

Türkçe Tweetler Üzerinden Yapay Sinir Ağları ile Cinsiyet Tahminlemesi

Gender Prediction from Turkish Tweets with Neural Networks

Erhan Sezerer
İzmir Yüksek Teknoloji Enstitüsü
Bilgisayar Mühendisliği
erhansezerer@iyte.edu.tr

Ozan Polatbilek
İzmir Yüksek Teknoloji Enstitüsü
Bilgisayar Mühendisliği
ozanpolatbilek@iyte.edu.tr

Selma Tekir
İzmir Yüksek Teknoloji Enstitüsü
Bilgisayar Mühendisliği
selmatekir@iyte.edu.tr

Özetçe—Yazar ayrılması, yazarı bilinmeyen bir metin üzerinden yazarına dair cinsiyet, yaş ve dil gibi bazı anahtar özelliklerin belirlenmesidir. Özellikle güvenlik ve pazarlama alanında önem arz etmektedir. Bu çalışmada, kullanıcıların tweetleri kullanılarak cinsiyetleri tahminlenmektedir. Yinelemeli Sinir Ağı (YSA) ve ilgi mekanizmasının birleşiminden oluşan bir model önerilmiştir¹. Bildiğimiz kadarıyla bu çalışma Twitter veri kümesi ile Türkçe’de ilk defa yapılmıştır. Önerilen model Türkçe, İngilizce, İspanyolca ve Arapça dillerinde sınanmış ve sırasıyla 80.63, 81.73, 78.22, 78.5 doğruluk değerlerine ulaşmıştır. Elde edilen doğruluk değerleri Türkçe’de en gelişkin, diğer dillerde ise rekabetçi bir başarıya ortaya koymaktadır.

Anahtar Kelimeler—Yazar Ayrılması, Cinsiyet Tahminlemesi, Derin Öğrenme, Yinelemeli Sinir Ağı, Yapay Sinir Ağları, İlgi Mekanizması, Twitter Veri Kümesi.

Abstract—Author profiling is the characterization of an author through some key attributes such as gender, age, and language. It’s an indispensable task especially in security and marketing. In this work, the gender of a Twitter user is predicted using his/her tweets. A model combining a recurrent neural network (RNN) with an attention mechanism is proposed. As far as we know such a predictive analytics is performed in Turkish Twitter dataset for the first time, and the proposed model is tested in Turkish, English, Spanish, and Arabic with accuracy scores of 80.63, 81.73, 78.22, 78.5 respectively. The accuracy values obtained exhibit state-of-the-art in Turkish and competitive performance in the other languages.

Keywords—Author Profiling, Gender Prediction, Deep Learning, Recurrent Neural Networks, Neural Networks, Attention Mechanism, Twitter Dataset.

I. GİRİŞ

Yazar ayrılması, yazarı bilinmeyen bir metin üzerinden yazarına dair yaş, cinsiyet, etnik köken gibi temel nüfus bilgileri yanında kişilik ve kültürel arka plan özelliklerinin belirlenmesidir [1]. Bu iş, tehdit oluşturabilecek kişilerin önceden tespit edilmesi, adli bilişim vakalarında suçun birine yüklenmesine dayanak oluşturma ve reklam hedef kitlesinin belirlenmesi gibi farklı disiplinlerdeki ihtiyaçlara cevap verebilir.

¹https://github.com/Darg-Iztech/Turkce-cinsiyet-tahminlemesi

TABLO I: Cinsiyet Tahminlemesinde Yüksek İlgi Kazanan Kelimeler - İlgi mekanizması YSA yöntemi ile birlikte kullanıldığında her bir kelime için üretilen ilgi değerleri. Parlak renk tonları yüksek ilgi değerlerine, sönük renkler düşük ilgi değerlerine karşılık gelmektedir.

