu modelin, kişi merkezli bakım, hasta güvenliği ve kalite iyileştirme açısından katkı sunduğu belirtilmiştir.

<i>Yazar</i>	<i>Yıl</i>	<i>Yöntem</i>	<i>Veri Seti</i>	<i>Metrikler</i>	<i>Sonuç</i>
Steele et al.	2025	BERTopic, Duygu Analizi	Kanada hasta deneyim anketleri (2016–2020)	Sentiment Score, Comment Rate, Comment Lenght	BERTopic, yeni ve anlamlı hasta deneyimi temaları belirleyerek klasik anketleri tamamladı

2.8 GPT-4 Turbo ile Hasta Yorumlarının Çok Etiketli Sınıflandırması

Sakai ve ekibi(Sakai, Lam, Mikaeili, Bosire ve Jovin, 2025), HCAHPS anketinden elde edilen 1.089 hasta yorumunu çok etiketli olarak sınıflandırmak amacıyla GPT-4 Turbo modelini kullanmıştır. Sıfır örnekli (zero-shot) öğrenme yöntemiyle %76.12 örnek bazlı F1 skoru ve %73.61 ağırlıklı F1 skoru elde edilmiştir. Bu başarı oranları, geleneksel makine öğrenimi yöntemleri ve diğer dil modellerine kıyasla anlamlı düzeyde yüksektir. Ayrıca, yorumlardan kişisel sağlık bilgilerini ayıklamak üzere bir PHI tespit çerçevesi geliştirilmiştir.

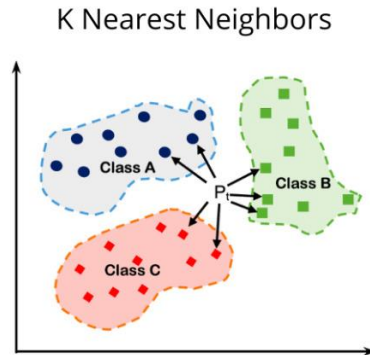
<i>Yazar</i>	<i>Yıl</i>	<i>Yöntem</i>	<i>Veri Seti</i>	<i>Metrikler</i>	<i>Sonuç</i>
--------------	------------	---------------	------------------	------------------	--------------

Sakai et al.	2024	GPT-4 Turbo (0-shot)	HCAHPS	F1 (76.12%), Weighted F1 (73.61%)	GPT-4 Turbo en yüksek sınıflandırma başarımını gösterdi
--------------	------	----------------------	--------	--------------------------------------	---

3. Yöntemler

3.1 Kullanılan Makine ve Derin Öğrenme Modelleri

3.1.1 KNN: KNN, sınıflandırma yaparken basit mantık ile hareket eder. Yeni bir veri geldiğinde, daha önce sınıfı bilinen verilere bakar ve ona en çok benzeyen K tanesini bulur. Bu benzer verilerin çoğunlukta olan sınıfı neyse, yeni veriye de o sınıfı atar Geeks For Geeks (2025) K-Nearest Neighbor(KNN) Algorithm,) (30.06.2025 tarihinde <https://www.geeksforgeeks.org/machine-learning/k-nearest-neighbours/> adresinden ulaşılmıştır).



Şekil 1. KNN temsili(Sachinsoni, 2023)K Nearest Neighbours — Introduction to Machine Learning Algorithms (01.07.2025 tarihinde <https://medium.com/@sachinsoni600517/k-nearest-neighbours-introduction-to-machine-learning-algorithms-9dbc9d9fb3b2> adresinden ulaşılmıştır).

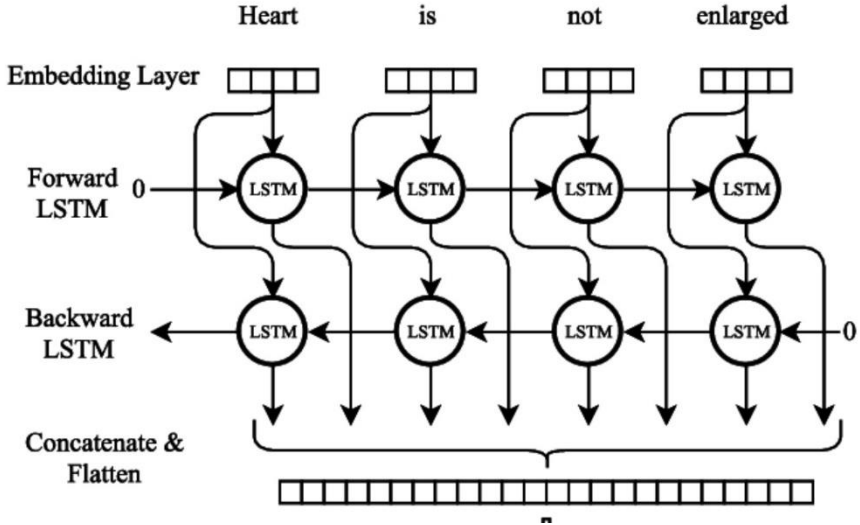
Veriler arasındaki benzerlik, genellikle iki nokta arasındaki uzaklık ölçülerek bulunur. Bu ölçüm için Öklid uzaklığı, cosine benzerliği veya Hamming uzaklığı gibi yöntemler tercih edilir.

Öklid (Euclidean) Uzaklığı, iki nokta arasındaki düz çizgi mesafesini hesaplar. Değerler ne kadar yakınsa, veriler o kadar benzer kabul edilir.

Cosine Benzerliği, vektörler arasındaki açısal benzerliği ölçer. Burada amaç, iki vektörün aynı yöne ne kadar baktığını anlamaktır. Aynı değerlere sahip olmasalar bile, benzer yönelimleri varsa yüksek benzerlik puanı elde edilir.

Hamming Uzaklığı, iki vektör arasında kaç elemanın farklı olduğunu sayar. Eğer veriler 0 ve 1 gibi ikili biçimlerdeyse, bu yöntem veriler arasındaki farkları basit ama etkili bir şekilde hesaplamak için uygundur

3.1.2 Bi-LSTM: Metin gibi verileri işlerken hem geçmiş hem de gelecek kelimeleri aynı anda göz önüne alan bir derin öğrenme modelidir. Normal LSTM modelleri sadece metni soldan sağa okurken, Bi-LSTM iki yönlü olarak çalışır. Bu sayede bir kelimenin tam anlamını çevresindeki tüm bağlamı kullanılarak daha iyi anlar(Anishnama, 2023) *Understanding Bidirectional LSTM for Sequential Data Processing* (30.06.2025 tarihinde <https://medium.com/@anishnama20/understanding-bidirectional-lstm-for-sequential-data-processing-b83d6283befe> adresinden ulaşılmıştır).

