

Sosyal Medyada “Clickbait” Haber Başlıklarının Makine Öğrenmesi Algoritmaları ile Tespit Edilmesi

Detecting “Clickbait” News on Social Media Using Machine Learning Algorithms

Şura GENÇ

Enformatik Enstitüsü
Orta Doğu Teknik Üniversitesi
Ankara, Türkiye
sura.genc@metu.edu.tr

Elif SURER

Enformatik Enstitüsü
Orta Doğu Teknik Üniversitesi
Ankara, Türkiye
elifs@metu.edu.tr

Özetçe—“Tık tuzağı” olarak adlandırılan ve son yıllarda sosyal medyada sıkça kullanılan clickbait, insanları manipüle etmek ve hedeflenen websitelerine çekmek amacıyla abartılmış ve içerikle uyumlayan başlıklar kullanılmasıdır. Bu tür başlıklarda söz edilen içerik ana metinde yer almadığı ve ana metin çoğunlukla düşük kaliteli olduğu için tıklanan linkler insanları hayal kırıklığına uğratmaktadır. Bu çalışmanın amacı, Türkçe haber kaynaklarının Twitter paylaşımlarını kullanarak haber başlıklarının clickbait olup olmadığını tespit etmektir. Bu amaçla clickbait başlıklar için Limon Haber¹ ve Spoiler Haber’in² ve clickbait olmayan başlıklar için Evrensel Gazetesi³ ve Diken Gazetesi’nin⁴ Twitter hesapları veri seti oluşturmak için kullanılmıştır. Yapay Sinir Ağları kullanılarak veri seti üzerinde yapılan çalışmalar sonucunda, 0.91 doğruluk ve 0.91 F1 skoruna sahip olan ve Türkçe veri setlerinde en iyi sonucu veren bir model geliştirilmiştir.

Anahtar Kelimeler—clickbait tespiti; yapay sinir ağları; haber başlıkları; sosyal medya

Abstract—Clickbait, which has become very common in social media in recent years, is a technique which uses exaggerated and unreal headlines in order to manipulate people and attract them to their websites. Since the content mentioned in the title is not presented in the main text or the content of text is low-quality, clicked on links often disappoint people. In this study, we attempt to detect clickbaits in Turkish news using Twitter posts. For this purpose, headlines of news were collected from Twitter accounts of Limon Haber¹ and Spoiler Haber² for clickbait data and from Twitter accounts of Evrensel Newspaper³ and Diken Newspaper⁴ for non-clickbait data. Experimental results on news headlines show that using an artificial neural network, our model performs for clickbait detection with an accuracy of 0.91 with an F1-score of 0.91—which is the highest score in Turkish data sets.

Keywords—clickbait detection; neural networks; news headlines; social media

I. GİRİŞ

Clickbait, içeriği objektif olarak yansıtmayan ve insanların dikkatini çekerek onları içeriğe yönlendiren başlıkları ya da tanıtım mesajlarını ifade etmektedir. İnsan davranışını belirleyen ve yönlendiren temel faktörlerden birisi olduğu psikoloji çalışmaları [6] tarafından gösterilmiş olan merak duygusu, son zamanlarda ticari amaçlar ile manipüle edilmektedir. Loewenstein merak kelimesini “özellikli bilgi için duyulan istek” olarak tanımlamış ve insanların kendilerinde bir bilgi boşluğu farketmelerine neden olacak bir durumla karşılaştıklarında bu boşluğu dolduracak biçimde davrandıklarını ileri sürmüştür [6]. Bu bağlamda, insanların clickbait başlıklarla karşılaştıkları zaman kendilerinde böyle bir bilgi boşluğu olduğunu hissettikleri ve eksik bilgiyi edinerek bu boşluğu doldurmaya çalışıyor olabilecekleri düşünülmüştür. İnsanlar clickbait başlıklar tarafından yönlendirildikleri içeriği okurken genellikle bildikleri şeyler ile bilmek istedikleri şeyler arasındaki fark bekledikleri gibi azalmamakta ve bu durum insanların hayal kırıklığı yaşamalarına sebep olmaktadır.