Tweet	Cinsiyet
YUHH Diyorum Guntekin Onay . Bu kadar GS dusmanligi olmaz . @**** @**** @**** @**** @**** PAD PAD...	Erkek
Ikizler Haluk un kizlari yaa inanamıyorum :o :o PAD PAD...	Kadın
@**** iç abi iç öyle daha doğal olur . PAD PAD...	Erkek
@**** Fenerbahçem için kurtulalım kuzum ;) PAD PAD...	Kadın

Günümüzde sosyal medya platformları, özellikle Twitter, çok sayıda kullanıcıya ait çok kaynaklı (metin, görüntü, video vb.) veriyi erişilebilir hale getirmesi nedeniyle yazar ayrılması probleminin çözümünde önemli bir veri potansiyeli sunmaktadır.

Literatürde metinlerden yazarın cinsiyetinin belirlenmesinde kullanılacak özellikler içerik tabanlı ve biçim tabanlı olarak gruplanmaktadır. Biçim tabanlı özellikler arasında belirteç ve edat kullanımı erkekler için belirleyici rol oynarken, zamir kullanımı kadınlarda öne çıkmaktadır. İçerik tabanlı özellikler ele alındığında ise erkeklerin teknoloji ile ilgili terimleri yoğun olarak kullandığı, kadınların ise çoğunlukla özel yaşama ve ilişkilere dair kelimeleri tercih ettiği gözlemlenmektedir [1]. Yazarların cinsiyet sınıflandırmasının yapılmasında özelliklerin açık bir şekilde çıkarılmasına alternatif olarak, ilgili özelliklerin örtülü olarak öğrenilmesini sağlayabilecek derin öğrenme yaklaşımları önemli bir potansiyel sunmaktadır.

Önerilen çalışmada derin öğrenme yöntemleri kullanılarak, atıkları tweet’ler üzerinden kullanıcıların cinsiyetlerinin tahminlenmesi hedeflenmektedir. Bu hedef doğrultusunda,

Yinelemeli Sinir Ağı (YSA) (ing. RNN) tabanlı bir yöntem geliştirilmiştir. Ayrıca daha önce geliştirilen evrişimsel sinir ağı (ESA) (ing. CNN) temelli yöntemin [2] deneysel değerlendirmesi Türkçe bir veri kümesi dahil edilerek genişletilmiş, bu sayede her iki yöntemin başarımı dört farklı dilde sınanmıştır. YSA ve ESA yöntemlerinin çıktısı olarak her kullanıcı için öznitelik vektörleri elde edilmiştir. Bu vektörler, ilgi mekanizmasından (ing. attention mechanism) geçirilerek, daha çok bilgi içeren yapılarının öne çıkarılması amaçlanmıştır. İlgi mekanizmasının çıktısı da Tamamen Bağlı Sinir Ağına (TBSA) (ing. Fully Connected Neural Network) verilerek kullanıcının cinsiyeti hakkında tahminleme yapılmıştır. Tablo I’de de görüldüğü üzere, ilgi mekanizması YSA ve ESA çıktılarındaki öznitelikleri, cinsiyet tahminlemedeki önemlerine göre daha çok öne çıkarmış ya da sönmüştür.

Önerilen çalışmanın temel bilimsel katkısı Twitter üzerinden cinsiyet tahminlemesi problemini çözmek üzere geliştirilen YSA tabanlı derin öğrenme modelidir. Model, Türkçe’de en gelişkin olmak üzere, İngilizce, Arapça ve İspanyolca dillerinde yüksek doğruluk değerini yakalamaktadır.

Bildirinin geri kalan kısmında ilk olarak özet bir literatür taraması (Bölüm II) verilmektedir. Sonrasında (Bölüm III), önerilen yöntem; veri kümesi, önışleme, gömülmeler ve model bileşenleri ile ayrıntılı bir şekilde açıklanmaktadır. Bölüm IV’te deneysel değerlendirme yapılmakta, ardından sonuç ve gelecek çalışmalar ortaya konmaktadır (Bölüm V).

II. LİTERATÜR TARAMASI

Twitter verisi üzerinden cinsiyet tahminlemesi konusunda CLEF 2018 kapsamında PAN 2018 çalıştayında ortak bir çalışma [3] düzenlenmiştir. Çalışma dahilinde Arapça, İngilizce ve İspanyolca dillerinde hem kullanıcı tweetlerini hem de kullanıcı tarafından paylaşılan imgeleri içeren cinsiyet etiketli bir derlem paylaşılmıştır.

