
GAZİ ÜNİVERSİTESİ
Gazi
1926

SOSYAL AĞLARA YÖNELİK ÖĞRENMEYE DAYALI BİR SPAM HESAP TESPİT MODELİ VE UYGULAMASI

Oğuzhan ÇITLAK

YÜKSEK LİSANS TEZİ
BİLGİSAYAR MÜHENDİSLİĞİ ANA BİLİM DALI

GAZİ ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

EKİM 2018

Oğuzhan ÇITLAK tarafından hazırlanan "SOSYAL AĞLARA YÖNELİK ÖĞRENMEYE DAYALI BİR SPAM HESAP TESPİT MODELİ VE UYGULAMASI" adlı tez çalışması aşağıdaki jüri tarafından OY BİRLİĞİ ile Gazi Üniversitesi Bilgisayar Mühendisliği Ana Bilim Dalında YÜKSEK LİSANS TEZİ olarak kabul edilmiştir.

Danışman: Dr. Öğr. Üyesi İbrahim Alper DOĞRU

Bilgisayar Mühendisliği Ana Bilim Dalı, Gazi Üniversitesi

Bu tezin, kapsam ve kalite olarak Yüksek Lisans Tezi olduğunu onaylıyorum. ……

Başkan: Doç. Dr. Ali Hakan IŞIK

Bilgisayar Mühendisliği Ana Bilim Dalı, Burdur Mehmet Akif Ersoy Üniversitesi ……

Bu tezin, kapsam ve kalite olarak Yüksek Lisans Tezi olduğunu onaylıyorum.

Üye: Doç. Dr. Nursal ARICI

Bilgisayar Mühendisliği Ana Bilim Dalı, Gazi Üniversitesi

Bu tezin, kapsam ve kalite olarak Yüksek Lisans Tezi olduğunu onaylıyorum. ……

Tez Savunma Tarihi: 12/10/2018

Jüri tarafından kabul edilen bu tezin Yüksek Lisans Tezi olması için gerekli şartları yerine getirdiğini onaylıyorum.

…….

Prof. Dr. Sena YAŞYERLİ

Fen Bilimleri Enstitüsü Müdürü

ETİK BEYAN

Gazi Üniversitesi Fen Bilimleri Enstitüsü Tez Yazım Kurallarına uygun olarak hazırladığım bu tez çalışmasında;

- Tez içinde sunduğum verileri, bilgileri ve dokümanları akademik ve etik kurallar çerçevesinde elde ettiğimi,
- Tüm bilgi, belge, değerlendirme ve sonuçları bilimsel etik ve ahlak kurallarına uygun olarak sunduğumu,
- Tez çalışmasında yararlandığım eserlerin tümüne uygun atıfta bulunarak kaynak gösterdiğimi,
- Kullanılan verilerde herhangi bir değişiklik yapmadığımı,
- Bu tezde sunduğum çalışmanın özgün olduğunu,

Bildirir, aksi bir durumda aleyhime doğabilecek tüm hak kayıplarını kabullendiğimi beyan ederim.

Oğuzhan ÇITLAK

12/10/2018

SOSYAL AĞLARA YÖNELİK ÖĞRENMEYE DAYALI
BİR SPAM HESAP TESPİT MODELİ VE UYGULAMASI
(Yüksek Lisans Tezi)

Oğuzhan ÇITLAK

GAZİ ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

Ekim 2018

ÖZET

Sosyal ağlar günümüzde çok yaygın olarak kullanılmaktadır. Bu yüzden kötü amaçlı kişilerin hedefi haline gelmesi kaçınılmaz bir hal almaktadır. Spam içerik ve mesajlar bu ağları kullananların güvenliğini ve performansını tehdit etmektedir. Her ne kadar sosyal ağların kendi spam hesap politikaları olsa da, bu politikalar çoğu zaman yetersiz kalmaktadır. Bu çalışmada, sosyal ağlarda spam hesapların tespiti için üç bileşenden oluşan öğrenmeye dayalı yeni bir model önerilmektedir. Bu bileşenler link analizi, makine öğrenmesi ve metin analizi olarak yer almaktadır. Link analizinde, sosyal ağ kullanıcısının attığı iletilerdeki linkler incelenmektedir. Kötücül link paylaşan sosyal ağ kullanıcıları tespit edilip spam hesap olarak belirlenmektedir. Link paylaşımı yapmayan ya da masum link paylaşan hesaplardaki spam hesapları bulmak için ise makine öğrenmesi yöntemine başvurulmaktadır. Bu yöntem için sosyal ağ üzerindeki paylaşımlardan bir veri kümesi oluşturulmuştur. Bu veri kümesi kullanılarak spam hesapların öznitelikleri tespit edilmiştir. Öznitelikler vasıtasıyla makine öğrenmesi bileşeni modellenmiştir. Metin analizi yönteminde ise sosyal ağ kullanıcılarının iletilerindeki metinler incelenmektedir. Hassas içerikli kelimelerin bu metinler içerisinde bulunup bulunmadıkları incelenmektedir. Önerilen modelin, ayrıca web tabanlı bir uygulaması gerçekleştirilmiştir. Geliştirilen uygulama vasıtasıyla yapılan deneysel çalışmalar neticesinde, önerilen modelin başarımı %96,23 olarak tespit edilmiştir. Bu çalışmada, sosyal ağ üzerindeki spam hesapları tespit etmek ve sosyal ağ spam tespit politikasına destekte bulunmak amaçlanmaktadır.

Bilim Kodu	:	92431
Anahtar Sözcükler	:	Sosyal ağlar, Twitter, Spam tespiti, Link analizi, Makine öğrenmesi, Metin analizi
Sayfa Adedi	:	97
Danışman	:	Dr. Öğr. Üyesi İbrahim Alper DOĞRU
İkinci Danışman	:	Öğr. Gör. Dr. Murat DÖRTERLER

A SPAM ACCOUNT DETECTION MODEL BASED ON LEARNING AND ITS APPLICATION FOR SOCIAL NETWORKS

(M. Sc. Thesis)

Oğuzhan ÇITLAK

GAZİ UNIVERSITY

GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES

October 2018

ABSTRACT

Social networks are widely used today. Therefore, they inevitably become the target of a malicious person. Spam contents and messages threaten the safety and performance of those networks used. Although social networks have their own spam account policies, these policies are often insufficient. In this study, a new model based on learning is proposed for the detection of spam accounts in social networks consisting of three components. These components include link analysis, machine learning and text analysis. In the link analysis, the links in the messages of the social network user are analyzed. Social network users who share a malicious link are determined and identified as a spam account. The machine learning method is used to find spam accounts sharing innocent link or not any link. A dataset has been created from the shares on the social network for this method. The features of the spam accounts are determined by using this dataset. The compound of machine learning is designed by means of these features. In the text analysis method, the texts in the messages of the social network users are analyzed. It is viewed whether sensitive contents are in these texts of messages or not. A web based application is indicated for the proposed model. As a result of the experimental studies carried out by the developed application, the performance of the proposed model is determinated as 96.23%. In this study, it is aimed to detect spam accounts on social network and the spam detection policy of social network is intended to support.

Science Code : 92431
Key Words : Social networks, Twitter, Spam detection, Link analysis, Machine learning, Text analysis
Page Number : 97
Supervisor : Assist Prof. Dr. İbrahim Alper DOĞRU
Co-Supervisor : Lec. Dr. Murat DÖRTERLER

TEŞEKKÜR

Çalışma süresince tez danışmanlığımı üstlenerek bana yol gösteren, çalışmanın planlanması, yürütülmesi, raporlanması ve sonuçlanmasında katkı ve destek sunan değerli sayın hocalarım; Dr. Öğr. Üyesi İbrahim Alper Doğru ve Öğr. Gör. Dr. Murat Dörterler'e teşekkür ederim. Desteklerinden dolayı halen üyesi olduğum Tuprag Metal Madencilik AŞ ailesine, yaşamımın her döneminde ve her türlü kararımda yanımda olan, maddi ve manevi desteklerini her zaman yanımda hissettiğim babam Veli Çıtlak'a, annem Fatma Çıtlak'a teşekkür ederim. Ayrıca, tez yazma sürecimde bana zorluk çıkarmayan küçük kızım Ber Sena Çıtlak'a, Araştırma sürecim boyunca her türlü desteğini esirgemeyen eşim Kübra Çıtlak'a en içten dileklerimle teşekkür ederim.

İÇİNDEKİLER

Sayfa

Sayfa

Sayfa

ÇİZELGELERİN LİSTESİ

Çizelge **Sayfa**

ŞEKİLLERİN LİSTESİ

Şekil **Sayfa**

SİMGELER VE KISALTMALAR

Bu çalışmada kullanılmış kısaltmalar, açıklamaları ile birlikte aşağıda sunulmuştur.

Kısaltmalar	Açıklamalar
ADMM	Alternating Direction
API	Uygulama Programlama Arayüzü
ARFF	Attribute File Format
BOT	Robot
CMD	Command Promt
DGA	Web Tabanlı Genetik Algortima
EIES	Electronic Information Exchange System
FOS	Fuzzy Based Information Decomposition Oversampling
JSON	Javascript Object Notation
KNN	K-Nearest Neighbour
MCC	Matthews Correlation Coefficient
NODEXL	Network Overview, Discovery of Exploration for Excel
PRC	Precision Recall
PROFIL	Account
ROC	Receiver Operating Characteristic
SMS	Short Message
SS	Social Status
SVM	Support Vektor Machine
TSP	Triple Significance Profile
TÜİK	Türkiye İstatistik Kurumu
URL	Uniform Resource Locator
XML	Extensible Markup Language

1. GİRİŞ

Günümüzde zaman ve mekândan bağımsız olarak internet yaygın bir şekilde kullanılabilmektedir. İnternet kullanıcıları için, sosyal ağlar kişilerin vakitlerinin büyük bir kısmını geçirdikleri mecralardır. Bireylerin kendilerini rahatlıkla ifade edebildikleri ve bilgi paylaşımında bulundukları bu sosyal platformlar günlük yaşamında vaz geçilmez bir parçası haline gelmektedir [1]. Son yıllarda internet kullanımında özellikle de Facebook, Twitter, Instagram, Google+ ve LinkedIn gibi servisler insanlar arasındaki sosyal iletişim ve haberleşme alanında önemli bir yere sahiptirler. Bu sosyal servisler kullanılarak internet üzerinden birbirleriyle anlık etkileşimde bulunulmaktadır. Twitter, sosyal servisler içerisinde en çok kullanılan platformlardan bir tanesidir. Bu kadar yaygın olan bir sosyal ağın kötü amaçlı kişilerin hedefi haline gelmesi kaçınılmazdır.

Twitter'daki spam içerik ve mesajlar bu ağı kullananların güvenliğini ve performansını tehdit etmektedir. Bu çalışmada, Twitter'da spam hesapların tespiti için üç bileşenden oluşan öğrenmeye dayalı yeni bir model önerilmektedir. Bu bileşenler link analizi, makine öğrenmesi ve metin analizidir.

Link analizinde, Twitter kullanıcısının attığı iletilerdeki linkler incelenmektedir. Bu linklerin, kötücül web sitelerinin sürekli güncellendiği bir veri havuzunda bulunup bulunmadığına bakılmaktadır. Analiz edilen hesap eğer kötücül link paylaşıyor ise Twitter spam politikasına [2] göre bu hesap %100 spamdır denilebilmektedir. Bu dinamik havuzda bulunmayan kötücül linkleri paylaşan hesapları belirlemek ya da link paylaşımı yapmayan spam hesapları bulmak için ise makine öğrenmesi yöntemine başvurulmaktadır.

Twitter'dan iletilerininin tamamında en az bir hassas içerikli kelime bulunan 250 bine yakın kullanıcının verisi makine öğrenmesi yönteminde kullanılmak üzere çekilmiştir. Twitter'ın bu çekilen veri kümesindeki askıya aldığı hesapların özellikleri incelenmiştir. Veri kümesindeki spam ve spam olmayan hesapların sergilediği özellikler sayesinde makine öğrenmesi bileşenlerinin öznitelikleri oluşturulmuştur. Topluluktan faydalanma yöntemi, makine öğrenmesi özniteliklerinde kullanılan Twitter hesaplarının analizinde etiketleme için kullanılmıştır [3]. Makine öğrenmesi yönteminin başarımında hesaplanmasında farklı algoritmaların sonuçları değerlendirilmiştir [4].

Metin analizi yöntemimizde ise Twitter kullanıcılarının iletilerindeki metinler incelenmiştir. Önceden belirlediğimiz hassas içerikli kelimelerin bu hesaplardan iletilip iletilmediklerine bakılmaktadır. Metin analizi sonucunda analiz edilen hesaplardaki olası hassas içerikler makine öğrenmesinde aldığımız başarılı sonucu destekleyici nitelikte olmaktadır. Yapılan değerlendirmeler sonucunda önerilen modelinin başarımının %96,23 olduğu tespit edilmiştir.