Bir internet sayfasının gördüğü ilgi o sayfaya yapılan tıklamaların miktarı ile ölçülmekte ve bu gösterge o sayfanın reklamlardan ne kadar faydalanacağını da belirlemektedir [8]. Bu gerçeklik son zamanlarda sosyal medyada ve özellikle Twitter’da clickbait oranının artmasına neden olmuştur. Aynı zamanda gazetecilik için bir etik ihlal olan clickbait sadece sosyal medyada değil, tüm alanlarda gazetecilik için bir tehdit oluşturmaktadır ve bilginin güvenilirliği ile ilgili endişeleri arttırmaktadır. Clickbait örnekleri aşağıda görülebilir:

- NASA’dan flaş açıklama: Hızla yok oluyor...
- SON DAKİKA | Ünlü iş insanı yaşamını yitirdi.
- CD’niz varsa çöpe atmayın! Bakın ne işinize yarayacak?

¹ <https://twitter.com/LimonHaber>

² <https://twitter.com/spoilerhaber>

³ <https://twitter.com/evrenselgzt>

⁴ <https://twitter.com/DikenComTr>

2016'dan beri, clickbait tespit etmek amacıyla makine öğrenmesi tekniklerini kullanan çalışmalarda bir artış olmuştur. İlk çalışmalardan birinde, Lojistik Regresyon, Naif Bayes ve Rastgele Orman Algoritması kullanarak sosyal medya haberlerinde clickbait tespiti yapılmıştır [8]. Başka bir çalışmada, özel karakterlerin etkileri arasında karşılaştırma yaparak clickbait tespit etmek için Karar Ağacı ve Destek Vektör Makinesi gibi algoritmalarla Tay Dilinde başlıklar incelenmiş ve Karar Ağacı % 99.90 doğruluk skoruna ulaşmıştır [10]. Yakın zamanda yapılmış olan ve clickbait Facebook gönderilerini % 98.3 doğrulukla tespit eden bir çalışmada derin öğrenme yöntemleri kullanılmıştır [9]. Çin haber sitelerinde clickbait tespiti için de Evrişimli Sinir Ağları kullanılmıştır [11]. Özyineli Sinir Ağları kullanarak BuzzFeed ve ViralStories gibi kaynakların başlıklarında clickbait tespit eden bir çalışma da bulunmaktadır [1].

Haber başlıklarında kullanılan özel karakterler, başlıkların clickbait olup olmadığını tespit etmek için önemli özellikler olarak görülmüş ve birçok çalışmada incelenmiştir [5][8][10]. Hardalov ve arkadaşları güvenilir haberleri sahte haberlerden ayırmak için söz konusu özellikleri kapsayan bir model geliştirmişlerdir [5]. Bu çalışmanın özellik seçim sürecinde büyük harf sayısının, çift tırnak (") sayısından sonra ikinci önemli özellik olduğu tespit edilmiştir. Bunun yanında, sosyal medya etiketinin (#) öneminin soru işareti ve ünlem işaretinden daha fazla olduğu gösterilmiştir. Potthast ve arkadaşlarının çalışmasında özel karakterlerden sosyal medya etiketinin kuyruklu a'dan (@) daha önemli bir özellik olduğu ve her iki özelliğin de nokta sayısından daha önemli olduğu tespit edilmiştir. Özel karakterlerden daha az öneme sahip olan ortalama kelime uzunluğunun ortalama başlık uzunluğundan daha belirleyici bir özellik olduğu gösterilmiştir [8].

Bu projenin amacı, yapay sinir ağları kullanarak Twitter platformundaki Türkçe haber başlıklarında clickbait tespit edilmesidir. Clickbait tespiti için Türkçe veri üzerinde çalışılan yalnızca bir çalışma bulunmaktadır ve söz konusu çalışmada yapay sinir ağları kullanılmamıştır [4]. Bunun yanında, söz konusu çalışmanın örnek sayısı ve clickbait olmayan kategorideki örneklerin uygunluğu konusundaki eksiklikleri bu çalışmada giderilmeye çalışılmıştır. Bu çalışmada, derin öğrenme sınıflandırıcısı olan Çok Katmanlı Algılayıcı [7], Türkçe veri seti üzerinde kullanılmıştır.

