**Düzce Üniversitesi Bilim ve Teknoloji Dergisi**

*Araştırma Makalesi*

# Makine Öğrenmesi Algoritmaları ile Trol Hesapların Tespiti

Bengisu ERDİ [a], Eylül Aleyna ŞAHİN [b], Muzaffer Su TOYDEMİR [c], Tansel DÖKEROĞLU [d]

[a] *Bilgisayar Mühendisliği Bölümü, Mühendislik Fakültesi, Türk Eğitim Derneği Üniversitesi, Ankara, TÜRKİYE*
[b] *Bilgisayar Mühendisliği Bölümü, Mühendislik Fakültesi, Türk Eğitim Derneği Üniversitesi, Ankara, TÜRKİYE*
[c] *Bilgisayar Mühendisliği Bölümü, Mühendislik Fakültesi, Türk Eğitim Derneği Üniversitesi, Ankara, TÜRKİYE*
[d] *Bilgisayar Mühendisliği Bölümü, Mühendislik Fakültesi, Türk Eğitim Derneği Üniversitesi, Ankara, TÜRKİYE*
* *Sorumlu yazarın e-posta adresi: tansel.dokeroglu@tedu.edu.tr*

**DOI : 10.29130/dubited.748366**

## ÖZET

Sosyal medya kullanımı beraberinde birçok yeni problemi de getirmektedir. Kişilerin düşünce, duygu ve fikirlerini kolaylıkla paylaşabildiği bu ortamlarda insanlara saldırılarda bulunan hesaplara son zamanlarda sıkça rastlanmaktadır. Siber zorbalık olarak adlandırılan bu eylemi yapan trol hesapların insanların sosyal yaşantılarına verdiği zararların engellenmesi bir ihtiyaç haline gelmektedir. Bu tip kullanıcıların sayıları takip edilemeyecek miktarlara ulaştığı durumlarda yazılımlar ile tespit edilmesi, engellenmesi ve sınırlandırılması gerekebilmektedir. Biz bu çalışma ile Twitter'daki trol hesapları tespit etmek için makine öğrenmesi destekli bir yazılım geliştirdik. Support Vector Machine, Logistic Regression ve Random Forest Regression yöntemleri ile Twitter'dan elde ettiğimiz veriler ile trol kullanıcıların mesajları üzerinden çıkardığımız özellikler üzerinde detaylı deneyler gerçekleştirdik. Elde ettiğimiz sonuçlarda %93.93'lere varan oranlarda trol hesapları tespit etmeyi ve engellemeyi başardık.

***Anahtar Kelimeler:*** *Trol, Makine öğrenmesi, Logistic regression, Support Vector Machine*

# Detection of Troll Accounts using Machine Learning Algorithms

## ABSTRACT

The use of social media is increasing day by day and introduces new problems. In these environments where people can easily share their thoughts, feelings and ideas, accounts that have made humiliating and offensive attacks have been frequently encountered. It becomes a necessity to prevent troll accounts on the social media. The messages of such users disturb people, and in cases where their number reaches unreachable amounts, they must be detected with software and blocked when necessary. With this study, we use machine learning methods to detect user accounts exhibiting trolling behaviors on Twitter. With the Support Vector Machine, Logistic Regression and Random Forest Regression, we conduct extensive experiments with the data we gathered on Twitter and the features we extracted from messages. In the results we obtained, we managed to detect and prevent troll accounts up to 93.93%.

***Keywords:*** *Troll, Machine learning, Logistic Regression, Support Vector Machine*

Geliş: 05/06/2020, Düzeltme: 04/11/2020, Kabul: 13/11/2020

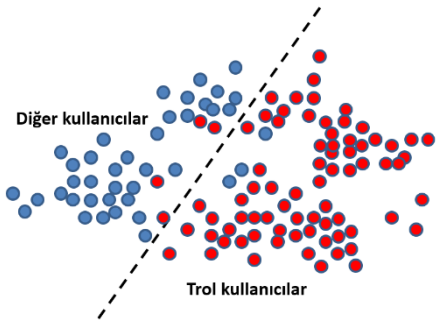
# I. GİRİŞ

Sosyal medya kullanımının son yıllarda büyük bir hızla artması ile birlikte, hesapların kullanım amaçları da değişiklik göstermeye başlamıştır. Bu değişikliğin gerçekleşmesiyle kullanıcılar, kimliği belirsiz hesaplardan siber zorbalık veya duygusal saldırılara maruz kalabilmektedir. Zorbalığa maruz kalan hesapların büyük çoğunluğunu ünlü kişiler ve siyasi liderler oluşturmaktadır. Dünya genelinde siber zorbalığa uğrayanların sayısı gün geçtikçe artmaktadır [1]. Kişisel hesapların yanı sıra, belirli bir konu veya topluluğa da saldırılar yapılabilmektedir. Siber zorbalık gösteren bireyler, sosyal medya hesapları açarken eksik veya farklı bilgilerle çoğu zaman kimliklerini gizli tutmaktadırlar. Bu sayede, negatif ve insanların psikolojisini kötü yönde etkileyebilecek yorumlarda bulunabilmek daha kolay bir şekilde yapılabilmekte ve hukuksal sorumluluklardan kurtulabilmenin de yolu açılmış olmaktadır. Neredeyse tüm sosyal medya araçlarında bu durum söz konusu olmaktadır.

Trol olarak adlandırılan bu kullanıcı türlerinin tespit edilmesi ve engellenmesine yönelik son dönemde çalışmalar yapılmasına ihtiyaç duyulduğu görülmektedir. Sosyal medya kuruluşları, kullanıcılarından gelen şikayetler üzerine, bu alanda bir dizi çalışmaya da öncülük etmektedir. Bu gelişmeler ile birlikte, trol kullanıcıların sosyal medya mesajlarında görülme ve rahatsız etme ihtimali azalmaktadır.

Biz bu çalışmamız ile birlikte, Türkçe mesajlaşma yapan trol hesaplarına yönelik olarak makine öğrenmesi destekli algoritmaların performanslarını test ederek bu alandaki etkin algoritmaları tespit etmeye çalıştık. Türkiye'de en çok kullanılan sosyal medya araçlarından birisi olması ve çok sayıda trol hesabın bulunmasından dolayı Twitter'ı deney ortamımız olarak seçtik. Twitter'ın 2019 verilerine göre 8,8 milyon kullanıcısı ile Türkiye dünya sıralamasında 7. sırada bulunmaktadır [2]. Bu anlamda Twitter bu çalışma için en uygun sosyal medya alanlarından birini oluşturmaktadır.