Bu çalışmada, sosyal ağlara yönelik öğrenmeye dayalı bir spam hesap tespit modelini geliştirilmeye çalışılmaktadır. Önerilen modelin sahip olduğu üç bileşen link analizi, makine öğrenmesi ve metin analizi yöntemleridir. İkinci bölümde, literatürde yapılan çalışmalar incelenmektedir ve en çok kullanılan spam tespit yöntemleri hakkında bilgi verilmektedir. Üçüncü bölümde ise ayrıntılı bir şekilde kullanılan yöntem ve araçlardan bahsedilmektedir. Dördüncü bölümde önerilen modelin uygulaması çalıştırılmaktadır ve kazanılan başarımlar gösterilmektedir. Beşinci bölümde ise elde edilen bulgular değerlendirilmektedir ve modelin başarısı hesaplanmaktadır. Son bölümde ise elde edilen sonuçlar değerlendirilmektedir.

2. LİTERATÜRDE YER ALAN ÖNCEKİ ÇALIŞMALAR

Sosyal ağ servisleri, kullanıcılarına ait verilerin güvenliği sağlamak ve kullanıcılarının karşılaşabilecekleri tehlikeleri göz önünde bulundurmak zorundadırlar. Ancak, son yıllarda hızla artan spam içerik kullanıcı güvenliğini ciddi şekilde tehdit etmektedir. İnternette bol miktarda yer alan ve sosyal medya alanlarını da bolca kullanan bu tehditlerine karşı alınabilecek en iyi önlem, spamların hangi yollarla kullanıcıları tehdit ettiklerini bilmek ve buna karşı kişisel tedbirleri almaktan geçmektedir [5]. Facebook, Twitter, Myspace, Linkedin, Google+ ve Instagram internette farklı amaçlar için kullanılan en yaygın sosyal ağlar olarak sayılabilmektedir. Ülkemizde gün geçtikçe sosyal ağların kullanımı artmaktadır.

Sosyal ağ üzerinde etkileşime geçen kullanıcılar, sanal bir ortamda yeni bir kimlik oluşturarak, birbirleriyle haberleşmektelerdir, paylaşımda bulunmaktalardır ve sosyal ilişkilerini arkadaş çevrelerini oluşturup geliştirmektelerdir. Dünyanın her yerinde, her yaştan insanın ve özellikle genç kullanıcıların zamanlarının büyük bölümünü geçirdikleri sosyal ağlar çok sıkça ziyaret edilmektedir [5, 6].

Türkiye İstatistik Kurumu'nun (TÜİK) 2018 Ocak-Mart ayları arası için yaptığı, "Hane Halkı Bilişim Teknolojileri Kullanım Araştırması" çalışmasına göre internete erişim sağlayabilen 16-74 yaş grubu kullanıcıların %82,4'ü sosyal medya üzerinde mesaj göndermişler veya fotoğraflı içerik paylaşmışlardır [7].

Sosyal ağların atası kabul edilen ve ilk olarak 1978 yılında New Jersey Teknoloji Enstitüsünde, Freeman ve ekibi tarafından geliştirilen Elektronik Enformasyon Değiş Tokuş Sistemi'nde (Electronic Information Exchange System, EIES) üyeler, kendi aralarında e-posta gönderebilmekte, duyuru panoları oluşturabilmekte ve kendilerine ait görev listeleri hazırlayabilmektelerdi. Günümüzde ki sosyal ağlara benzeyen ilk ağ ise, 1997 yılında hizmete giren Sixdegrees.com'dur [8]. 'Sixdegrees' ismi 1960 yılında sosyolog Stanley Milgram'ın, insanlar arasındaki iletişim yollarını belirlemek için yaptığı bir çalışmadan gelmektedir. Bu oluşturulan site, kullanıcılarına profil oluşturma ve arkadaşlarını listeleme imkanı vermektedir. Profil oluşturma gibi bazı özellikler, diğer arkadaşlık sitelerinde ve sanal gruplarda da olmasına rağmen, Sixdegrees.com bunları birleştiren ilk site olma

özelliğine sahiptir. Bu kabul edilen ilk sosyal ağ servisi, milyonlarca kullanıcıya hizmet vermesine rağmen 2000 yılında kapatılmıştır [8].

Bireylerin sosyal medyayı kullanım amacı kişiden kişiye değişmektedir. Sosyal medya araçlarından her bir bireyin beklentileri farklı olmakta, buda değişik kullanımlara sebep olabilmektedir. Sosyal medya kimi kullanıcılar için sosyalleşmeden kaçtığı, yalnız kalabildiği, daha çok izleyici olduğu bir ortam iken, diğer kullanıcılar için ise sosyalleşebildiği, topluluklar içinde takdir edilip, takip edilebildiği, kendini rahat ifade edebildiği alanlar ortaya çıkabilmektedir. Bu açıdan bakıldığında sosyal medya teknoloji üzerine kurulan, çok derin bir sosyal etkileşim, grup oluşumu ve işbirliğini arttırmayı sağlayan bir yapı olarak tanımlanmaktadır [9].

Çizelge 2.1.'de Ocak 2017 den Ocak 2018 e kadar olan internet kullanıcı sayısı ve bunların sosyal medya kullanım sayısındaki artış ile mobil cihaz kullanıcı sayısı ve bunlardan sosyal medya kullanım oranındaki artış görülmektedir [10].

Çizelge 2.1. Aktif sosyal medya kullanıcılarının artış oranı [10]

İnternet Kullanıcı Sayısı (milyon)	Aktif Sosyal Media Kullanıcı Sayısı (milyon)	Mobil Cihaz Kullanıcı Sayısı (milyon)	Aktif Mobil Sosyal Media Kullanıcı Sayısı (milyon)
3,419	2,307	3,790	1,968
Bir önceki yıla göre artış		Bir önceki yıla göre artış	
332	219	141	283
%9,71	%9,49	%3,72	%9,29

Çizelge 2.1 incelendiği zaman internet kullanımı arttıkça, sosyal ağların kullanımı da artmaktadır. Sosyal ağ servisleri kullanarak yapılan internet üzerindeki iletişimde olması gereken bazı güvenlik gereksinimleri vardır. Gizlilik, bilginin 3. şahıslardan gizlenmesini demektir. Esas itibariyle gizlilik, kişisel ve hassas bilgilere doğru kişilerin erişimini sağlarken, yanlış kişilerin erişimlerini engellemektir. Bütünlük, sahip olunan veya paylaşılan bilginin üçüncü şahıslar tarafından değiştirilmemesi ve bütünlüğünün bozulmaması demektir [11]. Uygunluk, bilgiye erişimde istenilen ve güvenilir kişilere erişim garantisinin sağlanmasıdır. Kullanıcı haberleşmesi gizliliği, sosyal ağ servislerinin de sağladığı bazı güvenlik tedbirleri de olmak zorundadır.

Sosyal ağ kullanıcının bilgilerine, ağ operatörlerince erişimin sağlanmıyor olması bunlardan birisidir. Kullanıcının gizlilik gereksinimlerine göre, IP adresine, mesajlarına, gizli kalmasını istediği görüntü bilgilerine ağ operatörleri tarafından da erişilememelidir. Sosyal ağların bu kadar hızlı bir şekilde büyümesi, kötü amaçlı yazılımların sayısında ve yayılmasında, çok hızlı bir artışa neden olmuştur [12]. Yoğun kullanım alanına sahip bir sanal ortamda kullanıcıya düşen bazı sorumluluklar ve dikkat edilmesi gereken durumlar söz konusudur.

Tüm internet kullanıcılarının dikkat etmesi gereken bazı güvenlik etkenleri sosyal ağlar içinde geçerlidir. Ayrıca sosyal ağlarda bir web sitesinden farklı olarak, anlık etkileşimin en yüksek seviyede olmasından kaynaklanan bazı güvenlik açıklarını da beraberinde getirmektedir. Örneğin, kullanıcıların gizlilik ilkelerine dikkat etmemesi, hesaplarının kontrolünü ve yönetimini tam olarak yapamamaları, en önemlisi de kişisel bilgilerini bu ortamlarda rahatlıkla paylaşarak kendi kendilerini hedef haline getirmeleridir [13].

Sosyal medyanın etkin bir şekilde bireylerin eğitiminde kullanılmaya başlanması ve bu ağları kullanarak attığı mesajlarla birlikte yapılan paylaşımlar bunun bir göstergesidir [14]. Sosyal medyada yapılan paylaşımların yansıttığı duyguların gerçek dünyada ki sonuçlarını tahmin etmek için nasıl kullanılabileceği ve hatta buradaki duygu, düşünce ve davranışlarının bireyin gerçek dünyada ki itibarının değerini belirlemek gibi farklı alanlarda kullanılması bunun en büyük göstergesidir [13].

eBizMBA sitesinin 1 Nisan 2018 tarihli verilerine göre en sık kullanılan sosyal ağ siteleri ve aylık ziyaretçi sayıları incelenmektedir. Bu araştırmaya göre Facebook'un 1 500 milyon aylık tekil kullanıcı sayısı ile birinci sırada, YouTube 1 499 milyon ile ikinci, Twitter'ın 400 milyon ile üçüncü, Instagram 275 milyon ile dördüncü, Linkedln ise 250 milyon tekil kullanıcı sayısı ile beşinci sırada yer aldığı görülmektedir [15].

Sosyal medya kullanımının ne kadar yüksek olduğunu gösteren bir çalışmaya göre aşağıdaki çizelgede belirtilen sonuçlar çıkartılmıştır [5]. Çizelge 2.2'de kullanıcısının aylık ortalama kaç gününü ve bir günde kaç saatini bir sosyal medya platformunda geçirdiği görülmektedir.

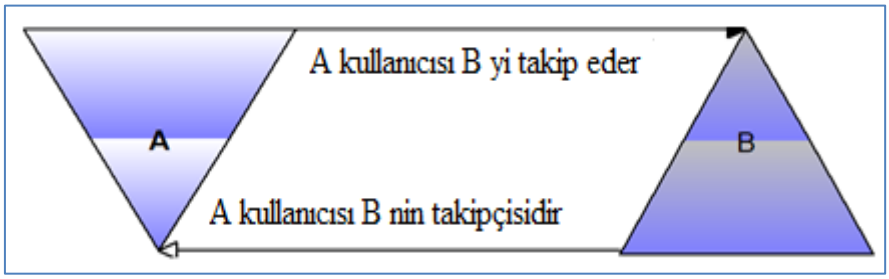
Çizelge 2.2. Sosyal medyada aylık ortalama harcanan zaman [15]

Sosyal Ağ	Aylık harcanan ortalama zaman (gün)	Günlük harcanan ortalama zaman (saat)	Ziyaretçi başına günlük görüntülenen sayfa sayısı
FaceBook	15	14,41	5,48
Instagram	11	5,44	3,80
Twitter	7,5	3,86	3,86
Google+	3,2	8,55	8,82

Çizelge 2.2 incelendiği zaman sosyal ağlar içerisinde en çok vakti Facebook platformu almaktadır. Instagram ikinci sırada gelmekte ve Twitter ise üçüncü sırada gelmektedir. Cisco'nun 2016 yılı için hazırladığı "yıllık güvenlik raporuna" göre zararlı yazılımların en çok karşılaşıldığı, en çok güvenlik tehdidinin alındığı yer olan sitelerin başında, en yüksek sayıda kullanıcıya sahip olan Facebook gelmektedir [16].

