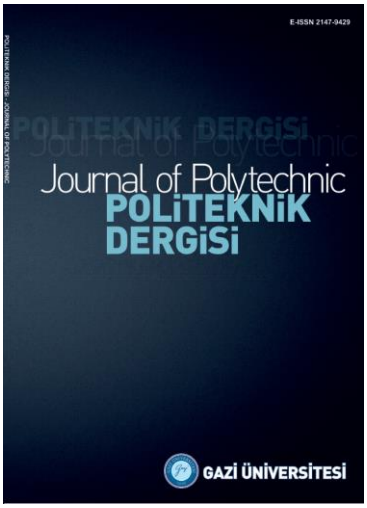


POLİTEKNİK DERGİSİ

JOURNAL of POLYTECHNIC

Graf sinir ağları ile ilişkisel Türkçe metin sınıflandırma

Relational Turkish text classification with graph neural networks

***Yazar(lar) (Author(s))**: Halil İbrahim OKUR[1], Kadir TOHMA[2], Ahmet SERTBAŞ[3]*

ORCID[1]: 0000-0003-0339-4626

ORCID[2]: 0000-0002-2631-7810

ORCID[3]: 0000-0001-8166-1211

To cite to this article: Okur H. İ., Tohma K. ve Sertbaş A., "Graf Sinir Ağları ile İlişkisel Türkçe Metin Sınıflandırma", ***Journal of Polytechnic***, 28(3): 867-879, (2025).

Bu makaleye şu şekilde atıfta bulunabilirsiniz: Okur H. İ., Tohma K. ve Sertbaş A., "Graf Sinir Ağları ile İlişkisel Türkçe Metin Sınıflandırma", ***Politeknik Dergisi***, 28(3): 867-879, (2025).

Erişim linki (To link to this article): http://dergipark.org.tr/politeknik/archive

DOI: 10.2339/politeknik.1423293

Graf Sinir Ağları ile İlişkisel Türkçe Metin Sınıflandırma

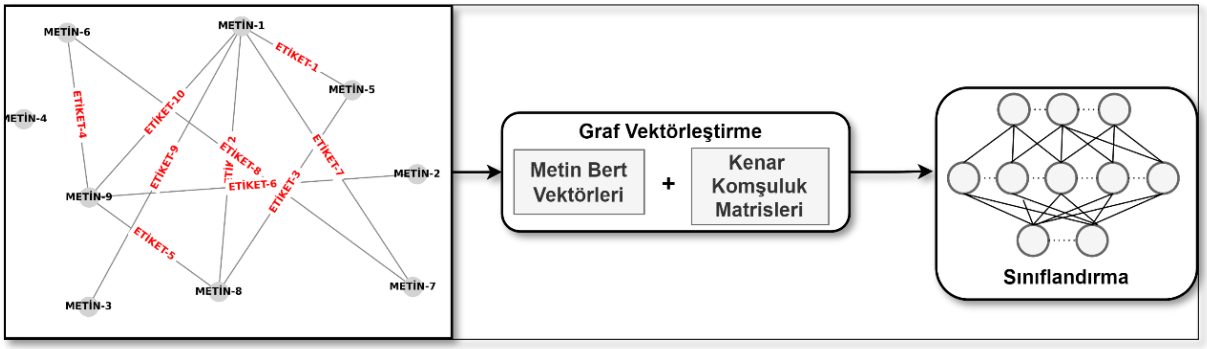
Relational Turkish Text Classification with Graph Neural Networks

Önemli noktalar (Highlights)

- *Türkçe newsTRT2023 veri kümesi oluşturulmuştur. / Turkish newsTRT2023 dataset was created.*
- *Etiket bazlı ilişkisel metin graf yapısı modellenmiştir. / Tag-based relational text graph structure is modeled.*
- *Graf Sinir Ağı (GNN) ile metinler sınıflandırılmıştır. / Texts were classified with the graph neural network.*

Grafik Özet (Graphical Abstract)

Oluşturulan haber metin veri kümesi, içerdikleri etiket bazlı ilişkiler sayesinde graf yapısına dönüştürülmüştür. Graf Sinir Ağları (GNN) kullanılarak metinler arası ilişkiler ve sınıflandırma görevleri etkili bir şekilde gerçekleştirilmiştir. (Şekil A) / The created news text dataset was converted into a graph structure thanks to the tag-based relationships it contains. Intertextual relations and classification tasks have been carried out effectively using Graph Neural Networks (GNN). (Figure A)



Şekil A. Graf Sinir Ağı ile İlişkisel Metin Sınıflandırması / **Figure A.** Relational Text Classification with Graph Neural Network

Amaç (Aim)

Türkçe metinler arasındaki ilişkilerin kullanılmasıyla graf sinir ağı modelinin elde edilmesi ve sınıflandırma performansının etkinliğinin gösterilmesi amaçlanmıştır. / The aim is to develop a Graph Neural Network model using the relationships between Turkish texts and to demonstrate the effectiveness of its classification performance.

Tasarım ve Yöntem (Design & Methodology)

Metinler arası ilişkisel bağlar (graf yapısı) sayısal bir formata dönüştürülmüş ve metinler sınıflandırılmıştır. / Relational links (graph structure) between texts were converted into a numerical format and texts were classified.

Özgünlük (Originality)

Özgün Türkçe haber veri kümesi üzerinde graf sinir ağı ile ilişkisel metin sınıflandırılması gerçekleştirilmiştir. / Relational text classification was performed with graph neural network on the original Turkish news dataset.

Sonuç (Conclusion)

GNN modelinin etkinliği, makine örenmesi modelleri ve bir derin sinir ağı modeli ile karşılaştırmalı olarak gösterilmiştir. / The effectiveness of the GNN model is demonstrated in comparison with machine learning models and a deep neural network model.

Etik Standartların Beyanı (Declaration of Ethical Standards)

Bu makalenin yazarları çalışmalarında kullandıkları materyal ve yöntemlerin etik kurul izni ve/veya yasal-özel bir izin gerektirmediğini beyan ederler. / The author(s) of this article declare that the materials and methods used in this study do not require ethical committee permission and/or legal-special permission.

Graf Sinir Ağları ile İlişkisel Türkçe Metin Sınıflandırma

Araştırma Makalesi / Research Article

Halil İbrahim OKUR[1*], Kadir TOHMA[2], Ahmet SERTBAŞ[3]

[1,2]Mühendislik ve Doğa Bilimleri Fakültesi, Bilgisayar Mühendisliği Bölümü, İskenderun Teknik Üniversitesi, Türkiye
[3]Mühendislik Fakültesi, Bilgisayar Mühendisliği Bölümü, İstanbul Üniversitesi-Cerrahpaşa, Türkiye
(Geliş/Received : 21.01.2024 ; Kabul/Accepted : 05.09.2024 ; Erken Görünüm/Early View : 13.10.2024)

ÖZ

Türkçe metin sınıflandırması ve ilişkisel analiz, dilin karmaşık yapısını anlamada ve doğal dil işleme süreçlerini geliştirmede kritik bir rol oynar. Bu çalışma, Türkçe metinlerin sınıflandırılması ve aralarındaki ilişkilerin derinlemesine analiz edilmesine odaklanmaktadır. Çalışmanın amacı, Türkçe'nin zengin morfolojik yapısını ve metinler arası ilişkileri etkin bir şekilde ele alarak, bu yapıyı yansıtan ileri düzey bir sınıflandırma modeli geliştirmektir. TRT-Haber web sayfasından elde edilen veri kümesi üzerinde graf tabanlı derin öğrenme teknikleri kullanılarak, yüksek performanslı bir model oluşturulmuştur. Metinlerin semantik vektör gösterimleri için BERT (BertTurk) modeli kullanılmış ve metinler arası ilişkileri gösteren kenar komşuluk matrisleri ile birleştirilmiştir. Bu veriler, graf sinir ağı (GNN) tabanlı sınıflandırma modeline beslenmiştir. Elde edilen sonuçlar, GNN modelinin %97.93 doğruluk oranı ile metinleri sınıflandırabildiğini ve ilişkisel yapıları başarıyla çözümleyebildiğini göstermektedir. Bu bulgular, metin sınıflandırması ve ilişkisel analizde graf tabanlı yaklaşımların etkinliğini ve potansiyelini ortaya koyarak, Türkçe metinlerin daha iyi anlaşılmasını ve işlenmesini sağlayacak yenilikçi yöntemlerin geliştirilmesine katkı sağlamaktadır.
Anahtar kelimeler: Türkçe doğal dil işleme, ilişkisel metin sınıflandırma, derin öğrenme, graf sinir ağları

Relational Turkish Text Classification with Graph Neural Networks

ABSTRACT

Text classification and relational analysis in Turkish play a critical role in understanding the language's complex structure and enhancing natural language processing (NLP) procedures. This study focuses on the classification of Turkish texts and the in-depth analysis of the relationships between them. The aim of the study is to develop an advanced classification model that effectively captures the rich morphological structure of Turkish and the intertextual relationships. Using a dataset obtained from the TRT-Haber website, graph-based deep learning techniques were employed to create a high-performance model. The BERT (BertTurk) model was used for semantic vector representations of texts, and adjacency matrices representing intertextual relationships were integrated. These data were then fed into a graph neural network (GNN) based classification model. The results demonstrate that the GNN model can classify texts with a remarkable accuracy rate of 97.93% and successfully resolve relational structures. These findings highlight the effectiveness and potential of graph-based approaches in text classification and relational analysis, contributing to the development of innovative methods for better understanding and processing Turkish texts.
Keywords: Turkish NLP, relational textclassification, deep learning, graph neural networks

1. GİRİŞ (INTRODUCTION)

Metin sınıflandırması, Doğal Dil İşleme (DDİ) alanının temel görevlerinden biri olarak kabul edilmektedir. Yıllar boyunca, farklı potansiyel uygulamaları nedeniyle geniş bir şekilde incelenmiştir [1-3]. Genel olarak, metin sınıflandırma modelleri, metinsel belgelere önceden tanımlanmış etiketler/sınıflar tahmin etmek üzere tasarlanmaktadır. Bu sınıflandırıcılar daha sonra metin anlama [4], konu tanımlama [5], duygu analizi [6], sahte haber tespiti [7], spam tespiti [8] gibi başlıca metin analizi ve madenciliği problemlerine katkıda bulunmak için kullanılmaktadır [9]. Son zamanlarda, İnternet'in hızlı gelişimi, Facebook, Twitter gibi sosyal ağ platformlarından [10], YouTube, Instagram gibi sosyal medya platformlarına [11], Wikipedia, YAGO gibi açık ansiklopedi/bilgi grafiklerine kadar İnternet'in büyük ölçekli çevrimiçi platformlarından kaynaklanan büyük metin kaynaklarının dramatik bir şekilde ortaya çıkmasına yol açmıştır. Gerçekten de, bu devasa metin kaynaklarından değerli bilgileri etkili bir şekilde çıkarabilecek daha güçlü araçlar geliştirme ihtiyacı artmaktadır. Özellikle, DDİ alanında makine öğrenimi / derin öğrenme konusunda yapılan günümüz çalışmaları ile birlikte, metin temsil öğrenimi, kümeleme ve sınıflandırma görevlerinde üst düzey performanslar sergileyen pek çok dikkate değer ve öne çıkan çalışma bulunmaktadır [12-14]. Özellikle zengin morfolojik özellikleriyle dikkat çeken Türkçe gibi diller, DDİ'nin karmaşıklığını ve geniş olanaklarını ön plana çıkarmaktadır [15].