Bildiğimiz kadarı ile literatürde Türkçe tabanlı ve Twitter trol hesaplarının sınıflandırılması için kullanılabilecek bir veri seti bulunmamaktadır. Bu yüzden, Twitter'dan ünlü bir siyasi parti liderinin hesabını hedef alan trol mesajlarını kullanarak 100 kullanıcı hesabından oluşan 238.925 adet mesajın incelemesi ile 2.102 adet özellik (feature) içeren bir veri seti hazırladık. Bu veri seti üzerinde güncel en iyi makine öğrenmesi algoritmaları arasında olan Support Vector Machine (SVM), Logistic Regression (LR) ve Random Forest Regression (RFR) ile deneyler gerçekleştirdik. Sonuçlarımızda %93.93 seviyesinde tahmin başarısı elde ettiğimizi gözlemledik. Şekil 1'de trol ve diğer kullanıcıları sınıflandırılması ile ilgili bir görsel sunulmuştur.




***Şekil 1.*** *Trol ve diğer kullanıcılar olarak sınıflandırılmış Twitter hesapları*

Makalemizin 2. Bölümünde daha önce yapılmış çalışmalar ile ilgili kapsamlı bilgi sunulmuştur. 3. Bölümde önerilen algoritmalar hakkında kısa bilgiler verilmektedir. 4. Bölümde hazırladığımız veri seti ve gerçekleştirilen deneyler ile ilgili sonuçlar verilmektedir. Son bölümde ise çalışmanın genel bir değerlendirmesi ile birlikte sonuçları ve gelecekte yapılacak çalışmalar özetlenmektedir.

# II. ÖNCEKİ ÇALIŞMALAR

Bu bölümde, trol sınıflandırma (tespit) problemi üzerinde yapılmış olan güncel çalışmalar ile ilgili detaylı bilgi verilmektedir. García ve arkadaşları son zamanlarda yaptıkları bir çalışmada, Twitter sosyal ağında iftira niteliğinde faaliyetler için kullanılan sahte profilleri tespit etmek için yorumların içeriğini analiz ederek profilleri tespit etmeye çalışmaktadır [3]. Bu yaklaşım ile, bir okulda siber zorbalık durumunu tespit etmek için geliştirdikleri bu metodoloji kullanılmıştır. Kızılkaya yazdığı doktora tezinde Twitter hesaplarında yaptığı duygu analizleri ile en yüksek oy alması beklenen aday ve partilere ilişkin bir ay boyunca atılan tweetleri ele almıştır [4]. Sriram ve arkadaşları 2010 yılında yaptıkları bir çalışmada Twitter'daki kısa mesajları sınıflandırarak haberler, etkinlikler, görüşler, fırsatlar ve özel mesajlar gibi önceden tanımlanmış genel sınıflara ayırmışlardır [5].

Xiang ve arkadaşları Twitter'da küfürlü mesajlar ile ilgili rahatsız edici içeriği olan tweetleri tespit etmek için yeni bir yarı denetimli yaklaşım önerdiler [6]. LR kullanarak 4.029 tweet mesajı üzerinden %75,1'lik bir gerçek pozitif orana (TP) ulaşmayı başarmışlardır. Fornacciari ve arkadaşları Twitter'daki trol ve gerçek kullanıcıların farklı özelliklerinin tespit edilmesini amaçlayan bir çalışma gerçekleştirdiler [7]. Bu özelliklerin tespit edilebilmesi için araştırmacılar TrollPacifier isimli bir sistem geliştirdiler. Yazma stili, duyarlılık, davranış biçimi, sosyal etkileşimler, profillerinde paylaştıkları medya türleri ve paylaşma zamanları olmak üzere altı kullanıcının grup özelliklerini belirlemeyi başardılar.

Tsantarliotis ve arkadaşları sosyal ağlardaki trol saldırılarını önceden belirlemek amacıyla bir çalışma gerçekleştirdiler [8]. Bu çalışmalarında araştırmacılar, trol saldırılarına maruz kalabilecek paylaşımları önceden tahmin edip bunları önlemeyi planlamışlardır. Bu tahmini yapabilmek için, kullanıcılar tarafından sosyal medyada yapılan paylaşımların içeriği ve tarihi özellik olarak belirlenip trol saldırılarına karşı zayıflıkları incelenmiştir. Silva ve Engelin Twitter gibi sosyal medya platformlarındaki işlenmemiş verinin çokluğu sebebiyle bu platformlarda paylaşılan kısa metinlerin sınıflandırılması gerekliliğinden yola çıkarak bir çalışma sunmuşlardır [9]. Bu sınıflandırmanın yapılabilmesi için Bag-of-Words (Kelimeler Çantası) modeli kullanılmıştır. Bu model çok kullanılan bir sınıflandırma tekniği olmasına rağmen belirli sınırlamaları bulunmaktadır. Bu sebeple araştırmacılar Twitter'da kullanıcılara ait profilleri ve metinleri incelemişlerdir. Her kullanıcının paylaştığı metinlerin içeriklerini, haber, olay, düşünce, pazarlama ve özel mesaj olarak sınıflara ayırmışlardır. Mihaylov ve Nakov yaptıkları çalışma ile Twitter'daki sahte hesapları saptamayı amaçlamışlardır [10]. Bunun için 'kümeleme analizi' algoritmalarından faydalandılar. Elde edilen kullanıcıların tweetleri hangi saat dilimi, gün ve sıklıkla attığına göre tespit edilerek, DBSCAN ve K-means kümeleme analizleri kullanılarak sınıflandırılmıştır. Hong ve arkadaşları yaptıkları çalışmada gelecekteki re-tweetlerin sayısıyla ölçülen mesajların popülerliğini tahmin etme problemini araştırdılar ve Twitter'da bilgi yayılımını etkileyen faktörlere ışık tutmaya çalıştılar [11]. Bulut ve Yörük çalışmalarında popülizm objektifiyle Türkiye'deki siyasi trollemeyi analiz etmiştir [12].

