



Multi-Label Text Analysis with GRU and CNN Based Hybrid Deep Learning Model

Halit Çetiner

EasyChair preprints are intended for rapid dissemination of research results and are integrated with the rest of EasyChair.

July 10, 2022

GRU ve CNN Tabanlı Hibrit Derin Öğrenme Modeli ile Çok Etiketli Metin Analizi

Halit Çetiner^{1*}

¹Teknik Bilimler MYO, Isparta Uygulamalı Bilimler Üniversitesi, Türkiye

*(halitcetiner@isparta.edu.tr)

Özet – Twitter, Facebook, Instagram, WhatsApp ve Wikipedia gibi sosyal ağlarda paylaşılan yazılı paylaşımlar her geçen gün artmaktadır. Sosyal ağlarda veri içerikleri herhangi bir kontrol olmaksızın artış göstermektedir. Artan bu yazılı sosyal paylaşım verileri içerisinden anlamlı, olumlu ve olumsuz veri içeriklerinin otomatik olarak analiz edilmesi önemlidir. Bu maksatla Wikipedia sosyal paylaşım sistemine ait yazılı paylaşımların içerik analizlerini gerçekleştirmek için metin analizi çalışması yapılmıştır. İçerik analizlerinde metin analizi ile verilerin ön işlenmesi ve sayısallaştırılması mümkün olsa da öznel haritasının elde edilmesinde derin öğrenme modellerine ihtiyaç bulunmaktadır. Metin analizinin gerçekleştirilmesinde doğal dil işleme yöntemleri kullanılmıştır. Analizi yapılan metinlerin içeriklerine göre çok etiketli metin sınıflandırılmasında Gated Recurrent Unit (GRU) ve Convolutional Neural Network (CNN) tabanlı bir hibrit model ile gerçekleştirilmiştir. Doğal dil işleme teknikleri, Wikipedia kullanıcı yorumları girdi olarak alındıktan sonra ön işlenmiştir. Ön işlenen metin verileri GloVe gömme katmanı ile sayısallaştırılmıştır. Sayısallaşan metinler, çok katmanlı derin öğrenme mimarisine girdi olarak verilerek işlenmiştir. Binlerce Wikipedia yorumunun makine öğrenmenin aktif alt alanlarından biri olan derin öğrenme ile sınıflandırılması belirtilen adımlar ile gerçekleştirilmiştir. Önerilen yöntem ile gerçekleştirilen deneysel çalışmalarda elde edilen performans metrikleri sunulmuştur.

Anahtar Kelimeler – GRU, CNN, derin öğrenme, çok etiketli metin analizi, Wikipedia

I. GİRİŞ

İnsanlar öğrenci, komşu, iş gibi farklı koşul ve ortamlarda tanıştığı arkadaşlıklar ile bağlantılarını koparmamak ve yeni bağlantılar sağlamak için Twitter, Instagram, Facebook, Wikipedia gibi sosyal ağlarda hesaplar açmaktadır. Hatta farklı amaçlara göre birden fazla hesap bile açmaktadır. Belirtilen sosyal ağlar başta olmak üzere günlük yaşantımızdaki olaylar, durumlar ve düşünceler ucuz ve hızlı bir şekilde paylaşılmaktadır. Bu sebepten dolayı belirtilen sosyal paylaşım ağlarına halk tarafından yoğun talep görülmektedir. Bunun sonucunda da her geçen gün kullanıcı sayısı artış göstermektedir [1]. Artan kullanıcı sayısı ile birlikte yapılan yorum ve mesaj sayısında da hızlı bir artış bulunmaktadır [2].

Wikipedia başta olmak üzere kullanıcısı her geçen gün artan sosyal ağlarda veri ve haberler

yalan olabildiği gibi olumsuz ve içeriği olumsuz haberler kullanıcıları negatif olarak etkilemektedir. Bu tür algı içerikli içeriklerin artmasının ana sebebi ise sosyal ağların kontrol dışında olmasıdır. Herhangi bir kontrol olmayan olumsuz düşünce, davranış ve eylemlerin sosyal ağlar üzerinde paylaşılması, bu tür olumsuz bilgi içeriklerinin her geçen gün artmasına sebep olmaktadır. Gerçekleştirilen bu çalışma, internet ortamında Wikipedia sosyal ağlarında yazılmış olan binlerce yorumun toxic, severe toxic, obscene, threat, insult ve identity hate adlı 6 farklı sınıfa otomatik sınıflandırmasını sağlamaktadır. Yapay zeka destekli olarak gerçekleştirilen otomatik çok etiketli metin analizi ile olumsuz sınıfta bulunan kullanıcı yorumları otomatik engellenebilecektir. Belirtilen bu hedefler doğrultusunda kullanıcı yorumları içerisinde birden fazla hedef etiketine

giren içerik olabileceği düşünülerek çok etiketli metin analizi gerçekleştirilmiştir.