Metin sınıflandırması, DDİ''nin temel alt görevlerinden biridir ve dili anlama ve işleme sürecinde kritik bir role sahiptir. Bu işlem, metinlerin doğru bir şekilde temsil edilmesiyle başlar ve genellikle kapsamlı veri kümeleri

*Sorumlu Yazar (Corresponding Author)
e-posta: hibrahim.okur@iste.edu.tr

ile derin öğrenme modellerinin kullanımını gerektirir. Dil anlama görevlerinde, BERT gibi güçlü modellerin başarısı, bu alandaki ilerlemenin bir göstergesidir. Büyük veri kümelerinde önceden eğitilmiş modellerin kullanımı, yeni model geliştirmenin karşılaştığı zorlukların üstesinden gelmede etkili bir yöntem sunar. Metin sınıflandırmasının uygulama alanları oldukça geniş ve çeşitlidir [16, 17]. Sosyal medya analizinden sağlık hizmetlerine kadar birçok alanda kullanılır [18-20]. Geleneksel sınıflandırma yöntemleri, manuel olarak hazırlanan özelliklere ve sınıflandırma algoritmalarına odaklanırken, yeni yaklaşımlar metinlerdeki tokenler arasındaki ilişkileri ve dizilimleri daha ayrıntılı bir şekilde incelemektedir [21]. Graf yapıları, metinler arasındaki karmaşık etkileşimleri modellemede büyük bir etkinlik gösterir. Bu nedenle, graf tabanlı yaklaşımların DDİ alanında kullanımı artmakta ve metin sınıflandırma ile ilişkisel analiz görevlerinde önemli bir rol oynamaktadır [22].

Son yıllarda, Doğal Dil İşleme (DDİ) için özellikle derin öğrenme alanında önemli başarılar elde edilmiştir [23]. Bu gelişmeler, bilgisayarların metinlerden özellikleri otomatik olarak öğrenip çıkarmada, metinsel özellik mühendisliği sürecinde insan çabasını en aza indirgeyebileceklerini göstermektedir. Metinler arası ilişkilerin anlaşılması ve analizi, DDİ'nin en zorlayıcı yönlerinden biridir ve bu, özellikle ilişkisel metin sınıflandırma görevlerinde daha da belirgindir [24]. Metinler arası ilişkiler, dilin doğasından kaynaklanan karmaşık ve çok katmanlı yapıları içerir; bu ilişkiler, metin içindeki ve metinler arasındaki bağlam, anlam ve niyeti ortaya çıkarır [25]. Graf Sinir Ağları (GNN), bu tür karmaşık yapıları modellemek için giderek daha fazla tercih edilen bir yöntem haline gelmiştir. GNN'ler, metinlerdeki ve metinler arasındaki ilişkileri, graf tabanlı bir yapıda etkin bir şekilde temsil edebilir. Bu yaklaşım, metinler arasındaki doğrudan ve dolaylı bağlantıları, ayrıca kelime, cümle ve paragraf düzeyindeki ilişkileri görselleştirmekte ve analiz etmekte son derece başarılıdır [26].

Çalışmamızda, haber metinlerinin sınıflandırılması ve ilişkisel yapılarının derinlemesine analizi üzerine odaklanılmıştır. Bu bağlamda, dilin zengin ve karmaşık yapısını detaylı bir şekilde yansıtabilmek için özel olarak oluşturulan bir Türkçe haber metin veri kümesi kullanılmıştır. Temel olarak, haber metinleri arasındaki etiket bazlı ilişkiler, detaylı bir graf yapısı içinde modellendirilmiştir. Bu modellemede, düğümler haber metinlerini temsil ederken, kenarlar ise bu metinler arasındaki ortak etiketler temelinde kurulan ilişkileri ifade etmektedir. Çalışmada, metinlerin semantik gösterimleri için gelişmiş bir doğal dil işleme modeli olan Türkçe BERT kullanılmıştır. Bu sayede, metinlerin yüksek boyutlu ve anlamlı gösterimleri elde edilerek, doğal dil anlama yeteneği geliştirilmiştir.

Ayrıca, metinler arası ilişkilerin sayısal bir formata dönüştürülmesi için, metin çiftlerinin kenar komşuluk matrisleri hesaplanmıştır. Bu yöntem, metinler arasındaki bağlantı sıklığını ve gücünü vektörler halinde kodlayarak, ilişkisel analizin daha detaylı yapılmasını sağlamıştır. Elde edilen zengin ve çok boyutlu graf vektör yapısı, bir graf sinir ağı (GNN) tabanlı sınıflandırma modeline beslenmiştir. Bu model, graf teorisini ve derin öğrenme tekniklerini bütünleştirerek, metinlerin hem sınıflandırılmasını hem de ilişkisel bağlamda anlamlandırılmasını hedeflemiştir. Modelin eğitimi çeşitli hiperparametre ayarlarıyla optimize edilirken, test işlemleri modelin genelleştirme kabiliyetini doğrulamak için yapılmıştır. Performans değerlendirmesi için doğruluk, F1 skoru ve AUC gibi metrikler kullanılmış, ayrıca makine öğrenimi algoritmalarının karşılaştırmalı analizleri gerçekleştirilmiştir. Bu kapsamlı ve çok yönlü yaklaşım, haber metinlerinin sınıflandırılması ve ilişkisel yapıların çözümlenmesi konusunda önemli bir potansiyel ortaya koymuştur.

Katkılar: Bu çalışma, Türkçe metin sınıflandırma ve ilişkisel analiz alanına önemli katkılarda bulunmuştur:

Türkçe İlişkisel Metin Veri Kümesinin Geliştirilmesi: Çalışmada Türkçe'nin zengin morfolojik yapısını ve metinler arası ilişkileri kapsayan yeni bir veri kümesi geliştirilmiştir. Bu veri kümesi ("newsTRT2023"), Türkçe doğal dil işleme alanındaki araştırmalar ve uygulamalar için değerli bir kaynak sunmaktadır.

Graf Sinir Ağları (GNN) Kullanımının İncelenmesi: ,Çalışma, metin sınıflandırma ve ilişkisel analizde Graf Sinir Ağlarının (GNN) etkinliğini ve potansiyelini ayrıntılı bir şekilde incelemiştir. Bu inceleme, GNN'nin Türkçe dil işleme bağlamında nasıl uygulanabileceğini ve optimize edilebileceğini ortaya koymaktadır, bu da Türkçe DDİ alanında önemli bir yenilik sağlamaktadır.

Yüksek Başarım Sunan Yöntemler: Çalışma, Türkçe metin sınıflandırma ve ilişkisel analiz alanında yüksek başarım sunan yöntemler geliştirmiştir. Bu yöntemler, Türkçe gibi zengin morfolojik yapısı olan bir dilin karmaşık yapısını daha etkin bir şekilde anlamak ve işlemek için geliştirilmiş olup, Türkçe DDİ alanında önemli bir ilerleme olarak değerlendirilmektedir. Elde edilen sonuçlara göre GNN modeli, %97.93 gibi dikkate değer bir doğruluk oranı elde etmiştir.

Çalışmanın 2. bölümü, ilgili literatürün kapsamlı bir incelemesini ve ilgili çalışmaların detaylarını içermektedir. 3. bölümde, bu çalışmada kullanılan metodoloji açıklanmakta ve önerilen metin sınıflandırma modelinin detayları anlatılmaktadır. 4. bölümde, deneysel çalışmalarımız ve bunların sonuçları ele alınmaktadır. 5. bölüm, bulguların ve bunların etkilerinin yanı sıra gelecekteki çalışmalar için potansiyeller üzerine bir tartışmaya ayrılmıştır. Son olarak, 6. bölüm çalışmayı sonuçlandırmakta, ana noktaları ve bulguları özetlemektedir.

2. İLGİLİ ÇALIŞMALAR (RELATED WORKS)

Bu bölümde, literatürdeki çalışmalar kapsamlı bir şekilde incelenmiş, çalışmanın odaklandığı temel

konular olan metin sınıflandırması, ilişkisel metin sınıflandırması ve Graf Sinir Ağları (GNN) alanlarındaki önemli araştırmalar değerlendirilmiştir.

Uslu ve ark. [27] çalışmasında, Türkçe haber metinlerinin makine öğrenmesi yöntemleri kullanılarak sınıflandırılması incelenmiştir. Bu çalışma, çeşitli makine öğrenme algoritmalarının Türkçe haber metinleri üzerindeki performansını karşılaştırmaktadır. Çelik ve Koç [28], TF-IDF, Word2Vec ve FastText gibi vektör model yöntemleri ile Türkçe haber metinlerinin sınıflandırılması üzerine bir çalışma yapmıştır. Bu çalışmada, bu yöntemlerin farklı makine öğrenmesi algoritmaları ile kombinasyonlarının performansı değerlendirilmiştir. Acı ve Çırak [29] tarafından yürütülen araştırma, Türkçe haber metinlerinin konvolüsyonel sinir ağları (CNN) ve Word2Vec kullanılarak sınıflandırılması üzerine odaklanmıştır. Karakurt ve Karcı [30], GloVe kelime gömmeleri ve sinir ağları ile Türkçe haber metinlerinin sınıflandırılması üzerine bir çalışma yapmıştır.

Parlak [31] tarafından yapılan araştırmada, metin sınıflandırmasının temelini oluşturan ön işleme tekniklerinin Türkçe ve İngilizce haber metinleri üzerindeki etkileri detaylı bir şekilde incelenmiştir. Bu çalışma, stop-word'lerin çıkarılması, kök indirgeme, harflerin küçültülmesi ve tokenleştirme gibi temel ön işleme adımlarının, metin sınıflandırma performansına olan katkılarını ve dil özelliklerine göre değişen performans farklılıklarını ortaya koymuştur.

Okur ve ark. [32] ise, Türkçe metin sınıflandırmada kullanılan ön eğitimli sinirsel modellerin etkinliğini ve performansını derinlemesine analiz etmiştir. Çalışmada, Word2Vec ve GloVe gibi kelime gömme yöntemleri ile BERT, Electra ve FastText gibi gelişmiş modellerin Türkçe metin sınıflandırma bağlamında nasıl kullanılabileceği ve bu modellerin avantajları üzerinde durulmuştur.

Demir [33] tarafından yürütülen araştırma, BERT tabanlı dil modelleri ve makine öğrenimi algoritmalarını kullanarak Türkçe haber metinleri üzerinde duygu analizi gerçekleştirmiştir. Bu çalışma, Türkçe metinlerdeki duygusal içeriklerin sınıflandırılmasında dil modellerinin ve algoritmaların başarısını değerlendirmiş ve dil işleme ile duygu analizi alanlarına metodolojik katkılar sunmuştur.

Tohma ve ark. [34] araştırmasında, Türkçe soru-cevap sistemleri için duygu analizi tekniklerinin kullanımı ve bu tekniklerin insan-robot etkileşimlerindeki etkisine odaklanılmıştır. Bu çalışma, Türkçe soru-cevap-duygu veri kümesinde gerçekleştirilen ön işleme adımları ve Ön Eğitimli Türkçe BERT Modeli gibi teknikleri kullanarak, duygu analizinde %91.05'e varan doğruluk değeri elde eden yeni bir hibrit model önermiştir.