PAN 2018 cinsiyet belirleme ortak çalışmasında Takahashi vd. [4] derin öğrenme tabanlı bir yaklaşım geliştirmişlerdir. Önerdikleri derin öğrenme modeli metin, imge ve metin ve imgeyi birleştirdikleri füzyon bileşeninden meydana gelmektedir. Metin bileşeni kelime gömülmelerini YSA ile, imge bileşeni imgeleri ESA ile işlemekte, füzyon bileşeninde ise metin ve imgeden gelen vektörler doğrudan çarpım ile birleştirilmekte ve sütun ve satır-bazlı havuzlama işleminden geçirilmektedir. Model Arapça, İngilizce ve İspanyolca dillerinde ortalama %81.98 doğruluk performansı ile ilk sırada yer almıştır.

Yine aynı ortak çalışmada, Daneshvar ve İnken [5] yalnızca metin verisi kullanarak destek vektör makinesi ile sınıflandırma yapmışlardır. Tweetler sırasıyla önışlemeden geçirilmekte, kelime ve karakter n-gramları oluşturulmakta ve LSA ile boyut indirgemenin ardından çıkarılan öznitelikler sınıflandırıcıya verilmektedir. Çalışma Arapça ve İspanyolca’da ilk sırada iken ortalama performansta %81.70 doğruluk oranı ile ikinci sırada yer almıştır.

Türkçe üzerinde cinsiyet tahminlemesi ile ilgili benzer çalışmalar mevcuttur. Doğan ve Diri [6], kendilerinin

topladıkları çeşitli haber makaleleri üzerinde kelime 2-, 3- ve 4-gramlarını kullanarak yazarın cinsiyetini tahminlemeye çalışmış ve diğer yöntemlerden (Naive Bayes, Destek Vektör Makinesi, Rastgele Orman, K-En Yakın Komşuluk) daha yüksek başarımlar elde etmişlerdir.

Öte yandan Gülşen vd. [7], kişilerin cinsiyetlerini kendi yazdıkları metinler yerine, ziyaret ettikleri web sayfalarının metinlerini kullanarak lojistik regresyon ile bulmaya çalışmışlardır.

Talebi ve Köse [8] ise kullanıcıların Facebook’tan yaptığı yorumlarını kullanarak Naive Bayes, Destek Vektör Makinesi ve K-en yakın komşu yöntemlerini sınamış, bu yöntemler arasında cinsiyet tahminlemesi konusunda en başarılı sonucu verenin ise Naive Bayes olduğunu göstermişlerdir.

Önerilen çalışma, bahsi geçen çalışmalardan iki yönüyle farklıdır. Bahsi geçen makalelerdeki metinler daha uzun ve yapısal iken Twitter verisi karakter sınırı dolayısıyla daha kısadır ve yapısalıktan uzaktır. Bu da problemin doğasını değiştirmektedir. İkinci ve son fark ise, bahsedilen yöntemler elyapımı öznitelik vektörleri kullanırken, bu çalışmada öznitelik vektörlerini kendi kendine öğrenebilen bir derin öğrenme yöntemi önerilmiştir.

III. YÖNTEM

Bu bölümde veri kümesi, önerilen YSA tabanlı yöntem ve bu veri üzerinde ilk defa sınanan daha önce geliştirilen ESA tabanlı yöntem [2] detayları ile açıklanmıştır.

A. Veri Kümesi

Twitter üzerinden, 4076 tane kullanıcıya ait hesaplardan, her bir kullanıcıya 100 tane tweet düşecek şekilde rastgele tweetler çekilmiş ve cinsiyet bilgisi ile işaretlenmiştir. Veri kümesindeki kullanıcılar eğitime ve test kümeleri olmak üzere ikiye ayrılmıştır. Bu kümelerde sırası ile 2500 ve 1576 adet kullanıcı bulunmaktadır: Eğitim kümesinde 1249 erkek ve 1256 kadın, test kümesinde ise 774 erkek ve 808 kadın kullanıcıya ait 100’er tweet bulunmaktadır. Kullanıcıları kümelere bölme işlemi rastgele yapılmış olup, her bir kümede yakın sayıda erkek ve kadın kullanıcı olması gözletilmiştir. Sonrasında ise eğitim kümesinin içerisinde 500 adet kullanıcı hiperparametre eniyilemesi için sağlama kümesi olarak ayrılmıştır.