Wikipedia kullanıcı yorumlarının toxic, severe toxic, obscene, threat, insult ve identity hate adlı 6 farklı sınıfa otomatik olarak kategorize edilmesini sağlayan bir doğal dil işleme konusudur [3]. Çok etiketli metin analizinde, aynı anda çıktıda çoklu etiketlerden birisi değil birden fazlası da çıktı olarak elde edilebilmektedir. İkili sınıflandırmada ise çoklu etiketleme benzer şekilde iki etiketten bir tanesinin belirlenmesi sağlanmaktadır. Bu anlamda çok etiketli sınıflandırma, çoklu sınıflandırma ve ikili sınıflandırmadan çok daha zordur [4]. Çok etiketli metin sınıflandırmasında, veri setinde her bir hedef etiketi ile ilişkili Wikipedia yorumları bulunmaktadır. Doğrudan hedef etiketlerine odaklanmadan kelimeler arasındaki bağlamı keşfedebilmek için metin analizi ile birlikte derin öğrenme teknolojisi kullanılmıştır.

Derin öğrenme teknolojisi [5] ise makine öğrenmenin alt alanı olarak bilgisayarlı görme [6], metin sınıflandırma [7], tarım [8] ve enerji [9] alanları gibi birçok alanda kayda değer başarılar elde etmiştir. Bu çalışmada, çok etiketli olarak metin sınıflandırma yapmak için metinlerin sayısallaştırılmasını sağlayan kelime vektörlerinin öğrenilmesine imkan tanıyan derin öğrenme modeli geliştirilmiştir [10]. Derin öğrenme modellerinde ise çok etiketli metin sınıflandırma yapmak için iki ayrı model tipi kullanılmaktadır. Bunlardan birincisi Konvolüsyonel Sinir Ağı (CNN) tabanlı modellerdir [11]. CNN tabanlı yöntemlerde kelimelerin özniteliklerini çıkartmayı sağlayacak filtreler bulunmaktadır. Farklı pencere boyutlarında, sayısallaştırılmış kelime vektörleri üzerinde gezinerek ayırt edici özniteliklerin elde edilmesini sağlayabilmektedir. İkincisi ise Tekrarlayan Sinir Ağı (RNN) tabanlı yöntemlerdir [12]. Kelime serileri üzerinde gezinerek sınıflandırma yaptırmaktadır. Uzun, kısa süreli bellek gibi farklı tipleri vardır. Farklı yapıları bulunan yapılar yardımıyla kelimelerin hafıza tarafından hatırlanmasını sağlayan model tipleridir.

Liao vd. twitter metin bilgilerinden bir özellik alanı çıkararak, çıkartılan özellik alanları arasındaki ilişkiyi bulmaya çalışan bir CNN yöntemi geliştirmiştir [13]. Twitter veri tabanı üzerinde duygu analizi gerçekleştirirken, metnin bir kısmı üzerinden duygu analizi yapmak yeterli olup olmadığını tartışmıştır. Normalde CNN algoritmaları görüntüden öznitelik çıkarma ve sınıflandırmada oldukça başarılıdır. Eğer ki metin