Malekzadeh ve ark. [35] çalışmasında, Graf Sinir Ağları (GNN) kullanımının literatürdeki yerini ve bu yaklaşımın metin sınıflandırmasındaki performansını ayrıntılı bir şekilde incelemiştir. Bu çalışma, GNN tabanlı yöntemlerin geleneksel derin öğrenme yaklaşımlarına kıyasla metin sınıflandırması ve çeşitli DDİ görevlerinde nasıl etkili olduğunu gözler önüne sermiştir. Çizelge 1' de ise, literatürde graf tabanlı metodolojileri kullanan çeşitli araştırmalar derlenmiştir.

Bu araştırmalar, metin sınıflandırma görevlerinde graf tabanlı yaklaşımların çeşitli veri kümeleri üzerindeki etkinliklerini ve performanslarını sergilemektedir. Çizelge 1' de bu konuda bulunan mevcut çalışmaların geniş bir yelpazesini sunmakta ve graf tabanlı metodolojilerin metin sınıflandırma alanındaki uygulamalarının kapsamlı bir özetini sağlamaktadır. Bu çizelgedeki çalışmaların detayları ve metodolojik yaklaşımları, bu bölümün devamında ayrıntılı bir şekilde ele alınmaktadır.

Huang ve ark. [36] tarafından yapılan araştırmada, GNN tekniklerinin metin sınıflandırmadaki kullanımı ve bu tekniklerin karmaşık yapıları üzerinde yoğunlaşılmıştır. Yapılan deneyler, GNN modelinin mevcut modellere kıyasla daha iyi performans gösterdiğini ve daha az bellek tükettiğini ortaya koymuştur.

Yao ve ark. [37], metin sınıflandırması için Graf Konvolüsyonel Ağları (GCN) kullanarak, kelime eşzamanlılığı ve doküman-kelime ilişkilerine dayalı bir Text GCN modeli eğitmişlerdir. Bu modelin, harici kelime gömme veya bilgi gerektirmeden, basit bir Text GCN'nin mevcut metin sınıflandırma yöntemlerine göre daha üstün olduğunu göstermiştir.Kumar ve ark. [38] çalışması, sosyal medya ağlarındaki metinlerin sınıflandırılması için Graf Sinir Ağı (GNN) kullanımını incelemiştir. Araştırma, GNN'nin iki boyutlu vektörler ve metinler için uygun olduğunu ve metin sınıflandırmasında mevcut yöntemlere göre daha etkili olduğunu gözlemlemiştir.

C. Liu ve ark. [39], metin sınıflandırması için Graf Konvolüsyonel Ağları (GCN) kullanımını incelerken, dokümanlar arası ilişkileri içeren karmaşık ve zengin ilişki tabanlı matris grafikleri oluşturarak yüksek doğruluk elde etmiştir.

Chen ve ark. [40] çalışması, kazaların nedenlerini ve süreçlerini analiz etmek için kaza araştırma raporlarının metinlerinin otomatik analizine odaklanmıştır. Çalışma, ilişkisel graf konvolüsyonel ağ (R-GCN) ve önceden eğitilmiş BERT'e dayalı bir metin madenciliği tabanlı kaza nedeni sınıflandırma yöntemi önermiştir.

X. Liu ve ark. [41] çalışması, doğal dil işleme görevlerinde ilişki sınıflandırmasının önemini vurgulamış ve ilişkisel metin sınıflandırma modeli önermiştir. Bu model, etiketlenmiş Çince metin veri kümesi üzerinde yapılan deneylerle ve SemEval-2010 ve KBP37 gibi genel İngilizce veri kümeleri üzerinde gerçekleştirilen deneylerle test edilmiştir.

Literatür incelemesi, metin sınıflandırması ve ilişkisel analiz alanlarındaki mevcut araştırmaların geniş kapsamlı bir özetini sunmuştur, bu da bu alanlardaki bilgi birikimini ve çeşitli araştırma yöntemlerini ayrıntılı bir şekilde gözler önüne sermiştir.

Çizelge 1. Literatürdeki graf tabanlı metin sınıflandırma çalışmaları (Graph-based text classification studies in the literature)

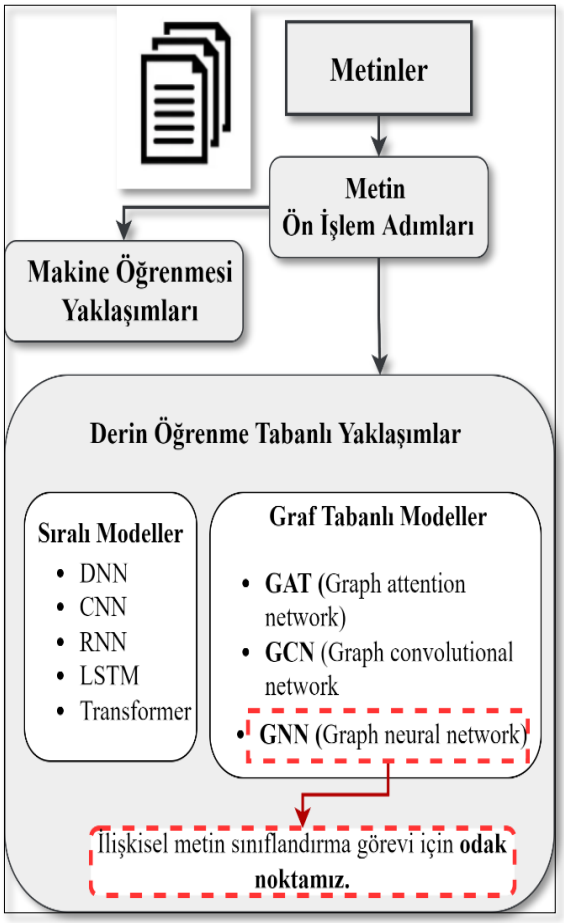
Çalışmalar	***Model***	***Karşılştırılan Modeller***	***Veri Kümesi***	***Sonuçlar*** *(Veri Kümesi Sırası ile Doğruluk / F1-Skor %)*
Huang ve ark. [36]	GCN+Softmax	CNN, LSTM, fastText, Graph-CNN, Text-GCN	R8, R52, Ohsumed	Doğruluk: 97.8, 94.6, 69.4
Yao ve ark. [37]	GCN	TF-IDF + LR LSTM, Bi-LSTM, PV-DBOW, PV-DM, PTE, FastText, SWEM, LEAM,Graph-CNN-C, Graph-CNN-S, Graph-CNN-F	20NG R8, R52, Ohsumed, MR	Doğruluk: 86.3, 97.1, 93.6, 68.4, 76,7
C. Liu ve ark. [39]	Document-relational GCN	Text GCN, Synonym augmented text GCN	20NG R8, R52, Ohsumed, MR	Doğruluk: 86.8, 96.9, 93.2, 65.9, 76.6
Chen ve ark. [40]	Relational-GCN	Decision trees, KNN, Naive Bayes, Logistic regression, LSTM, GRU, SGRU	OSHA dataset	F1-Score: (0.77)
X. Liu ve ark. [41]	Feature-based GCN	RNN, CNN, GCN, RGCN,R-BERT	SemEval-2010 Task 8, KBP37	Doğruluk: 89.75, 70.28

Ancak, bu inceleme sırasında, özellikle Türkçe metinlerin sınıflandırılmasında ilişkisel graf sinir ağları kullanımı açısından belirgin bir boşluk fark edilmiştir. Bu çalışma, bu boşluğu doldurmayı ve Türkçe metinler için yenilikçi ve etkili metodolojik yaklaşımlar geliştirmeyi hedeflemektedir. İlerleyen bölümlerde, bu hedef doğrultusunda geliştirilen yöntemler, gerçekleştirilen deneyler ve ulaşılan sonuçlar, derinlemesine ve ayrıntılı bir şekilde ele alınacak ve tartışılacaktır.

3. MATERYAL VE METOD (MATERIAL AND METHOD)

Bu çalışma, Doğal Dil İşleme (DDİ) alanında, özellikle Türkçe metin sınıflandırma ve ilişkisel analizi üzerinedir. Araştırmanın temel odak noktası, Türkçe metin ilişkisel olarak graf tabanlı özgün bir yöntem ve bu yöntemin elde ettiği yüksek başarım oranlarıdır. Türkçe'nin zengin morfolojik yapısı ve metinler arasındaki karmaşık ilişkileri göz önünde bulundurarak, bu çalışma Graf Sinir Ağları (GNN) ve Derin Sinir Ağları (DNN) gibi teknolojileri entegre eden bir yaklaşımı benimsemiştir. Buna göre özgün, "newsTRT2023" adlı, kapsamlı bir haber metni veri kümesinin detaylı analizi yapılmıştır. Şekil 1' de Türkçe metin işleme ve sınıflandırma çalışmalarında kullanılan bir dizi makine öğrenimi ve derin öğrenme metodolojisini özetlemektedir.

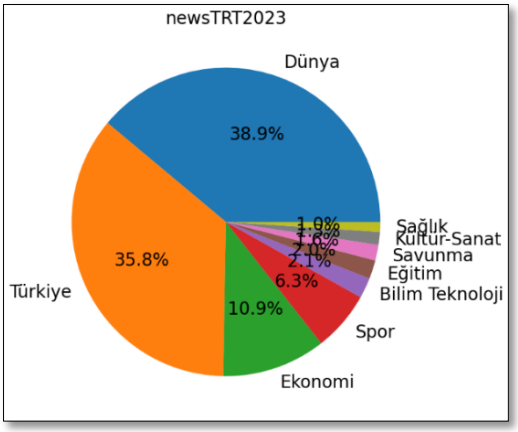
Metinler, öncelikle çeşitli ön işleme adımlarından geçirilerek, Lojistik Regresyon, Gradyan Yükseltme Makineleri gibi geleneksel makine öğrenimi algoritmaları ile sınıflandırılmak üzere işlenir. Ayrıca, Derin Sinir Ağları (DNN) [42], Evrişimsel Sinir Ağları (CNN) [43] gibi sıralı modeller ve Uzun-Kısa Süreli Bellek (LSTM) [44], Transformer [45] gibi gelişmiş derin öğrenme modelleri de metinleri dizisel olarak işlemek için kullanılır. Bunun yanı sıra, Graf Dikkat Ağları (GAT) [46], Graf Evrişimsel Ağlar (GCN) [47] ve özellikle araştırmanın odağı olan ve kesikli kırmızı çizgilerle vurgulanan Graf Sinir Ağları (GNN) [48], metinler arasındaki karmaşık ilişkileri graf tabanlı yapılar kullanarak modellemek için tercih edilir. Bu yaklaşımlar, Türkçe'nin zengin morfolojik yapısını ve metinler arası ilişkileri derinlemesine analiz etme kapasitesini geliştirmeyi amaçlar.