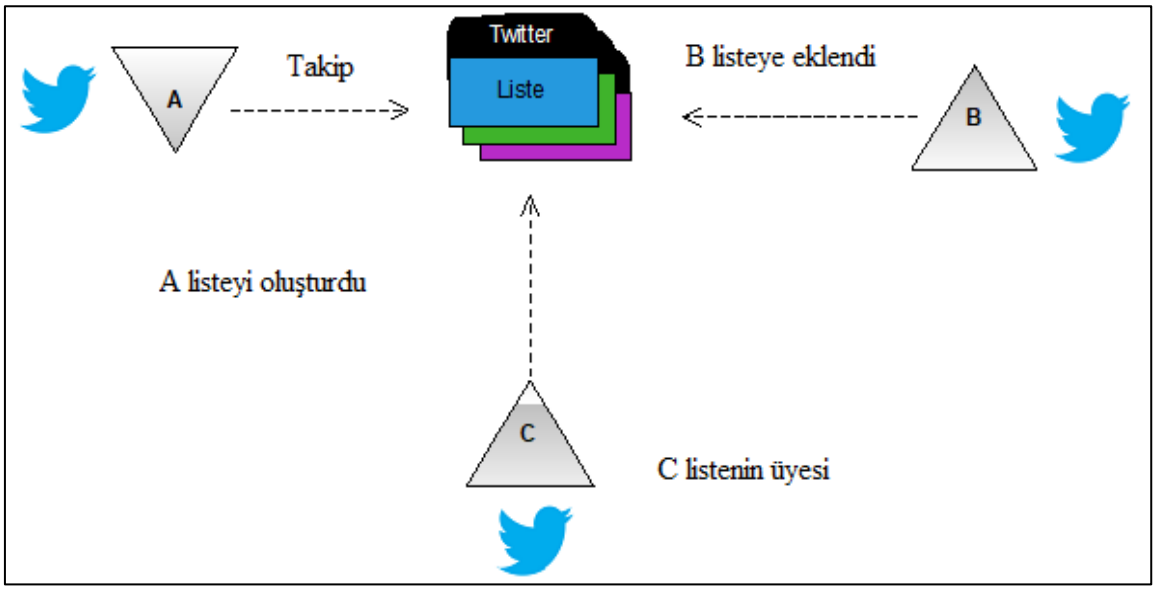
2.1. Twitter'in Mimari Yapısı

Twitter sosyal platformu, kullanıcılarının ilgilendikleri diğer kullanıcı hesaplarını takip edebilmesini sağlamaktadır. Diğer sosyal medya platformlarından farklı olarak, kullanıcılarının arasında çift yönlü ilişki bulunmaktadır. Bu çift yönlü ilişkiden kasıt, bir kullanıcı takipçilerinden birini takip edebilir ya da takip etmeyebilir. Bu durum, tek yönlü bağlantı kurulması yerine çift yönlü ilişki olarak düşünülebilir. Kullanıcı eğer bir iletiyi, beğen (like) veya ileti (ReTweet-RT) işlemi yaparsa bu iletiyi takipçileri (followers) ile paylaşması anlamına gelmektedir. Twitter'daki kullanıcılar arasındaki ilişki Şekil 2.1'de gösterilmektedir. A kullanıcısı B kullanıcısını takip etmektedir ya da A kullanıcısı B kullanıcısının bir takipçisi durumundadır.



Şekil 2.1. Twitter'daki kullanıcılar arasındaki ilişki [17]

Şekilde 2.1'de A kullanıcısı B'yi takip etmektedir. A kullanıcısı aynı zamanda B kullanıcısının takipçisidir. Burada ikili bir durum söz konusudur. Her kullanıcının benzersiz kendisine ait tek bir Twitter kullanıcı adı vardır. Kullanıcılar, Twitter'da bahset(mention) olarak adlandırılan @ karakterini kullanarak ve sonuna kullanıcı adlarını ekleyerek iletileri başka kullanıcılara yönlendirebilir. Kullanıcılar, Twitter'daki iletilerin birisinde bahsedildiklerinde veya ReTweet oldukları zaman bildirimlerle anında bilgi almaktalardır.



Şekil 2.2. Twitter'daki listeler ve kullanıcılar arasındaki ilişkiler [17]

Şekil 2.2'de gösterilen A kullanıcısı, listenin üyesi(member) olarak sınıflandırılmaktadır. Twitter'ın bir diğer özelliği, kullanıcıların ilgileri aynı veya benzer olan kullanıcıları gruplandırarak ilgi alanlarını düzenlemek için kullanıcılara özel listeler oluşturmasına izin vermesidir [18, 19]. Benzer şekilde, bir kullanıcının oluşturduğu listeyi başka oluşturduğu listelere ekleyerek hepsini bir arada yönetebilmesi mümkündür. Abone olunan listeler abone olunan (member of) olarak sınıflandırılırken, kullanıcı tarafından sahip olunan olarak listelenir.

2.2. Literatürde Sıklıkla Geçen Spam Tespit Yöntemleri

Spamlar sosyal medya ağlarında kullanıcıların en çok dikkat etmesi gereken tehditlerdir ve spam hesapları tespit etmek için birçok araştırma yapılmıştır. Bu tehditlerin nasıl fark edileceği konusunda yapılan akademik çalışmalar kategorize edilerek bu başlık altında sunulmuştur. Bu yöntemler sınıflandırılırken en güncel ve en çok kullanılan yöntemler

dikkate alınmıştır. Twitter'da en sık kullanılan spam tespit yöntemlerinden aşağıda bahsedilmektedir.

2.2.1. Anomali tespit yöntemi

Anomali, normalden farklı tutum ve davranış sergilemek anlamına gelmektedir [20]. Bir davranış normal olandan farklılık göstermesi durumunda tehlikeli davranış olarak kabul edilmektedir [21]. Twitter'da normal bir kullanıcının hareketleri, istatiksel olarak tahmin edilebilen matematiksel değerlere benzemektedir. Burada önemli olan normal hareketin bilinmesi ve anormal olan hareketin normal olandan ayırt edilebilmesidir. Normal davranış sonucunda elde ettiğimiz örüntüler esas alınarak davranışlar gözlemlenmektedir ve bir anormallik varsa normal davranışla kıyaslanarak bu hareket tespit edilip ayrıştırılmaktadır [22]. Anormallik tespit yönteminde kullanıcıların temel özellikleri ele alınarak normal veya anormal davranışları kontrol edilmektedir. Sosyal ağlarda sadece farklı kullanıcılar arasındaki arkadaşlık bağlarının tespitinde kullanılmaktadır. Kullanıcının katıldığı bazı grupları ve ağları test ederek sonraki davranışlarını tespit etmeye çalışmaktadır.

Anormallik tespit yönteminin avantajı daha önceden bilinmeyen spam hareketlerinin keşfedilmesi olasılığıdır. Anormallik tespit yönteminde karşılaşılan en büyük sorun spam tespitinde ki yanlış alarmın doğru alarma bölünme sayısının yüksek olmasıdır. Twitter'da spam üreten hesaplar sosyal ağ üzerinde, arkadaş takip gruplarında ya da oluşturulan popüler başlıkta anormal bir davranış göstermektelerdir [21].

Rashidi ve Fung yaptıkları bir çalışmada, mobil Android kullanıcılarına yönelik tehdit içeren spamlar için etkili bir anormallik tespit sistemi geliştirmiştir [23]. Bu sistemde, makine öğrenmesi yöntemleri kullanılarak Android mobil işletim sisteminin çekirdek seviyesi ile kullanıcı seviyesi izlenerek spamların normal davranışlardan ayırt edilmesi sağlanmıştır. Önerdikleri bu yöntemin başarısını, elde ettikleri sonuçlar doğrultusunda ortaya koymuşlardır. Anormallik spam tespit yönteminde Twitter kullanıcısının sergilediği davranış teknikleri temel alınmaktadır. Kullanıcının gönderdiği mesajların içerik ve sayısı, aldığı beğeni sayısı, mesaja yapılan yorumlar ve burada geçirilen sürenin değerlendirilmesi söz konusudur [23].

Sosyal ağlar, katılımcılar arasında çeşitli ilişkiler yakalayabilir. Örneğin aile üyelerinin oluşturduğu bir ağ ve gösterdikleri davranışlar önemli birer özelliktir. Bu nedenle

kullanıcıların, sosyal ağda dostluk ve aile bağlarının kendi içinde ki bağlantı ve ilişkileri birbirleriyle örtüştürülerek bir davranış öngörü oluşturmak her yönden yararlı bir çalışma olacaktır [24]. Bütün bunları dikkate alan bir çalışmada, Sosyal ağlarda ki arkadaşlık ve aile bağlarının arasındaki ilişkilerin, geleneksel özelliklere göre %15-%30 arasında daha yüksek bir oranda doğru öngörü çıkarılmasını sağladığı görülmektedir [24].

Egele ve diğ.'nin 2011 yılında yaptıkları bir çalışmada, Compa adında Anomaly spam tespit aracı geliştirmişlerdir. Compa aracı oluşturulurken 1,4 milyondan fazla herkese açık Twitter iletisi ve 106 milyondan fazla Facebook iletisinin olduğu bir veri kümesi kullanılmıştır [25]. Bu aracı oluşturmalarındaki amaç; Anormal davranış tespiti ve istatiksel bir model oluşturmak, ani davranış değişikliği olan sosyal ağ hesaplarını tespit edebilmektir. Twitter ve Facebook'taki veri kümeleri Weka yazılımının [26, 27] Sıralı Minimal Optimizasyon (Sequential Minimal Optimization-SMO)'u ile etiketlenmiştir. Çizelge 2.3'de Weka içerisinde etiketlenen Compa aracı kullanılarak, rasgele seçilen bireysel kullanıcıların davranış profilleri oluşturulmuştur.

Çizelge 2.3. Twitter ileti ve link değerlendirme sonuçları [25]

Ağ ve Benzer Özellik	Twitter İletisi		Twitter Paylaşılan Link	
	Gruplar	Hesaplar	Gruplar	Hesaplar
#Tehlike tespit edilen grup ve hesap sayısı	9 362	343 229	1 236	54 907
# Tehlike tespit edilen popüler grup uygulama sayısı	1 647	178 557	251	8 254
# Tehlike tespit edilen istek uygulama sayısı	7 715	164 672	985	46 653

Davranış profilleri bireysel hesaplardan ve popüler toplu uygulamalardan gönderilerek tehlikeli olanlar tespit edilmeye çalışılmıştır. Tespit edilen bu değerler Çizelge 2.3'de, İleti ve Paylaşılan Link bazında verilmiştir.

2.2.2. Kısa link analizli yaklaşım

Kötü amaçlı kişiler tarafından çeşitli saldırılara maruz kalarak ele geçirilen bir web sitesi "zombi" olarak adlandırılır. Bu ele geçirilen site başka saldırılar için kullanılır ve saldırganların kötü amaçlarına hizmet ettirilir. Örneğin, URL yeniden yönlendirme

mekanizmaları, web tabanlı saldırıları gizlice gerçekleştirmenin bir yolu olarak yaygın şekilde kullanılmaktadır. Bu yönlendirme mekanizmaları, ele geçirilen bir web sitesine yönlendirme kodu ekleyerek ziyaretçiyi, otomatik olarak kötü amaçlı bir yazılım dağıtım sitesine yönlendirebilmektedir. Bu şekilde çalışan kötü niyetli web sitelerine karşı birçok savunma işlemi geliştirilmiş olmasına rağmen bugün hala çok sayıda aktif kötü amaçlı web sitesiyle karşılaşılabilmektedir.

Bir çalışmada Akiyama ve diğ.'nin, URL izleme sistemini hayat geçirip dört yıl boyunca veri elde etmişlerdir. Bu süre zarfında, 776 farklı web sitesinden çıkarılan 100 000'den fazla kötü amaçlı yönlendirme URL'si toplanmıştır [28]. Tıklama ile dolandırıcılık yapan saldırganların büyük çoğunluğu URL yeniden yönlendirmeyi kullanmaktadır. URL izleme sistemi, Web tabanlı genetik algoritmalar (Domain Genetic Algorithm-DGA) kullanılarak yönlendirme URL'lerini engellemek için onları kara listeye almaktadır. Web yönlendirme zincirindeki sitelerin, domain ve IP değerlerinin eşzamanlı olarak kullanılması yönlendirme zincirinin sağlamlığını göstermektedir [28].

Bu sonuçlar temel alınarak, kötü amaçlı URL yönlendirmelerine karşı en pratik önlem; Güvenlik veya ağ operatörlerinin, bal küpü tabanlı izleme sisteminden edinilen faydalı bilgileri ulaşılır olmaktan çıkarılmasıdır. Böylece web tabanlı saldırıların altyapısı bozularak önü alınmış olmaktadır. Buna ek olarak, saldırı yapanları tanımlamak için web reklamcılık bilgilerinin izleme kimliklerini toplayarak engellemektir [29].

Dünyada ileti atan milyonlarca kullanıcı, gerçek zamanlı arama sistemleri ve farklı türdeki veri madenciliği araçları ile Twitter'daki olayların ve haberlerin yankılarını izleyebilmektedir [30]. Bununla birlikte, kısa sürede yayılan haberler, anlık durum bildirimlerine olanak tanıyan sosyal ağlar, yeni spam türleri için de bazı uygun ortamlar sunmaktadırlar. Örneğin, Twitter'da en çok konuşulan (trend topic) öğeler, yeni bir gelir fırsatı yaratma olarak görülmektedir. Spam gönderenler, trend alan bir konunun tipik sözcüklerini içeren iletileri URL gibi kısaltıp, kullanıcıları birbiriyle ilişkili olmayan web sitelerine yönlendirirler. Bu tür spam göndericilere karşı kullanıcı tarafında herhangi bir engelleme veya güvenlik tedbiri olmadığı takdirde, istemeden de olsa gerçek zamanlı arama hizmetlerinin verimini düşürmeye katkıda bulunmuş olmaktalardır [31]. Şekil 2.3'de bir popüler hashtag'e gönderilmiş ve görünürde bir reklam gibi görülen ancak zararlı bir link içeren spam bir ileti örneği görülmektedir.