II. YÖNTEM

A. Veri Seti

Haber başlıkları haber kanallarının Twitter hesaplarından günlük olarak paylaşıldığı için Twitter'daki verilerin miktarının ve yapısının clickbait tespiti için oldukça uygun olduğu düşünülmüştür. Clickbait kategorisindeki veriler, Limon Haber ve Spoiler Haber'in Twitter hesaplarından alınmıştır. Bu hesaplar çeşitli haber kaynaklarının clickbait içeren haber başlıklarını alıntılarında ve haberin başlıkta belirtilmeyen içeriğini paylaşmaktadır. Limon Haber'in hesabı 21369 ve Spoiler Haber'in hesabı 14600 tane gönderi içermektedir (16 Kasım 2018 tarihinde). Bu iki hesabın gönderileri Türkçe haber kaynaklarının birçoğunu içermektedir (Hürriyet, Sabah, Odatv, CNN Türk, Sol Haber, Spor Arena, NTV, Cumhuriyet, T24,

BBC Türkçe, MedyaTava, Gerçek Gündem, ABC, Sputnik Türkiye, Sözcü, Demirören Haber Ajansı, Milliyet, Posta, Aydınlık, İhlas Haber Ajansı, A Spor, Star, Halk TV, Tele1 TV, Yeni Akit, Mynet, Anadolu Ajansı, BirGün, Vatan, Yeniçağ). Söz konusu gönderiler bu hesaplar tarafından halihazırda clickbait olarak etiketlendiği için analize oldukça uygundur. Clickbait içermeyen gönderiler Evrensel ve Diken Gazetelerinin Twitter hesaplarından alınmıştır. Evrensel Gazetesi'nin 281000 ve Diken Gazetesi'nin 195000 tane paylaşımı mevcuttur (27 Aralık 2018 tarihinde). Tablo 1 clickbait olan ve olmayan başlıkların alındığı Twitter hesaplarını gösterirken, Tablo 2 ham ve işlenmiş tweet sayısını göstermektedir.

Twitter API her bir Twitter hesabında sınırlı sayıda veri alınmasına izin verdiği için (yaklaşık 3200 tweet), seçilen hesaplardan Twitter verilerini paylaşımları istenmiştir. Limon Haber ve Evrensel Gazetesi verilerini paylaşmış ve Limon Haber tarafından alıntılanan gönderilere erişmek için Tweepy kütüphanesi⁵ kullanılmıştır. Diken Gazetesi ve Spoiler Haber'e yapılan ricanın yanıtsız kalması nedeniyle bu hesapların verileri manuel olarak alınmıştır. Çalışmada clickbait olan (18204) ve olmayan (20997) örnek sayısı neredeyse eşit sayıda tutulmuştur.

B. Veri İşleme

Aynı anlama sahip fakat farklı çekim ekleri alan kelimelerin veri setinde ayrı kelimeler olarak temsil edilmesi eğitim sürecinin doğruluk oranını azaltacağı için veri setindeki kelimelerin çekim eklerinden ayrılarak yalnızca kökleri ile temsil edilmesine karar verilmiş ve bu işlem için Ahmet Aksoy tarafından geliştirilen kalbur projesinin "kelime_bol" kütüphanesi kullanılmıştır⁶. Veri setinde yer alan kelimelerin çekim ekleri kaldırılmış, fakat farklı anlamlara gelen kelimeler üreten yapım ekleri olduğu gibi bırakılmıştır. Bunun yanında, karışıklığa yol açacağı düşünüldüğü için tweetlerin içerdiği bağlantı adresleri "linkvar" kelimesi ile değiştirilmiştir. Oluşturulan veri setini analize hazırlamak için toplanan verideki farklı kelimeleri içeren bir vektör oluşturulmuştur. Daha sonra kelime frekansları sayılarak, her örnek oluşturulan vektör üzerinden tekil olarak temsil edilmiştir.