Özsoy çalışmasında, korkuyu, politik iktidarın kurulması ve sürdürülmesinde önemli bir enstrüman olarak görmekte ve Türkiye'de Twitter trollerinin ürettikleri hashtagler üzerinden Türkiye'de egemen olan seküler korkuları aktive etmeyi hedeflediklerini belirtmektedir [13]. Bu kapsamda, Türkiye'deki Twitter trolleri, korku ve iktidar bağlamında ele almıştır. Chavoshi ve arkadaşları Twitter'da, insan tarafından işletilmesi pek olası olmayan anormal olarak ilişkili kullanıcı hesaplarını tanımlamak için bir teknik geliştirmişlerdir [14]. Bu yeni bot algılama yaklaşımı, çapraz korelasyon kullanıcı faaliyetlerini dikkate alınmakta ve kullanıcıları bağımsız olarak dikkate alan ve son zamanlarda büyük miktarda etiketlenmiş veri gerektiren mevcut bot algılama tekniklerinin aksine etiketli veri gerektirmemektedir. Jane ve arkadaşları Rusya'nın İnternet Araştırma Ajansı'nın Twitter'da sahte hesaplar çalıştırarak 2016 ABD seçimlerine müdahale etmeye çalıştığına dair kanıtlar olduğunu ve bu hesapların genellikle "Rus trolleri" olarak anıldığını belirtiyorlar [15]. Bilim adamları çalışmalarında bir Twitter hesabının 170 bin kontrol hesabı kümesi içinde bir Rus trolü olup olmadığını tahmin eden makine öğrenme modelleri geliştirerek ve Twitter'da halen Rus devleti adına hareket eden aktif hesaplar bulmak için bu modelin kullanılabileceğini göstermişlerdir. Badawy ve arkadaşları Rus trol

hesapları tarafından Twitter'da üretilen yayınları yeniden paylaşan kullanıcılara daha yakından göz atarak bu manipülasyon kampanyasının etkilerini araştırmışlardır [16]. 16 Eylül - 9 Kasım 2016 tarihleri arasında Twitter'da yaklaşık 5.7 milyon ayrı kullanıcı tarafından paylaşılan 43 milyondan fazla seçimle ilgili gönderi içeren bir veri kümesi incelenmiştir. Gelişmiş bot tespit tekniklerini kullanılarak, liberal ve muhafazakar kullanıcıların sırasıyla %4.9 ve %6.2'sının bot olduğu tahmin edilmiştir.

Cheng ve arkadaşları trol davranışlarının bir azınlıkla sınırlı olmadığını, sıradan insanların da bu tür davranışlarda bulunabileceğini gösteren bir çalışma sunmaktadır [17]. Seah ve arkadaşları trollerin çevrimiçi forumlardaki metinsel içerik hissinden saptanması için bir yaklaşım önermektedir [18]. Araştırmacılara göre, troller genellikle gönderilerinde negatif duyguları ifade ettikleri için, duygu analizinden özellikler türetilebilir ve trollerin ikili ve sıralı sınıflandırmasını yapmak için SVM kullanılabilmektedir.

# III. KULLANILAN SINIFLANDIRMA ALGORİTMALARI

Bu bölümde, çalışmamızda trol hesapların sınıflandırılması için kullanılan makine öğrenmesi teknikleri LR, SVM ve RFR kısaca tanıtılmıştır [19].

## A. LOGISTIC REGRESSION (LR)

LR, sonucu belirleyen bir veya daha fazla bağımsız değişken bulunan veri kümesini incelemek için kullanılan istatistiksel bir yöntemdir [12]. Sonuç, ikili bir değişkenle değerlendirilir. LR, bağımlı değişken ikili yani yalnızca 1 (doğru vb.) veya 0 (yanlış vb.) olarak gösterilen verileri içeriyor. LR, iki yönlü karakteristiği ile ilgili bir dizi bağımsız değişken arasındaki ilişkiyi tanımlamak için en uygun modeli bulmayı amaçlar. LR, ilgi karakteristiklerinin varlığının olasılığını logit dönüşümünü tahmin etmek için bir formülün katsayılarını (ve standart hatalarını ve önem seviyelerini) üretir:

$$\text{log}it(p) = b_0 + b_1 X_1 + b_2 X_2 + b_3 X_3 + \ldots + b_k X_k \quad (1)$$

Burada *p*, karakteristik özelliğinin var olma olasılığıdır.

$$\frac{p}{p-1} = \frac{\text{karakteristik özelliğinin var olma olasılığı}}{\text{karakteristik özelliğinin var olmama olasılığı}} \quad (2)$$

ve

$$\text{log}it(p) = \ln\left(\frac{p}{1-p}\right) \quad (3)$$

Karekök hataların toplamını en aza indirgeyen parametreleri seçmek yerine, LR ile tahmin, örnek değerlerin gözlem olasılığını en yükseğe çıkaran parametreleri seçer. LR, bağımlı değişken ikili olduğunda yürütülecek uygun regresyon analizidir. Tüm regresyon analizlerinde olduğu gibi, LR'da bir tahmini analizdir. LR, veriyi tanımlamak ve bir bağımlı ikili değişken ile bir veya daha fazla nominal, sıra arası, aralık veya oran seviyesinde bağımsız değişkenler arasındaki ilişkiyi açıklamak için kullanılır.

### B. SUPPORT VECTOR MACHINES (SVM)

**S**VM, ayrı kategorilere ait verilerin birbirlerinden bir hiper düzlem aracılığıyla ayırmak için kullanılan bir makine öğrenmesi tekniğidir [13]. Aynı zamanda sınıflandırma ve ya regresyon problemleri için kullanılabilen denetimli bir algoritmadır. LR ile benzer bir sınıflandırma algoritmasıdır. Her iki teknik de iki sınıfı ayıran en iyi çizgiyi bulmaya çalışırlar. Algoritma çizilecek doğrunun iki sınıfında elemanlarına en uzak yerden geçecek şekilde ayarlanmasını sağlar. Hiçbir parametre almayan bir tekniktir. SVM aynı zamanda doğrusal ve doğrusal olmayan verileri de sınıflandırabilir ancak genellikle verileri doğrusal olarak sınıflandırmaya çalışır. Doğrusal sınıflandırma gerçekleştirmenin yanı sıra, SVM'ler, çekirdek numarası diye adlandırılanları kullanarak doğrusal olmayan sınıflandırmayı verimli bir şekilde gerçekleştirilebilir ve girişlerini yüksek boyutlu özellik alanlarına örtülü olarak eşlerler.

### C. RANDOM FOREST REGRESSION (RFR)

Rastgele ormanlar, birden çok karar ağaçlarından oluşan yaklaşık tahminlerde bulunan denetimli bir öğrenme tekniğidir. Hiper parametre kestirimi yapılmadan da iyi sonuçlar vermesi hem regresyon hem de sınıflandırma problemlerine uygulanabilir olmasından dolayı popüler makine öğrenmesi modellerinden bir tanesidir [14]. RFR, birden fazla karar ağacını oluşturur ve daha doğru ve istikrarlı bir tahmin elde etmek için onları birleştirir, bir karar ağacı veya bir bagging (torbalama) sınıflandırıcısı olarak aynı hiper parametreye sahiptir. Bir düğümü parçalara ayırırken en önemli özelliği aramak yerine, rastgele bir özellik alt kümesi arasında en iyi özelliği arar. Bu, genellikle daha iyi bir modelle sonuçlanan geniş bir çeşitlilikle sonuçlanır. Bu nedenle, RFR, bir düğümün bölünmesi için algoritma tarafından özelliklerin sadece rastgele bir alt kümesi dikkate alınır.