Şekil 2.3. Spam örneği popüler bir hashtag [31]

Benevenuto ve diğ.'nin bu konuda yaptıkları çalışmalarında, spam göndericilerin Twitter'da tespit edilebilmesi yollarını ele almaktadırlar [32]. Veri seti olarak, 54 milyondan fazla kullanıcıyı, 1,9 milyar bağlantıyı ve yaklaşık 1,8 milyar tweeti içeren geniş bir Twitter verisi toplamışlardır. 2009'dan itibaren üç ünlü trend-topic konuyla ilgili tweetler kullanılmış, spam olan ve olmayan kişilere göre sınıflandırılmış geniş bir etiketli kullanıcı koleksiyonu oluşturulmuştur. Daha sonra bu koleksiyondan bir dizi özellik çıkarmışlardır. Bu özellikleri, makine öğrenmesinin öznitelikleri olarak kullanarak, spam göndericileri algılamak ve onların potansiyel olarak kullanılabilecekleri ileti içerikleri ve kullanıcı sosyal davranışıyla ilgili bir sonuç üretmek için kullanılmışlardır [33]. Önerdikleri yöntem ile spam gönderenlerin çoğunu tespit etmeyi başarırken, spam olmayan kullanıcıların da sadece küçük bir oranda yanlış sınıflandırılmışlardır. Spam göndericilerinin yaklaşık %70'i ve spam olmayan gönderilerin % 96'sı doğru olarak sınıflandırmışlardır [32].

2.2.3. Karşılaştırma ve kıyaslama yaklaşımlar

Sosyal ağlardaki zararlı yazılımları yönetenler, kullanıcının kişisel bilgilerini sızdırmak, yanlış bilgi yayma işlemleri için insan gibi davranan özel programlar (BOT) kullanırlar. Botlar, arkadaş istekleri gönderme, mesaj gönderme ve web sitelerinde ki sahte bilgileri ileterek kullanıcıyı etkileyebilir ve bunu son derece hızlı bir şekilde ve otomatik olarak yapılabilir [34, 35].

Bilgisayarların insanlar gibi davranış göstermeleri ilk kez Reeves ve Nass, tarafından 1996 yılındaki çalışmalarında dikkat çekilmiştir [36]. Bu insan davranışları sergileyen botların kullanıcılar tarafından güvenilir, cazibeli, yetkin olarak algılandıklarını fark edilmiştir [34]. Bu botlar, rastgele çalışırlar ve isteklerini kabul edenlerle, iletişim kurarlar. Bu yolları

kullanılan botlar çoğu kez gerçek kullanıcıların duygusal ve zayıf yönlerini su istimal ederek sosyal mühendislik sergilerler ve bu şekilde birçok kez amaçlarına ulaşmaktalardır [36].

"Karşıtlık karşılaştırmaları" diye adlandırabileceğimiz bu yaklaşımda gerçek kullanıcılarla gerçek olmayan kullanıcıların makine öğrenmesi sisteminde Destek Vektör Makinesi (Support Vektor Machine-SVM) ile kullanılan sınıflandırma metodu ile analizlerinin yapılmasıdır. Otomatik (Bot) hesaplardan atılan mesajlar ile gerçek mesajların karşılaştırılarak analizlerinin yapılması spam belirlemede kullanılan etkili bir tespit yöntemidir.

Fernandes ve diğ.'nin 2015 yılında yapılan bir çalışmada, karşılaştırma ve kıyaslama yaklaşım yöntemini kullanılarak, %90 oranında benzer F1 doğruluk puanı (F1 puanı; hassaslığın harmonik ortalamasıdır) elde edilmiştir [37]. Ancak, bu F1 doğruluk puanı gerçek kullanıcıların sergiledikleri anormal davranışları sınıflandırmada sorun teşkil etmiştir. Bunu önlemek için, ikinci bir aşamalı sınıflandırma yöntemi gerçekleştirilmiş ve %74'lük bir ortalama ile F1 doğruluğu elde edilmiştir. Bu doğruluklar, kademeli özellik seçimi ve kategori sonuçlarının tek tek kontrol edilmesiyle oluşan kategori dengelenmesi kullanılarak özellik alanının boyutu azaltılarak elde edilmektedir [37].

Clark ve diğ.'nin geliştirdikleri yöntemde, mevcut algılama algoritmaları ve robotik hesapları tanımlamak için bazı tipik özellikler (tweetler arasındaki süre, takipçi sayısı, vb.) kullanmışlardır [38]. Burada, otomatik mesajlar atan hesapları tanımlamak için bir kıstas ile gerçek kullanıcıların doğal dil yapısını kullanan güçlü bir sınıflandırma şeması sunmuşlardır. Bu şema, yalnızca metin üzerinde çalıştığı için esnektir ve sadece Twitter'da değil herhangi bir metin içeren diğer sosyal medya servislerinde de uygulanabilmektedir.

Wu ve diğ.'nin yaptıkları bir çalışmada, sosyal medya kullanıcısı ve spam mesajının birlikte tespit edilmesinde "mikroblogging" yönteminin kullanımı önerilmektedir [39]. Bu yaklaşımda, sosyal medyada spam gönderenleri belirleme ve spam mesajı belirleme birleştirilerek kullanıcılar ve mesajlar arasındaki ilişkiler incelenmiştir [39]. Buna ek olarak, belirleme sonuçlarını hassaslaştırmak için kullanıcılar arasındaki sosyal ilişkileri ve mesajlar arasındaki bağlantılar çıkarılmıştır. Ayrıca bu yöntemin verimli bir algoritması çıkarılarak, en kısa zamanda ve en fazla nasıl adım atılacağı ve bunun üstesinden nasıl gelineceğinin belirtildiği hızlandırılmış bir yöntem önerilmiştir [29]. Gerçek dünyadaki bir "mikroblog"

veri setinde yapılan kapsamlı denemeler, önerilen yaklaşımın, hem sosyal spam gönderenleri algılamayı hem de spam mesajlarını tespit etmeyi başarılı ve etkili bir şekilde yapabildiğini göstermiştir [39].

2.2.4. Sahte bilgi tespiti yöntemi

Sosyal ağlardaki yaygın kullanılan bir spam örneği de aldatıcı spamlardır. Bu spamlar genelde kullanıcıları aldatıcı, yanlış bilgi ve içerik yaymaktalardır [40]. Kullanıcıyı cezbeden, ilgi çekici ve görünüşte hiçbir zararlı öğe bulunmayan sahte mesajlarla kullanıcılar zararlı site veya adreslere yönlendirmektelerdir Bu aldatıcı mesajlara verilen yanıtların bölgesel analizi, hangi sitelere yönlendirildiği, kullanıcıdan ne tür bilgiler istendiği analiz edilerek spam tespiti yapılması sağlanmaktadır. Aşağıda Şekil 2.4'de bu mesajlara örnek gösterilmektedir. Kullanıcıya gönderilen sahte bilgiyi tespit edilerek ve bu sahte bilgi analiz yapılarak spam adrese ulaşılmaktadır [40, 41].

Şekil 2.4. Aldatıcı ve sahte bilgi spam örneği

Twitter üzerinden iletilen bağlantılar masum gözükebilmektedir. Kötücül linklerin kendilerini saklama amaçları bulunmaktadır. Kötücül bir link kendisini masum bir link olarak algılattığı zaman, birçok Twitter kullanıcısını üzerine çekebilmektedir.

2.2.5. Trend başlıklar analiz yöntemi

Twitter'da spam algılamanın yapıldığı önceki çalışmaların birçoğunun, kötü niyetli kullanıcı hesaplarının ve Bal Küpü tabanlı yaklaşımların tanımlanması üzerine yoğunlaştığı görülmektedir. Bununla birlikte, kullanılmaya başlayan iki yeni yöntem daha vardır. Bunlar; spam tespitlerin kullanıcının bilgisi olmadan izole edilmesi ve trend konularında spam tespit etmek için dilin istatistiksel analizinin uygulanmasıdır. Trend-topic olan konuları, ortaya çıkan internet eğilimlerini ve tartışma konularını herkesin sözün özel yapısına dikkat ederek yakalamaktalardır.

Romo ve Araujo tarafından önerilen bir yaklaşımda; dili birincil araç olarak kullanarak gerçek zamanlı olarak spam iletilerini tespit etmeye çalışmışlardır [43]. Deneysel çalışma için, 34 bin trend-topic konu ve 20 milyon tweet'e sahip geniş bir veri seti toplamışlardır. Bu setin yanında, spam gönderenler tarafından değiştirilmemiş ve azaltılmış belli özellikler tablosu da hazırlamışlardır. Ayrıca, sosyal ağlarda spamların sergilediği özellikleri analiz etmek amacıyla ve diğer özelliklerle birleştirilebilen bazı özelliklere sahip bir makine öğrenme sistemi geliştirmişlerdir. Bunlara ek olarak, kurulan bu sistemin, spam hesaplarının tespitine dayanan en gelişmiş teknoloji sistemleriyle aynı seviyede, F-Ölçütü metriğine göre başarılı bir performans elde edebildiğini gösteren kapsamlı bir değerlendirme yapmışlardır. Bu değerlendirme sonucunda önerilen bu sistemin, kullanıcı hesapları yerine tweetlerin analiz edilmesiyle gerçek zamanlı trend-topic konularında spamların algılanmasına yarar sağladığı görülmektedir [43].

2.2.6. Takip ve takipçi kıyaslama yönetimi

Son yıllarda çok hızlı büyüyen sosyal ağlardan biri olan Twitter, içerisinde birçok spam kullanıcısını da barındırmaktadır. Birçok araştırmacı, Twitter'daki bu şüpheli kullanıcıları tanımlamak için spam algılama yöntemleri önermişlerdir. Bu yöntemlerden birisi de, kullanıcıların arasındaki takipçi (follower) ve arkadaş (friend) ilişkilerini iyi analiz etmektir. Twitter'ın var olan spam politikasına dayalı olarak, yeni içerik temelli özellikler ve grafik temelli özellikler de spam algılamayı kolaylaştırmaktadır.

Wang'nın yaptığı bir çalışmada, Twitter tarafından sağlanan bir API ve bir web tarayıcısı ile yeni bir yöntem geliştirilmiştir [44]. Twitter üzerinde kamuya açık olan verilerden toplam 25 bin kullanıcı, 500 bin Tweet ve 49 milyon takipçi ve arkadaş ilişkisi toplanmıştır. Şüpheli davranışları normal olanlardan ayırmak için makine öğrenme sistemi içerisin de Navie Bayes sınıflandırma algoritması uygulanmıştır. Veri seti analiz edilip algılama sisteminin performansı, geleneksel değerlendirme metrikleri ve çeşitli sınıflandırma yöntemlerine ile karşılaştırılmıştır [44]. Elde edilen sonuçlar, Navie Bayes sınıflandırma algoritmasının F-Ölçütü metriğinin en iyi genel performansa sahip olduğunu göstermektedir. Eğitimli tüm veri kümesi test edildiğinde ise; sonuç spam algılama sisteminin %89 hassasiyet elde edebildiğini göstermektedir [44].

Hacıefendioğlu'nun bir araştırmasında, sosyal ağlar çok sıklıkla reklam ortamı olarak kullanılmaktadır, paylaşılan reklamlar sosyal paylaşım siteleri kullanıcıları tarafından %75,8 gibi bir oran ile baktıkları ayrıca %59,2 oranında reklamdaki ürünleri diğer kullanıcılara önerdikleri görülmüştür [45].

Jeong ve diğ.'nin yaptıkları bir çalışmada, Twitter'da spamları tespit etme yollarını ele almaktalardır. Zararlı bu mesajları tespit için, spam gönderen kullanıcıları takip ederken, aynı zamanda meşru bir kullanıcıyı da takip etmektelerdir [46]. Böylece spam göndericilerinin aktif ilişkileri ile meşru kullanıcıların aktif ilişkilerini karşılaştırma ve analizine dayanan bir sınıflandırma şeması önermişlerdir. Bu özellikler için basamaklı toplumsal ilişkilere odaklanılmış ve her biri merkezi olan iki aşamalı bir alt ağda Üçlü Önem Profili (Triple Significance Profile-TSP) ve Sosyal Statü (Social Status-SS) kullanan TSP Filtreleme ve SS Filtreleme olmak üzere iki plan geliştirilmiştir [46]. Ayrıca, hem TSP hem de SS özelliklerini birleştiren ve bir "topluluk yöntemi" olan Kademeli Filtreleme yöntemini önermişlerdir. Çalışmalarında gerçek Twitter veri setleri kullanılmış, yapılan deneysel çalışmalarda önerilen üç yaklaşımın, kullanımının çok kolay olduğu görülmüştür. Bu yöntemin avantajı önerilen şemaların ölçeklenebilir olması, tüm şebekeyi analiz etmek yerine, kullanıcı merkezli olarak ağları incelemesidir [46].