C. Özellik Seçimi

Potthast ve arkadaşlarının çalışmasına [8] dayanarak, sosyal medya etiketinin (#), soru işaretinin (?), ünlem işaretinin (!), noktanın (.), kuyruklu a'nın (@) ve diğer özel karakterlerin

TABLO I. CLICKBAIT OLAN VE OLMAYAN VERİ KAYNAKLARI

Twitter Hesapları	Tweet Sayıları	
	Clickbait içeren	Clickbait içermeyen
Limon Haber	16306	0
Spoiler Haber	1898	0
Evrensel Gazetesi	0	10937
Diken Gazetesi	0	10060
Toplam	18204	20997

TABLO II. HAM VE İŞLENMİŞ VERİ MİKTARLARI

Veri Tipi	Tweet Sayıları		
	Clickbait içeren	Clickbait içermeyen	Toplam
Ham Veri	23198	291937	315135
İşlenmiş Veri	18204	20997	39201

⁵ <http://www.tweepy.org/>

⁶ <https://github.com/ahmetax/kalbur>

miktarı bu çalışmada kullanılacak özellikler olarak seçilmiştir. Clickbait haber başlıklarında noktalama işaretlerinin formal kullanımı ihlal edildiği için [2], clickbait tespit etmede özel karakterlerin sayısı önemli bir göstergedir. Bu bilgilere dayanarak analizde "Diğer Özel Karakterler" adlı bir özelliğe de yer verilmiş ve yukarıda bahsedilen özel karakterler dışındaki tüm özel karakterler sayılmıştır.

Hardalov ve arkadaşlarının çalışması [5] incelendiğinde, clickbait tespiti için en uygun özelliklerin büyük harf, sosyal medya etiketi, soru işareti ve ünlem işareti sayısı olduğu görülmüştür. Bunun yanında, clickbait haber başlıkları "SONDAKİKA", "FLAŞ" gibi büyük harflerden oluşan kelimeler içerdiğinden, bu başlıkların clickbait olmayan başlıklardan daha fazla büyük harf içerdiği düşünülmüştür. Bu tespit sonucunda büyük harfler de analize eklenmiştir. Hardalov ve Potthast'ın çalışmalarındaki [5][8] önemli özelliklerden ortalama kelime uzunluğu ve ortalama tweet uzunluğu da çalışmaya dahil edilmiştir

D. Model

Bu çalışmada, haber başlıklarının içerdiği kelimeleri temsil eden özellik vektörünün yanısıra önemi literatürde belirtilen 9 özellik kullanılmıştır. Veri setinin içerdiği kelimelerin köklerini içeren bir özellik vektörü oluşturulmuştur. Böylece her bir tweet, clickbait olan ve clickbait olmayan haber başlıklarındaki kelimelerle ilgili bilgiler kullanılarak bu kelime uzayı yardımıyla temsil edilmiştir. Bu işlemde sonra, her bir haber başlığını temsil eden 10329 boyutlu bir özellik vektörü elde edilmiştir. Son olarak, her bir haber başlığının yukarıda bahsedilen 9 özelliği de özellik vektörüne eklendikten sonra 10338 boyutlu vektörle temsil edilen örnekler yapay sinir ağı eğitmek için kullanılmıştır.