analizinde cümlenin tamamındaki duygu anlayamadığı takdirde CNN yönteminin tek başına duygu analizinde başarı olamayacağını ifade etmektedir. Elde edilen CNN ile duygu analizi sonuçlarına bakılınca CNN yönteminin SVM ve Naive Bayes gibi klasik yöntemlerden daha iyi sonuç verdiği görülmektedir [13]. RNN mimarileri ise değişken uzunluktaki cümlelerin sınıflandırma problemleri ile geniş bir şekilde kullanılmaktadır. Herhangi bir metin diziliminin öncesi, sonrası ve yerel bağlamlarını hafızaya alabilmek için RNN tabanlı bir model önermiştir [14]. Bu yöntemde CNN mimarilerinin temelini oluşturan konvolüsyon katmanı ile yerel bilgileri yakalamaya çalışırken, iki yönlü RNN yapısı ile de geçmiş ve gelecek bilgilerine erişmek amacıyla kullanmıştır. Standart RNN modellerinin bir eksikliği olarak yalnızca tarihsel bağlamdan yararlanmasıdır. İki yönlü RNN yapısı da LSTM ve GRU yapılarının birleşiminden oluşan bir model ile değiştirilmiştir. Bu şekilde kapsamlar bilgileri yakalamanın standart yinelenen modeller ile karşılaştırıldığında sınıflandırma doğruluğunu önemli ölçüde artırdığını beyan etmektedir [14].

Bu çalışmada kullanılan çok etiketli metin sınıflandırma probleminin çözümünde doğal dil işleme, yinelemeli sinir ağlarından GRU yapısı ve yerel bilgileri yakalayarak ayırt edici özniteliklerin elde edilmesini sağlayan CNN yapısı kullanılmıştır. Veri setindeki tüm Wikipedia cümleleri aynı standarda getirilmiştir. Eğitim ve test işlemlerinde kullanılacak metinleri aynı standartta olması zorunludur. Bunun için de küçük harfe çevirme, özel karakterlerin kaldırılması, sayısal ifadeler ve boşlukların yok edilmesi gerekmektedir. Bunlara benzer yapılan tüm işlemlere ön işleme denilmektedir. Wikipedia cümle yorumları önceden eğitilmiş GloVe gömme katmanı kullanarak kelime yerleştirme uygulanmıştır. Sonrasında ise çok katmanlı derin öğrenme modellerine girdi olarak verilmiştir. Bu çalışmanın literatüre sağladığı ana katkılar aşağıda sunulmuştur.

- Wikipedia başta olmak üzere sosyal ağlardaki yorumların otomatik analizini sağlayan bir model önerilmiştir.
- Hafif bir model önerilerek büyük bir veri setinde eğitimin daha uzun sürmesinin önüne geçilmiştir.
- GRU yapısının RNN yapısının yerine kullanılarak metin analizlerindeki bir

önceki ve sonraki bağlamları hafızada tutmada fayda sağladığı görülmüştür.

Makalenin bu adımından sonraki adımları üç bölümden oluşmaktadır. Birinci bölümde makalede kullanılan Wikipedia yorumlarını içeren veri seti detaylandırılmıştır. Aynı zamanda makale kapsamında önerilen modeli oluşturan yapılar anlatılmıştır. İkinci Bölümde yapılan deneysel çalışmalardaki parametreler hakkında bilgi verilmiştir. Buna ek olarak araştırma bulguları ve sonuçlar sunulmuştur. Son bölümde ise elde edilen son bilgiler ile çalışma sonuçlandırılmıştır.

II. MATERYAL VE YÖNTEM

Bu bölümde, makalede deneysel çalışmalara girdi olarak kullanılan ham veri seti hakkında bilgi verilmektedir. Makale deneysel çalışmalarına girdi olarak verilen ham veri seti detaylandırıldıktan sonra önerilen derin öğrenme modelinin temel yapısı hakkında bilgi verilecektir.

A. Materyal

Kapsamlı ve detaylı veri seti hazırlama görevleri uzman insanlar tarafından yapılmaktadır. Uzman insanlar Wikipedia yorumlarının hangi sınıfına gireceğini daha kolay bir şekilde tahmin edebilmektedir [25]. Wikipedia yorumları da 5000 farklı uzman tarafından hazırlanmış bir veri setidir [15]. Veri setindeki kullanıcı yorumlarının %90'ı gibi büyük bir oranı herhangi bir etiket sınıfına ait değildir. 20.000 Wikipedia yorumundan oluşan veri seti altı farklı sınıfa sahiptir [16]. Bu altı farklı sınıf toxic, severe toxic, obscene, threat, insult ve identity hate adlı sınıflardır. Bu sınıfların aynı anda bir veya daha fazlasının hedef etiketi olarak çıktığı olarak elde edilebilmesi problemi, çalışmayı çok sınıflı problem olmaktan çıkartıp çok etiketli sınıfa dönüştürmektedir.