Yöntem Türkçeye ek olarak PAN çalışması üzerinden paylaşılmış olan üç dilde daha denenmiştir: İngilizce veri kümesinde 3000 eğitim, 1900 test kümesinde olmak üzere toplam 4900 kullanıcı, İspanyolca veri kümesinde 3000 eğitim, 2200 test olmak üzere toplam 5200 kullanıcı, son olarak da Arapça veri kümesinde 1500 eğitim, 1000 test kümesinde olmak üzere toplam 2500 kullanıcı bulunmaktadır.

Bu çalışmada kullanılan veri kümesi² açık olarak yayınlanmıştır.

B. Önışleme

Sınanan her bir yöntem için farklı önışlemeler kullanılmış olup aşağıda detaylı olarak açıklanmıştır.

²<https://cloud.iyte.edu.tr/index.php/s/5DhqdlUCCdB60qG>

1) *YSA Modeli İçin Önışleme:* Her bir tweet kelimelerine ayrılmış ve bütün karakterleri küçük hale getirilmiş şekliyle sisteme kelimelerden oluşan bir vektör şeklinde girilmiştir. Vektörün uzunluğu ise veri kümesindeki en uzun tweete göre ayarlanıp daha kısa olan tweetlere doldurma (ing. padding) eklenmiştir.

2) *ESA Modeli İçin Önışleme:* Her bir tweet harflerine ayrılıp, büyük harfler küçük hale getirilmiştir. Noktalama işaretleri ise kullanıcılar tarafından duygularını ifade etmek için (gülümseme, somurtma işaretleri (ing. emoji) vb.) kullanıldığından metinden çıkartılmamıştır. Her bir tweetin uzunluğu maksimum karakter sayısına eşit olacak şekilde ayarlanıp daha kısa olan tweetlere doldurma eklenmiştir. Metinler sisteme, harflerden oluşan bir vektör şeklinde girilmiştir.

C. Gömülmeler

YSA modelinin kelime gömülmelerinin eğitilmesi için çok fazla veriye ihtiyaç duyması, dolayısıyla gömülmelerin model ile birlikte eğitilememesi nedeniyle Glove [9] kelime gömülmeleri kullanılmıştır. Glove kelime gömülmelerinin seçilmesindeki ilk etken Glove'un birçok dilin yanı sıra Türkçe kelime gömülmelerini de barındırmasıdır. İkinci etken ise gömülmelerin Twitter verisi ile eğitilmiş olmasından dolayı Twitter veri kümesinin doğasına daha uygun olacağına düşünülmüştür.

ESA modeli için 25'lik karakter gömülmeleri kullanılmıştır. Veri kümesi bu karakter gömülmelerini eğitmek için yeterli büyüklükte olduğundan gömülmeler model ile birlikte eğitilmiştir. İlk değer atamaları düzgün dağılım içerisinde rastgele örnekleme ile yapılmıştır.

D. Yöntem

Çalışmada YSA ve ESA temelli iki farklı yöntem sınanmıştır. Bu yöntemler, bir kullanıcının ayrışması için gerekli olan biçim bazlı ve içerik bazlı öznitelikleri karşılayacak şekilde seçilmiştir [1]. Yerel öznitelikleri çıkarma konusunda başarılı olan ESA kullanılarak biçim bazlı özniteliklerin bulunması hedeflenirken, cümlenin anlamını kavrama konusunda başarılı olan YSA yöntemi ile içerik bazlı öznitelikler karşılanmaya çalışılmıştır [10].