## IV. DENEYLER VE ELDE EDİLEN SONUÇLARIN DEĞERLENDİRİLMESİ

Bu bölümde, deneylerde kullandığımız veri setinin hazırlanması, bu veri setinden elde edilen alt veri setleri ile yapılan deney sonuçları sunularak değerlendirilmiştir.

### A. TROL VERİ SETİNİN HAZIRLANMASI

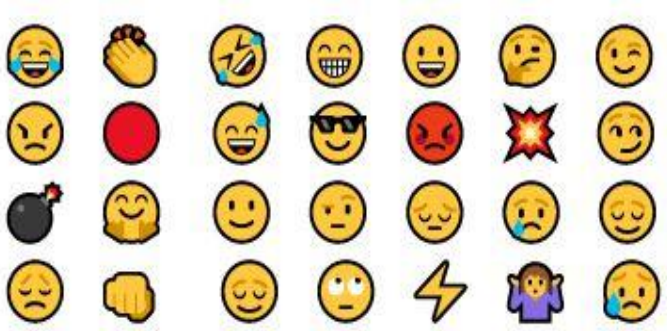
Makalemizdeki algoritmaları test edebilmek için yeni bir Türkçe Twitter veri seti oluşturmamız gerekti. Bu amaç için 50 sıradan kullanıcı ve 50 trol hesap olmak üzere toplamda 100 adet Twitter hesabını inceledik ve bu hesaplardan toplam 238.925 adet Türkçe tweet indirdik. Ayrıca bu kullanıcıların Twitter'a kayıt olurken girdiği kişisel bilgileri de tablo halinde kaydettik. Bu bilgiler; kullanıcı adı, biyografi, konum, link, doğum tarihi, Twitter'a katıldığı tarih, takipçi, takip edilen ve tweet sayıları bilgilerini içermektedir. Kullanıcıların adı, katıldığı tarih, tweet sayısı, takipçi ve takip edilen sayısı dışındaki diğer bilgiler opsiyonel olarak doldurulmaktadır. Bu nedenle bazı kullanıcıların bu bilgileri tam olarak elde edilememiştir. Hesapların kişisel bilgileri kaydedildikten sonra, Twitter'da yazılımcı (developer) olarak bir hesap açarak, Twitter API erişim talebinde bulunduk, erişim anahtarına sahip olduktan sonra Java programa dili yardımı ile projemiz ve Twitter API arasında bir bağlantı sağladık. Böylelikle Twitter API, bize saptadığımız tüm kullanıcı hesaplarının tweetlerine erişmemizi ve elde etmemizi sağlamış oldu. Bu sayede 100 kullanıcının adına özel not defteri oluşturulmuş ve tüm tweetleri kaydedilmiştir. Ardından, trol ve sıradan hesap olmak üzere iki yeni kayıt defteri açılmış ve tüm sıradan ve trol hesaplar kendi aralarında yeni kayıt defterlerine kaydedilmiştir. Bunu düzenlemedeki amacımız, sıradan ve trol hesapların kendi aralarında en çok kullanılan kelimeleri saptamaktır.

Python program dilini kullanılarak bu iki farklı kayıt defterindeki tweetler analiz edilip, kullanılan her kelimenin tekrar etme değeri tespit edildi. Python'daki sözlük yapısını kullanarak key(=kelime),

value(=sıklık) modelinde eşledikten sonra, kelimeleri sıklık değerlerine göre azalan bir sırayla kayıt ettik.

Bu aşamada hazırladığımız veriler makine öğrenmesinde kullanacağımız algoritmalara verilecek özellikleri belirlememizi sağlamış oldu. Sıradan ve trol hesapların kelime sıklığının sıralanmış halini incelemek bize 2.102 adet özellik (feature) tespit ettik. Bu özelliklerin 1406 tanesi sıradan hesaplar, 696 tanesi de trol hesaplar tarafından kullanılmaktadır. Trol ve sıradan hesapların özelliklerini belirledikten sonra her bir kullanıcının hangi özellikleri kullandıklarını saptadık. Eğer bir kullanıcı 2.102 adet özellikten herhangi birisini kullandıysa o özelliğin kullanıcı-özellikler matrisindeki değerini 1, eğer kullanmıyorsa o haneye 0 olarak belirledik. Bu kodlama sayesinde 100 kullanıcıya ait 238.925 adet veriyi dosya formatında yazdık.

Son olarak bu bilgilerin yanında tablonun ikinci sütununda, kullanıcının trol mü yoksa sıradan bir kullanıcı mı olduğunu kaydettik. Eğer kullanıcı trol ise 1, değilse o haneye 0 yazdık. Bu bilgi sayesinde, Makine Öğrenmesi algoritmaları ile deneylere başladığımız zaman, algoritmaların sonuçlarının doğruluklarını öğrenebildik. Bütün verileri bir arada topladığımız bu dosyada; sütun bilgileri bütün özellikleri, satır bilgileri ise tüm kullanıcı isimleri olacak şekilde yazdık. Ayrıca yukarıda belirtilen adımları tekrarlayarak yeni bir veri seti oluşturduk. Bunu önceki veri setinden ayıran fark 0/1 gösterimi yerine kullanıcı-özellikler matrisi, kullanıcıların her özelliği kaç kere kullandıklarına göre doldurulmuş olmasıdır. Bu veri setlerimizde ayrıca, kullanıcıların tweetlerinde geçen emojileri de belirledik. Şekil 1'de Twitter'da sıkça kullanılan emoji setinden bazıları görülmektedir. Verilerimizi diğer araştırmacıların da faydalanabilecekleri şekilde kendi web sayfamız üzerinden erişime açtık[1].



***Şekil 2.*** *Twitter'da kullanılan emojiler*

## B. TÜM ÖZELLİKLER (FEATURES) İLE YAPILAN DENEY SONUÇLARI

Deneylerin bu aşamasında kullanıcıların en çok kullanılan ilk 50 özellik ve tweetlerindeki emojilerin 0/1 formatında tutulduğu veri setleri ile her özelliği kaç kere kullandıklarına göre oluşturulan veri setleri ile yapılan deneylerin sonuçları 3 algoritma için Tablo 1, 2, 3, 4, 5 ve 6 da sunulmuştur.