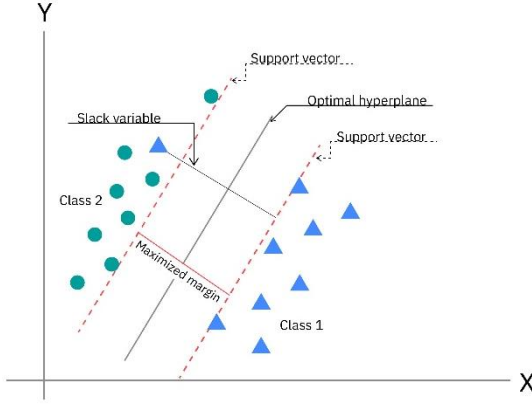


Şekil 2. Bi-LSTM şeması(Preserving Privacy in Arabic Judgments: AI-Powered Anonymization For Enhanced Legal Data Privacy”, t.y.)

Bi-LSTM özellikle duygu analizi gibi cümle yapısının önemli olduğu görevlerde başarılı sonuçlar verir. Bu çift yönlü yapı, modelin bağlamı daha derinlemesine kavramasını sağlar. Uzun ve karmaşık cümlelerde bile anlamı yakalayabilir.

3.1.3 SVM : Destek Vektör Makineleri (SVM) sınıflandırma, regresyon ve aykırı değerleri bulmak için kullanılan denetimli bir öğrenme tekniğidir. Bir düzlemdeki 2 veya daha fazla nokta kümelerini ayırmak için doğrular çizer. Veri kümeleri düşünüldüğünde bu doğrunun, iki kümenin noktaları için de maksimum uzaklıkta olması amaçlanır(Türkoğlu, 2021)

Support Vector Machine (SVM) Algoritması-Makine Öğrenmesi- (30.06.2025 tarihinde <https://mfatihto.medium.com/support-vector-machine-algoritması-makine-öğrenmesi-8020176898d8> adresinden ulaşılmıştır).



Şekil 3. SVM şeması(“What Is Support Vector Machine?”, 2023)

Bu yaklaşım, modelin yeni verileri daha doğru sınıflandırmasını sağlar ve öğrenilen sınırın genellenebilirliğini artırır.

3.1.4 CNN : Convolutional Neural Network (CNN), özellikle görüntü tanıma ve doğal dil işleme görevlerinde yaygın olarak kullanılan bir mimaridir. Özelliklerin çıkarımı için konvolüsyon filtreleri kullanır, bu sayede işlenmemiş veriden anlamlı sonuçlar üretebilir.

$$G[i, j] = \sum_{u=-k}^k \sum_{v=-k}^k H[u, v] F[i - u, j - v]$$

Şekil 4. Konvolüsyon Formülü(Awaits, 2020)

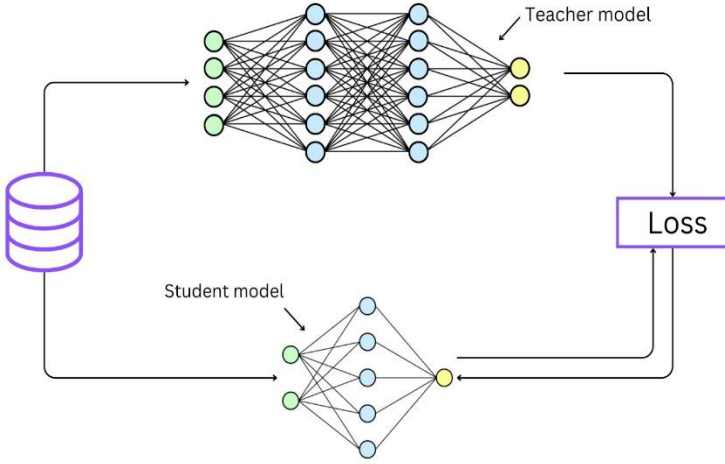
Konvolüsyon İşlemi Formülü, CNN'in temel yapı taşı olan konvolüsyon işlemini tanımlar. X, giriş görüntüsü veya özelliğini,

K ise filtreyi temsil eder. K filtresi X üzerinde kaydırılarak çarpım ve toplam işlemleri uygulanır. Bu sayede filtre, giriş verisindeki belirli bir özelliği tespit etmeye çalışır ve filtreler farklı özellikleri öğrenerek anlamlı temsiller oluşturur.

3.1.5 RoBERTa : RoBERTa, Facebook AI araştırmacıları tarafından geliştirilen, BERT modelinin gelişmiş bir versiyonudur. Doğal dil işleme alanında kullanılan güçlü bir derin öğrenme modelidir. Transformer mimarisi üzerine kurulu olan bu model, metinleri anlamak ve bağlama göre yorumlamak için kelimeler arasındaki ilişkileri öğrenir.

3.1.6 DistilBert : DistilBERT, BERT'in daha küçük, daha hızlı ve daha verimli bir versiyonudur. Hugging Face tarafından geliştirilen bu model, BERT'in doğruluk seviyesini büyük ölçüde korurken, %40 daha az parametre ve %60 daha hızlı çalışma gibi önemli performans özellikleriyle öne çıkar.

“knowledge distillation” yani bilgi damıtımı yöntemiyle eğitilmiştir. BERT ile aralarında öğretmen öğrenci ilişkisi vardır. Öğrenci model, öğretmenin çıktılarından öğrenerek benzer doğruluk performansına ulaşmaya çalışır.

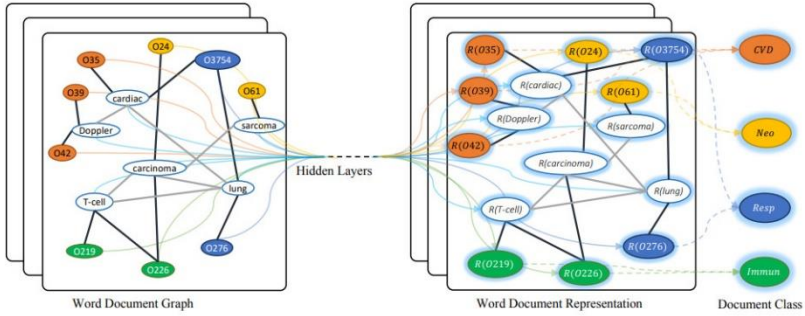


Şekil 5. Öğretmen-öğrenci modelinde bilgi aktarımı (“Distilbert: A Smaller, Faster, and Distilled BERT - Zilliz Learn”, t.y.)

3.1.7 textRNN : Text RNN, sıralı olan verilerle çalışmak için tasarlanmış bir yapay sinir ağı türüdür. Bir metindeki kelimeleri sırayla işler ve her kelimenin önceki kelimelerle olan bağlamını öğrenmeye çalışır. Bu sayede metin içinde anlam ilişkilerini kurabilir.