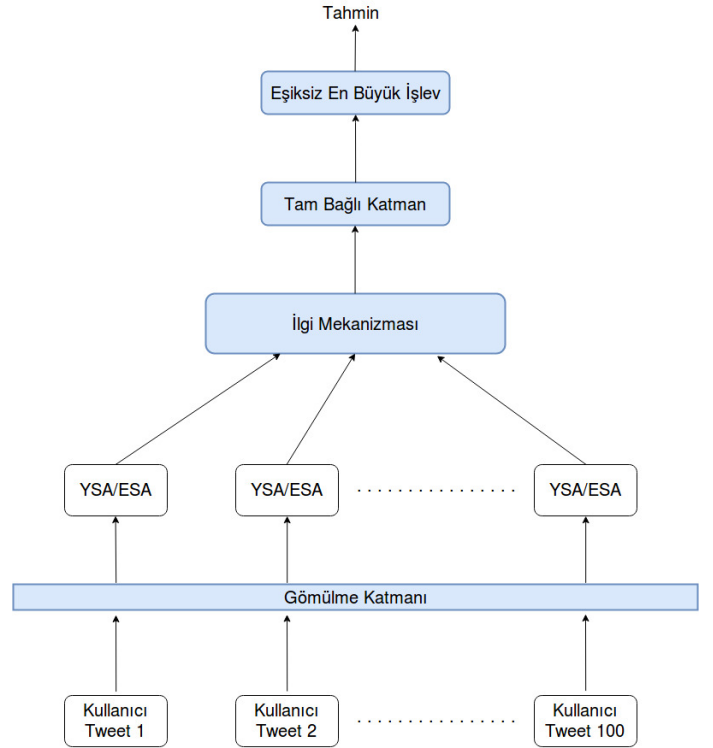
1) *YSA yöntemi:* İleri ve geri yönlü olmak üzere 150'şer GRU [11] hücrelerine sahip bir çift yönlü YSA kullanılmıştır. Her bir tweet için bir öznitelik vektörü üretilmiş olup daha sonra bu öznitelik vektörleri ilgi mekanizması ile her bir kullanıcı için bire indirgenmiştir.

2) *ESA yöntemi:* ESA yönteminde 3×3 , 6×6 , 9×9 boyutlarında olmak üzere 100'er filtre kullanılmıştır. Yöntem belirli bir kullanıcıya ait olan 100 adet tweet'i eşzamanlı olarak alıp, her biri için bir öznitelik vektörü üretecek şekilde tasarlanmıştır. Bu öznitelik vektörleri daha sonra ilgi mekanizmasından geçirilerek her bir kullanıcıya ait tek bir öznitelik vektörü üretilmesi sağlanmıştır.

3) *İlgi Mekanizması:* Kullanıcıların attıkları her bir tweet'in cinsiyet belirlemesi için aynı miktarda bilgiyi yansıtmayacağı fikrinden yola çıkılarak, her bir tweet'ten çıkan öznitelik

vektörleri birleştirilmek ya da ortalaması alınmak yerine Bahdanau ilgi mekanizması [12] ile birleştirilmiştir. Her bir yöntemin (YSA, ESA), her bir tweet için ürettiği öznitelik vektörleri bu ilgi mekanizmasından geçirilerek içerdikleri bilginin derecesi öğrenilip bu değer ile doğru orantılı olacak şekilde son sınıflandırmanın yapılması sağlanmıştır. Literatürde de daha önce gösterildiği üzere [2], bu mekanizma "nereye bakması" gerektiğini öğrenerek doğruluğu artırmaktadır. İlgi mekanizması YSA için hem kelime bazında hem de tweet bazında ESA yöntemi için ise sadece tweet bazında uygulanmıştır. Tablo I'de ilgi mekanizmasının YSA ile birlikte kullanıldığında kelimeler üzerinde nasıl bir ağırlıklandırma yaptığı görülebilir.

İlgi mekanizmasının çıktısı olan, kullanıcılara ait öznitelik vektörleri, boyutun sınıf sayısına indirgenmesi için tamamen bağlı katmana (ing. fully connected layer) sokulmuş ve bu katmanın çıktısı eşiksiz en büyük işlevden (ing. softmax function) geçirilmiş ve bu işlevin çıktısı ise tahminleme olarak kullanılmıştır. Yöntemin şeması genel hatları ile Şekil 1'de görülebilir.



Şekil 1: Önerilen Yöntemler.

4) *Hiperparametreler ve Eğitime:* Hiperparametrelerin belirlenmesi için ızgara arama (ing. grid search) algoritması kullanılmıştır. En iyi sonuçları veren parametreler Tablo II'de gösterilmiştir.