***Tablo 1.*** *Algoritmaların tüm özelliklerin 0/1 gösterimi ile elde edilen veri setindeki doğruluk yüzdeleri*

| | ***LR*** | ***SVM*** | ***RFR*** |
|---|---|---|---|
| *True positive* | *48* | *47* | *46* |
| *True negative* | *40* | *42* | *40* |
| *False positive* | *2* | *3* | *4* |
| *False negative* | *10* | *8* | *10* |

***Tablo 2.*** *Algoritmaların en çok kullanılan ilk 50 özelliğin 0/1 gösterimi ile elde edilen veri setindeki doğruluk yüzdeleri*

| | ***LR*** | ***SVM*** | ***RFR*** |
|---|---|---|---|
| *True positive* | *48* | *47* | *48* |
| *True negative* | *45* | *44* | *43* |
| *False positive* | *2* | *3* | *2* |
| *False negative* | *5* | *6* | *7* |

***Tablo 3.*** *Algoritmaların emojilerin 0/1 gösterimi ile elde edilen veri setindeki doğruluk yüzdeleri*

| | ***LR*** | ***SVM*** | ***RFR*** |
|---|---|---|---|
| *True positive* | *44* | *40* | *42* |
| *True negative* | *37* | *36* | *33* |
| *False positive* | *6* | *10* | *8* |
| *False negative* | *13* | *14* | *17* |

***Tablo 4.*** *Algoritmaların tüm özelliklerin kullanım sıklığı ile elde edilen veri setindeki doğruluk yüzdeleri*

| | ***LR*** | ***SVM*** | ***RFR*** |
|---|---|---|---|
| *True positive* | *46* | *45* | *45* |
| *True negative* | *38* | *39* | *39* |
| *False positive* | *4* | *5* | *5* |
| *False negative* | *12* | *11* | *11* |

***Tablo 5.*** *Algoritmaların en çok kullanılan ilk 50 özelliğin kullanım sıklığı ile elde edilen veri setindeki doğruluk yüzdeleri*

| | ***LR*** | ***SVM*** | ***RFR*** |
|---|---|---|---|
| *True positive* | *47* | *48* | *47* |
| *True negative* | *37* | *39* | *39* |
| *False positive* | *3* | *2* | *3* |
| *False negative* | *13* | *11* | *11* |

***Tablo 6.*** *Algoritmaların emojilerin kullanım sıklığı ile elde edilen veri setindeki doğruluk yüzdeleri*

| | ***LR*** | ***SVM*** | ***RFR*** |
|---|---|---|---|
| *True positive* | *47* | *42* | *43* |
| *True negative* | *33* | *31* | *35* |
| *False positive* | *3* | *8* | *7* |
| *False negative* | *17* | *19* | *15* |

Tablo 7'de, algoritmaların çalışma zamanları saniye olarak sunulmuştur. Çalışma zamanlarının 4 saniyeyi geçmediği ve oldukça kısa sürelerde tahmin yapabilme kabiliyetine sahip olduğu görülmektedir.

***Tablo 7.*** *Algoritmaların çalışma zamanları (saniye).*

| | *0/1 gösterimli veri seti* | | |
|---|---|---|---|
| | *Tüm özellikler ile* | *En sık 50 özellik ile* | *Emojiler ile* |
| **LR** | 2.93 | 2.72 | 2.33 |
| **SVM** | 2.90 | 2.62 | 2.52 |
| **RFR** | 3.0 | 2.33 | 2.40 |
| *ortalama* | 2.94 | 2.56 | 2.42 |
| | | | |
| | *Sıklık gösterimli veri seti* | | |
| | *Tüm özellikler ile* | *En sık 50 özellik ile* | *Emojiler ile* |
| **LR** | 3.53 | 2.76 | 2.69 |
| **SVM** | 2.49 | 2.34 | 2.39 |
| **RFR** | 2.87 | 2.31 | 2.05 |
| *ortalama* | 2.96 | 2.47 | 2.38 |

## C. LOGISTIC REGRESSION İLE YAPILAN DENEY SONUÇLARI

Altı adet veri seti ile LR deneyleri gerçekleştirilmiştir. İlki kullanıcılar tarafından herhangi bir özellik ya da özellikler kullanılmış ise, özelliğin bulunduğu o kullanıcıya ait satırdaki hücreye 1, kullanılmamış ise 0 yazılan; ikincisi ise bir özelliğin ya da özelliklerin bir kullanıcı tarafından hangi sıklık ile kullanıldığının hücrelere yazıldığı veri setidir.

100 Twitter kullanıcısının (50 adedi trol, 50 adedi gerçek) %80'ini eğitim verisi, %20'ini de test verisi olarak ayrılmıştır. Bu test verilerinin çıkaracağı sonuçlardaki yanılma payını görmek için Python programlama dili ile yazdığımız koda, test verilerinin tahmini (Scoring) sonuçlarını yüzde olarak hesaplayan bir kod parçası eklenmiş ve bu veri seti için test verilerinin tahmini sonucu %89.43 olarak hesaplanmıştır. Tahmini sonucun hesaplanmasından sonra tahmini ve gerçek değerlerin karşılaştırılmasını görmek için hata matrisi (confusion matrix) oluşturulmuştur. Bu hata matrisindeki kombinasyonlardan doğru pozitif (true positive) 48, doğru negatif (true negative) 40, yanlış pozitif (false positive) 2 ve yanlış negatif (false negative) 10 sonuçlarını elde edilmiştir.

Kullanıcıların her özelliği kaç kere kullandıklarını tutan ikinci veri seti içinde ise ilk deneyde yaptıklarımızın aynısını uygulayarak, bu ikinci deneyimizde test verilerinin tahmini sonucu %85.08 olarak hesaplanmıştır. Hata matrisinde tahmini ve gerçek değerlerin karşılaştırılması sonucunda ise doğru pozitif (true positive) 46, doğru negatif (true negative) 38, yanlış pozitif (false positive) 4 ve yanlış negatif (false negative) 12 değerleri bulunmuştur.

2.102 adet olan özelliği 50 özelliğe indirerek oluşturduğumuz 0/1 kullanıcı-özellikli üçüncü veri setimiz ile ilk deneyde yaptıklarımızın aynısını uygulayarak %93.93 tahmini sonucuna ulaşılmıştır. Hata matrisinde tahmini ve gerçek değerlerin karşılaştırılması sonucunda doğru pozitif (true positive) 48, doğru negatif (true negative) 45, yanlış pozitif (false positive) 2 ve yanlış negatif (false negative) 5 değerleri elde edilmiştir.