Geçmiş bilgileri belleğinde tutar ve bu belleği her adımda güncelleyerek sonraki kelime işlenirken kullanır ancak uzun cümlelerde eski bilgileri unutabilme gibi sorunları vardır.

3.1.8 textGCN : Metin sınıflandırma için geliştirilmiş graf tabanlı bir derin öğrenme yöntemidir. Metinleri sırayla değil, graf yapısı olarak ele alır; belgeler ve kelimeler düğüm, aralarındaki ilişkiler ise kenarlarla temsil edilir. Kelime eş-oluşumları ve belge-kelime bağlantıları bu graf üzerinden modellenir.

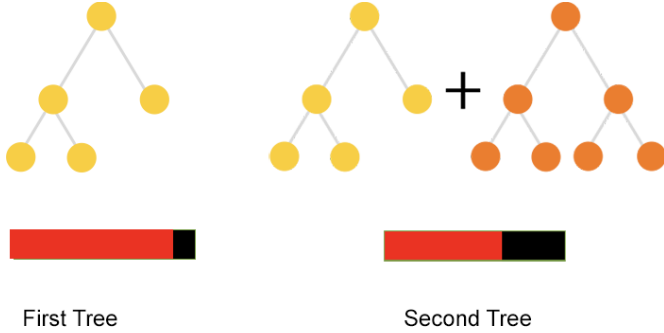


Şekil 6. Metin GCN'nin Şematiği (Yao, Mao ve Luo, 2018)

Model, Graph Convolutional Network (GCN) işlemleriyle her düğüm için bağlamsal temsiller üretir. Böylece sıralı yapıya ihtiyaç duymadan metin içi ilişkiler öğrenilir. Özellikle etiketli verinin az olduğu senaryolarda etkili sonuçlar verir.

3.1.9 XGBoost : Gradient Boosting olarak bilinen algoritmanın geliştirilmiş halidir. Karar ağaçlarının birbirlerinin hatalarını giderecek şekilde ard arda dizilmesi prensibine dayanır. Öncelikle bir karar ağacı eldeki veri ile eğitilir. Daha sonra bir sonraki karar ağacı ilk karar ağacının hataları ile eğitilir ve onu düzeltmeye çalışır. Üçüncü karar ağacı ikinci karar ağacının hatalarını düzeltmeye çalışır. Bu şekilde karar ağaçları eklenerek süreç devam eder. En sonunda bütün ağaçların sonuçları toplanır.

3.1.10 CatBoost : CatBoost, karar ağaçlarını ardı ardına kullanarak hataların minimize edildiği ve tahmin oranının iyileştirilmesini sağlayan bir gradient boosting tekniğidir.



Şekil 7. Birinci ve ikinci ağaçlar genelde yüksek hata içerir(“CatBoost—Open-source gradient boosting library”, t.y.)

Bu teknik çalıştırılırken önce karar ağacı oluşturulur sonrasında tahmin içerisinde ne kadarlık hata oranı var hesaplanır. Her bir oluşturulan karar ağacının amacı bir önceki karar ağacı tarafından meydana gelen hataların düzeltilip, stabil hâle getirilmesini sağlamaktır ve bu bir döngü halinde çalışmaya devam eder.

4. Deney Ortamı, Sonuçlar ve Değerlendirme

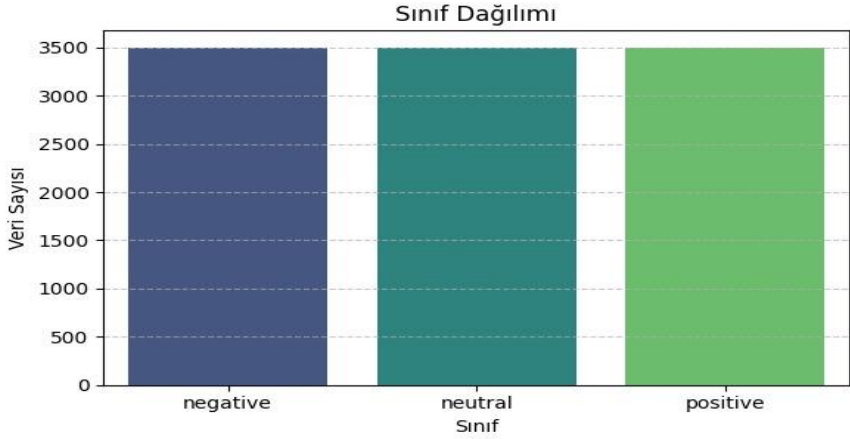
4.1 Veri Seti

4.1.1 NHS Verisi

National Health Service (NHS)(“NHS website developer portal”, t.y.), Birleşik Krallık’ta 1948 yılında kurulmuş olan kamuya ait sağlık hizmetleri sistemidir. NHS resmî sitesinde hastaların yorumları ve verdikleri puanların (Rating) verilerini çekebilmemizi sağlayan bir api hizmeti mevcuttur [kaynak1]. Rating etiketinin ortalamaları alınarak 0 ile 5 arasında değerler elde edilmiştir. Bu değerler 0 – 2.5 arası negatif, 2.5 – 3.5 arası nötr ve 3.5 – 5 arası pozitif olarak etiketlenmiş ve 3 sınıf elde edilmiştir. Bu şekilde veri

seti daha sade ve analiz edilebilir bir hale getirilmiş, sınıflandırma modelleri için uygun bir yapı oluşturulmuştur.

Toplam 10500 veri bulunun bu veri setinde her sınıfın veri sayısı eşitlenmiştir. Her sınıfta 3500 veri bulunmaktadır.