Şekil 1. Metin sınıflandırmasına yönelik genel taksonominin gösterimi (Representation of general taxonomy for text classification)

3.1. Veri Kümesi (Dataset)

Bu çalışmada, Türkiye ve dünya gündeminden toplanan haber içeriklerini kapsamlı bir şekilde analiz etmek için "newsTRT2023" adında detaylı bir haber metni veri kümesi oluşturulmuştur. Dijital çağın hızlı bilgi akışı ve çeşitliliği göz önünde bulundurularak hazırlanan bu veri kümesi, TRT Haber'in resmi dijital platformlarından elde edilen haber içeriklerinden derlenmiştir [49]. Belirli bir tarih aralığını kapsayan bu veri kümesinde, haber başlıkları, metinleri, yayın tarihleri, etiketler ve haberlerde bahsedilen özgün entiteler, gelişmiş web kazıma teknikleri kullanılarak sistematik bir biçimde toplanmıştır. Bu veri kümesi, Türkçe haber metinlerinin sınıflandırılması ve ilişkisel analizi konusunda yapılacak araştırmalara zemin hazırlamak amacıyla tasarlanmıştır.



Şekil 2. "newstrt2023" veri kümesi kategori oranları ("newstrt2023" dataset category rates)

"newsTRT2023" veri kümesi, Şekil 2' de gösterildiği üzere, TRT-Haber web sayfasında belirli bir tarih aralığında (06.04.2023 ile 13.11.2023 tarihleri arası) yayınlanan haberleri içermektedir ve her bir haber metni titizlikle incelenmiştir. Bu veri kümesinde yer alan haberler, yayın tarihi ve saati, başlığı, web adresi, kategorisi, kısa özeti, tam metni, ilişkili etiketler ve haberde bahsedilen özgün entityler gibi detaylı özelliklerle zenginleştirilmiştir. Toplam 4027 haberden oluşan bu veri kümesi, 'Dünya', 'Türkiye', 'Ekonomi' gibi 9 farklı kategoriye ayrılmıştır, her kategori haberlerin çeşitli yönlerini temsil etmektedir. 'Dünya' kategorisi, 1567 haber ile en geniş kapsamlı segmenti oluştururken, 'Türkiye' ve 'Ekonomi' kategorileri sırasıyla 1442 ve 438 haber ile önemli birer bölümü temsil etmektedir. Ek olarak, bu veri kümesi, 2713 benzersiz etiket içermekte ve bu etiketler, haberlerin detaylı ilişkisel analizi ve sınıflandırılması için kritik bir rol oynamaktadır. Bu kapsamlı ve detaylı veri kümesi, Türkçe haber metinlerinin sınıflandırılması ve analizi konusundaki araştırmalara önemli bir katkı sağlamaktadır.

Açık kaynak olarak literatüre sunulacak olan 'newsTRT2023' veri kümesi, akademik araştırmalardan uygulamalı projelere kadar geniş bir kullanım alanına hitap eden zengin bir kaynak olarak öne çıkmaktadır. Medya içerik analizinden haber trendlerini izlemeye, kamuoyu araştırmalarından yapay zeka destekli haber sınıflandırma sistemlerine kadar çeşitli uygulamalar için idealdir. Bu veri kümesi, metin madenciliği ve doğal dil işleme tekniklerini kullanarak haber metinlerindeki yaygın kelimeleri, ana konuları analiz etme imkanı sunar. Ayrıca, bu veri kümesi, makine öğrenmesi modellerinin eğitilmesi ve test edilmesi için de değerli bir kaynak olarak kullanılabilir. Veri kümesinden seçilmiş örnekler Çizelge 2' de sunulmuştur, bu örnekler veri kümesinin çeşitliliğini ve zenginliğini göstermektedir. Çizelge 2' de, 'newsTRT2023' veri kümesinden seçilen örnekler yer almaktadır.

Bu örnekler, ilişkisel metin sınıflandırma görevlerinde kullanılmak üzere, veri kümesini metin (text), varlık (entity) ve etiket (label) bilgilerini içermektedir. Varlıklar, metinleri graf olarak temsil etmek ve metinler arasındaki ilişkileri görselleştirmek için kullanılmıştır. Bu yaklaşım, veri kümesindeki metinler arasındaki ilişkisel bağlantıları anlamak ve sınıflandırmak için etkili bir metod sunmaktadır. Çizelge 2, veri kümesinin nasıl işlendiğini ve ilişkisel metin sınıflandırmada kullanılacak temel öğeleri örneklendirerek göstermektedir.

3.2. Vektör ve Graf Temsili (Vector And Graph Representatıon)

Metinlerin sınıflandırılması için, ilk adım olarak metinleri uygun biçimde işleyerek temizlemek gerekmektedir.

Çizelge 3' te örneği gösterildiği üzere çeşitli metin temizleme adımları gerçekleştirilmiştir [50]. (harfleri küçültme (converting lowercase), noktalama işaretlerini çıkarma (removing punctuation), sayıları temizleme (removing numbers), fazla boşlukları kaldırma (removing extra spaces), özel karakterleri silme (removing special characters), durdurma kelimelerini (removing stopwords) çıkarma ve kök indirgeme ya da lemmatizasyon (performing stemming or lemmatization) gibi)

Bu işlemler tamamlandıktan sonra, metinler sınıflandırma algoritmalarının işleyebileceği bir formata, yani vektörel bir temsile dönüştürülür. Metinlerin vektörel hale getirilmesi sürecinde TF-IDF [51], Word2Vec [52], GloVe [53], FastText [54], BERT [55] gibi çeşitli vektör temsil modelleri kullanılabilir.

Bu çalışmada ise, özellikle Türkçe metinler için optimize edilmiş olan Turkish BERT [55] vektör modeli tercih edilmiştir. Bu yaklaşım, metinlerin daha etkin bir şekilde işlenmesini ve sınıflandırılmasını sağlayarak, metin sınıflandırma algoritmalarının daha doğru ve etkili sonuçlar üretmesine katkıda bulunmaktadır.

Metinlerin graf temsilleri, metinleri matematiksel graf olarak modelleme yöntemleridir ve bu temsiller, metin içerisindeki kelimeler, kavramlar veya cümleler arasındaki ilişkileri ve bağlantıları görselleştirmek için kullanılır.

Çizelge 2. "newstrt2023" veri kümesine ait örnekler (Examples of the "newstrt2023" dataset)

Haber Başlığı	İçerik (Metin)	Etiketler	Kategori
AKSUNGUR 30 bin feet yükseklikte uçuş gerçekleştirdi	Türk Havacılık Uzay Sanayii AŞ (TUSAŞ) tarafından yerli ve milli imkanlarla geliştirilen Aksungur silahlı insansız hava aracı (SİHA), TUSAŞ Motor Sanayii (TEI) tarafından üretilen … (haber içerik devami)	'TUSAŞ', 'AKSUNGUR', 'SİHA', 'Yerli ve Milli Teknolojiler'	Savunma
Geçen yıl en çok turist ağırlayan 5. ülke Türkiye oldu	Söz konusu raporda, turizm endüstrisi alanında değişen dinamiklerin ve trendlerin öncelikleri etkilediği belirtilirken … (haber içerik devami)	'gemi', 'Şampiyonlar Ligi', 'Muğla', 'dolar', 'Antalya', 'İran','turizm', 'Turizm', 'EY web sitesinden', 'Dubai'	Ekonomi

Çizelge 3. Veri kümesine ait metin ön işleme öncesi ve sonrası örnek (Example of before and after text preprocessing of dataset)

Metin Ön işleme Öncesi ve Sonrası	
Önce	Gençlik ve Spor Bakanı Mehmet Muharrem Kasapoğlu, Türkiye'nin 2028 ve 2032'de gerçekleştirilecek Avrupa Futbol Şampiyonalarına ev sahipliği yapmak için adaylık dosyasını imzaladı. Adaylık dosyası, 12 Nisan'da Türkiye Futbol Federasyonu (TFF) tarafından UEFA'ya sunulacak.
Sonra	Gençlik ve spor bakanı mehmet muharrem kasapoğlu türkiye'nin ve gerçekleştirilecek avrupa futbol şampiyonalarına ev sahipliği yapmak için adaylık dosyasını imzaladı adaylık dosyası türkiye futbol federasyonu tff tarafından uefa'ya sunulacak
Etiketler	'mehmet muharrem kasapoğlu', 'tff'

Graf tabanlı yöntemler, sosyal medya gönderileri, forum içerikleri, URL'ler gibi çeşitli kaynaklardan elde edilen geniş kapsamlı metin veri kümelerinin işlenmesi için özel olarak geliştirilmiştir. Bu metodolojiler, metin verilerini analiz etmek ve sistemlerdeki ilgili detayları tespit etmek için oldukça etkilidir. Metinler, içerdikleri kelimeler, anlam yapıları veya varlık ağlarına göre grafiksel olarak temsil edilebilir [56].

3.3. Sınıflandırma Yöntemleri Ve Graf Sinir Ağları (Classıfıcatıon Methods And Graph Neural Networks)

Metin sınıflandırma sürecinde, metinlerin öncelikle ön işlemeden geçirilmesi ve ardından vektörel bir forma dönüştürülmesi gerekmektedir. Bu vektör dönüşümün ardından, elde edilen veriler bir sınıflandırma modeline girdi olarak sunulur. Bu sınıflandırma modelleri, makine öğrenmesi algoritmaları veya sinir ağı tabanlı modeller olabilir.

Bu çalışmada, ilişkisel metin graf yapısının vektör formuna dönüştürülmesi süreci takip edilmiş ve sonrasında bu veriler bir sinir ağı yapısında eğitim ve test işlemlerine tabi tutulmuştur. Bu işlemle, bir graf sinir ağı yapısı (GNN) elde edilmiştir. Elde edilen bu GNN modeli, farklı sınıflandırıcılarla karşılaştırmalı olarak değerlendirilmiş ve performans analizi yapılmıştır. Bu yaklaşım, Türkçe ilişkisel metin sınıflandırma alanında yeni ve etkili bir metodoloji olarak öne çıkmakta ve bu alandaki mevcut çalışmalara önemli bir katkı sunmaktadır.

Graf Sinir Ağları (GNN), yapısal verilerin işlenmesi ve analizi için kullanılan güçlü bir yapay zeka yaklaşımıdır. GNN' 'ler, veri noktaları arasındaki ilişkileri ve bağlantıları modelleyerek, bu bağlantıların veri üzerindeki etkilerini anlamak için graf teorisine dayalı yöntemler kullanır. Bu yaklaşım, düğümler (noktalar) ve kenarlar (bağlantılar) arasındaki etkileşimleri analiz eder ve özellikle sosyal ağlar, moleküler yapılar ve ulaşım ağları gibi karmaşık yapıların modellemesi için idealdir. GNN' nin çift yönlü işleme özelliği, düğümlerin hem gelen hem de giden bağlantılarına bakarak grafın topolojik yapısını anlama konusunda etkilidir. Ayrıca, her düğüm komşu düğümlerden bilgi toplayarak yerel ve küresel öğrenme sağlar, bu da GNN'leri düğümlerin ve kenarların bağlamsal özelliklerini anlamada ve yorumlamada yüksek başarıya ulaştırır.