Kullanıcıların her özelliği kaç kere kullandıklarını tutan ve 50 adet özellik bulunduran dördüncü veri setinde önceki veri setlerinde yaptığımız adımları uygulayarak %86.24 tahmini sonucuna ulaşılmıştır. Hata matrisinde tahmini ve gerçek değerlerin karşılaştırılması sonucunda doğru pozitif (true positive) 47, doğru negatif (true negative) 37, yanlış pozitif (false positive) 3 ve yanlış negatif (false negative) 13 değerleri elde edilmiştir.

2.102 adet olan özelliğin içinden sadece emoji özellikleri ile oluşturduğumuz 0/1 kullanıcı-özellikli beşinci veri setimiz ile ilk deneyde yaptıklarımızın aynısını uygulayarak %82.46 tahmini sonucuna ulaşılmış, hata matrisinde tahmini ve gerçek değerlerin karşılaştırılması sonucunda doğru pozitif (true positive) 44, doğru negatif (true negative) 37, yanlış pozitif (false positive) 6 ve yanlış negatif (false negative) 13 değerleri elde edilmiştir.

Kullanıcıların her özelliği içindeki emojileri kaç kere kullandıklarını tutan veri setinde önceki veri setlerinde yaptığımız adımları uygulayarak %84.66 tahmini sonucuna ulaşılmıştır. Hata matrisinde tahmini ve gerçek değerlerin karşılaştırılması sonucunda doğru pozitif (true positive) 47, doğru negatif (true negative) 33, yanlış pozitif (false positive) 3 ve yanlış negatif (false negative) 17 değerleri elde edilmiştir.

Ayrıca her bir veri seti için çıkan sonuçların başarı değerlendirmesini yapabilmek için 'K-Fold Cross Validation' (K-Katlamalı Çapraz Doğrulama) uygulanmıştır. Bunu yapmaktaki amacımız yukarıda anlattığımız ve elde ettiğimiz sonuçları doğrulamaktır. Bunun için veri setlerimizi 5 eşit parçaya bölecek şekilde k değeri 5'e eşitlenmiş ve 5 farklı sonuç elde edilmiştir. Tablo 8, her bir veri seti için, elde edilen 5-Fold Cross Validation sonuçlarının ortalama değerlerini göstermektedir.

***Tablo 8.*** *LR ile 6 veri seti ile elde edilen 5-Fold Cross Validation sonuçlarının ortalama değerleri*

| ***2.102 özellik (0/1) veriseti*** | ***2.102 özellik (sıklık) veriseti*** | ***50 özellik (0/1) veriseti*** | ***50 özellik (sıklık) veriseti*** | ***Emoji özellik (0/1) veriseti*** | ***Emoji özellik (sıklık) veriseti*** |
|---|---|---|---|---|---|
| *89.43* | *85.08* | ***93.93*** | *86.24* | *82.46* | *84.66* |

## D. SUPPORT VECTOR MACHINE İLE YAPILAN DENEYLERİN SONUÇLARI

LR deneyimizde kullandığımız aynı veri setlerini bu deneyimiz için de kullanılmış ve deneydeki aynı adımlar izlenmiştir. Fakat SVM deneyleri uygulanırken, LR deneylerinden farklı olarak doğrusal çekirdek (linear kernel) ile kullanılmıştır. X ekseni 100 Twitter kullanıcısının 2.012 adet özelliği kullanıp kullanmadıklarını; y ekseni ise bu kullanıcıların trol olup olmadıklarını içeren ilk veri seti üzerinden 100 Twitter kullanıcısının (50 adedi trol, 50 adedi gerçek) %80'ini eğitim verisi, %20'ini de test verisi olarak ayırarak SVM çalıştırılmış ve %89.57 tahmini sonucuna ulaşılmıştır. Hata matrisinde tahmini ve gerçek değerlerin karşılaştırılması sonucunda doğru pozitif (true positive) 47, doğru negatif (true negative) 42, yanlış pozitif (false positive) 3 ve yanlış negatif (false negative) 8 değerleri elde edilmiştir.

Kullanıcıların her özelliği kaç kere kullandıklarını tutan ikinci veri seti içinde ise ilk deneyde yaptıklarımızın aynısı uygulanmıştır. Bu ikinci deneyimizde test verilerinin tahmini sonucu %85.60 olarak hesaplanmıştır. Hata matrisinde tahmini ve gerçek değerlerin karşılaştırılması sonucunda doğru pozitif (true positive) 45, doğru negatif (true negative) 39, yanlış pozitif (false positive) 5 ve yanlış negatif (false negative) 1 değerleri bulunmuştur.

İlk veri setindeki özellik sayısını 50'ye indirerek oluşturulan 0/1 kullanıcı-özellikli üçüncü veri setinde önceki deneylerimizde izlediğimiz adımların aynısını uygulayarak %92.42 sonucuna ulaşılmıştır. Hata matrisinde tahmini ve gerçek değerlerin karşılaştırılması sonucunda doğru pozitif (true positive) 47, doğru negatif (true negative) 44, yanlış pozitif (false positive) 3 ve yanlış negatif (false negative) 6 değerleri elde edilmiştir.

Kullanıcıların her özelliği kaç kere kullandıklarını tutan ve 50 adet özellik bulunduran dördüncü veri setinde önceki veri setlerinde yaptığımız adımları uygulayarak %88.50 tahmini sonucuna ulaşmıştır. Hata matrisinde tahmini ve gerçek değerlerin karşılaştırılması sonucunda doğru pozitif (true positive)

48, doğru negatif (true negative) 39, yanlış pozitif (false positive) 2 ve yanlış negatif (false negative) 11 değerleri oluşmuştur.

2.102 adet olan özelliğin içinden sadece emoji özellikleri ile oluşturduğumuz 0/1 kullanıcı-özellikli beşinci veri setimiz ile ilk deneyde yaptıklarımızın aynısını uygulayarak %77.41 tahmini sonucuna ulaşılmıştır. Hata matrisinde tahmini ve gerçek değerlerin karşılaştırılması sonucunda doğru pozitif (true positive) 40, doğru negatif (true negative) 36, yanlış pozitif (false positive) 10 ve yanlış negatif (false negative) 14 değerleri oluşmuştur.

Kullanıcıların her özelliği içindeki emojileri kaç kere kullandıklarını tutan altıncı veri setinde önceki veri setlerinde yaptığımız adımları uygulayarak %76.37 tahmini sonucuna ulaşılmış ve hata matrisinde tahmini ve gerçek değerlerin karşılaştırılması sonucunda doğru pozitif (true positive) 42, doğru negatif (true negative) 31, yanlış pozitif (false positive) 8 ve yanlış negatif (false negative) 19 değerleri elde edilmiştir.