B. CNN

Sınıflandırma problemlerinde sıklıkla kullanılan klasik tam bağlantılı sinir ağları, bağlantı problemine neden olabilmektedir. Girdi olarak alınan Wikipedia yorumlarının uzaklık ve yakınlık durumuna göre incelenmesi gerekmektedir. Belirtilen sorundan dolayı CNN tabanlı yöntemler çok etiketli metin analizinde kullanılmıştır. Aynı zamanda CNN tabanlı yöntemler tüm gizli nöronlar diğer katmanlar ile bağlantı sağlamak yerine, küçük bir bölgesine sağlanan bağlantı ile hesaplama maliyetini azaltmaktadır. CNN yerel bilgilerin yakalanmasında, RNN tabanlı yöntemlere göre daha başarılıdır [17]. RNN tabanlı

yöntemler ise ileri ve geri bilgileri hafızada tutma özelliğine sahiptir.

CNN yöntemleri ilk olarak görüntüler üzerinde öznetelik çıkartmak için geliştirildiği için temeli filtrelerin gezinmesine dayanmaktadır. Filtreleme işlemi sırasında satır ve sütun bazında gezinerek ayırt edici özneteliklerin elde edilmesi sağlanmaktadır. Burada konvolüsyon çekirdeğindeki değerler ile gezdirilen nesnenin penceresine karşılık gelen değerlerin çarpılmasına konvolüsyon denilmektedir. Konvolüsyon ifadesi Denklem 1'de gösterilmiştir. Denklem 1'de gösterilen formülde bulunan I yapısı veri matrisini göstermektedir. I veri matrisi üzerinde K adında olan $[i,j]$ boyutlarında bulunan kare matrisi gezdirilmektedir. Belirtilen gezinme işlemi sonrasında $I * K$ adlı $[x,y]$ boyutunda bir veri matrisi elde edilmiştir.

$$(I * K)_{xy} = \sum_{i=1}^h \sum_{j=1}^w K_{ij} \cdot I_{x+i-1,y+j-1} \quad (1)$$

C. GRU

Geçit Yinelemeli hücreler olarak bilinen GRU yapısı RNN tabanlı bir mimaridir. RNN temelli yöntemlerde yaşanan uzun vadeli bağımlılıklar gizli durumda değiştirilen bilgi kontrol edilerek kaldırmıştır. GRU mimarisinde sıfırlama ve güncelleme kapıları bulunmaktadır [8].

$$\text{Güncelleme}(u_t) = \sigma(W^u x_t + U^{(u)} h_{t-1}), \quad (2)$$

$$\text{Sıfırlama}(s_t) = \sigma(W^s x_t + U^{(s)} h_{t-1}), \quad (3)$$

Gizli durum değerleri ile giriş değeri fonksiyonları olan sıfırlama ve güncelleme kapıları GRU mimarisinin önemli bir fonksiyonudur. Gizli durumu temsil eden h_t sembolünü tespit etmek için RNN adlı temel mimariden farklı bir yol izlemektedir. Denklem 2 ve 3 arasında her GRU kapı formülünün farklı bir ağırlık değeri kullandığı görülmektedir.

D. Önerilen Model

Öncelikli olarak ham veriler ön işlemlerden geçirilmiştir. Ön işlemde geçirilen veriler kelime gömme katmanına sokularak sayısal vektörler elde edilmiştir. Gömme katmanında GloVe metodu kullanılmıştır. GloVe kelime yerleştirme metodu, matris çarpanı ve yerel bağlam özelliklerini birleştiren bir regresyon modelidir [18], [19]. Cümle girdilerinden oluşan büyük kelime

havuzlarındaki kelime bağlamlarından anlamlı vektör uzayları oluşturmaktadır. GloVe [20] sayım tabanlı olarak kelime çiftleri oluşturmaktadır. Kelimelerin sayısallaştırmasını öznitelik matrisine dönüştürerek gerçekleştirmektedir. Bu işlem stokastik gradyan inişi (SGD) olarak bilinen yinelenen bir sayısal yöntem ile yapılmaktadır. Denklem 4'teki R ifadesi metin matrisidir. P, Q ise rastgele değerlerden meydana gelmektedir. Bu rastgele değerler çarpılarak R' değeri elde edilmektedir. Sonrasında ise R' ile R arasındaki fark azaltılmaya çalışılmaktadır. Bu şekilde P, Q değerleri ayarlanmaktadır.