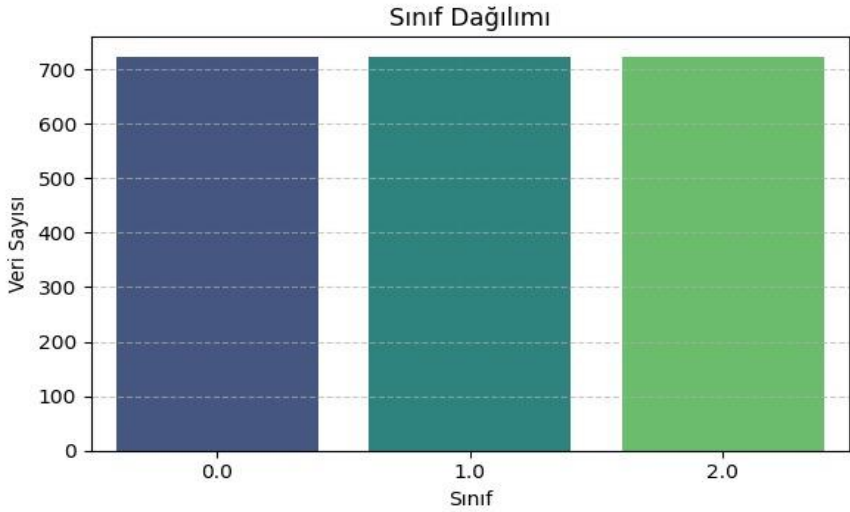
Şekil 8. NHS veri kümesi sınıf dağılımı

4.1.2 RateMD Verisi

Avijit Thawani'nin "Are Online Reviews of Physicians Biased Against Female Providers?" adlı makalesinde kullandığı verisetidir (Thawani, Paul, Sarkar ve Wallace, 2019). RateMDs.com sitesinden açık kaynaklı bir kod kullanılarak elde edilmiştir. Veri setinde doktorlar, hastalar tarafından dört farklı başlık altında 1 ile 5 arasında puanlanarak değerlendirilmiştir. Her başlık için ayrı ayrı puanlama yapılmış olup bu çoklu değerlendirmeler sadeleştirilerek tek bir sınıfa (label) dönüştürülmüştür. Dört başlıktan elde edilen puanlar bir araya getirilerek genel bir sınıflandırma yapılmış, bu sınıflandırma sürecinde 0 negatif, 1 nötr ve 2 pozitif olmak üzere üç farklı sınıf tanımlanmıştır. Etiketleme aşamasında, Yüksek puanlardan oluşan yorumlar pozitif sınıfa dahil edilmiştir. Aynı şekilde, düşük puanlardan oluşan yorumlar negatif, düşük ve

yüksek puanlardan oluşan çoklu değerlendirmeler ise nötr olarak etiketlenmiştir. Bu şekilde veri seti daha sade ve analiz edilebilir bir hale getirilmiş, sınıflandırma modelleri için uygun bir yapı oluşturulmuştur.

Bu veride de bütün sınıfların veri sayısı eşitlenmiştir. Nötr verilerin sayısı görece az olduğu için diğer sınıflardaki veri sayıları azaltılmıştır. Toplam 2169 veri vardır ve her sınıf 723 veriye sahiptir.



Şekil 9. RateMD veri kümesi sınıf dağılımı

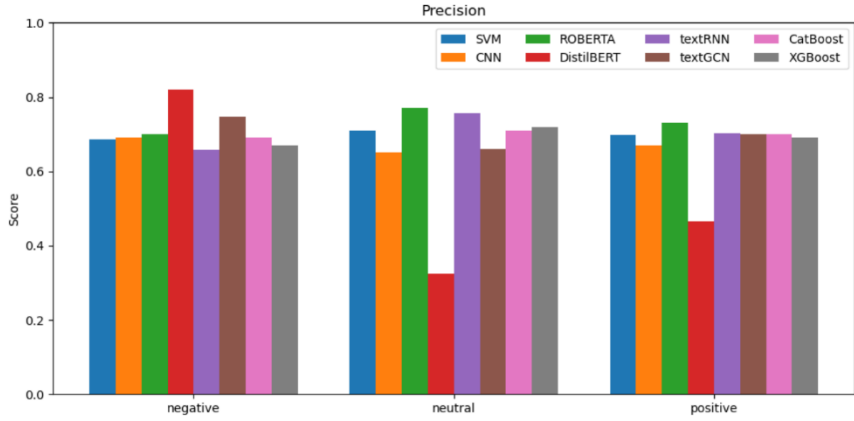
4.2 Sonuçlar ve Değerlendirme

4.2.1 NHS Veri Kümesi

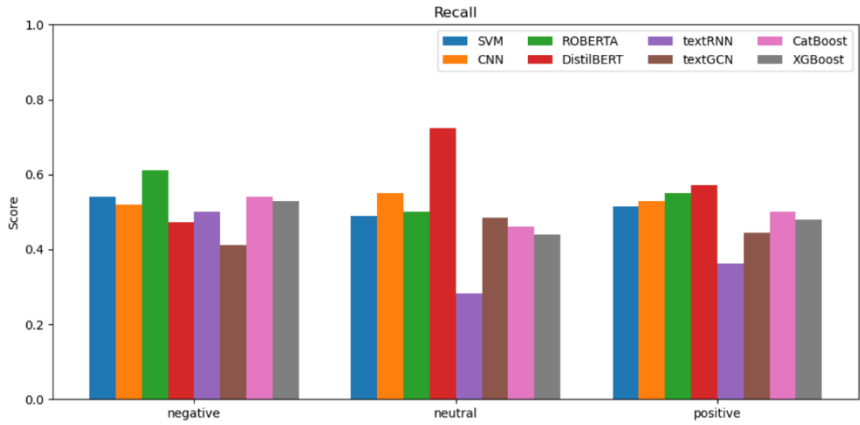
Tablo 1. NHS veri kümesi

Algoritma Adı	%70 Eğitim %30 Test	%80 Eğitim %20 Test

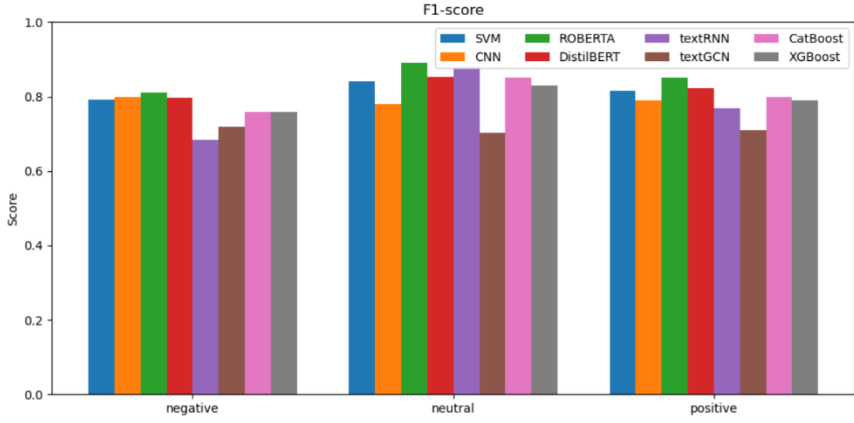
SVM	Accuracy: 0.6804 F1 score: 0.6762 Recall: 0.6804 Precision: 0.6736	Accuracy: 0.6838 F1 score: 0.6785 Recall: 0.6838 Precision: 0.6755
CNN	Accuracy: 0.66 F1 score: 0.66 Recall: 0.66 Precision: 0.67	Accuracy: 0.65 F1 score: 0.65 Recall: 0.65 Precision: 0.65
RoBERTa	Accuracy: 0.72 F1 score: 0.71 Recall: 0.72 Precision: 0.71	Accuracy: 0.73 F1 score: 0.72 Recall: 0.73 Precision: 0.72
DistilBERT	Accuracy: 0.6341 F1 score: 0.6203 Recall: 0.6340 Precision: 0.6963	Accuracy: 0.6890 F1 score: 0.6843 Recall: 0.6890 Precision: 0.6816
textRNN	Accuracy: 0.6390 F1 score: 0.6111 Recall: 0.6390 Precision: 0.6138	Accuracy: 0.6467 F1 score: 0.6420 Recall: 0.6467 Precision: 0.6388
textGCN	Accuracy: 0.6252 F1 score: 0.6186 Recall: 0.6252 Precision: 0.6159	Accuracy: 0.6452 F1 score: 0.6378 Recall: 0.6455 Precision: 0.6380
XGBoost	Accuracy: 0.66 F1 score: 0.66 Recall: 0.66 Precision: 0.65	Accuracy: 0.67 F1 score: 0.66 Recall: 0.67 Precision: 0.66
CatBoost	Accuracy: 0.6737 F1 score: 0.6669 Recall: 0.6737 Precision: 0.6639	Accuracy: 0.6771 F1 score: 0.6706 Recall: 0.6771 Precision: 0.6678