LR ile yapılan deneylerde de yaptığımız gibi, SVM için de her bir veri seti için çıkan sonuçların başarı değerlendirmesini yapabilmek için 'K-Fold Cross Validation' (K-Katlamalı Çapraz Doğrulama) uygulanmıştır. Bunu yapmaktaki amacımız yukarıda anlattığımız ve elde ettiğimiz sonuçları doğrulamaktır. Bunun için veri setlerimizi 5 eşit parçaya bölecek şekilde k değeri 5'e eşitlenmiş ve 5 farklı sonuç elde edilmiştir. Tablo 9, her bir veri seti için, elde edilen 5-Fold Cross Validation sonuçlarının ortalama değerlerini göstermektedir.

***Tablo 9.*** *SVM ile 6 veri seti ile elde edilen 5-Fold Cross Validation sonuçlarının ortalama değerleri*

| ***2.102 özellik (0/1) veriseti*** | ***2.102 özellik (sıklık) veriseti*** | ***50 özellik (0/1) veriseti*** | ***50 özellik (sıklık) veriseti*** | ***Emoji özellik (0/1) veriseti*** | ***Emoji özellik (sıklık) veriseti*** |
|---|---|---|---|---|---|
| *89.57* | *85.60* | ***92.42*** | *88.50* | *77.41* | *76.37* |

## E. RANDOM FOREST REGRESSION İLE YAPILAN DENEYLERİN SONUÇLARI

Rastgele orman regresyonunun üstünde deney yapabilmek için tekrardan SVM ve LR deneylerinde kullanmış olduğumuz 6 veri seti kullanılmıştır. Veri setlerinin x ekseni kullanıcılar tarafından kullanılan ya da kullanılmayan 2.102, 50 ya da emoji özelliklerinin bilgilerini; y ekseni ise bu 100 kullanıcının troll olup olmadıkları bilgisini tutmaktadır. 100 kullanıcının %80'i eğitim verisini, %20'si de test veri setini oluşturmaktadır. 6 adet virgülle ayrılmış veri dosyasının test verilerinin tahmini sonucu, kullanıcıların özellikleri kullanıp kullanmadığını 1 (kullanıyorsa) ve 0 (kullanmıyorsa) ile belirleyen dosya için %86.66; kullanıcıların özellikleri kaç kere tweetlerinde kullandığını tutan dosya için %84.90 sonucu kaydedilmiştir. %86.66 tahmini sonucunu veren birinci veri seti için hata matrisi, doğru pozitif (true positive) 46, doğru negatif (true negative) 40, yanlış pozitif (false positive) 4 ve yanlış negatif (false negative) 10 sonuçlarını vermiştir.

Özelliğin ya da özelliklerin bir kullanıcı tarafından kaç kere kullanıldığını gösteren ikinci veri setinde ise ilk deneyde yaptıklarımızın aynısı uygulanmıştır ve %84.90 tahmini sonucunu elde edilmiştir. Bu %84.90 tahmini sonucunu veren veri seti için ise hata matrisi, doğru pozitif (true positive) 45, doğru negatif (true negative) 39, yanlış pozitif (false positive) 5 ve yanlış negatif (false negative) 1 olarak elde edilmiştir

2.102 adet olan özelliği 50 özelliğe indirerek oluşturduğumuz 0/1 kullanıcı-özellikli üçüncü veri setimiz ile ilk deneyde yaptıklarımızın aynısını uygulayarak %92.69 tahmini sonucuna ulaşılmıştır. Hata matrisinde tahmini ve gerçek değerlerin karşılaştırılması sonucunda doğru pozitif (true positive) 48, doğru negatif (true negative) 43, yanlış pozitif (false positive) 2 ve yanlış negatif (false negative) 7 değerleri oluşmuştur.

50 özelliğin ya da özelliklerin bir kullanıcı tarafından kaç kere kullanıldığını gösteren dördüncü veri setinden önceki deneylerimizde izlediğimiz adımların aynısını uygulayarak %88.23 tahmini sonucu elde edilmiştir. Bu %88.23 tahmini sonucunu veren veri seti için ise hata matrisi, doğru pozitif (true positive) 47, doğru negatif (true negative) 39, yanlış pozitif (false positive) 3 ve yanlış negatif (false negative) 11 sonuçlarına ulaşılmıştır.

2.102 adet olan özelliğin içinden sadece emoji özellikleri ile oluşturduğumuz 0/1 kullanıcı-özellikli beşinci veri setimiz ile ilk deneyde yaptıklarımızın aynısını uygulayarak %76.60 tahmini sonucuna ulaşılmıştır. Hata matrisinde tahmini ve gerçek değerlerin karşılaştırılması sonucunda doğru pozitif (true positive) 42, doğru negatif (true negative) 33, yanlış pozitif (false positive) 8 ve yanlış negatif (false negative) 17 değerleri elde edilmiştir.

Kullanıcıların her özelliği içindeki emojileri kaç kere kullandıklarını tutan son veri setinde önceki veri setlerinde yaptığımız adımları uygulayarak %79.05 tahmini sonucuna ulaşılmıştır. Hata matrisinde tahmini ve gerçek değerlerin karşılaştırılması sonucunda doğru pozitif (true positive) 43, doğru negatif (true negative) 35, yanlış pozitif (false positive) 7 ve yanlış negatif (false negative) 15 değerleri oluştu.

LR ve SVM ile yapılan deneylerde de yaptığımız gibi, RFR için de her bir veri seti için çıkan sonuçların başarı değerlendirmesini yapabilmek için 'K-Fold Cross Validation' (K-Katlamalı Çapraz Doğrulama) uygulanmıştır. Bunu yapmaktaki amacımız yukarıda anlattığımız ve elde ettiğimiz sonuçları doğrulamaktır. Bunun için veri setlerimizi 5 eşit parçaya bölecek şekilde k değeri 5'e eşitlenmiş ve 5 farklı sonuç elde edilmiştir. Tablo 10, her bir veri seti için, elde edilen 5-Fold Cross Validation sonuçlarının ortalama değerlerini göstermektedir.

| ***2.102 özellik (0/1) veriseti*** | ***2.102 özellik (sıklık) veriseti*** | ***50 özellik (0/1) veriseti*** | ***50 özellik (sıklık) veriseti*** | ***Emoji özellik (0/1) veriseti*** | ***Emoji özellik (sıklık) veriseti*** |
|---|---|---|---|---|---|
| *86.66* | *84.90* | ***92.69*** | *88.23* | *76.60* | *79.05* |