Sinir Ağlarının eğitilmesi için Adam [13] eniyileme metodu çapraz entropi kayıp fonksiyonu ile birlikte kullanılmıştır. Aktivasyon fonksiyonu olarak ReLu seçilmiş olup, aşırı uymanın (ing. overfitting) engellenmesi için L2 Düzenleştirme kaybı kullanılmıştır.

TABLO II: Türkçe Veri Kümesi İçin Hiperparametreler

Yöntem	Öğrenme Oranı	Düzenleştirme Param.
İlgi Mekanizmalı YSA	1×10^{-3}	5×10^{-5}
İlgi Mekanizmalı ESA	1×10^{-3}	1×10^{-6}

IV. DENEYSEL DEĞERLENDİRMELER

Türkçe veri kümesinin ilk defa bu çalışmada denenmesi nedeniyle literatürde, önerilen yöntemle karşılaştırılabilecek bir çalışma bulunmamaktadır. Önerilen yöntemin Türkçe ve diğer dillerdeki başarımlarını karşılaştırmak ve Türkçe veri kümesinin zorlayıcılığını sınamak amacıyla PAN 2018 veri kümesinde bulunan diğer dillerle karşılaştırmalar yapılmıştır. Elde edilen doğruluk değerleri Tablo IV'te görülmektedir.

Görülebildiği üzere Türkçe, doğruluk değeri olarak İngilizce'nin altında, Arapça ve İspanyolca'nın üstünde yer almaktadır. Bu da hazırlanmış olan Türkçe veri kümesinin İngilizce'den daha zorlayıcı fakat diğer dillerden daha az zorlayıcı olacağını göstermektedir. Sonuçlarla ilgili göze çarpan bir başka konu ise, Türkçe harici dillerde YSA yönteminin ESA'ya göre istatistiksel olarak belirgin bir farkla daha iyi olmasına karşın, Türkçe'de bu farkın öneminin çok daha azalmış olmasıdır. Bu durum, toplanmış olan verinin yapısına bağlı olabileceği gibi Türkçe'de cinsiyeti betimleyen özneliklerin ağırlıklı olarak biçim bazlı olmasından da kaynaklanabilir.

Önerilen yöntem İngilizce'de en gelişkin yönteme doğruluk değeri olarak çok yaklaşmış olup, İngilizce, İspanyolca ve Arapça dillerinin ortalamasında ise en gelişkin yöntemin gerisinde kalmıştır. Sadece metin bazlı değerlendirme göz önüne alındığında PAN 2018'de düzenlenen ortak çalışmadaki en iyi yöntem olan Takahashi vd.'nin [4] 78,47'lik ortalamasını geride bırakarak, derin öğrenme modeli kullanan yöntemler arasında en gelişkin olmayı başarmış, fakat Daneshvar ve İnkpın'in [5] destek vektör makinesinin gerisinde kalmıştır.

TABLO III: Önerilen Yöntemlerin Doğrulukları

Yöntem	Türkçe	İngilizce	İspanyolca	Arapça	Ort.
İlgi Mek. ESA	80.318	78.474	75.000	71.800	75.091
İlgi Mek. YSA	80.637	81.737	78.227	78.500	79.488

TABLO IV: Önerilen Yöntemin Doğruluk Karşılaştırması

Yöntem	Türkçe	İngilizce	İspanyolca	Arapça
Önerilen Yöntem	80.637	81.737	78.227	78.500
Daneshvar vd.[5]	-	0.8221	0.8200	0.8090
Takahashi vd.[4]	-	0.7968	0.7864	0.7710

V. SONUÇ

Bu çalışmada, kullanıcının tweetleri üzerinden cinsiyet tahminlemesi yaparken YSA tabanlı bir yöntem önerilmiştir. Önerilen yöntem, hazırlanan cinsiyet etiketli Türkçe Twitter veri kümesi üzerinde denenmiştir. Önerilen YSA yöntemi ile birlikte daha önce geliştirilen ESA tabanlı yöntem [2], Türkçe'nin yanısıra İngilizce, İspanyolca ve Arapça Twitter veri kümeleri üzerinde sınanmıştır. Yöntemlerin çıktıları, her

kullanıcı için bir vektör olmak üzere, ilgi mekanizmasından geçirilerek, bulunan öznelikler arasından daha etkili olanların ön plana çıkarılması amaçlanmıştır. Bu sayede özellikle YSA tabanlı yöntemde, her dil için yüksek doğruluk değeri elde edilebilmiştir. Türkçe için bu çalışma Twitter veri kümesinde ilk defa yapılmış olup, İngilizce dili için güncel yöntemlerin en gelişkinine çok yakın sonuçlar elde edilmiştir.