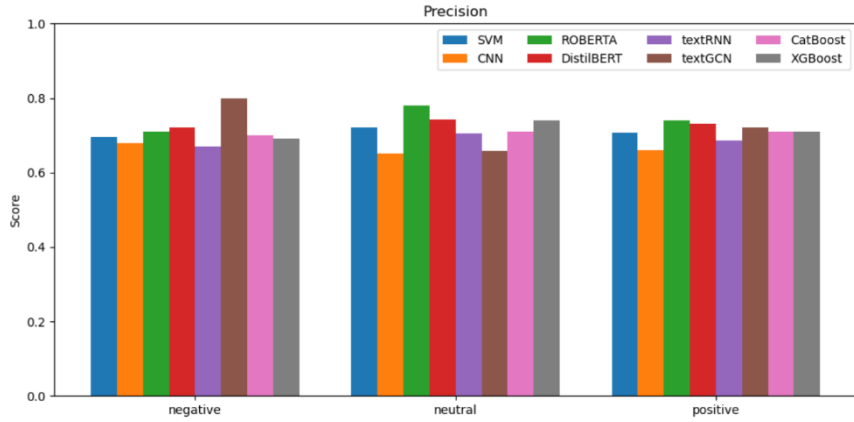
Şekil 9. NHS veri kümesi için yüzde 70 precision sonucu



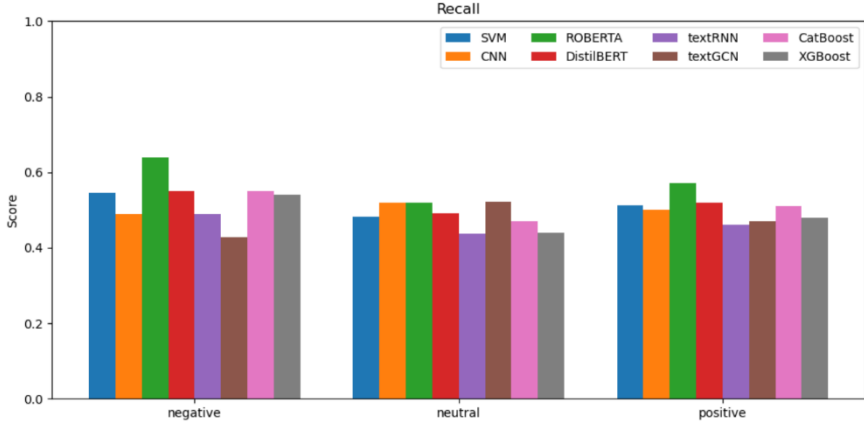
Şekil 10. NHS veri kümesi için yüzde 70 recall sonucu



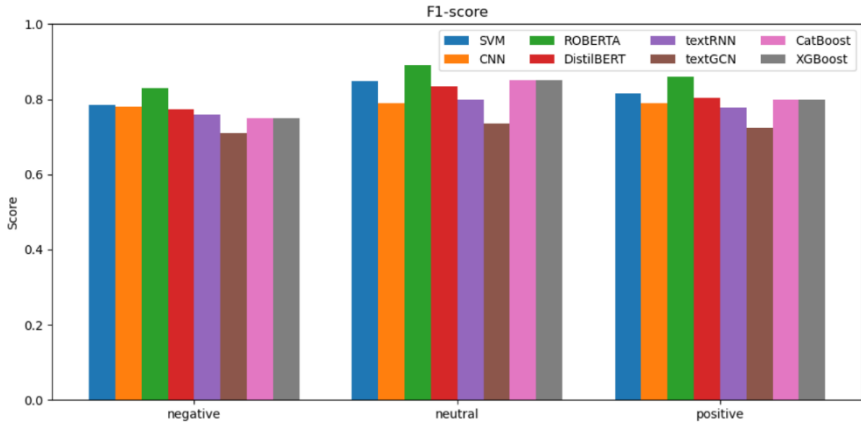
Şekil 11. NHS veri kümesi için yüzde 70 F1-score sonucu



Şekil 12. NHS veri kümesi için yüzde 80 precision sonucu



Şekil 13. NHS veri kümesi için yüzde 80 recall sonucu



Şekil 14. NHS veri kümesi için yüzde 80 F1-score sonucu

NHS veri kümesi göz önünde bulundurulduğunda en başarılı sonuçları veren algoritmanın hem %70 eğitim – %30 test hem de %80 eğitim – %20 test oranlarında RoBERTa olduğu görülmektedir. Eğitim yüzdesi arttıkça performansı da artan RoBERTa modeli özellikle yüksek veriyle daha güçlü genelleme yeteneği sergilemektedir. Benzer şekilde DistilBERT de düşük