III. DEĞERLENDİRME

Bu çalışma için iki model oluşturulmuştur. Birincisi, 10338 özellik ile temsil edilen veri seti tarafından eğitilmiştir. Öte yandan, haber başlıklarını kelime uzayıyla temsil etmenin etkisini görmek için, başka bir yapay sinir ağı yalnızca seçilen 9 özellik ile eğitilmişti ve ikinci modelin söz konusu 9 özelliği karşılaştırmak için bir gösterge olacağı düşünülmüştür. Her iki model için de Scikit-learn'de⁷ yer alan Çok Katmanlı Algılayıcı kullanılmıştır. Scikit-learn kütüphanesinin GridSearchCV yöntemi yardımıyla farklı parametrelerin modellere katkısı araştırılarak, kullanılan parametrelere göre önerilen en iyi iki model belirlenmiştir. Her model için en iyi performans gösteren hiperparametreler kullanılmıştır. Tablo 3 modeller için kullanılan hiperparametreleri göstermektedir. Eğitim ve test setinin doğruluk derecesini uzun bir zaman aralığında görmek için, model eğitilirken yüksek bir dönem (epok) sayısı seçilmiştir. Her bir modelin eğitiminden önce, 39201 örnekten oluşan veri seti, eğitim ve test seti olarak rastgele biçimde bölünmüştür (%80-%20). Oluşturulan modeller 31360 örnekten oluşan eğitim seti ile eğitilmiştir. Eğitimden önce, yukarıda bahsedilen 9 özelliğin ağırlıkları her bir giriş nöronunun ilk gizli katmandaki nöronlarla olan bağlantılarının mutlak değerleri toplanarak hesaplanmış ve bu ağırlıklar Tablo 4'te gösterilmiştir. Clickbait olmayan başlıkların daha fazla büyük harf içermesinin, spesifik bilgi verme amacıyla kişi ve yer isimlerinin bu

TABLO III. KULLANILAN HIPERPARAMETRELER

Parametreler	Kullanılan Değerler	
	Model 1	Model 2
Epok Sayısı (Number of Epochs)	12	100
Yığın Sayısı (Batch Size)	120	150
Gizli Katman Sayısı (Hidden Layer Size)	10	5, 2
Maksimum İterasyon (Maximum Iteration)	5	5
Alfa (Alpha)	1e-4	0.05
Optimizasyon Algoritması (Solver)	SGD	ADAM
Tolerans (Tolerance)	1e-8	1e-4
Öğrenme Oranı (Learning Rate)	Sabit	Sabit
Aktivasyon Fonksiyonu (Activation Function)	tanh	relu

TABLO IV. KULLANILAN ÖZELLİKLERİN AĞIRLIKLARI

Özellikler	Haber Başlıkları	
	Clickbait olan	Clickbait olmayan
Ortalama Kelime Uzunluğu	6.01	6.45
Ortalama Tweet Uzunluğu	7.97	11.65
Ortalama Büyük Harf Sayısı	4.48	6.51
Sosyal Medya Etiketleri (#)	0.08	0.05
Soru İşareti (?)	0.11	0.04
Ünlem İşareti (!)	0.35	0.03
Nokta (.)	0.60	0.38
Kuyruklu a (@)	0.00	0.06
Diğer Özel Karakterler	0.45	1.40

başlıklarda sıkça kullanılmasıyla ilgili olduğu düşünülmektedir. Clickbait olmayan başlıkların bu özelliği, bu başlıkların uzunluğunun fazla olmasını da açıklamaktadır. Sosyal medya etiketi, soru işareti, ünlem işareti ve nokta sayısının clickbait olan başlıklarda fazla olduğu da tabloda görülmektedir ve bu durum, clickbait başlıklarda noktalama işaretlerinin formal kullanımının ihlal edildiği tespitiyle tutarlı görünmektedir.

İlk modelin eğitim ve test skorları Şekil 1'de, verilen özelliklerin önemi Şekil 2'de görülebilir. İkinci modelin eğitim ve test skorları Şekil 3'te, verilen özelliklerin önemi ise Şekil 4'te yer almaktadır. Şekil 1 ve Şekil 2 karşılaştırıldığında, veri setini temsil etmek için kelime frekanslarının kullanılmasının doğruluk skorunu % 5 artırdığı görülebilmektedir. Tablo 5, bu katkının hassasiyet, geri çağırma ve F1 skorları için de önemli olduğunu göstermektedir. Model 1 ile clickbait başlıkları ayırt etmek için kritik olan "flaş", "o", "belli", "şok", "açıkla" ve "şoke" kelimeleri gibi en önemli 25 özellik tespit edilmiştir.