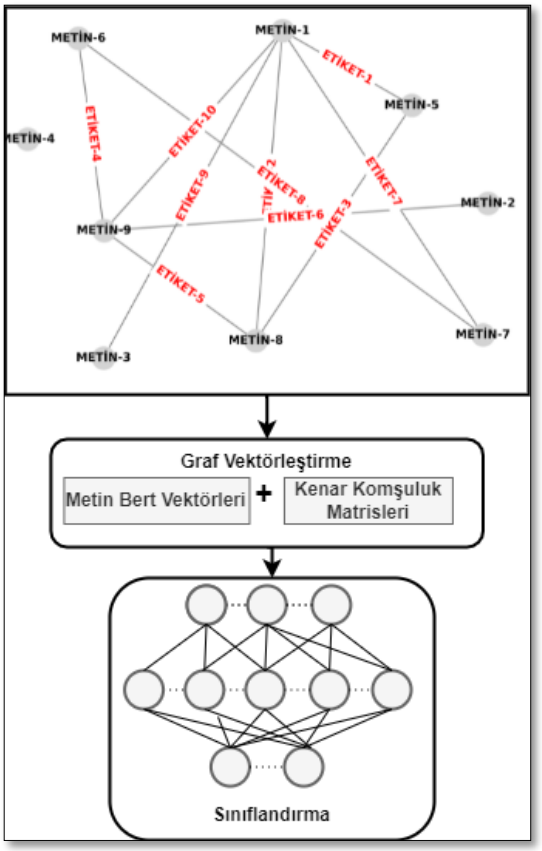
Derin Sinir Ağları (DNN), çok katmanlı yapay sinir ağlarından oluşur ve bu yapısı sayesinde, verilerdeki karmaşık özellikleri ve ilişkileri öğrenme kapasitesine sahiptir. Bu yaklaşım, girdi olarak alınan verilerin birden fazla katmandan geçirilerek daha derinlemesine işlenmesine dayanır ve bu süreç, verilerin daha karmaşık ve soyut temsillerinin oluşturulmasını sağlar. Bu çalışmada graf yapısı bir DNN modeline girdi olarak verilerek bir GNN yapısı elde edilmektedir.

GNN, haber metinlerinin sınıflandırılması ve ilişkisel analizlerinde kullanılmıştır.

Haber metinleri, çeşitli öğeler (anahtar kelimeler, konular, etiketler, varlıklar) ve bu öğeler arasındaki ilişkiler açısından bir graf yapısı olarak modellenmiştir. GNN'nin bu yapının analizi için kullanılması, haber metinlerindeki farklı öğeler arasındaki ilişkilerin derinlemesine anlaşılmasını sağlamıştır. Ayrıca, haber öğelerinin ve ilişkilerinin daha ayrıntılı bir şekilde sınıflandırılmasını mümkün kılmış ve haber metinlerinin yapısal özelliklerinin daha iyi anlaşılmasına katkıda bulunmuştur. Bu yaklaşım, haber metinlerinin kategorizasyonu ve tematik analizinde önemli bir ilerleme sağlayarak, metin sınıflandırma sürecini daha etkin ve kesin bir hale getirmiştir.

3.4. Önerilen Yöntem (Suggested Method)

Bu çalışma, Türkçe metin sınıflandırma ve ilişkisel analizi alanında, metin graf yapısı ve Derin Sinir Ağları (DNN) teknolojilerini içeren bir yaklaşımı benimsemiştir. Bu önerilen metodoloji, haber metinlerinin analizi için graf yapısının ve DNN' nin güçlü yönlerini birleştirmektedir. GNN ile haber metinleri anahtar kelimeler (etiketler), varlıklar ve bu öğeler arasındaki ilişkileri yardımı ile graf tabanlı bir sinir ağı modeli elde edilmektedir.



Şekil 3. Graf sinir ağı ile ilişkisel metin sınıflandırması (Relational text classification with graph neural network)

Bu yaklaşım, haber metinlerinin yapısal ve bağlamsal özelliklerinin derinlemesine analizini sağlar ve karmaşık ilişkileri aydınlatır. Ayrıca metinlerin kelime seçiminden cümle yapısına ve genel tonuna kadar olan çok boyutlu özelliklerini işleyerek, daha derinlemesine ve detaylı bir sınıflandırma yapılmasına imkân tanır.

Çalışmada kullanılan "newsTRT2023" veri kümesi, 4027 adet Türkçe haber metni içermektedir. Her haber metni, yayın tarihi, başlık, web adresi, kategori, kısa özet, tam metin, ilişkili etiketler ve haberde bahsedilen varlıklar gibi çeşitli özellikleri içerir. Bu zengin veri kümesi, modelin geliştirilmesi ve test edilmesi için ideal bir zemin sunar. Veri kümesi, %70 eğitim ve %30 test olmak üzere ikiye ayrılmıştır (2818 adet eğitim, 1209 adet test verisi). Bu yapı, modelin performansını değerlendirmek için dengeli bir yaklaşım sağlar. Ayrıca, GNN (Grafik Sinir Ağları), bu veri kümesindeki haber metinlerinin daha ayrıntılı analizini ve etkili bir şekilde sınıflandırılmasını sağlar. Çalışmada kullanılan graf modelin temsili gösterimi, Şekil 3'te sunulmuştur.

Haber metinlerinin ilişkisel etiketler temelinde modellendiği bir graf yapısını gösteren bu yapıda, düğümler haber metinlerini, kenarlar ise bu metinler arasındaki etiket bazlı ilişkileri (ortak etiketlerin varlığı durumunda) temsil etmektedir. Buna göre, Haber makalelerine etiket olarak eklenmiş olan özgün varlıklar (entityler) temel alınarak bir metin graf yapısı oluşturulmuştur. Bu yaklaşım, haber metinlerinin içerdiği bilgilerin grafiksel temsili üzerinden daha detaylı ve derinlemesine analizler yapılmasına olanak tanımaktadır. Bu graf temsiller, metin içerisindeki ilişkisel bağlantıları ve yapıları anlamak için güçlü bir araç olarak öne çıkmaktadır.

Haber metinlerini ilişkisel etiketler yardımıyla modellemede kullanılan graf yapısı için matematiksel bir model oluşturulmuştur. Oluşturulan graf yapısının matematiksel gösterimi Formül 1'deki gibidir.

$$f(G) = GNN(X, A) \quad (1)$$

Bu formül, Matematiksel işlem, Graf notasyonu, düğüm özellikleri, kenar ağırlıkları, kenar komşuluk matrisi ve graf vektörleştirme adımlarını içermektedir. İlgili adımlar maddeler halinde gösterilmiştir:

Graf Notasyonu:

$$G = (V, E) \quad (2)$$

Formül 2'de tanımlanan graf G, düğümler kümesi V ve kenarlar kümesi E olmak üzere.

$$V = \{v_1, v_2, \ldots, v_n\} \quad (3)$$

Formül 3'de her v_i bir haber metnini temsil eder.

$$E \subseteq \{ (v_i, v_j) \mid v_i, v_j \in V \} \quad (4)$$

Formül 4'de (v_i, v_j) düğümleri arasındaki bir kenarı temsil eder.

Düğüm Özellikleri:

Her düğüm v_i için, BERT modeli kullanılarak elde edilen semantik vektör x_i 'dir.

Kenar Ağırlıkları:

W_{ij}, düğüm çiftleri v_i ve v_j arasındaki ilişkinin gücünü ve sıklığını temsil eder. (W_{ij}, iki haber metni arasındaki ilişkinin gücünü ve sıklığını belirlemek için kullanılır. Bu değer, varlık benzerliği (iki haber metninde geçen ortak kişi, yer veya nesne gibi varlıkların oranı), etiket benzerliği (ortak etiketlerin oranı) ve metin benzerliği (metinlerin vektör temsilleri arasındaki benzerlik) bileşenlerinin belirli ağırlıklarla birleştirilmesiyle hesaplanır. Her bir bileşene verilen ağırlıklar, modelin performansını optimize etmek için ayarlanır ve bu şekilde, W_{ij} değeri iki haber metni arasındaki bağlantının gücünü ve sıklığını temsil eder.)

Kenar Komşuluk Matrisi:

A grafın topolojik yapısını temsil eden kenar komşuluk matrisidir. Eğer $(v_i, v_j) \in$ E ise $A_{ij} = w_{ij}$, aksi halde $A_{ij} = 0$.

Graf Vektörleştirme:
Grafın vektörleştirilmesi için, her düğüm için semantik vektörler ve kenar komşuluk matrisi kullanılarak özellik matrisi X oluşturulur.

Graf Sinir Ağı (GNN):
GNN modeli, grafı ve onun özelliklerini kullanarak sınıflandırma ya da diğer tahminler yapar ve f(G) olarak ifade edilir.

Model Eğitimi ve Testi:
Model eğitimi, belirlenen kayıp fonksiyonu L ve hiperparametrelerle optimize edilir. Test işlemleri, modelin performansını ölçmek için çeşitli metriklerle değerlendirilir.
Bu adımlara göre; G, haber metinlerini ve bunlar arasındaki ilişkileri temsil eden bir graf yapısıdır. X, haber metinlerinin özelliklerini temsil eden matristir. Bu özellikler, BERT modeli kullanılarak elde edilen semantik vektörlerdir. Her bir vektör, bir haber metninin içeriğini yüksek boyutlu bir uzayda temsil eder. A, grafın kenar komşuluk matrisidir ve grafın topolojik yapısını temsil eder.
Bu matris, haber metinleri arasındaki ilişkilerin varlığını ve gücünü belirten sayısal değerler içerir. Matristeki her bir değer (Aij), i ve j düğümleri (haber metinleri) arasındaki bağlantının varlığını ve şiddetini gösterir. Bu matematiksel model, haber metinlerinin karmaşık yapılarını çözümlemek ve anlamak için kullanılmıştır.
Elde edilen zengin ve çok boyutlu graf vektör yapısı, graf sinir ağı (GNN) tabanlı sınıflandırma modeline entegre edilmiştir. Bu model, graf temsillerini ve derin öğrenme tekniklerini bir araya getirerek, metinlerin sınıflandırılmasını ve ilişkisel bağlamda anlamlandırılmasını hedeflemektedir.
Çalışmada kullanılan GNN modeli, giriş özelliklerinin boyutuna göre yapılandırılmış üç tam bağlantılı (fully connected) katmandan oluşmaktadır; bu katmanlardan ilki ve ikincisi ReLU aktivasyon fonksiyonunu kullanırken, son katmanda sınıf sayısına uygun olarak nöronlar ve logaritmik softmax aktivasyon fonksiyonu uygulanmıştır. Modelin optimizasyon süreci, öğrenme oranı 0.005 ve weight_decay değeri 5e-4 olan Adam optimizörü ile yürütülmüştür.
Eğitim süreci 50 epoch boyunca devam etmiş, her epoch sonunda modelin eğitim ve test kayıpları ile doğruluk oranları kaydedilmiştir. Bu süreç boyunca, modelin performansı eğitim ve test setleri üzerinde değerlendirilen kayıp ve doğruluk metrikleri ile sürekli olarak izlenmiştir. Ayrıca, modelin hiperparametreleri optimize edilmiş ve genelleştirme kabiliyetini doğrulamak amacıyla test işlemleri gerçekleştirilmiştir.
Bu süreçte model performansı; Doğruluk (Accuracy), F1 skoru, Kesinlik (Precision) ve Duyarlılık (Recall) gibi çeşitli metriklerle değerlendirilmiştir. Bu kapsamda bu metriklere ait formüller içerisinde belirtilen hesaplamalar için Doğru Pozitif (TP), modelin gerçekten pozitif olan örnekleri doğru bir şekilde pozitif olarak sınıflandırmasıdır. Yanlış Pozitif (FP), aslında negatif olan örneklerin pozitif olarak sınıflandırılmasıdır. Doğru Negatif (TN), gerçekten negatif olan örneklerin doğru bir şekilde negatif olarak sınıflandırılmasıdır. Yanlış Negatif (FN), aslında pozitif olan örneklerin negatif olarak sınıflandırılmasıdır.
Bunun yanı sıra, modelin sonuçları, farklı makine öğrenimi algoritmaları ile karşılaştırmalı olarak analiz edilmiştir. Bu sayede, modelin sınıflandırma performansı derinlemesine incelenmiş ve karşılaştırmalı sonuçlar elde edilmiştir. Performans ölçütleri ve formüller:

Kesinlik (Precision): Kesinlik, Pozitif olarak tahmin edilen değerlerin aslında ne kadarının Pozitif olduğunu gösterir (Formül 5).

$$\text{Kesinlik} = \frac{\text{TP}}{TP+FN} \quad (5)$$

Duyarlılık (Recall): Duyarlılık (Recall), aslında Pozitif olması gereken örneklerin ne kadarının model tarafından doğru bir şekilde Pozitif olarak tahmin edildiğini gösteren bir metriktir (Formül 6).

$$\text{Duyarlılık} = \frac{\text{TP}}{TP+FP} \quad (6)$$

F1 Skor: F1 Skoru, Kesinlik ve Duyarlılık değerlerinin harmonik ortalamasını gösterir ve bu metrik, özellikle dengesiz veri kümelerinde model performansını değerlendirmek için kullanılır (Formül 7).

$$\text{F1 Skor} = \frac{2 \times \text{Duyarlılık} \times \text{Kesinlik}}{Duyarlılık + Kesinlik} \quad (7)$$

Doğruluk (Accuracy): Doğruluk değeri, modelde doğru tahmin edilen alanların toplam veri kümesine oranıyla hesaplanır (Formül 8).

$$\text{Doğruluk} = \frac{\text{TP} + \text{TN}}{TP+TN+FP+FN} \quad (8)$$

Karşıtlık Matrisi (Confusion Matrix): Karşıtlık Matrisi, sınıflandırma modelinin performansını değerlendirmek için kullanılan bir tablodur. Her sınıf için gerçek ve tahmin edilen sınıf sayısını gösterir. Bu matris, modelin hangi sınıfları doğru veya yanlış tahmin ettiğini anlamak için önemli bir araçtır.

4. DENEYSEL SONUÇLAR VE PERFORMANS ÖLÇÜTLERİ (EXPERIMENTAL RESULTS AND PERFORMANCE METRICS)

Bu çalışmada, Türkçe haber metinlerinin sınıflandırılması ve ilişkisel analizi için graf tabanlı sinir ağları ve derin öğrenme tekniklerinin entegrasyonu üzerine yoğunlaşılmıştır. Araştırmamızda, çeşitli makine öğrenimi modelleri, sinir ağı modeli (DNN) ve gelişmiş graf sinir ağı (GNN) yöntemi karşılaştırmalı olarak değerlendirilmiştir.

Çalışmada kullanılan çeşitli kütüphaneler, ilişkisel graf tabanlı sinir ağı sınıflandırması ve metin ön işleme işlemleri için gerekli araçları sağlamaktadır. Bu kütüphaneler arasında Türkçe metinlerde kök bulma işlemleri için TurkishStemmer [57], doğal dil işleme görevleri için NLTK[58], SPARQL sorgularını yürütmek için sparqlwrapper [59], BERT gibi önceden eğitilmiş dil modelleriyle çalışmak için transformers ve grafik sinir ağları için kullanılan PyTorch tabanlı torch_geometric [60] bulunmaktadır.

Bu kütüphaneler, metin ön işleme, doğal dil işleme ve graf tabanlı sinir ağı modellerinin geliştirilmesi sürecinde önemli bir rol oynamıştır. Ayrıca, bu süreçte çeşitli metin ön işleme adımları ve grafik verilerinin işlenmesine yönelik kodlar da kullanılmıştır. Elde edilen sonuçlar, Çizelge 4'te özetlenmiştir. Çizelge 4' teki sonuçlar, Graf Sinir Ağı (GNN) modelinin, diğer modellere göre önemli ölçüde daha yüksek doğruluk, geri çağırma, kesinlik ve F1 skoru değerlerine sahip olduğunu göstermektedir.

Çizelge 4. Deneysel sonuçlar (Experimental results)

Model	**Accuracy**	**Recall**	**Precision**	**F1-Score**
Logistic Regression	89.71	89.71	89.54	89.43
Light Gradient Boosting Machine	86.98	86.98	85.74	85.74
Extreme Gradient Boosting	86.66	86.66	86.24	85.69
Extra Trees Classifier	83.68	83.68	82.22	80.85
Random Forest Classifier	83.85	83.85	82.30	81.08
Gradient Boosting Classifier	83.82	83.82	82.09	82.29
Linear Discriminant Analysis	85.20	85.20	85.27	85.12
K Neighbors Classifier	86.62	86.62	86.45	85.97
SVM - Linear Kernel	89.11	89.11	88.87	88.59
Ridge Classifier	86.05	86.05	85.72	85.39
Deep Neural Network (DNN)	92.14	72.18	85.44	77.00
Graph Neural Network (GNN)	**97.93**	**92.22**	**95.08**	**93.59**

GNN, haber metinlerinin yapısal ve bağlamsal özelliklerini derinlemesine analiz ederek karmaşık ilişkileri başarıyla çözümlemiştir. Bu, GNN'nin, metin sınıflandırmada ve ilişkisel analizde son derece etkili bir araç olduğunu kanıtlamaktadır. Ayrıca, Deep Neural Network (DNN) modelinin de yüksek doğruluk oranına sahip olduğu görülmektedir. Ancak, GNN modeline göre daha düşük metrik değerlerine sahip olması, DNN'nin belirli bağlamlarda GNN kadar etkili olmadığını göstermektedir.

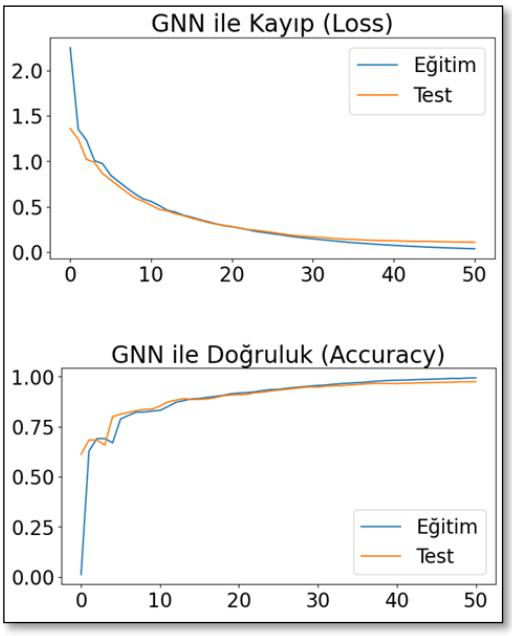
Sistemin performansını anlamak ve modelin sağlıklı bir şekilde eğitildiğinden emin olmak için performans metrikleri kritik bir öneme sahiptir. Bu metrikler, modelin eğitim ve test aşamalarında ne ölçüde başarılı olduğunu anlamamızı sağlar ve potansiyel problemleri önceden görmemize olanak tanır. Performans analizi, modelin eğitim verilerini ne kadar iyi öğrendiği ve öğrenilen bilgileri genel veri kümeleri üzerinde ne kadar iyi uygulayabildiği konusunda bize net bilgiler verir. Şekil 4'te gösterilen kayıp ve doğruluk gibi metrikler GNN'nin öğrenme evresi boyunca nasıl geliştiğini ayrıntılı bir şekilde ortaya koyar.

Şekil 4'teki GNN ile Doğruluk grafiği, modelin eğitim ve test kümelerindeki sınıflandırma doğruluğunu gösterir. Her iki veri kümesi için de doğruluk, eğitim sürecinin başında hızla artmış ve 10. epoch civarında yüksek bir değere ulaşmıştır.

Eğitim ve test doğruluk değerleri arasında belirgin bir fark olmaması, modelin iyi bir genelleme yapabildiğini ve görmediği veriler üzerinde de yüksek performans gösterebildiğini göstermektedir.

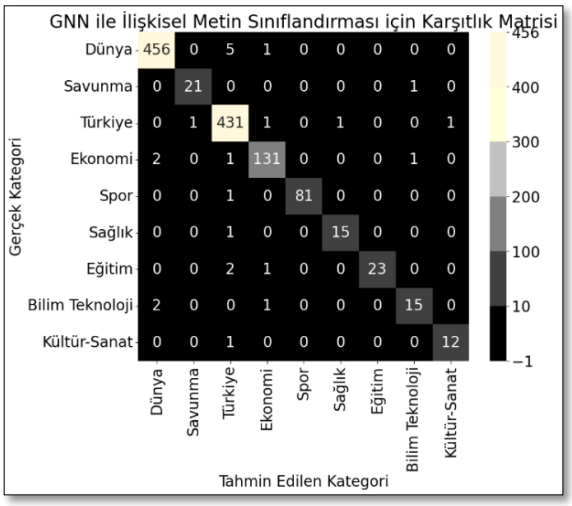
Şekil 4' de gözlemlenen yüksek doğruluk ve düşük kayıp değerleri, GNN modelinin veri kümesi üzerinde etkili bir şekilde öğrendiğini ve iyi bir genelleme kapasitesine sahip olduğunu göstermektedir. Bu, özellikle 50 epoch'a kadar devam eden istikrarlı performansla daha da pekişmektedir. Bu durum, modelin karmaşık haber metni veri kümesini anlama ve sınıflandırma konusunda başarılı olduğunu ve güvenilir sonuçlar üretebildiğini gösterir.

Eğitim ve test kümele arasındaki performansın benzerliği, modelin overfitting'e (aşırı uyum) maruz kalmadığını işaret eder. Bu, modelin yalnızca eğitim verilerini ezberlemek yerine, genel bir öğrenme ve genelleme yapabildiğini gösterir, bu da gerçek dünya verilerine uygulandığında da yüksek performans gösterebileceği anlamına gelir.



Şekil 4. GNN model eğitim-test sırasındaki kayıp ve doğruluk eğrisi (Loss and accuracy curve during GNN model training-testing)

Ayrıca, modelin performansını analiz etme bağlamında, karşıtlık matrisi gibi araçlar da büyük önem taşımaktadır. Şekil Y'de sunulan karşıtlık matrisi, modelin sınıflandırma kabiliyetini ve tahminlerinin hassasiyetini detaylı bir biçimde göstererek, kayıp ve doğruluk metriklerinin yanı sıra, modelin performansını çok boyutlu bir perspektiften değerlendirmemize olanak tanır.