2.2.7. Topluluk öğrenimi yöntemi

İnternetin hayatımıza tamamen girdiği günümüzde sosyal medya kullanımı oldukça artmış ve tüm insanlar küreselleşen dünyada birbirileriyle iletişime geçebilme imkânına kavuşmuş durumdalardır. İstatistikler göstermektedir ki sosyal ağların ortalama kullanım oranı diğer sitelerin kullanım oranlarını geçmiş bulunmaktadır [47, 48].

Sosyal ağlarda kullanıcıya yönelik tehlikeleri azaltmak için yapılan birçok yeni çalışma bulunmaktadır. Makine öğrenme teknikleri bu çalışmalarda Twitter spamını sınıflandırmak için uygulanmaktadır. Ayrıca, Twitter'da sınıflandırılan spamlar Twitter verilerindeki dengesizliğini ortadan kaldırmaktalardır [36, 49].

Liu ve diğ.'nin yaptıkları çalışmada, spam ve spam olmayan sınıflar arasındaki dengesiz dağılımın spam algılama oranı üzerinde büyük bir etkisi olduğunu göstermişlerdir [50]. Bu

problemin çözümünde, bulanık tabanlı bilgi ayrıştırma (Fuzzy-based Information Decomposition Oversampling-FOS) algoritmasını uygulamışlardır. Sınırlı gözlemlenen örneklerden sentetik bir veri seti üreten bulanık tabanlı yeni bir yöntem önermişlerdir. Ayrıca, üç aşamada dengesiz gibi görülen verilerden daha doğru bir sınıflandırıcı ile öğrenen, bir topluluk öğrenme yaklaşımı geliştirmişlerdir. Bu yöntemde ilk önce, dengesiz veri kümesindeki sınıf dağılımı ayarlanmıştır. İkinci olarak, yeniden sınıflandırılmış bir veri kümesinin her biri üzerine bir sınıflandırma modeli oluşturulmuştur. Son aşamada ise, tüm sınıflandırma modellerinden gelen sonuçları birleştirmek için çoğunluklu bir oylama sistemi geliştirilmiştir. Değerlendirme amacıyla, Twitter verileriyle ilgili yapılan deneylerde elde edilen sonuçlar, önerilen öğrenme yaklaşımının spam, sınıf dağılımı dengesiz olan veri kümelerindeki spam algılama oranını önemli ölçüde artırabileceğini göstermektedir [50, 51].

2.2.8. Hesap oluşturma zamanı temelli yöntem

İnternet ortamında aktif olan kötü niyetli kişiler, sosyal ağlarda spam dağıtımı gibi büyük ölçekli basit saldırıları gerçekleştirmek için çok sayıda kısa süreli ve kötü niyetli hesaplardan yararlanırlar. Bununla birlikte, hesap veya mesaj bilgilerine dayanan geleneksel algılama yöntemleri, algılama algoritmalarını çalıştırmadan önce bu tür bilgileri toplamak için çok fazla zaman harcamaktalardır, böylece kötü niyetli kişiler, hesapları askıya alınana kadar sürekli spam hesaplarını kullanmaya çalışmaktalardır [52].

Lee ve Kim yaptıklar çalışmalarında, potansiyel olarak kötü amaçlı hesap gruplarını oluşturma zamanı çevresinde filtreleme yapmak için yeni bir algılama şeması önermişlerdir [53]. Bu amaçla, benzer algoritmalar kullanılarak, kötü amaçlı hesapları tanımlamak için “algoritmik” olarak oluşturulan hesap adları ile gerçek hesap adları arasındaki farklar kullanılmaktadır. Kısa bir süre içinde oluşturulan hesaplar için, benzer kullanıcı adı özelliklerini paylaşan grup hesaplarını ve kötü amaçlı hesap kümelerini sınıflandırmak için ayrı bir sınıflandırma algoritması uygulamışlardır. Çalışmalarında veri seti olarak, Twitter'dan toplanan 4,7 milyon kullanıcı hesabı kullanmışlardır. Oluşturulan bu şema sadece kullanıcı hesap isimlerine ve oluşturulma sürelerine dayanmasına rağmen kabul edilebilir bir doğruluk değeri elde etmektedir. Bu yöntem, kötü amaçlı hesap gruplarına karşı, ayrıntılı bir analiz gerçekleştirmek için hızlı bir filtre olarak kullanılabilecek yapıdadır [53].

2.2.9. Kısa mesaj analiz yöntemi

Sosyal ağlarda, çokça karşılaşılan durumlardan biriside çok fazla sayıda gelen toplu mesajların varlığıdır. İsteğe bağlı olmayan bu toplu mesajlar mevcut spam filtreleri tarafından etkili bir şekilde ayırt edilebilse de, mesaj örnekleri değiştirilerek spam filtrelerini yanıltmakta ve görevlerine devam etmektelerdir. Ancak, kısa mesajlar (SMS) için mevcut spam filtreleme tekniklerinin zayıf noktaları pek araştırılmamıştır. Diğer spam uygulamalardan farklı olarak, kısa mesaj uygulamalarında yalnızca bir kaç anahtar kelime mevcuttur ve uzunluğunun genellikle bir üst sınırı vardır. Bu durumda mevcut karşıtlık öğrenme algoritmaları kısa mesaj spam filtrelemede etkin bir şekilde çalışmayabilmektedir [54].

Kısa mesaj analizi hakkında 2015 yılında yapılmış bir çalışmada, kısa mesaj spam filtrelemede, iyi bir kelime saldırısı ve karşı saldırı metodu ile uzunluğu kısıtlı bir mesajın verimli bir şekilde nasıl çalıştıkları ve bunların birbirleriyle olan ilişkilerinin ne ölçüde olduğu araştırılmıştır [54]. Yapılan bu çalışmada, ağırlık değerlerine ve ayrıca sözcüklerin uzunluğuna dayalı bir sınıflandırıcının etkisini en üst düzeye çıkaran iyi bir sözcük saldırısı stratejisi önerilmektedir. Diğer yandan, özelliği yeniden değerlendirilmiş, başarılı bir kaçınma için sayısı arttırılmış, karakterler gerektirecek şekilde kısa bir kelimeyi temsil eden özelliğin önemini en aza indirgeyen yeni bir ölçeklendirme işlevi de sunulmaktadır [54]. Yöntemin başarısı, SMS ve yorum spamlarından oluşan bir veri seti kullanılarak değerlendirilmiştir. Elde edilen sonuçlar, kısa kelime spam filtrelemesinin, sözcük saldırısına karşı sağlamlığının kritik bir faktörü olduğunu teyit etmektedir.

Twitter'ın hızlı bir şekilde büyümesi, spam kapasitesinde ve karmaşıklığında çarpıcı bir artışa neden olmaktadır. “başlık etiketleri”, “bahset” ve kısaltılmış URL'ler gibi belirli Twitter bileşenlerini spam göndericileri etkin bir şekilde kötüye kullanılmaktadır. Ancak, buna benzer özellikler, daha önceki çalışmalarda da gösterildiği gibi yeni spam hesaplarını tespit etmede de önemli bir faktör olarak ortaya çıkarılmaktadır [55].

Miller ve diğ.’nin yaptıkları çalışmalarında, ilk olarak daha önceki çalışmaların spam algılamayı bir sınıflandırma sorunu olarak gördüğünü, ancak bunu kendilerinin bir anomali tespit problemi olarak kabul ettiklerini belirtmişlerdir [56]. İkinci olarak, daha önceki çalışmalarda analiz edilen kullanıcı bilgileri ve tweet metinlerinin özelliklerini veri seti

olarak belirlemişlerdir. Son olarak, tweetlerin akış özelliklerini etkin bir şekilde kullanarak, spam kimliğini belirlemeyi de kolaylaştırmak için “StreamKM++” ve “DenStream” olmak üzere iki akış kümeleme algoritmasını değiştirerek kullanmışlardır. Her iki algoritma da, normal Twitter kullanıcılarını gruplandırarak, aykırı isimleri spam göndericiler olarak kullanılmaktadır. Bu algoritmalar ayrı ayrı denendiğinde iyi bir performans sergiledikleri görülmektedir. StreamKM++ ile %99 geri çağırma ve %6,4 yanlış pozitif oranı, DenStream ile de %99 geri çağırma ve %2,8 yanlış pozitif oranı sonuçları elde edilmiştir. Bu algoritmalar birlikte kullanıldığında ise, sistemin spam kullanıcıları tespit ederken, normal kullanıcıların yalnızca %2,2'sini yanlış tespit ettiği, buna karşılık spam göndericilerinin tamamını ise doğru tespit edebildiği görülmüştür [56, 57].

2.2.10. Balküpü tabanlı spam tespiti yöntemi

Bal Küpü yöntemleri, spam tespitinde ve özellikle veri toplamada çokça kullanılan, pratik ve kolay bir yöntemdir. Dagon ve diğ.’nin 2004 yılında yaptıkları çalışmada, sosyal ağların, küresel bazda izlenerek spamlar için erken tespit sağlanabileceğini söylemişlerdir [58]. Bu çalışmada yüksek tespit oranlarına ulaşmak ve doğru bir uyarı mekanizması oluşturmak için değiştirilmiş honeypotlar kullanan bir “Honeypot ağ’ı (HoneyStat)" oluşturmuşlardır. Bunun avantajı, geleneksel etkileşimli honeypotların aksine, HoneyStat düğümlerinin komut dosyası ile yönlendirilmesi ve büyük bir kullanıcı ağını kapsamasıdır [58]. HoneyStat düğümleri, üç ayrı uyarı kategorisi üretmektedir; Belllek uyarıları, Disk yazma uyarıları ve ağ uyarılarıdır. Bu düğümler sayesinde veri toplama işlemi otomatikleştirilir ve düğüme bir uyarı verildiğinde önceki trafiğin zamanı analiz edilebilir. Tutulan bir log kaydı analizi ile önceki ağ aktivitesini açıklayan durum belirlenir. Elde edilen sonuç, kullanıcının otomatik veya solucan saldırısının olup olmadığını göstermektedir. Bu çalışmada, HoneyStat oluşturmanın önceki zararlı yazılım algılama tekniklerine göre daha gelişmiş olduğu gösterilmiştir. İlk olarak, küçük ağlardaki zararlı saldırılarından gelen izleme dosyalarını kullanarak “sıfır gün solucanlarının” (yeni ortaya çıkmış zararlı yazılım) nasıl tespit edildiği. İkinci olarak, saldırı için bağlantı noktalarında, çoklu zararlı yazılımların nasıl algıladığı gösterilmiştir. Ayrıca HoneyStat'tan gelen uyarılar, saldırı bilgileri ve oranları gibi toplanan geleneksel bilgilerle birlikte kullanılabilmektelerdir [58].

Yang, ve diğ.’nin yaptıkları bir çalışmada, Twitter spam göndericilerinin beğenilerini daha etkili sosyal Honeypotları oluşturmak ve bunları sosyal spam göndericilerine karşı

savunmak için yeni yollar üretmeyi ve bu yollar için bazı, "ölçütler" koymayı önermişlerdir [59]. Spam göndericileri tuzağa düşürmek için çeşitli sosyal davranış kalıpları ile heyecan verici honeypot'lar oluşturmaktalardır. Beş aylık bir veri toplama aşamasından sonra, Twitter spamcılarının hedeflerini nasıl buldukları konusunda detaylı bir analiz yapılmaktadır. Analiz sonuçlarına göre, gelişmiş ve etkili bir sosyal Honeypot oluşturmak için neler gerekir konusu değerlendirilmiştir. Özellikle, aynı zaman periyodunda kullanılan, bu gelişmiş honeypot'lar, spam gönderenleri belirlemede, "geleneksel" honeypot'lara karşı yaklaşık 26 kat daha hızlı sonuç vermektelerdir. Yang ve diğ.'nin yaptıkları çalışmanın ikinci bölümünde, spam göndericileri cezbeden Honeypot'ların yeni bir veri toplama yaklaşımı incelenmiştir. Burada amaç, sınırlı kaynakları ve sınırlı zamanı göz önüne alarak, mümkün olan örneklemleri almak için tüm Twitter hesaplarını taramak yerine (daha sonraki spamcı analizler için) etkin tarama ve örneklemeye öncelik tanıma stratejisinin geliştirilmesidir. Geniş Twitter ağında bulunan spam hesaplarının zevkleri ile ilgili veri toplanarak iki yeni, etkin ve geçerli bir örnekleme yaklaşımı oluşturulmuştur [59].

2.2.11. Spam tespit araçları kullanma yöntemi

Sosyal ağlarda kullanıcıları tehdit eden kötü amaçlı hesap ve mesajlar birçok yöntem ile bulunabildiği gibi bazı hazır harici yazılımlar da spam kullanıcı ve mesajları tespitinde kullanılmışlardır [39]. Yazılımlardan bazıları şunlardır; Integro [42], SybilRank [41], NodeXL, (Network Overview, Discovery and Exploration for Excel) [60], Pajek [49], ReDites [24], Canary Honeypots [59].