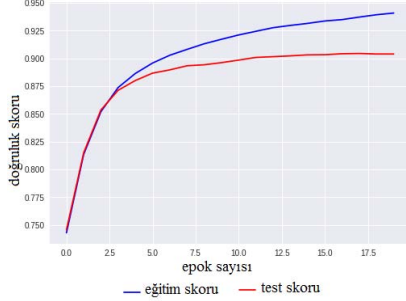
IV. TARTIŞMA

Bu çalışma clickbait olan ve olmayan Türkçe haber başlıklarının makine öğrenmesi yaklaşımları kullanılarak sınıflandırıldığı ikinci çalışmadır. Çalışma 0.91 doğruluk skoru ile Türkçe veri seti üzerinde clickbait tespiti yapmada en yüksek doğruluk oranına ulaşmıştır. Türkçe veri ile yapılan ilk projenin [4] örnek sayısı ve clickbait olmayan örneklerin uygunluğu konusundaki problemlerinin çözülmesi bu çalışmanın güçlü yanlarından. Söz konusu çalışmada BBC Türkçe ve Anadolu Haber Ajansı gibi clickbait içerikli haber kaynaklarından alınan veriler clickbait içermeyen kategoride değerlendirilmişken, bu çalışmada Limon Haber ve Spoiler Haber'den elde edilen verilere dayanarak bu haber kaynaklarının clickbait başlıklar içerdiği tespit edilmiş ve bu veriler clickbait kategorisinde

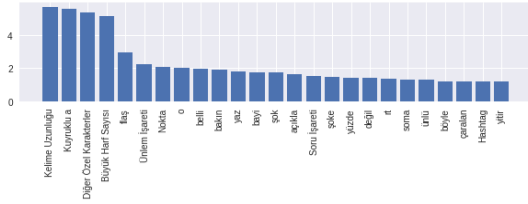
⁷ <http://scikit-learn.org/>

TABLO V. MODEL 1 VE MODEL 2 SONUÇLARI

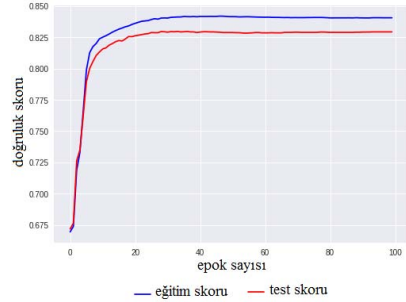
Metrikler	Sonuçlar	
	Model 1	Model 2
Özellik Vektörü Boyutu	10338	9
Eğitim-Test Seti Oranı	80%-20%	80%-20%
Örnek Sayısı	39201	39201
Doğruluk Skoru (Accuracy)	0.9125	0.8292
Hassasiyet Skoru (Precision)	0.91	0.83
Geri Çağırma Skoru (Recall)	0.91	0.83
F1 Skoru	0.91	0.83



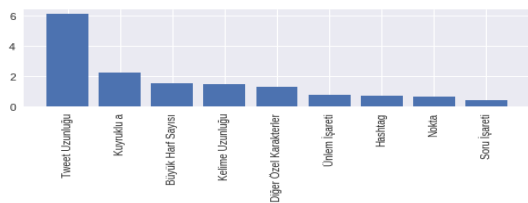
Şekil 1. İlk modelin eğitim ve test puanları.



Şekil 2. İlk modelde yer alan en önemli 25 özellik.



Şekil 3. İkinci modelin eğitim ve test puanları.



Şekil 4. İkinci modelde yer alan 9 özellik.

değerlendirilmiştir. Çalışmanın bir diğer güçlü yanı, örnek sayısının (39201) Geçkil ve arkadaşlarının çalışmasındaki [4] örnek sayısından (4000) ve —Rony, Hassan ve Yousuf'un çalışması [9] dışında— yukarıda bahsedilen diğer çalışmalardan daha fazla olmasıdır. Bununla birlikte, Geçkil ve arkadaşlarının çalışmasında [4] clickbait kategorisindeki başlıklar 3 haber