Şekil 5. GNN ile ilişkisel metin sınıflandırması için karşıtlık matrisi (Confusion matrix for relational text classification with GNN)

Şekil 5' te görüldüğü üzere karşıtlık matrisi, modelin farklı kategorilerdeki haber metinlerini ne kadar iyi sınıflandırdığını gösterir. Diyagonal üzerindeki büyük değerler (örneğin, 'Dünya' kategorisi için 458, 'Türkiye' için 430, 'Ekonomi' için 128 gibi), modelin çoğu durumda doğru tahminlerde bulunduğunu gösterir, yani bu kategorilerdeki haber metinlerini doğru bir şekilde sınıflandırabilmiştir. Diyagonal dışındaki değerler (örneğin, 'Dünya' kategorisindeki 2 'Ekonomi' yanlış sınıflandırması gibi), modelin bazı kategorilerde hatalı sınıflandırmalar yaptığını gösterir. Bununla birlikte, genel olarak modelin sınıflandırma performansı oldukça yüksek ve hata oranları düşüktür, bu da GNN'nin metin sınıflandırma konusunda etkili olduğunu göstermektedir. Özellikle yüksek hacimli (çok örnekli) kategorilerde ('Dünya' ve 'Türkiye') modelin iyi bir doğrulukla çalıştığı gözlemlenmiştir.

Buna göre, Türkçe "newsTRT2023" veri kümesi üzerindeki çeşitli makine öğrenimi modellerinin performansları, Graf Sinir Ağı (GNN) mimarisiyle kıyaslanarak değerlendirilmiştir. Çizelge 4' de, klasik sınıflandırma algoritmaları ve çeşitli derin öğrenme modelleri dikkate alınmış ve bunlarla önerilen GNN modelinin doğruluk oranları karşılaştırılmıştır. Elde edilen bulgular, GNN modelinin %97.93'lük etkileyici doğruluk oranı ile diğer algoritmaları geride bıraktığını göstermektedir. Bu sonuçlar, GNN'nin haber metinlerinin sınıflandırılması gibi zorlu görevlerde, metinler arası ilişkisel ve yapısal özellikleri derinlemesine analiz ederek bağlamsal anlayışı artırma yeteneğine işaret etmektedir. Önerilen GNN mimarisi, literatürde sunulan diğer çalışmalarla kıyaslandığında (Çizelge 1), Türkçe metin sınıflandırma problemlerine yönelik yenilikçi bir yaklaşım sergileyerek öne çıkmaktadır. Literatürdeki çok dilli ve İngilizce odaklı veri kümeleri üzerinde elde edilen sonuçlarla karşılaştırıldığında, "newsTRT2023" veri kümesinde yapılan deneyler benzer ya da daha yüksek başarımlar elde etmiştir. Bu, GNN mimarilerinin Türkçe gibi morfolojik olarak zengin dillerin karmaşıklığını ele alabilme kapasitesinin bir göstergesidir. Ayrıca, derin öğrenme ile graf tabanlı analiz tekniklerinin bütünleşik kullanımının, dilin bağlamsal ve yapısal özelliklerini başarıyla modelleyebildiğini ve böylece yüksek doğruluk oranlarına ulaşabildiğini kanıtlamaktadır.

5. TARTIŞMA (DISCUSSION)

Bu bölümde, "newsTRT2023" veri kümesi üzerinde uygulanan Graf Sinir Ağları (GNN) tabanlı modelin sonuçları, bulguları ve bu bulguların Türkçe metin sınıflandırma ve ilişkisel analiz alanına etkileri tartışılmaktadır.

Çalışmanın en önemli bulgularından biri, modelin Türkçe haber metinlerinin sınıflandırılmasında ve ilişkisel analizinde gösterdiği yüksek başarı oranıdır. Model, GNN' nin graf tabanlı analitik gücü ile derin öğrenme kabiliyetlerini başarılı bir şekilde birleştirmiş ve bu sayede haber metinlerinin daha detaylı ve doğru bir şekilde sınıflandırılmasını sağlamıştır. Bu yaklaşım, metinlerin daha kapsamlı bir şekilde incelenmesine ve karmaşık yapılarının daha iyi anlaşılmasına olanak tanımıştır. Model, Türkçe haber metinlerinin çeşitli özelliklerini ve bağlamsal ilişkilerini başarılı bir şekilde analiz etmiş ve bu analizler sonucunda elde edilen bilgiler, Türkçe doğal dil işleme alanındaki mevcut bilgi birikimine önemli katkılarda bulunmuştur. Modelin

başarısı, Türkçe metin analizi ve sınıflandırmasında yeni yaklaşımların potansiyelini ortaya koymaktadır. Bu model, medya analizi, kamuoyu araştırmaları ve yapay zeka destekli haber sınıflandırma sistemleri gibi alanlarda pratik uygulamalara sahip olabilir. Ayrıca, modelin sağladığı detaylı analizler, metin madenciliği ve doğal dil işleme tekniklerinin geliştirilmesinde yeni yollar açabilir.

Çalışmada kullanılan veri kümesi, sınıflar arasında bazı farklılıklar göstermektedir. Özellikle bazı sınıflarda (örneğin, Sağlık, Kültür-Sanat, Savunma) diğer sınıflara kıyasla daha az sayıda örnek bulunmuştur. Bu durum, modelin bu sınıflardaki tahmin performansını etkileyebilir. Ancak, mevcut çalışmada uygulanan modelleme ve optimizasyon teknikleri, bu durumun etkilerini en aza indirmeyi hedeflemiştir. Sonuçlar, veri kümesindeki bu farklılıklara rağmen, modelin genel performansının tatmin edici olduğunu ortaya koymaktadır. İlerleyen çalışmalarda, veri kümesinin daha dengeli hale getirilmesi için ek veri toplama süreçlerine odaklanılacak ve özellikle az temsil edilen sınıflar için daha fazla veri sağlanmaya çalışılacaktır. Bu sayede, modelin tüm sınıflar için daha tutarlı ve dengeli sonuçlar vermesi hedeflenmektedir.

Bu çalışma, Türkçe metin sınıflandırma ve ilişkisel analiz çalışmalarına yenilikçi bir katkı sunmaktadır. Bununla birlikte, modelin gücünü tam anlamıyla değerlendirebilmek için daha fazla sayıda ve çeşitlilik gösteren veri kümeleri üzerindeki testlerin yapılması, algoritmanın gücünü artırabilir ve onu çeşitli kullanım senaryolarına daha uygun hale getirebilir. Gelecekteki çalışmalar, mevcut modeli daha da genişletebilir ve Türkçe doğal dil işleme alanında daha geniş perspektifler sunabilir.

6. SONUÇ (CONCLUSION)

Gelişen teknoloji ve artan veri hacmi, dil işleme ve metin analizi alanlarında yenilikçi ve etkili çözümler geliştirmenin gerekliliğini ortaya koymaktadır. Bu bağlamda, Türkçe haber metinlerinin sınıflandırılması ve ilişkisel analizine yönelik bu çalışma, dil işleme alanındaki mevcut metodolojileri geliştirmek için Graf Sinir Ağları (GNN) kullanımını önermektedir. Çalışmada kullanılan "newsTRT2023" veri kümesi, Türkçe'nin zengin morfolojik yapısını ve metinler arasındaki ilişkileri kapsayarak, dil işleme tekniklerinin doğruluğunu ve kapsamını test etmek için ideal bir platform sunmaktadır. GNN modeli, haber metinlerinin yapısal ve bağlamsal özelliklerini derinlemesine analiz ederek, %97.93 gibi dikkate değer bir doğruluk oranı elde etmiştir. Bu sonuç, GNN'nin karmaşık ve ilişkisel metin verilerini işleme konusunda üstün bir başarıya sahip olduğunu göstermektedir. Ayrıca, bu araştırma, Türkçe metin sınıflandırma ve ilişkisel analizde kullanılan geleneksel ve derin öğrenme tabanlı modellerle karşılaştırmalı bir değerlendirme yaparak, GNN'nin bu modellere kıyasla üstün performansını kanıtlamaktadır. Bu bulgular, Türkçe metin sınıflandırma çalışmalarında GNN'nin önemli bir potansiyel taşıdığını vurgulamakta ve bu alandaki metodolojik yaklaşımları daha ileriye taşıma potansiyeline sahiptir. Araştırma, Türkçe metin analizinde derin öğrenme ve graf tabanlı yöntemlerin bütünleştirilmesinin, dilin zengin ve karmaşık yapısını başarılı bir şekilde anlamada ve işlemede ne kadar etkili olduğunu göstermektedir. Bu çalışma, dil işleme ve metin analizi alanlarında gelecekte yapılacak araştırmalara katkıda bulunarak, bu alanlardaki bilgi birikimini ve uygulama yelpazesini genişletmeyi hedeflemektedir.

ETİK STANDARTLARIN BEYANI (DECLARATION OF ETHICAL STANDARDS)

Bu makalenin yazarları çalışmalarında kullandıkları materyal ve yöntemlerin etik kurul izni ve/veya yasal-özel bir izin gerektirmediğini beyan ederler.

YAZARLARIN KATKILARI (AUTHORS' CONTRIBUTIONS)

Halil İbrahim OKUR: Çalışma konsepti ve tasarımına katkıda bulunmuş, veri toplamış, deneyler yapmış, sonuçların analiz ve yorumlanmasına katkıda bulunmuş ve makalenin hazırlanmasında rol almıştır.

Kadir TOHMA: Çalışma konsepti ve tasarımına katkıda bulunmuş, veri toplamış, deneyler yapmış, sonuçların analiz ve yorumlanmasına katkıda bulunmuş ve makalenin hazırlanmasında rol almıştır.

Ahmet SERTBAŞ: Çalışma konsepti ve tasarımına katkıda bulunmuş ve sonuçların analiz ve yorumlanmasına katkıda bulunmuştur.

ÇIKAR ÇATIŞMASI (CONFLICT OF INTEREST)

Bu çalışmada herhangi bir çıkar çatışması yoktur.

KAYNAKLAR (REFERENCES)

[1] Dhar, A., Mukherjee, H., Dash, N. S. ve Roy, K., "Text categorization: past and present.", ***Artificial Intelligence Review***, 54, 3007-3054, (2021).

[2] Kowsari, K., Jafari Meimandi, K., Heidarysafa, M., Mendu, S., Barnes, L., & Brown, D., "Text classification algorithms: A survey.", ***Information***, 10(4), 150, (2019).

[3] Minaee, S., Kalchbrenner, N., Cambria, E., Nikzad, N., Chenaghlu, M., & Gao, J., "Deep learning--based text classification: a comprehensive review.", ***ACM computing surveys (CSUR)***, 54(3), 1-40, (2021).

[4] Büyük, O., "Joint intent detection and slot filling for Turkish natural language understanding.", ***Turkish Journal of Electrical Engineering and Computer Sciences***, 31(5), 844-859, (2023).

[5] Pan, D., Yang, Z., Tan, H., Wu, J., & Lin, H., "Dialogue topic extraction as sentence sequence labeling.", ***In CCF International Conference on Natural Language Processing and Chinese Computing*** (pp. 252-262). Cham: Springer, (2022).

[6] Tohma, K., Okur, H. I., Kutlu, Y., & Sertbas, A., "Sentiment Analysis in Turkish Question Answering Systems: An Application of Human-Robot Interaction.", ***IEEE Access***, (2023).

[7] Nature Switzerland. Koru, G. K., & Uluyol, Ç., "Detection of Turkish Fake News from Tweets with BERT Models.", ***IEEE Access***, (2024).