Bu programlardan Integro yazılımı sahte hesaplara göre gerçek hesapların geçerliliklerini kıyaslayarak puan vermekte ve bu puanlara göre spam hesap tespiti sağlanmaktadır [42]. İnsan ilişkilerinde artık yüz yüze iletişim yerine, teknolojisinin kullanılarak sanal iletişimin tercih edilmeye başlaması sosyal medyanın her alanda kullanılmasına neden olmaktadır.

2.3. Spam Tespit Yöntemlerinin Karşılaştırılması

Spam tespit yöntemleri bu kısımda literatürde yer alan ve bu konuda yapılmış çalışmaların birbirleriyle karşılaştırılarak spamlar hangi yollarla kullanıcıların kişisel verilerini tehdit etmektelerdir ve bunu hangi methodlarla yapmaktalardır, detaylı bir şekilde incelenmiştir ve elde edilen sonuçlar sunulmuştur. Literatürde sosyal ağlarda spam tespiti için yapılmış

güncel çalışmaların birbirlerine göre farklılıklarına odaklanılmıştır. Çizelge 2.4'de bu çalışmalar ve özellikleri görülmektedir.

Çizelge 2.4. Literatürde yer alan spam tespit çalışmalarının karşılaştırılması

Makale	Teknik	Algoritma	Veri Kümesi	Değerlendirme Metriği
Akiyama ve ark. (2017)	İzleme Sistemi	Web Tabanlı Genetik Algortima	Yönlendirme Kodlarının Enjekte Edildiği URL Veri Kümesi	Performans Oranı
Fernandes ve ark. (2015)	Makine Öğrenme Sistemi	Sınıflandırma ve Kümeleme Algortima	Twitter Veri Kümesi	F- Ölçütü
Clark ve ark. (2016)	Makine Öğrenme Sistemi	Geleneksel Sınıflandırma Algoritma	Twitter Bot Veri Kümesi	Alıcı İşletim Karakteristiği
Wu ve ark. (2016)	Makine Öğrenme Sistemi	Çarpanların Alternatif Yön Metodu Tabanlı Algoritma	Gerçek Mikroblog Veri Kümesi	λ Parametre
Chen ve ark. (2016)	Makine Öğrenme Sistemi	Grafik Tabanlı Algoritma	Twitter ve URL Veri Kümeleri	Doğru Onaylanmış
Romo ve Araujo (2013)	Makine Öğrenme Sistemi	Geleneksel Sınıflandırma Algoritma	Twitter Veri Kümesi	F- Ölçütü
Jeong ve ark. (2016)	Makine Öğrenme Sistemi	Üçlü Önem Profili ve Sosyal Statü Filtreleme Basamaklı Filtreleme	Gerçek Twitter Veri Kümeleri	Doğru Onaylanmış
Shigang ve ark. (2016)	Makine Öğrenme Sistemi	Rastgele Aşırı Örnekleme, Rastgele Örnekleme, Bulanık Tabanlı Bilgi Ayrıştırması	Twitter ve URL Veri Kümeleri	F- Ölçütü
Lee ve Kim (2014)	Makine Öğrenme Sistemi	Oluşturma ve Destek Vektor Makinesi Algoritma	Kullanıcı Hesap Veri Kümesi	Hatalı Reddedilmiş
Miller ve ark. (2014)	Makine Öğrenme Sistemi	DenStream ve SteramKM++	Gerçek Twitter Veri Kümeleri	F- Ölçütü ve Duyarlılık
Boshmaf ve ark. (2015)	Üçüncü Parti Yazılım	Indigo ve Random Forest Algortima	Facebook Hesabı Veri Kümesi	Alıcı İşletim Karakteristiği
Wang (2010)	Makine Öğrenme Sistemi	Naive Bayes Algoritma ve Twitter API	Gerçek Twitter Veri Kümeleri	F- Ölçütü

Çizelge 2.4'de görüldüğü gibi sosyal ağlar üzerinde hızla yayılan zararlı yazılımlarda olan spamların tespit yöntemleri incelenmiştir. İncelenen çalışmalarda hangi yöntemlerin

kullanıldığı, ne gibi sonuçların elde edildiği ortaya konulmaktadır. Yapılan karşılaştırmalara göre sosyal ağlarda spam hesapların tespit edilmesinde farklı yöntemlerin kullanıldığını anlaşılmaktadır. Ele alınan çalışmaların büyük çoğunluğunun makine öğrenmesi tabanlı yöntemleri kullanmış oldukları görülmüştür. Makine öğrenmesi yöntemi önerdiğimiz sosyal ağlarda spam hesap tespit modelimizin önemli bir bileşenidir. Buna ek olarak az da olsa harici yazılımların kullanıldığı çalışmalarda mevcuttur. Çok yakın sonuçlar elde eden bu çalışmalardan en yüksek başarıyı yakalayan çalışmanın makine öğrenmesine dayalı ve akış tabanlı sınıflandırma algoritmalarının kullanıldığı, Miller ve diğ. tarafından 2014 yılında yapılan çalışma olduğu görülmektedir [56]. Bu çalışmanın en belirgin ve önemli özelliği iki sınıflandırma algoritmasının kombine edilerek kullanılmış olmasıdır. Yine %95 üzeri değer elde eden Clark ve Chen çalışmalarının da Miller'inkine çok benzediği, onların da makine öğrenmesine dayalı yöntem kullandığı, tek farkın standart sınıflandırma algoritmaları kullanmaları olduğu belirlenmiştir [38, 40]. Elde edilen sonuçların birbirlerine çok yakın oldukları, Çizelge 2.5'te özellikle %90 üzerinde çok yüksek değerlere sahip yedi çalışma görülmektedir.

Çizelge 2.5. Literatürde yer alan spam tespit çalışmalarının özellik tablosu

Sıra	Çalışmalar	Doğruluk Oranları
1	Akiyama ve ark. [28]	96,50
2	Fernandes ve ark. [37]	90,00
3	Clark ve ark. [38]	95,21
4	Wu ve ark. [39]	93,00
5	Chen ve ark. [40]	95,00
6	Boshmaf ve ark. [42]	76,00
7	Romo ve Araujo [43]	94,50
8	Wang [44]	89,00
9	Jeong ve ark. [46]	96,30
10	Liu ve ark. [50]	78,00
11	Lee ve Kim. [53]	86,53
12	Miller ve ark. [56]	97,10

Çizelge 2.5'deki çalışmalarda İleti mesajlarından oluşan veri kümelerinin kullanılmıştır. Tek istisnai durum, Akiyama'nın çalışmasında görülmekte ve mesajlar yerine veri kümesi olarak sadece URL bilgileriyle çalışması olmaktadır. Kullandığı sistem ve algoritma ile diğer çalışmalardan oldukça farklı bir çalışma ortaya koymaktadır ve elde ettiği sonuç ile tatmin edici bir başarıya ulaşmaktadır. Aslında bu çalışma bize ilerleyen yıllarda daha çeşitli veri kümeleri ve değişen algoritmalarla daha yüksek sonuçlar alınabileceğinin en büyük

göstergesidir. İncelenen çalışmalar arasında sadece iki çalışma %80'nin altında değer elde etmektedir. %80'nin altında kalan çalışmalardan en düşük sonucu elde eden çalışmanın harici bir yazılımla gerçekleştirilmiş olduğu görülmektedir. Bu yüzden harici yazılımların bu tür çalışmalarda kullanılması önerilmemektedir. Milyonlarca kullanıcısı ile çok yoğun bir trafiğe sahip olan sosyal ağlarda kullanılması en etkili sistemin makine öğrenmesine dayalı sistemler olduğu görülmektedir. Ayrıca yapılan çalışmalarda elde edilen sonuçlar ve başarı oranlarının kullanılan algoritmalara ve veri kümelerine göre değiştiği açıkça görülmektedir. Makine öğrenme yöntemlerinin kullanıldığı yöntemlerde sistemin eğitim yapılan kısmında ki veri kümesinin çok zengin ve çeşitli olması sistemin test ettiği veri kümesi üzerinde daha doğru sonuçlar alınmasını sağlamaktadır. Yine incelenen çalışmalarda, spam tespiti için gerçek Twitter veri kümeleri, URL bilgileri, sosyal medyadaki profil bilgileri de kullanılmaktadır. Bu bilgiler arasında gerçek Twitter veri kümelerinin kullanıldığı çalışmaların daha yüksek başarım oranları elde ettikleri görülmektedir. Sınıflandırma metoduyla yapılan çalışmalarda, spam mesajlara sahip hesaplar ile normalde spam üretmeyen gerçek kullanıcılara ait bazı mesajların da spam olarak kabul edildiği görülmektedir.

3. YÖNTEM VE ARAÇLAR

Önerilen öğrenmeye dayalı spam hesap tespit modeli, sosyal ağlar içerisin de çok önemli bir yere sahip olan Twitter üzerinde kullanılmak için tasarlanmıştır. Bu bölümde, bu modeli oluşturan araçlardan, yöntemlerden ve önerilen modeldeki kullanım yöntemlerinden bahsedilmektedir.

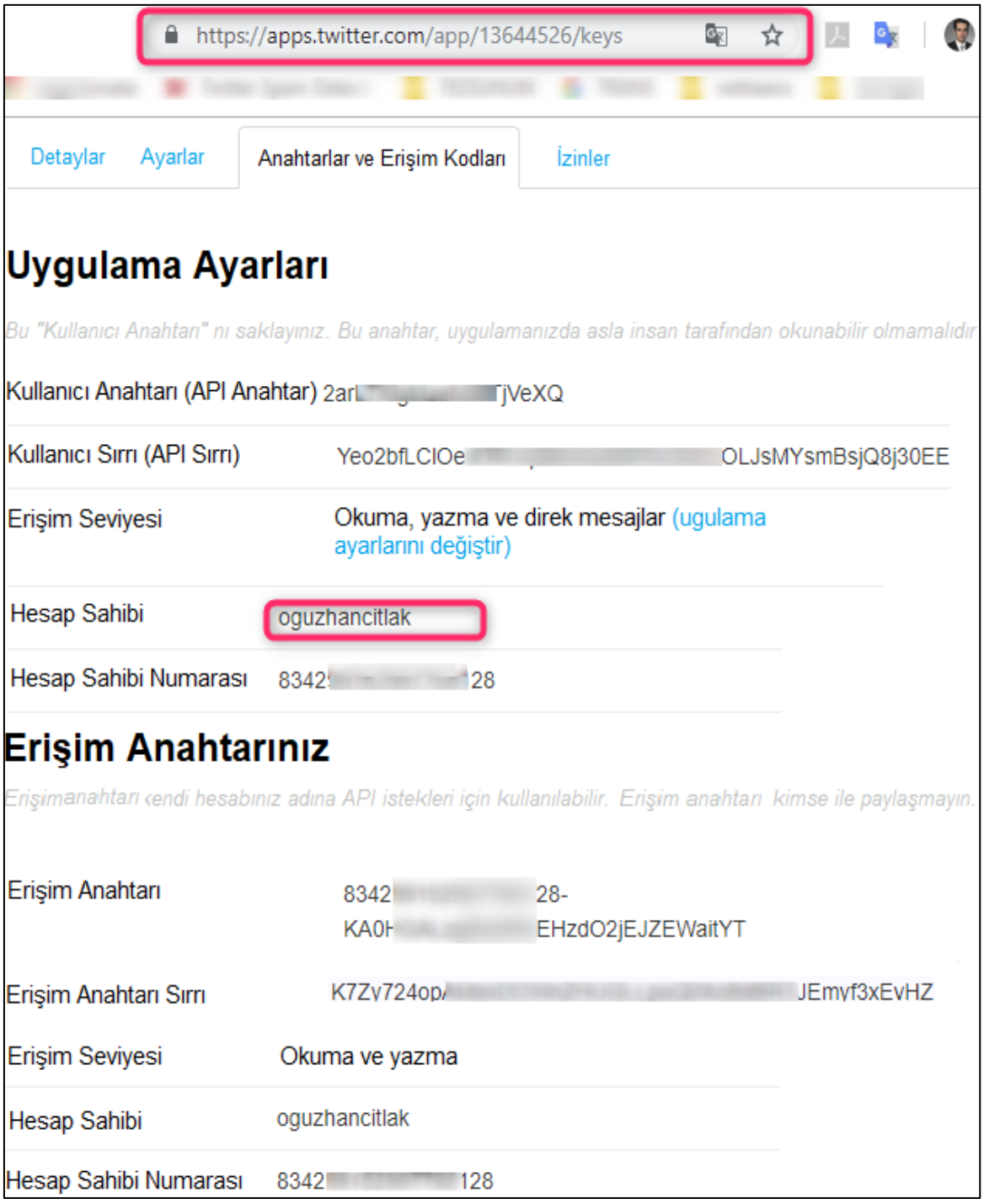
3.1. Araçlar

Kullanılan Twitter veri kümesi dosyası Twitter üzerinden oluşturulmuştur. Twitter'dan oluşturulan API sayesinde uygulama geliştiricileri, çalışmalarında Twitter verilerini kullanabilmektedir [61]. Twitter'in veri kümesi oluştumak için sağladığı ara yüzü sayesinde çalışma veri kümesi dosyası oluşturulmuştur [61]. Twitter veri kümesi içerisinde 221 756 adet kullanıcı hesabının detaylarının olduğu çok büyük bir veri kümesidir. Bu veri kümesi, Json formatında Python yazılımı kullanılarak Twitter üzerinden çekilmiştir. Literatür araştırmalarına bakıldığında Twitter üzerinde yapılan çalışmalarının çok büyük bir kısmında Twitter veri kümeleri kullanılmaktadır [55, 62]. Eğer sosyal ağlarda bir çalışma yapılacaksa ihtiyaç duyulan olan en önemli materyal veri kümesidir. Python aracılığı ile çekilen Twitter veri kümesi çekme kod parçacığı EK-3 içerisindedir.