kaynağından toplanmışken, bu çalışmada 30 haber kaynağının verileri kullanılmıştır. Diğer bir önemli konu, başlıkların önceden sınıflandırılmış olmasıdır. Bu çalışmada kullanılan clickbait başlıklar, Limon Haber ve Spoiler Haber tarafından etiketlenmiş ve clickbait içermeyen haber kaynakları Limon Haber ve Spoiler Haber'in verileri incelenerek seçilmiştir. Çalışmanın bu güçlü yanı, kullanılan veriler sınıflandırılırken yanlış davranılmış olma ihtimalini oldukça düşük kılmıştır. Son olarak, bu çalışma sayesinde içerisinde clickbait olan ve olmayan Türkçe haber başlıkları bulunan ve gelecek çalışmalar için önemli bir kaynak olabilecek bir veri seti oluşturulmuştur. Çalışmanın sınırlılığı ise clickbait içermeyen başlıkların yalnızca iki Twitter hesabından alınmasıdır. Eğitim aşamasında, belli bir haber kaynağının örüntülerine ait olduğu düşünülen kelimeler kaldırılmış olsa da, bu tür kelimelerin tamamından kaçınmak mümkün olmamıştır (örneğin; Soma ve bayi). Haber kanallarının Twitter paylaşımlarına özgü olan belirli kalıplardan ve özelliklerden eğitim aşamasında kaçınmak için daha fazla Twitter hesabından veri alınmalıdır.

Sosyal medya ağlarına yayılmakta olan clickbait stratejisi platformların kullanılabilirliğine ve güvenilirliğine zarar vermektedir [8]. Bu çalışma ile clickbait tespit etme doğruluk oranları artırılmış ve sosyal medya platformlarındaki doğru bilgiyi ayırt edebilecek bir araç geliştirilmiştir.

TEŞEKKÜR

Bu çalışmanın yürütülmesi sırasında Twitter verilerini bizimle paylaşan Limon Haber ve Evrensel Gazetesi'ne teşekkür ederiz.

KAYNAKLAR

- [1] Anand, A., Chakraborty, T., and Park, N., "We Used Neural Networks to Detect Clickbaits: You Won't Believe What Happened Next!" *Lecture Notes in Computer Science Advances in Information Retrieval*, 541–547, 2017.
- [2] Chakraborty, A., Paranjape, B., Kakarla, S., and Ganguly, N., "Stop Clickbait: Detecting and Preventing Clickbaits in Online News Media.," *IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*, 2016.
- [3] Fu, J., Liang, L., Zhou, X., and Zheng, J., "A Convolutional Neural Network for Clickbait Detection", *4th International Conference on Information Science and Control Engineering (ICISCE)*, 2017.
- [4] Geçkil, A., Müngen, A. A., Gündogan, E., and Kaya, M., "A Clickbait Detection Method on News Sites", *IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*, 2018.
- [5] Hardalov, M., Koychev, I., and Nakov, P., "In Search of Credible News", *Artificial Intelligence: Methodology, Systems, and Applications Lecture Notes in Computer Science*, 172–180, 2016.
- [6] Loewenstein, G., "The Psychology of Curiosity: A Review And Reinterpretation", *Psychological Bulletin*, 116(1), 75–98, 1994.
- [7] Pedregosa, F. et al., "Scikit-learn: Machine Learning in Python", *JMLR*, 12, 2825–2830, 2011.
- [8] Potthast, M., Köpsel, S., Benno, S., and Hagen, M., "Clickbait Detection", *38th European Conference on IR Research*, 810–817, 2016.
- [9] Rony, M. M., Hassan, N., and Yousuf, M., "Diving Deep into Clickbaits", *Proceedings of the 2017 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining ASONAM*, 2017.
- [10] Wongsap, N., Prapphan, T., Lou, L., Kongyoung, S., Jumun, S., and Kaothanthong, N., "Thai Clickbait Headline News Classification and its Characteristic", *ICESIT-ICICTES*, 2018.
- [11] Zheng, H. T., Chen, C.Y., Yao, X., Sangaiah, A.K., Jiang, Y., and Zhao, C. Z., "Clickbait Convolutional Neural Network", *Symmetry*, 10, 138, 2018.