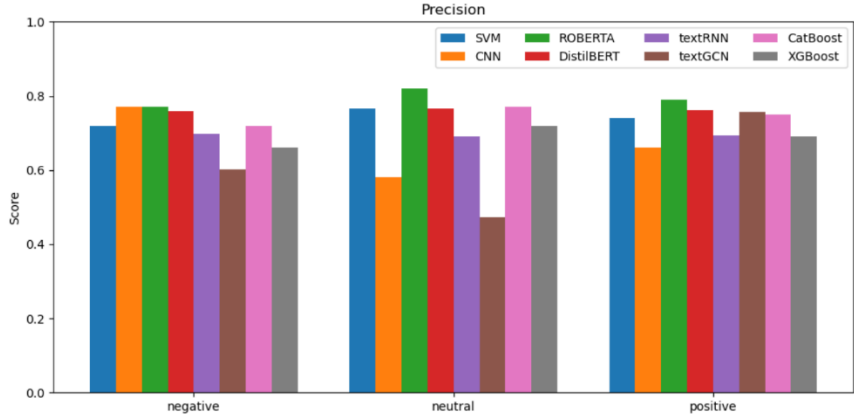
eğitim oranlarında nispeten zayıf sonuçlar vermesine rağmen eğitim oranı arttığında önemli bir performans artışı göstererek daha rekabetçi hale gelmektedir. CatBoost algoritması da benzer şekilde yüksek eğitim oranlarında daha iyi sonuç vermekte ve istikrarlı bir başarı profili çizmektedir. SVM modeli ise düşük ve yüksek eğitim oranlarında oldukça benzer sonuçlar vermekte, dolayısıyla eğitim oranından çok fazla etkilenmeyen dengeli bir algoritma olarak öne çıkmaktadır. Ancak bu modelin başarı düzeyi, RoBERTa gibi daha gelişmiş modellerin gerisinde kalmaktadır. Öte yandan CNN modeli dikkat çekici bir şekilde eğitim oranı arttıkça daha kötü sonuçlar vermektedir; bu da modelin yüksek veriyle overfitting ya da yeterli öğrenememe gibi sorunlar yaşadığını göstermektedir. textRNN ve textGCN algoritmaları genel olarak zayıf sonuçlar verirken, eğitim oranı artsa bile diğer modellerle aralarındaki farkı kapatamamaktadır. XGBoost ise ortalama bir başarı sergilemekte ve eğitim yüzdesi değişiminden çok fazla etkilenmemektedir.

4.2.2 RateMD Veri Kümesi

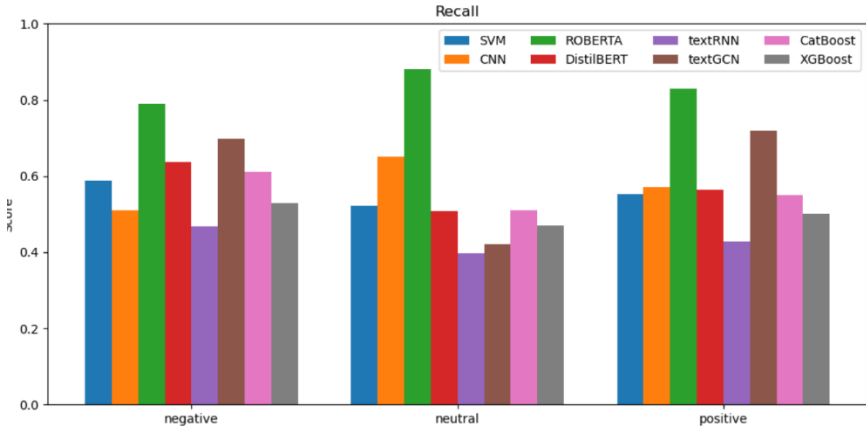
Tablo 2. RateMD veri kümesi

Algoritma Adı	%70 Eğitim %30 Test	%80 Eğitim %20 Test
SVM	Accuracy: 0.6897 F1 score: 0.6859 Recall: 0.6897 Precision: 0.6843	Accuracy: 0.7074 F1 score: 0.7028 Recall: 0.7076 Precision: 0.7017
CNN	Accuracy: 0.65 F1 score: 0.66 Recall: 0.65 Precision: 0.68	Accuracy: 0.63 F1 score: 0.64 Recall: 0.63 Precision: 0.66

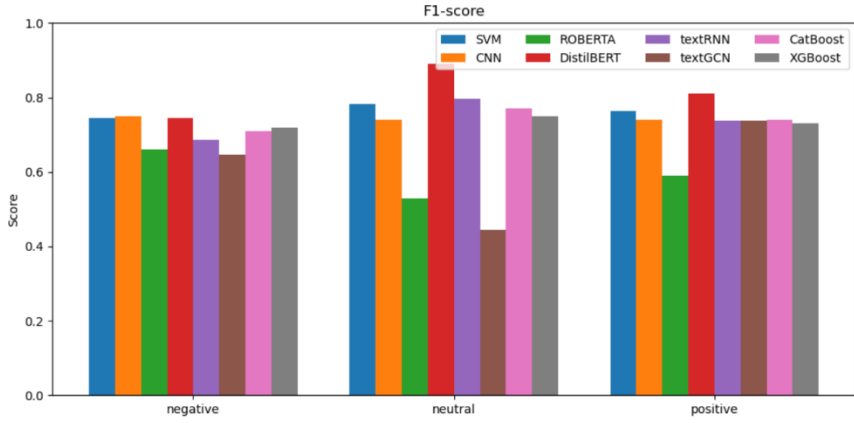
RoBERTa	Accuracy: 0.75 F1 score: 0.74 Recall: 0.75 Precision: 0.74	Accuracy: 0.76 F1 score: 0.75 Recall: 0.76 Precision: 0.75
DistilBERT	Accuracy: 0.7204 F1 score: 0.7122 Recall: 0.7204 Precision: 0.7130	Accuracy: 0.7488 F1 score: 0.7430 Recall: 0.7489 Precision: 0.7421
textRNN	Accuracy: 0.6283 F1 score: 0.6204 Recall: 0.6283 Precision: 0.6172	Accuracy: 0.6221 F1 score: 0.6066 Recall: 0.6221 Precision: 0.6154
textGCN	Accuracy: 0.6125 F1 score: 0.6095 Recall: 0.6125 Precision: 0.6102	Accuracy: 0.6205 F1 score: 0.6179 Recall: 0.6205 Precision: 0.6183
XGBoost	Accuracy: 0.64 F1 score: 0.63 Recall: 0.64 Precision: 0.63	Accuracy: 0.65 F1 score: 0.64 Recall: 0.65 Precision: 0.64
CatBoost	Accuracy: 0.6851 F1 score: 0.6798 Recall: 0.6851 Precision: 0.6793	Accuracy: 0.6959 F1 score: 0.6895 Recall: 0.6961 Precision: 0.6882



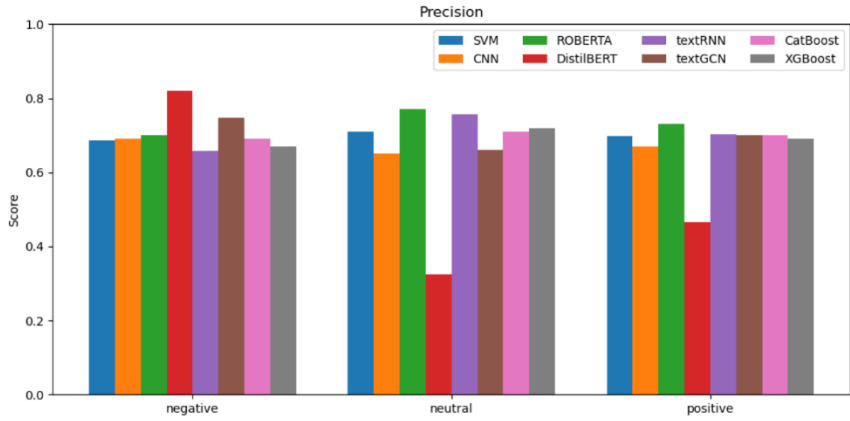
Şekil . RateMD veri kümesi için yüzde 70 precision sonucu