Twitter, gelişmiş bir mesajlaşma platformu sağlamasının yanı sıra yazılım geliştiricileri içinde bir geliştirme ara yüzüne sahiptir [63].

Twitter, kullanıcıların paylaşım durumlarını, takipçi sayısı, ileti sayısı ve kim ne kadar ileti yayımlamış gibi detaylı analiz bilgileri sunan bir ara yüze sahiptir [64]. Herkesin yaptığı paylaşımlar akışkan iletim diye de adlandırılabilen bir veri nehrine akmaktadır [57].

Twitter üzerinden veri kümesi oluşturulduğunda bir takım sifrelere ihtiyaç bulunmaktadır. Kullanıcı Anahtarı (Consumer Key), Kullanıcı Sırrı (Consumer Secret) , Erişim Anahtarı (Access Token) ve Erişim Anahtarı Sırrı (Access Token Secret) özellikleri @oguzhancitlak Twitter hesabının API anahtarlarını göstermektedir. API anahtarı, hesap açan her kullanıcı için farklıdır. Twitter üzerinden edinilen API anahtarına veri kümesini oluşturmak için oluşturduğumuz hesabın detayları Şekilde 3.1'de görülmektedir.



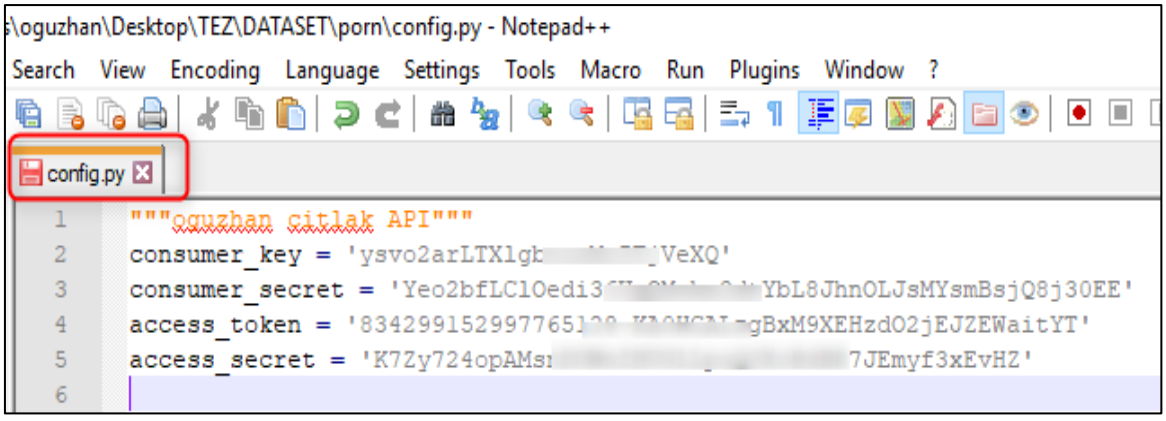
Şekil 3.1. Twitter API şifreleme sayfası [63]

Şekilde gösterildiği gibi Twitter API, kullanıcılara bağımsız bir platform sunmaktadır. Bu API nin sağladığı kolaylıklardan şu şekilde bahsedilmektedir [63, 65]. Twitter ana sayfasındaki yayın akışı takip edilebilmektedir. Geliştirilen bir uygulama üzerinden Twitter'a erişilebilmektedir. Bir kullanıcının profil bilgileri görüntülenebilmektedir. İstenilen kişiye direk mesaj atılabilmektedir. Takip etme ya da takipten çıkma işlemleri

yapılabilmektedir [65]. Twitter API, sahip olduğu özelliklerini modelimizde çalıştırabilmemiz için kullandığımız bir anahtardır. Sosyal medya sitelerinden otomatik bağlantı almak için veya bir site içerisinde sosyal medya üzerinden kullanıcının girişini yapabilmek için API anahtarlarına ihtiyaç duyulmaktadır. Modelde kullanılan veri kümesini Twitter üzerinden çekmek için @oguzhancitlak hesabı üzerinden bir API hesabı oluşturulmuştur [63]. Twitter uygulama yönetim sayfasında, ilgili hesabımız için şifreler oluşturulmaktadır. Bu çalışmada, Twitter üzerinden veri kümemizi çekebilmek için API anahtarına ihtiyaç duyulmaktadır.

Python, Twitter üzerinden veri kümesini oluşturabilmek için bu çalışmada kullanılan bir yazılımdır. Pyhton geliştirilmesine 1990 yılları başında başlayan Hollandalı bir yazılımcı tarafından geliştirilmiştir. Bir derleyiciye ihtiyaç duymaz ve yazım süreci oldukça hızlıdır [66]. Twitter'ın uygulama geliştiricileri için kütüphane önerileri bulunmaktadır. Bu öneri listesi [65] içerisinde Python çok önemli bir yer tutmaktadır. Aynı zamanda Ptyhon Kütüphanesi Twitter tarafından da veri kümesi oluşturulması amacı için önerilmektedir [63]. Twitter API üzerinden oluşturulan şifre kodları, Python için tasarlanış kod parçacığı içerisinde kullanılmıştır. Veri kümesi Python yazılımı aracılığı ile oluşturulmuştur. Veri kümesini oluşturmak için Python yazılımı tek başına yeterli olmamaktadır. Tweepy modülünün de Python yazılımına entegre edilmesi gerekmektedir. Twitter, uygulama geliştiricileri için önerdiği birkaç modülden bir tanesi de Tweepy modülüdür [67].

Python yazılımında, Şekil 3.2'de gösterildiği gibi içerisinde Twitter API hesap şifrelerinin olduğu bir config.py dosyası kullanılması gerekmektedir. Aşağıdaki config.py dosyasının içerisindeki kod parçacığı gözükmektedir.



Şekil 3.2. Kullanılan config.py kod parçacığı

Twittter API anahtarı, Python da çalışan veri kümesi çekme kod parçacığına dâhil edilmiştir ve gerekli ayarlamalar yapılmıştır. Nisan 2017 tarihinde içerisinde 221 756 Twitter kullanıcısının özniteliklerinin bulunduğu veri kümesi Twitter üzerinden çekilmiştir. Veri kümesinde bulunan kullanıcıların özelliklerinden EK-2'de bahsedilmektedir. Elde edilen öznitelikler arasından ihtiyaç duyulan kısımları makine öğrenmesi yöntemde kullanılmak üzere ayrılmıştır.

Twitter üzerinden çekilen veri kümesi için hassas içerikli kelime kullanılmaktadır. Spam davranışında bulunan hesapların hassas içerikli kelimeleri daha fazla kullandıkları spam tespit yöntemlerinin karşılaştırılması sonucu anlaşılmaktadır. Twitter'dan çekilen veri kümesinde, spam hesapların sıklıkla bulunması önerilen spam hesap tespit modelinin başarımını arttırmaktadır. Twitter'dan herhangi bir ileti paylaşan kullanıcıların verileri, içerisinde spam özellikli hesapların az olacağı kanısıyla çekilmemiştir. Hassas içerikli kelimelerden birisi olan *bet* kullanılarak veri kümesi oluşturulmuştur. Çizelge 3.1'de kullanılan verikümesinin özellikleri bir arada gösterilmektedir.

Çizelge 3.1. Kullanılan veri kümesinin özellikleri

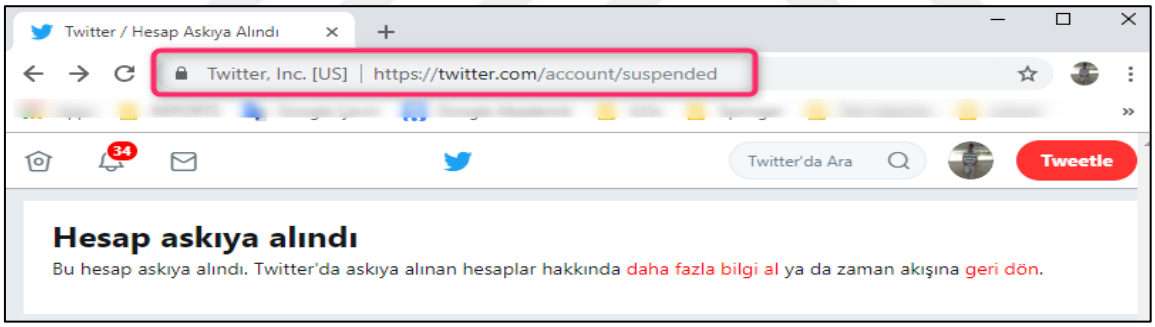
Sıra	Açıklama	Sonuç
1	Kullanılan hesap	@oguzhancitlak
2	Veri Çekimi Başlangıç Tarihi	13 Nisan 2017 13:20:43
3	Veri Çekimi Bitiş Tarihi	16 Nisan 2017 02:04:11
4	Çekilen Kullanıcı Sayısı	221 756
5	Dosya Uzunluğu (byte)	1 019 742 837
6	Karakter Sayısı (boşluklar hariç)	1 018 855 813
7	Çekilen Data Boyutu	972 MB
8	Kullanılan Hassas İçerik	bet
9	Çekilen Dosya Türü	JSON
10	Dosya Satır Sayısı	443 512
11	Kelime Sayısı	107 307 574
12	Kullanılan Yazılım	Python

Çizelge 3.1'de detayları verilen veri kümesini çekmek için kullanılan kod parçacığı EK-3'de verilmektedir. Çekilen veri kümesi içerisinden rastgele 81 317 tane Twitter hesabı makine

öğrenmesi yönteminde kullanmak için ayrılmıştır. Bölünme sonucu veri kümesinin boyutunda da küçülme olmaktadır ancak bu bir problem teşkil etmemektedir. Kalan 140 439 hesap içerisinden rastgele hesaplar seçilmiştir, aralarından 1 225 adet Twitter hesabı Weka da eğitim kümesinde kullanmak maksadıyla ayrılmaktadır. Ayrılan kullanıcı hesaplarının öznitelikleri çıkartılmıştır.

Spam hesapların belirlenmesinde izlenilen yöntem bekle ve kontrol et yöntemidir [68]. Veri kümesinin oluşturulduğu zamandan bir yıl sonra Nisan 2018 tarihinde veri kümesi içerisinden seçilen rastgele hesapların kullanıcı adları Twitter üzerinde teker teker kontrol edilmiştir. Veri kümesi ilk oluşturulurken etkin olan hesaplar bir yıl sonrasında Twitter tarafından askıya alınmışlardır.

Eğitim kümesinin öznitelikleri oluştururken veri madenciliği gereksiz bilginin ayıklanması yöntemi kullanılmıştır [68]. Askıya alınan hesaplar teker teker ayıklanarak makine öğrenmesi yöntemi yazılımlarından olan Weka'da kullanılan özellikleri çıkartılmıştır.



Şekil 3.3. Twitter'da askıya alınan bir hesap [69]

Şekil 3.3'de görüldüğü gibi askıya alınan hesap linki gösterilmektedir. Twitter'ın bu hesapları ne kadar süre sonrasında askıya aldığı şuan için tespit edilememektedir. EK-4'de @LoErA5Ngngo7EEr isimli spam bir kullanıcıya ait Twitter hesap detay özellikleri gösterilmektedir. Twitter tarafından askıya alınan iki adet örnek spam hesap; @LoErA5Ngngo7EEr [EK-4], @donnelllebla41 [69] kullanıcılarıdır. Twitter, bu iki hesabı askıya almıştır. Json (Javascript Object Notation), Twitter üzerinden çekilen ham dosyaların kayıt edildiği veri formatıdır [62]. Json, genellikle Javascript uygulamaları için kullanılmaktadır. Veri transferlerinde XML'den daha az yer kaplaması sağlanmaktadır. PHP, Java, .Net, Web servis uygulamaları ve mobil uygulamalarında veri transferlerinde sıklıkla kullanılmaktadır [23, 62].