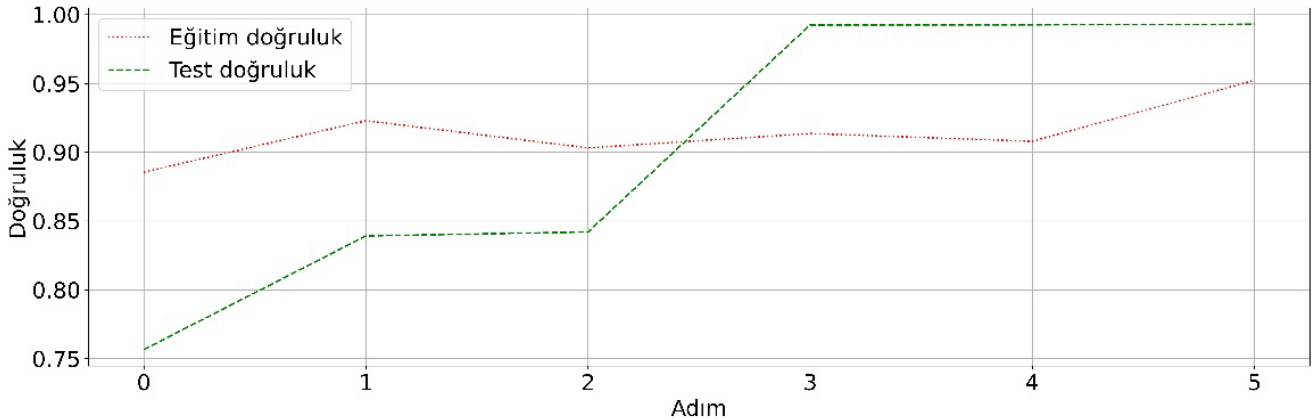
$$R = P * Q \approx R' \quad (4)$$

Sayısallaşan metin ifadeleri önerilen modele girdi olarak verilmiştir. Önerilen modelin ikinci adımında uzamsal unutma katmanı ile modelin aşırı öğrenmesinin önüne geçilmiştir. Üçüncü adımında 256 adet 8×8 boyutunda ReLU aktivasyon fonksiyonlu bir boyutlu konvolüsyon katmanı eklenmiştir. Dördüncü adımında 3×3 boyutunda bir boyutlu maksimum havuzlama eklenmiştir. Sayısallaştırılmış Wikipedia metin yorumları üzerinde farklı pencere boyutlarında yapılan filtreleme işlemleri sayesinde maksimum havuzlama gerçekleştirilmiştir. Beşinci adımında 256 adet 8×8 boyutunda ReLU aktivasyon fonksiyonlu bir boyutlu konvolüsyon katmanı

eklenmiştir. Altıncı adımında 3×3 boyutunda bir boyutlu maksimum havuzlama eklenmiştir. Yedinci adımında tekrar üçüncü ve dördüncü adımın tekrarı gerçekleştirilerek yerel ayırt edici özniteliklerin çıkartılması sağlanmıştır. Sekizinci adımda ise dördüncü ve altıncı adımda olduğu gibi 3×3 boyutunda bir boyutlu maksimum havuzlama ile etkisi büyük olan öznitelikler seçilmiştir. Dokuzuncu adımda katmanlar arası girdileri normalleştiren toplu normalleştirme katmanı eklenmiştir. Onuncu adımda iki yönlü 128 nöronlu GRU katmanı eklenmiştir. On birinci adımda dokuzuncu adımdaki gibi toplu normalleştirme katmanı eklenmiştir. On ikinci adımda 0.1 oranında nöron bırakma sağlayan unutma katmanı eklenmiştir. On üçüncü adımda ise 512 nöronlu ReLU aktivasyon fonksiyonuna sahip bir katman eklenmiştir. On dördüncü adımda ise sigmoid aktivasyon fonksiyonlu bir sınıflandırma katmanı eklenmiştir.