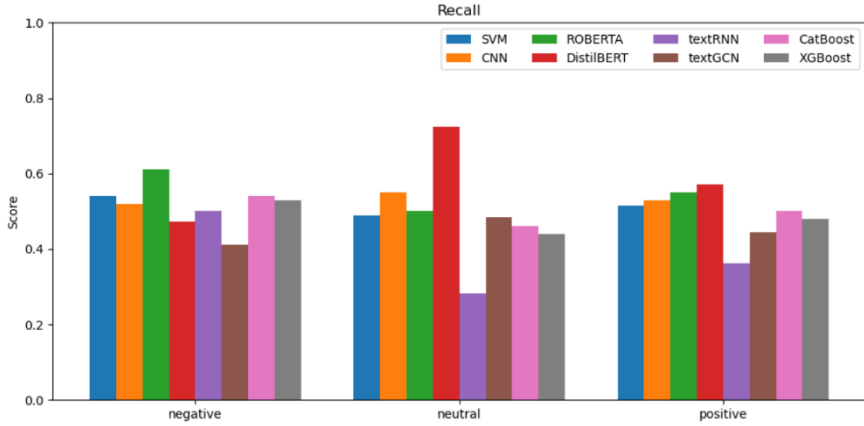
Şekil . RateMD veri kümesi için yüzde 70 recall sonucu



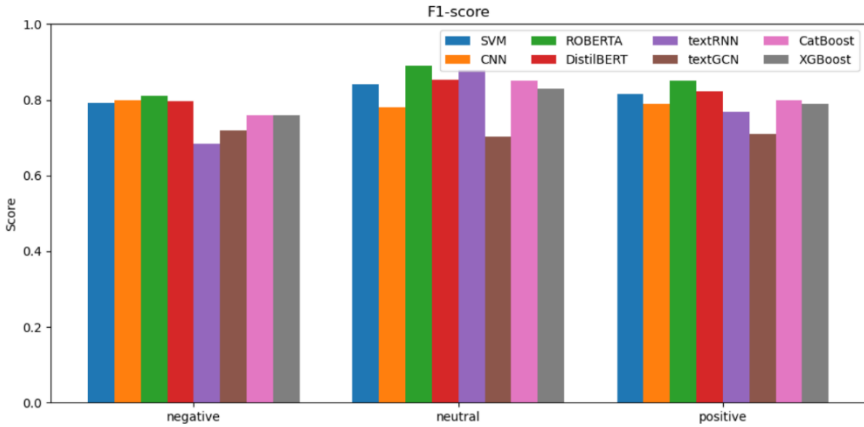
Şekil . RateMD veri kümesi için yüzde 70 F1-score sonucu



Şekil . RateMD veri kümesi için yüzde 80 precision sonucu



Şekil . RateMD veri kümesi için yüzde 80 recall sonucu



Şekil . RateMD veri kümesi için yüzde 80 F1-score sonucu

RateMD veri kümesi göz önünde bulundurulduğunda ise NHS veri kümesine benzer bir tablo görülmektedir. En yüksek başarıyı yine RoBERTa sağlamaktadır ve eğitim oranı yükseldikçe performansı artmaya devam etmektedir. DistilBERT modeli, NHS veri kümesinde olduğu gibi RateMD veri kümesinde de eğitim yüzdesi arttıkça belirgin bir başarı yükselişi göstermekte ve RoBERTa'ya yaklaşmaktadır. Bu durum, DistilBERT'in daha fazla veriye

eriştiğinde öğrenme kapasitesinin güçlendiğini ortaya koymaktadır. CatBoost yine güvenilir ve istikrarlı bir model olarak dikkat çekmekte ve eğitim oranı arttıkça küçük ama pozitif değişimler göstermektedir. SVM, NHS veri kümesindeki gibi burada da dengeli bir model olarak yer almakta ve eğitim oranı yükseldiğinde olumlu bir performans artışı göstermektedir. Ancak bu artış sınırlı kalmakta ve üst düzey modellerin gerisinde kalmaktadır. CNN algoritması NHS veri kümesindeki gibi burada da benzer bir şekilde düşük eğitim oranında daha iyi sonuç verirken yüksek eğitim oranında performans kaybı yaşamaktadır. Bu durum, CNN'in bu tür metin verilerine uygun olmadığını ya da aşırı öğrenme sorunları olduğunu göstermektedir. textRNN modeli ise eğitim oranı yükseldikçe performansı düşen nadir modellerden biridir. Bu, modelin eğitim verisini verimli kullanmadığını göstermektedir. textGCN ve XGBoost da genel olarak ortalama veya düşük başarılar sergilemiş, eğitim oranı değişimleri performanslarına belirgin bir etki yapmamıştır.

5. Sonuç

Bu çalışmada, dijitalleşmenin artmasıyla birlikte sağlık sistemine dair önemli bir geri bildirim kaynağı haline gelen çevrimiçi hasta yorumlarının duygu analizi gerçekleştirilmiştir. Temel problem, hasta deneyimlerinin sağlık hizmeti kalitesine yönelik anlamlı içgörülere dönüştürülmesidir. Bu doğrultuda yorumlar, doğal dil işleme ve makine öğrenmesi algoritmaları aracılığıyla pozitif, nötr ve negatif olarak sınıflandırılmıştır.

Çalışmada iki farklı sağlık değerlendirme platformundan alınan NHS ve RateMD veri kümeleri kullanılmış; KNN, SVM, CNN, textRNN, textGCN, XGBoost ve CatBoost gibi geleneksel makine öğrenmesi algoritmalarının yanı sıra RoBERTa, DistilBERT gibi Transformer tabanlı derin öğrenme modelleri de değerlendirmeye alınmıştır. Elde edilen sonuçlar göstermektedir ki Transformer mimarili modeller, özellikle RoBERTa, her iki veri kümesinde de en yüksek doğruluk, F1, precision ve recall değerlerine ulaşarak diğer algoritmaların önüne geçmiştir. Eğitim yüzdesi arttıkça RoBERTa ve DistilBERT gibi modellerin performansının belirgin

şekilde iyileştiği gözlemlenmiştir. CatBoost ve SVM ise istikrarlı fakat göreceli olarak daha sınırlı başarı sağlamıştır. CNN gibi bazı modellerin ise eğitim verisi arttıkça performansında düşüş yaşadığı görülmüştür.