***Tablo 10.*** *RFR ile 6 veri seti ile elde edilen 5-Fold Cross Validation sonuçlarının ortalama değerleri*

# IV. SONUÇLAR

Bu çalışma ile birlikte Twitter trol hesaplarının sınıflandırılarak tespit edilebilmesi için makine öğrenmesi algoritmalarının etkin bir şekilde kullanılabileceğini göstermiş olduk. Kendi hazırladığımız 100 Twitter kullanıcısının 238.925 adet tweet mesajı ve diğer kullanıcı bilgilerinden çıkarılan 2.102 adet özellik (feature) üzerinde yapılan deneyler ile LR, SVM ve RFR makine öğrenmesi teknikleri denendi.
LR algoritmasının en çok kullanılan 50 özelliğin 0/1 gösterimli veri seti ile SVM ve RFR'a göre daha iyi sonuçlar elde ettiği (%93.93) ve ikili sınıflandırma işlemini en iyi şekilde gerçekleştirdiği deneysel olarak gösterilmiş oldu. Kullanıcıların tüm özellikleri 0/1 olarak belirten veri seti ile yapılan deneylerde, %89.57 sonucuyla SVM en yüksek yüzdeye sahiptir. LR, SVM ve RFR tekniklerinin en iyi sonuçları 50 özellik (0/1) veriseti ile elde ettikleri görülmektedir. Emojili veri setlerinde LR algoritmasının (sıklık) veriseti ile %84.66'lara varan sonuçlar elde ettiği gözlemlenmiştir.

Gelecekte yapacağımız çalışmalarda, Extreme Learning Machines (ELM) ve Deep Learning (DL) algoritmaları ile sonuçları daha iyileştirmeyi hedeflemekteyiz. GPU (Graphics Processing Unit) desteği ile yapılacak deneyler ile hesaplamaların daha kısa sürelerde ve daha büyük veri üzerinde çalıştırılması düşünülmektedir. Özellik seçimi (feature selection) algoritmaları ile trol sınıflandırma sürecinin daha da iyileştirebileceğini tahmin etmekteyiz. Haddop spark gibi büyük veri araçları kullanarak çalışmalar yapılması düşünülmektedir.

# V. KAYNAKLAR


[1] K. K. Cole, "It's like she's eager to be verbally abused": Twitter, trolls, and (en) gendering disciplinary rhetoric. *Feminist Media Studies,* vol. 15(2), pp. 356-358, 2015,

[2] https://www.bbc.com/turkce/haberler-turkiye-53259275 (Erişim zamanı; Haziran, 20, 2020).

[3] P. Galán-García *et al.*, Supervised machine learning for the detection of troll profiles in twitter social network: Application to a real case of cyberbullying. *Logic Journal of the IGPL*, vol. 24(1), pp. 42-53, 2016.

[4] Y. M. Kızılkaya, "Duygu analizi ve sosyal medya alanında uygulama," Doktora Tezi, Sosyal Bilimler Enstitüsü / Ekonometri Anabilim Dalı / İstatistik Bilim Dalı, Uludağ Üniv.,Türkiye, 2018.

[5] B. Sriram, D. Fuhry, E. Demir, H. Ferhatosmanoglu, and M. Demirbas, "Short text classification in twitter to improve information filtering," in *Proc. 33rd Int. ACM SIGIR Conf. on Research and Development in İnformation Retrieval,* 2010, pp. 841-842,

[6] G. Xiang, B. Fan, L. Wang, J. Hong, and C. Rose, "Detecting offensive tweets via topical feature discovery over a large scale twitter corpus," in *Proc. 21st ACM Int. Conf. on Information and Knowledge Management,* 2012, pp. 1980-1984,

[7] P. Fornacciari et al., "A holistic system for troll detection on Twitter," *Computers in Human Behavior,* vol. 89, pp. 258-268, 2018,

[8] P. Tsantarliotis, E. Pitoura, and P. Tsaparas, "Troll Vulnerability in Online Social Networks," in *Proc. IEEE/ACM Int. Conf. on Advances in Social Networks Analysis and Mining (ASONAM),* Aug. 2016, pp. 1394-1396.

[9] M. Engelin, and F. De Silva, "Troll detection: a comparative study in detecting troll farms on Twitter using cluster analysis," KTH, Stockholm, Sweden, May 2016.

[10] T. Mihaylov, T., and P. Nakov, "Hunting for troll comments in news community forums," in *Proc. 54th Annual Meeting of the Association for Computational Linguistics,* Nov. 2019, pp. 399-405.

[11] L. Hong, O.Dan, and B. D. Davison, "Predicting Popular Messages in Twitter," in *Proc. 20th Int. Conf. Companion on World Wide Web,* Mar. 2011, pp. 57–58.

[12] E. Bulut, and E. Yörük, "Mediatized populisms Digital populism: Trolls and political polarization of Twitter in Turkey," *International Journal of Communication,* vol. 11, pp. 4093–4117, 2017.

[13] D. Özsoy, "Tweeting political fear: Trolls in Turkey," *Journal of History School (JOHS),* vol. 12, pp. 535-552, Jun. 2015.

[14] N. Chavoshi, H. Hamooni, and A. Mueen, "Identifying correlated bots in twitter," in *Proc. Int. Conf. on Social Informatics,* Nov. 2016, pp. 14-21.

[15] J. Im, *et al.*, "Still out there: Modeling and identifying Russian troll accounts on Twitter," Jan. 2019, arXiv:1901.11162.

[16] A. Badawy, E. Ferrara, and K. Lerman, "Analyzing the digital traces of political manipulation: The 2016 russian interference twitter campaign," in *Proc. IEEE/ACM Int. Conf. on Advances in Social Networks Analysis and Mining (ASONAM),* Aug. 2011, pp. 258-265,

[17] J. Cheng *et al.*, "Anyone can become a troll: Causes of trolling behavior in online discussions," in *Proc. ACM Conf. on Computer Supported Cooperative Work and Social Computing,* Feb.2017, pp. 1217-1230.

[18] C. Seah, et al., "Troll detection by domain-adapting sentiment analysis," in *Proc. 18th Int. Conf. on Information Fusion (Fusion)* , Jul. 2015, pp. 792-799.

[19] K. Simsek. "Makine Öğrenmesi Dersleri 4a: Lojistik Regresyon." https://medium.com/ data-science-tr/makine-öğrenmesi-dersleri-4-lojistik-regresyon-304fefab0a49 (Erişim zamanı; Haziran, 20, 2020).