III. ARAŞTIRMA BULGULARI

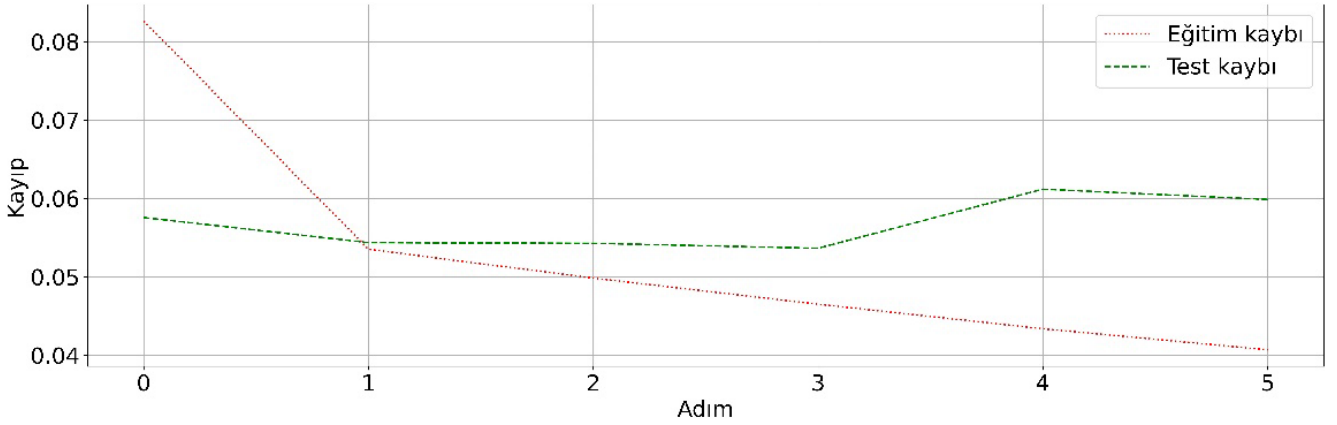
Şekil 1'de doğruluk grafiği verilen önerilen modelin altı iterasyonlu deneysel çalışma sonuçları gösterilmiştir. Test doğruluk %99.0 başarı oranına ulaşırken, eğitim doğruluk oranı %95.23 başarı oranında kalmıştır. Altı iterasyonlu deneysel çalışmalarda 0.01 eğitim oranında Adam optimizasyon yöntemi kullanılmıştır.



Şekil 1. Önerilen model eğitim ve test doğruluk grafiği

Şekil 2'de ise ikili çapraz entropiye göre elde edilen eğitim ve test kayıp grafikleri elde edilmiştir. Test kaybı %0.05 civarında iken, eğitim

kaybı %0.04 seviyesindedir. Elde edilen sonuçlar incelendiğinde GRU ve CNN mimarilerini temel olarak geliştirilen derin öğrenme modelinin iyi bir başarı sağladığı görülmektedir.



Şekil 2. Önerilen model eğitim ve test kayıp grafiği

IV. SONUÇLAR

Sosyal ağlarda hızlı mesajlaşma ve yoğun görüntülenme ile birlikte kullanıcı sayısında artış gözlenmektedir. Artış ile birlikte kullanıcı yorum içeriklerini doğru bir şekilde kontrol etmek gerekli olmuştur. Olumsuz ya da zarar verici içeriklerin kontrol edilerek otomatik sınıflandırılmasını sağlayan bir derin öğrenme modeli önerilmiştir. Önerilen derin öğrenme modeli GRU ve CNN yapılarının temel özelliklerini içermektedir. 3 adet konvolüsyon ve maksimum havuzlama katmanının ardı sıra kurgulanması ile birlikte metin bağlamlarına ilişkin yerel değerler yakalanmıştır. Bu değerler üzerinden ileri ve geri hafızada bilgilerin GRU yöntemi ile tutulması sağlanarak tam bağlantı katmanına ve sınıflandırma katmanına girdi oluşturulmuştur.