Gelecekte yapılacak çalışmalarda veri seti çeşitliliğinin artırılması, Türkçe gibi farklı dillerdeki yorumların da analiz edilmesi, yorumların daha ayrıntılı kategori etiketlerine ayrılması ve GPT benzeri büyük dil modelleriyle sıfır örnekli sınıflandırma yöntemlerinin entegrasyonu planlanmaktadır. Ayrıca, yorumlardan yalnızca duygu değil; aynı zamanda öneri, şikâyet ve memnuniyet temalarının otomatik çıkarımı üzerine yoğunlaşılması hedeflenmektedir. Bu doğrultuda geliştirilecek sistemler, sağlık politikalarının şekillenmesine ve hasta merkezli hizmet iyileştirmelerine önemli katkılar sunabilir.

Kaynaklar

Alexander, G., Bahja, M. ve Butt, G. F. (2022). Automating Large-scale Health Care Service Feedback Analysis: Sentiment Analysis and Topic Modeling Study. *JMIR Medical Informatics*, 10(4), e29385.

doi:10.2196/29385

Anishnama. (2023, 18 Mayıs). Understanding Bidirectional

LSTM for Sequential Data Processing. *Medium*.

<https://medium.com/@anishnama20/understanding-bidirectional-lstm-for-sequential-data-processing-b83d6283befc> adresinden erişildi.

Awaits. (2020, 17 Mart). [DL] 8. CNN 1(Convolutional Neural Network Basics). *Learning*. <https://medium.com/jun94-devpblog/dl-8-cnn-1-convolutional-neural-network-basics-2f60c93d7d22> adresinden erişildi.

CatBoost—Open-source gradient boosting library. (t.y.). 2 Temmuz 2025 tarihinde <https://catboost.ai/news/catboost-enables-fast-gradient-boosting-on-decision-trees-using-gpus> adresinden erişildi.

Distilbert: A Smaller, Faster, and Distilled BERT - Zilliz Learn. (t.y.). 2 Temmuz 2025 tarihinde <https://zilliz.com/learn/distilbert-distilled-version-of-bert> adresinden erişildi.

K-Nearest Neighbor(KNN) Algorithm. (02:01:05+00:00). *GeeksforGeeks*. 2 Temmuz 2025 tarihinde <https://www.geeksforgeeks.org/machine-learning/k-nearest-neighbours/> adresinden erişildi.

NHS website developer portal. (t.y.). 2 Temmuz 2025 tarihinde <https://www.nhs.uk> adresinden erişildi.

Osório, L. ve Fachada, N. (2024). Patient-Centered Healthcare: 15th Advanced Doctoral Conference on Computing, Electrical and Industrial Systems, DoCEIS 2024. L. M. Camarinha-Matos ve F. Ferrada (Ed.), *DoCEIS içinde* , IFIP Advances in Information and Communication Technology (ss. 152-163). Springer Science and Business Media Deutschland GmbH.
doi:10.1007/978-3-031-63851-0_10

Osváth, M., Yang, Z. G. ve Kósa, K. (2023). Analyzing Narratives of Patient Experiences: A BERT Topic Modeling Approach. *Acta Polytechnica Hungarica*, 20(7), 153-171. doi:10.12700/APH.20.7.2023.7.9

(PDF) Preserving Privacy in Arabic Judgments: AI-Powered Anonymization For Enhanced Legal Data Privacy. (t.y.). 2 Temmuz 2025 tarihinde https://www.researchgate.net/publication/374691308_Preserving_Privacy_in_Arabic_Judgments_AI-Powered_Anonymization_For_Enhanced_Legal_Data_Privacy adresinden erişildi.

Sachin soni. (2023, 11 Haziran). K Nearest Neighbours—
Introduction to Machine Learning Algorithms. *Medium*.
[https://medium.com/@sachinsoni600517/k-nearest-
neighbours-introduction-to-machine-learning-
algorithms-9dbc9d9fb3b2](https://medium.com/@sachinsoni600517/k-nearest-neighbours-introduction-to-machine-learning-algorithms-9dbc9d9fb3b2) adresinden erişildi.

Sagarika Suresh Thimmanayakanapalya, Pavankumar
Mulgund, ve Raj Sharman. (2022). Development of an
Automated Physician Review Classification System: A
Hybrid Machine Learning Approach. *ResearchGate*
içinde . International Conference of Information
Systems, sunulmuş bildirisi, Denmark.
[https://www.researchgate.net/publication/365960526_
Development_of_an_Automated_Physician_Review_
Classification_System_A_Hybrid_Machine_Learning_
Approach](https://www.researchgate.net/publication/365960526_Development_of_an_Automated_Physician_Review_Classification_System_A_Hybrid_Machine_Learning_Approach) adresinden erişildi.

Sakai, H., Lam, S. S., Mikaeili, M., Bosire, J. ve Jovin, F.
(2025, 20 Şubat). Large Language Models for Patient
Comments Multi-Label Classification. arXiv.
doi:10.48550/arXiv.2410.23528

Steele, B., Fairie, P., Kemp, K., D'Souza, A. G., Wilms, M. ve Santana, M. J. (2025). Identifying Patient-Reported Care Experiences in Free-Text Survey Comments: Topic Modeling Study. *JMIR Medical Informatics*, 13(1), e63466. doi:10.2196/63466

Tang, J. E., Arvind, V., Dominy, C., White, C. A., Cho, S. K. ve Kim, J. S. (2023). How Are Patients Reviewing Spine Surgeons Online? A Sentiment Analysis of Physician Review Website Written Comments. *Global Spine Journal*, 13(8), 2107-2114. doi:10.1177/21925682211069933

Thawani, A., Paul, M. J., Sarkar, U. ve Wallace, B. C. (2019). Are Online Reviews of Physicians Biased Against Female Providers? *Proceedings of the 4th Machine Learning for Healthcare Conference* içinde (ss. 406-423). Machine Learning for Healthcare Conference, sunulmuş bildiri, PMLR. <https://proceedings.mlr.press/v106/thawani19a.html> adresinden erişildi.

Türkoğlu, M. F. (2021, 14 Temmuz). Support Vector Machine Algoritması-Makine Öğrenmesi-. *Medium*. 2 Temmuz 2025 tarihinde <https://mfatih.to.medium.com/support-vector-machine-algoritmas%C4%B1-makine-%C3%B6%C4%9Frenmesi-8020176898d8> adresinden erişildi.

What Is Support Vector Machine? | IBM. (2023, 12 Aralık). 2 Temmuz 2025 tarihinde <https://www.ibm.com/think/topics/support-vector-machine> adresinden erişildi.

Yao, L., Mao, C. ve Luo, Y. (2018, 13 Kasım). Graph Convolutional Networks for Text Classification. *arXiv*. doi:10.48550/arXiv.1809.05679

Zhao, Y., Li, T., Yuan, Q. ve Deng, S. (2024). How to detect fake online physician reviews: A deep learning approach. *DIGITAL HEALTH*, 10, 20552076241277171. doi:10.1177/20552076241277171