[8] Karasoy, O., & Ballı, S., "Spam SMS detection for Turkish language with deep text analysis and deep learning methods***", Arabian Journal for Science and Engineering***, 47(8), 9361-9377, (2022).

[9] Çıtlak O., Dörterler M. ve Doğru İ. A., "A hybrid spam detection framework for social networks", ***Journal of Polytechnic***, 26(2): 823-837, (2023).

[10] Zucco, C., Calabrese, B., Agapito, G., Guzzi, P. H., & Cannataro, M., "Sentiment analysis for mining texts and social networks data: Methods and tools.", ***Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery***, 10(1), e1333, (2020).

[11] Shivakumara, P., Alaei, A., & Pal, U., "Mining text from natural scene and video images: A survey.", ***Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery***, 11(6), e1428, (2021).

[12] Pintas, J. T., Fernandes, L. A., & Garcia, A. C. B., "Feature selection methods for text classification: a systematic literature review.", ***Artificial Intelligence Review***, 54(8), 6149-6200, (2021).

[13] Okur, H.I., Tohma, K., Sertbas, A., "Relational turkish text classification using distant supervised entities and relations.", ***Computers, Materials & Continua***, 79(2), 2209-2228, (2024).

[14] Bilen B., Horasan F., "LSTM network based sentiment analysis for customer reviews", ***Politeknik Dergisi***, 25(3): 959-966, (2022).

[15] Tohma, K., & Kutlu, Y., "Challenges Encountered in Turkish Natural Language Processing Studies.", ***Natural and Engineering Sciences***, 5(3), 204-211, (2020).

[16] Gasparetto, A., Marcuzzo, M., Zangari, A., & Albarelli, A., "A survey on text classification algorithms: From text to predictions." ***Information***, 13(2), 83, (2022).

[17] [17] Li, Q., Peng, H., Li, J., Xia, C., Yang, R., Sun, L., ... & He, L. "A survey on text classification: From traditional to deep learning.", ***ACM Transactions on Intelligent Systems and Technology (TIST)***, 13(2), 1-41, (2022).

[18] Kumar, A. V., Gupta, D., & Venugopalan, M., "Cyberbullying Text Classification for Social Media Data Using Embedding and Deep Learning Approaches.", ***In 2023 14th International Conference on Computing Communication and Networking Technologies (ICCCNT)*** (pp. 1-6). IEEE, (2023).

[19] Türkmen, H., Dikenelli, O., Eraslan, C., Callı, M. C., & Özbek, S. S., "BioBERTurk: Exploring Turkish Biomedical Language Model Development Strategies in Low-Resource Setting.", ***Journal of Healthcare Informatics Research***, 7(4), 433-446, (2023).

[20] Umer, M., Imtiaz, Z., Ahmad, M., Nappi, M., Medaglia, C., Choi, G. S., & Mehmood, A., "Impact of convolutional neural network and FastText embedding on text classification.", ***Multimedia Tools and Applications***, 82(4), 5569-5585, (2023).

[21] Altinel Girgin, A. B., "Semantic text classification: A survey of past and recent advances.", (2018).

[22] Zhang, Y., Yu, X., Cui, Z., Wu, S., Wen, Z., & Wang, L., "Every document owns its structure: Inductive text classification via graph neural networks.", ***arXiv*** preprint arXiv:2004.13826, (2020).

[23] Nguyen, G., Dlugolinsky, S., Bobák, M., Tran, V., López García, Á., Heredia, I., ... & Hluchý, L., "Machine learning and deep learning frameworks and libraries for large-scale data mining: a survey.", ***Artificial Intelligence Review***, 52, 77-124, (2019).

[24] Cervetti, G. N., & Wright, T. S., "The role of knowledge in understanding and learning from text***", Handbook of reading research***, 5, (2020).

[25] Kowsari, K., Jafari Meimandi, K., Heidarysafa, M., Mendu, S., Barnes, L., & Brown, D., "Text classification algorithms: A survey.", ***Information***, 10(4), 150, (2019).

[26] Wu, L., Chen, Y., Shen, K., Guo, X., Gao, H., Li, S., ... & Long, B., "Graph neural networks for natural language processing: A survey.", ***Foundations and Trends® in Machine Learning***, 16(2), 119-328, (2023).

[27] Uslu, O., & Özmen-akyol, S., "Türkçe haber metinlerinin makine öğrenmesi yöntemleri kullanılarak sınıflandırılması.", ***Eskişehir Türk Dünyası Uygulama ve Araştırma Merkezi Bilişim Dergisi***, 2(1), 15-20, (2021).

[28] Çelik, Ö., & Koç, B. C., "TF-IDF, Word2vec ve Fasttext vektör model yöntemleri ile Türkçe haber metinlerinin sınıflandırılması.", ***Dokuz Eylül Üniversitesi Mühendislik Fakültesi Fen ve Mühendislik Dergisi***, 23(67), 121-127, (2021).

[29] Acı, Ç., & Çırak, A., "Türkçe haber metinlerinin konvolüsyonel sinir ağları ve Word2Vec kullanılarak sınıflandırılması.", ***Bilişim Teknolojileri Dergisi***, 12(3), 219-228, (2019).

[30] Karakurt, M., & KARCI, A., "GloVe Kelime Gömmeleri ve Sinir Ağları ile Haber Metinlerinin Sınıflandırılması.", ***International Journal of Pure and Applied Sciences***, 9(1), (2023).

[31] Parlak, B., "The effects of preprocessing on Turkish and English News Data.", ***Sakarya University Journal of Computer and Information Sciences***, 6(1), 59-66, (2023).

[32] Okur, H. I. ve Sertbaş, A., "Pretrained neural models for turkish text classification.", ***In 2021 6th International Conference on Computer Science and Engineering (UBMK)*** (pp. 174-179). IEEE, (2021).

[33] Demir, E., & Bilgin, M., "Sentiment Analysis from Turkish News Texts with BERT-Based Language Models and Machine Learning Algorithms.", ***In 2023 8th International Conference on Computer Science and Engineering (UBMK)*** (pp. 01-04). IEEE, (2023).

[34] Tohma, K., Okur, H. I., Kutlu, Y., & Sertbas, A., "Sentiment Analysis in Turkish Question Answering Systems: An Application of Human-Robot Interaction.", ***IEEE Access***, (2023).

[35] Malekzadeh, M., Hajibabaee, P., Heidari, M., Zad, S., Uzuner, O., & Jones, J. H., "Review of graph neural network in text classification.", ***In 2021 IEEE 12th annual ubiquitous computing, electronics & mobile communication conference (UEMCON)*** (pp. 0084-0091). IEEE, (2021).

[36] [36] Huang, L., Ma, D., Li, S., Zhang, X., & Wang, H., "Text level graph neural network for text classification.", ***arXiv*** preprint arXiv:1910.02356, (2019).

[37] Yao, L., Mao, C., & Luo, Y., "Graph convolutional networks for text classification.", ***In Proceedings of the AAAI conference on artificial intelligence*** (Vol. 33, No. 01, pp. 7370-7377), (2019).

[38] Kumar, V. S., Alemran, A., Karras, D. A., Gupta, S. K., Dixit, C. K., & Haralayya, B., "Natural Language Processing using Graph Neural Network for Text Classification.", ***In 2022 International Conference on Knowledge Engineering and Communication Systems (ICKES)*** (pp. 1-5). IEEE, (2022).

[39] Liu, C., Wang, X., & Xu, H., "Text Classification Using Document-Relational Graph Convolutional Networks.", ***IEEE Access***, 10, 123205-123211, (2022).

[40] Z. Chen et al., "Relational graph convolutional network for text-mining-based accident causal classification.", ***Applied Sciences*** 12.5, 2482. (2022).

[41] Liu, X., Tian, J., Niu, N., Li, J., & Han, J., "Standard Text" Relational Classification Model Based on Concatenated Word Vector Attention and Feature Concatenation.", ***Applied Sciences***, 13(12), 7119, (2023).

[42] LeCun, Y., Bengio, Y., & Hinton, G., "Deep learning.", ***Nature***, 521(7553), 436-444, (2015).

[43] LeCun, Y., Bottou, L., Bengio, Y., & Haffner, P., "Gradient-based learning applied to document recognition." ***Proceedings of the IEEE***, 86(11), 2278-2324, (1998).

[44] Hochreiter, S., & Schmidhuber, J., "Long short-term memory.", ***Neural computation***, 9(8), 1735-1780, (1997).

[45] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., ... & Polosukhin, I., "Attention is all you need.***", Advances in neural information processing systems***, 30, (2017).

[46] Veličković, P., Cucurull, G., Casanova, A., Romero, A., Lio, P., & Bengio, Y., "Graph attention networks.", ***arXiv*** preprint arXiv:1710.10903, (2017).

[47] Kipf, T. N., & Welling, M., "Semi-supervised classification with graph convolutional networks.", ***arXiv*** preprint arXiv:1609.02907, (2016).

[48] Scarselli, F., Gori, M., Tsoi, A. C., Hagenbuchner, M., & Monfardini, G., "The graph neural network model.", ***IEEE transactions on neural networks***, 20(1), 61-80, (2008).

[49] https://www.trthaber.com/tum-mansetler-sayfa-1.html, "TRT Haber Tüm Manşetler", (2023).

[50] Hickman, L., Thapa, S., Tay, L., Cao, M., & Srinivasan, P., "Text preprocessing for text mining in organizational research: Review and recommendations.", ***Organizational Research Methods***, 25(1), 114-146, (2022).

[51] Aizawa, A., "An information-theoretic perspective of tf–idf measures.", ***Information Processing & Management***, 39(1), 45-65, (2003).

[52] Mikolov, T., Grave, E., Bojanowski, P., Puhrsch, C., & Joulin, A., "Advances in pre-training distributed word representations.", ***arXiv*** preprint arXiv:1712.09405, (2017).

[53] Pennington, J., Socher, R., & Manning, C. D., "Glove: Global vectors for word representation.", ***In Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)*** (pp. 1532-1543), (2014).

[54] Joulin, A., Grave, E., Bojanowski, P., Douze, M., Jégou, H., & Mikolov, T., "Fasttext. zip: Compressing text classification models.", ***arXiv*** preprint arXiv:1612.03651, (2016).

[55] Devlin, J., Chang, M. W., Lee, K., & Toutanova, K., "Bert: Pre-training of deep bidirectional transformers for language understanding.", ***arXiv*** preprint arXiv:1810.04805, (2018).

[56] Hu, L., Zhang, M., Li, S., Shi, J., Shi, C., Yang, C., & Liu, Z., "Text-graph enhanced knowledge graph representation learning". ***Frontiers in Artificial Intelligence***, 4, 697, (2021).

[57] Çilden, E. K., "Stemming Turkish words using snowball.", (2006).

[58] [58] Bird, S., "NLTK: the natural language toolkit.", ***In Proceedings of the COLING/ACL 2006 Interactive Presentation Sessions*** (pp. 69-72), (2006).

[59] Herman, I., Fernández, S., & Tejo, C., "***SPARQL endpoint interface to python.",*** *URL: https://sparqlwrapper.readthedocs.io/en/latest/main.html* , (2012).

[60] Fey, M., & Lenssen, J. E., "Fast graph representation learning with PyTorch Geometric.", ***arXiv*** preprint arXiv:1903.02428, (2019).