KAYNAKLAR

- [1] S. R. Sahoo and B. B. Gupta, "Multiple features based approach for automatic fake news detection on social networks using deep learning," *Appl. Soft Comput.*, vol. 100, p. 106983, 2021, doi: 10.1016/j.asoc.2020.106983.
- [2] B. Horne and S. Adali, "This just in: Fake news packs a lot in title, uses simpler, repetitive content in text body, more similar to satire than real news," in *Proceedings of the international AAAI conference on web and social media*, 2017, vol. 11, no. 1, pp. 759–766.
- [3] B. Liu, X. Liu, H. Ren, J. Qian, and Y. Wang, "Text multi-label learning method based on label-aware attention and semantic dependency," *Multimed. Tools Appl.*, vol. 81, no. 5, pp. 7219–7237, 2022, doi: 10.1007/s11042-021-11663-9.
- [4] F. Gargiulo, S. Silvestri, and M. Ciampi, *Deep Convolution Neural Network for Extreme Multi-label Text Classification*. 2018. doi: 10.5220/0006730506410650.
- [5] J. Schmidhuber, "Deep learning in neural networks: An overview," *Neural networks*, vol. 61, pp. 85–117, 2015.
- [6] V. Campos, B. Jou, and X. Giro-i-Nieto, "From pixels to sentiment: Fine-tuning CNNs for visual sentiment prediction," *Image Vis. Comput.*, vol. 65, pp. 15–22, 2017.
- [7] K. S. Tai, R. Socher, and C. D. Manning, "Improved semantic representations from tree-structured long short-term memory networks," *arXiv Prepr. arXiv1503.00075*, 2015.
- [8] H. Çetiner and B. Kara, "Recurrent Neural Network Based Model Development for Wheat Yield Forecasting," *J. Eng. Sci. Adiyaman Univ.*, vol. 9, no. 16, pp. 204–218, 2022, doi: 10.54365/adyumbd.1075265.
- [9] H. Çetiner and İ. Çetiner, "Analysis of Different Regression Algorithms for the Estimate of Energy Consumption," *Eur. J. Sci. Technol.*, no. 31, pp. 23–33, Dec. 2021, doi: 10.31590/ejosat.969539.
- [10] B. Pang and L. Lee, "Seeing stars: Exploiting class relationships for sentiment categorization with respect to rating scales," *arXiv Prepr. cs/0506075*, 2005.
- [11] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," *Adv. Neural Inf. Process. Syst.*, vol. 25, 2012.
- [12] K. Funahashi and Y. Nakamura, "Approximation of dynamical systems by continuous time recurrent neural networks," *Neural networks*, vol. 6, no. 6, pp. 801–806, 1993.
- [13] S. Liao, J. Wang, R. Yu, K. Sato, and Z. Cheng, "CNN for situations understanding based on sentiment analysis of twitter data," *Procedia Comput. Sci.*, vol. 111, pp. 376–381, 2017, doi: https://doi.org/10.1016/j.procs.2017.06.037.
- [14] Y. Zhang, J. E. Meng, R. Venkatesan, N. Wang, and M. Pratama, "Sentiment classification using Comprehensive Attention Recurrent models," in *2016 International Joint Conference on Neural Networks (IJCNN)*, 2016, pp. 1562–1569. doi: 10.1109/IJCNN.2016.7727384.
- [15] E. Wulczyn, N. Thain, and L. Dixon, "Ex Machina: Personal Attacks Seen at Scale," in *Proceedings of the 26th International Conference on World Wide Web*, 2017, pp. 1391–1399. doi: 10.1145/3038912.3052591.
- [16] A. K. Jayaraman, A. Murugappan, T. E. Trueman, and E. Cambria, "Comment toxicity detection via a multichannel convolutional bidirectional gated recurrent unit," *Neurocomputing*, vol. 441, pp. 272–278, 2021, doi: 10.1016/j.neucom.2021.02.023.

- [17] Y. LeCun, Y. Bengio, and G. Hinton, “Deep learning,” *Nature*, vol. 521, no. 7553, pp. 436–444, 2015, doi: 10.1038/nature14539.
- [18] K. Lin and D. Pomerleano, “Global matrix factorizations,” *Math. Res. Lett.*, vol. 20, Jan. 2011, doi: 10.4310/MRL.2013.v20.n1.a9.
- [19] H.-C. Cho, N. Okazaki, and K. Inui, *Inducing Context Gazetteers from Encyclopedic Databases for Named Entity Recognition*, vol. 7818. 2013. doi: 10.1007/978-3-642-37453-1_31.
- [20] J. Pennington, R. Socher, and C. Manning, *Glove: Global Vectors for Word Representation*, vol. 14. 2014. doi: 10.3115/v1/D14-1162.