Çalışmanın devamında Twitter kullanıcılarının paylaştıkları fotoğrafların da işlenmesi düşünülmektedir. Bu sayede kullanıcıların cinsiyetlerini öne çıkarabilecek farklı bir örüntü bulunması hedeflenmekte ve bu örüntünün, metin tabanlı yöntemlerle birleştirilerek daha yüksek bir doğruluk oranına ulaşılabileceği düşünülmektedir.

TEŞEKKÜR

İYTE-CVRG araştırma grubuna, testlerimizi yaptığımız donanımı bize sağladıkları için çok teşekkür ederiz.

KAYNAKLAR

- [1] S. Argamon, M. Koppel, J. W. Pennebaker, and J. Schler, "Automatically profiling the author of an anonymous text," *Commun. ACM*, vol. 52, no. 2, pp. 119–123, Feb. 2009.
- [2] E. Sezerer, O. Polatbilek, O. Sevgili, and S. Tekir, "Gender prediction from tweets with convolutional neural networks," in *Proceedings of the Ninth International Conference of the CLEF Association (CLEF 2018)*, 2018.
- [3] F. M. R. Pardo, P. Rosso, M. M. y Gómez, M. Potthast, and B. Stein, "Overview of the 6th author profiling task at pan 2018: Multimodal gender identification in twitter," in *CLEF*, 2018.
- [4] T. Takahashi, T. Tahara, K. Nagatani, Y. Miura, T. Taniguchi, and T. Ohkuma, "Text and image synergy with feature cross technique for gender identification: Notebook for PAN at CLEF 2018," in *Working Notes of CLEF 2018 - Conference and Labs of the Evaluation Forum, Avignon, France, September 10-14, 2018.*, 2018.
- [5] S. Daneshvar and D. İnkpın, "Gender identification in twitter using n-grams and LSA: notebook for PAN at CLEF 2018," in *Working Notes of CLEF 2018 - Conference and Labs of the Evaluation Forum, Avignon, France, September 10-14, 2018.*, 2018.
- [6] S. Doğan and B. Diri, "Türkçe dokümanlar için n-gram tabanlı yeni bir sınıflandırma (ng-ind): Yazar, tür ve cinsiyet," *Türkiye Bilişim Vakfı Bilgisayar Bilimleri ve Mühendisliği Dergisi*, vol. 3, pp. 11 – 19, 2010.
- [7] E. Gülşen, H. Gündüz, Z. Cataltepe, and L. Serinol, "Big data feature selection and projection for gender prediction based on user web behaviour," in *2015 23rd Signal Processing and Communications Applications Conference (SIU)*, May 2015, pp. 1545–1548.
- [8] M. Talebi and C. Köse, "Identifying gender, age and education level by analyzing comments on facebook," in *2013 21st Signal Processing and Communications Applications Conference (SIU)*, April 2013, pp. 1–4.
- [9] J. Pennington, R. Socher, and C. D. Manning, "Glove: Global vectors for word representation," in *Empirical Methods in Natural Language Processing (EMNLP)*, 2014, pp. 1532–1543.
- [10] Y. Goldberg and G. Hirst, *Neural Network Methods in Natural Language Processing*. Morgan & Claypool Publishers, 2017.
- [11] J. Chung, Ç. Gülçehre, K. Cho, and Y. Bengio, "Empirical evaluation of gated recurrent neural networks on sequence modeling," *CoRR*, 2014.
- [12] D. Bahdanau, K. Cho, and Y. Bengio, "Neural machine translation by jointly learning to align and translate," in *Proceedings of the 3rd International Conference on Learning Representations*, 2014.
- [13] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *CoRR*, 2014.