Twitter üzerinde bulundurduğu ham verilerin işlenmesi için Json formatını kullanmaktadır [70]. Json iki yapı üzerine oluşturulmuştur. Bu yapı isim/değer çift koleksiyon ve sıralı değer listesi şeklindedir [62]. Aşağıda kullanılan veri kümesinin json formatında saklanması için kod parçacığı içerisindeki ilgili satırlar gösterilmektedir.

```
def __init__(self, data_dir, query):
query_fname = format_filename(query)
self.outfile = "%sstream_%s.json" % (data_dir, query_fname)

def parse(cls, api, raw):
status = cls.first_parse(api, raw)
setattr(status, 'json', json.dumps(raw))
return status
```

Bu satırlar sayesinde, Python aracılığı ile çekilen veri kümesini otomatik olarak json formatında kayıt edilebilmektedir. Çekilen veri kümesinin kayıt edilmesi ve işlenmesi için bir kayıt dosyasının oluşmasına ihtiyaç duyulmaktadır.

3.2. Yöntem

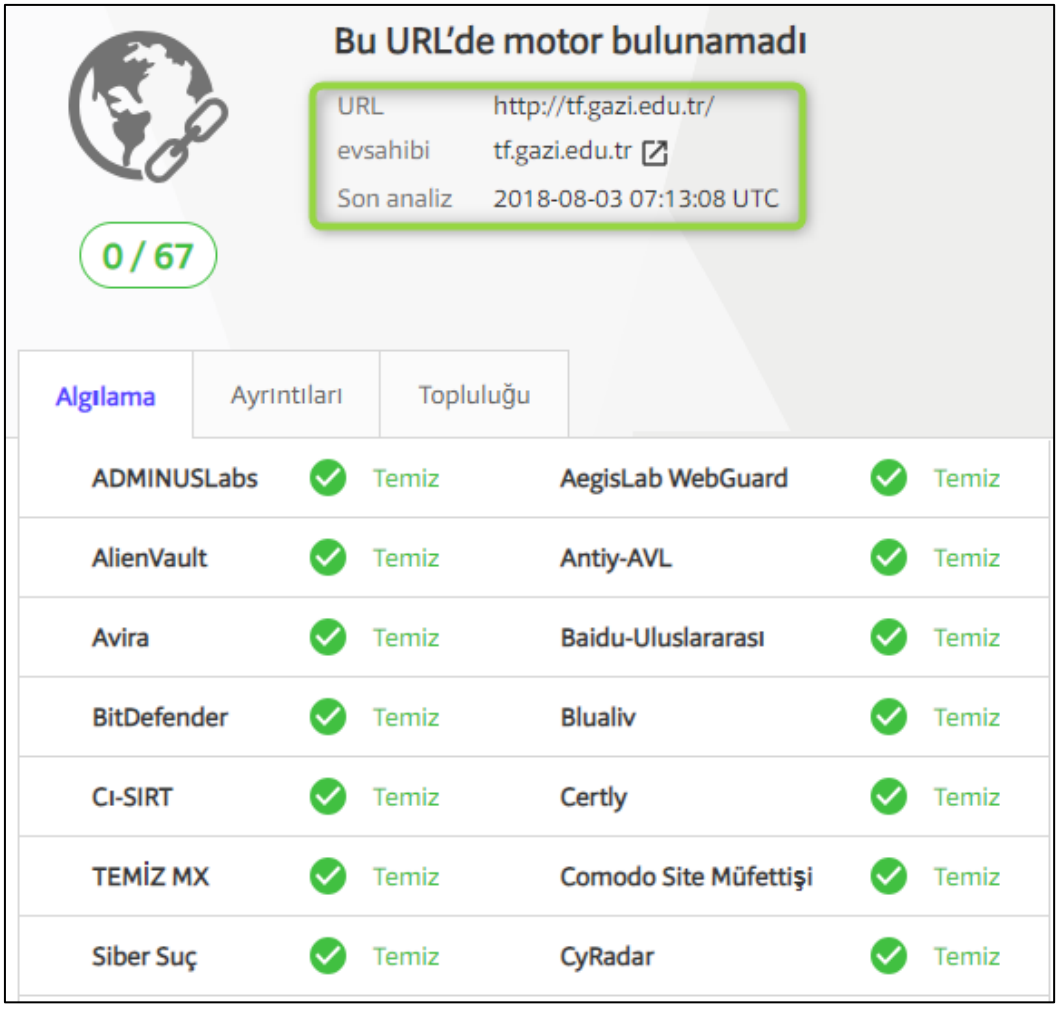
Önerilen model üç bileşenden oluşmaktadır. Bu bileşenler; Link analizi, Makine öğrenmesi ve Metin analizi olarak sıralanmaktadır. Her bir yöntemin sahip olduğu farklı öznitelikler bulunmaktadır. Bu yöntemlerin, modelde uygulanabilmeleri için ihtiyaç duydukları ayarlamalardan aşağıda bahsedilmektedir.

3.2.1. Kısa link analizi yöntemi

VirüsTotal veri kümesi, kötücül olduğundan şüphelenilen dosya ya da bağlantıları bir dünya virüs programına taratmak yerine çevrimiçi olarak analiz edilebildiği bir servistir [71]. Gönderilen linkler ya da dosyalar microsoft, symantec, kaspersky, mcafee gibi çok büyük firmaların yazılımları ile taranmaktadır. EK-8 içerisinde VirusTotal üzerinden oluşturulan API anahtarı bilgisi gösterilmektedir. VirusTotal, literatürde kendisine çok geniş yer bulan spam analiz veri kümelerini bir arada barındırmaktadır [71]. Spam ve kötücül link analizlerinde kullanılan büyük veri kümelerinin birkaç tanesi şu şekildedir; Google Safebrowsing [72, 73, 74], Kaspersky [75], OpenPhish [73], Comodo Site Inspector [76],

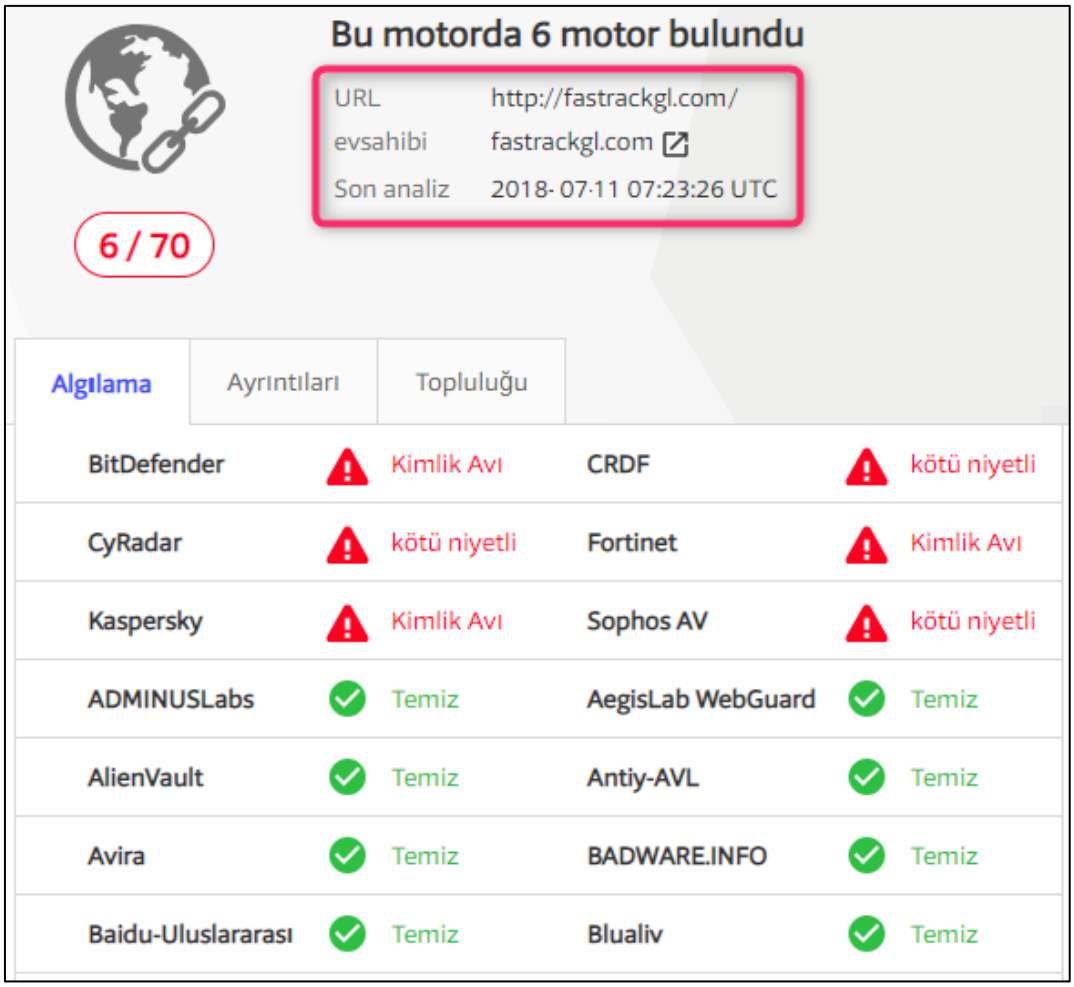
Forcepoint ThreatSeeker [77], Opera ve Yandex Safebrowsing [74, 78].

Aşağıdaki Şekil 3.4 ve Şekil 3.5‘de, VirusTotal sistemiyle yapılmış bir link analizi gözükmektedir.



Şekil 3.4. Kötücül olmayan link [79]

Şekilde 3.4’de VirusTotal sitesinde analiz edilen bir link için verilen doğrulama gösterilmektedir. Analiz edilen link için kötücül bir değer çıkmamaktadır. Bu tarz linkler masum link olarak kabul edilmektelerdir [71, 79]. VirusTotal bünyesinde bulundurduğu güvenlik veri kümeleri sayesinde, kendisine yönelndirilen bağlantıları test edebilmektedir. Atmıştan fazla güvenlik firmasının veri kümesini aynı anda çalıştırmaktadır [72, 73]. Dinamik bir yapısı bulunmaktadır. Kendisini sinamik yapısı syesinde sürekli güncellemektedir. Destek aldığı güvenlik firmalarında kötücül olarak belirtilen bir bağlantı aynı anda VirusTotal sisemindede kötücül olarka etiketlenmektedir [74]. VirusTotal’de yapılan diğer link analizi Şekil 3.5’teki gibidir.



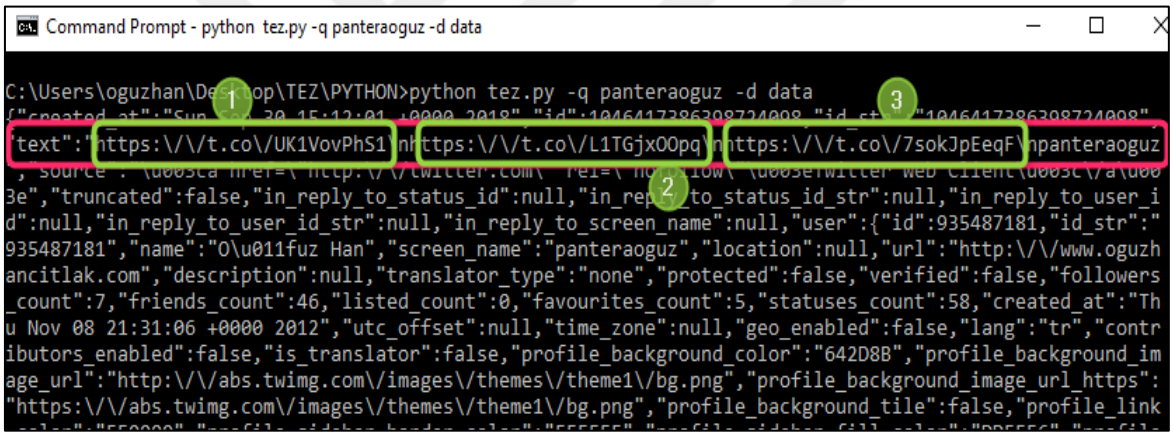
Şekil 3.5. Kötücül olan link [80]

Şekilde 3.5'de analiz edilen link kötücül olarak bulunmuştur. Altı adet farklı veri kümesinin sonuçları gözükmektedir. Önerilen modelin link analizi bölümünde VirusTotal veri seti kullanılmaktadır. Öncelikle bir API hesabına sahip olmak gerekmektedir. API hesabımız oguzhan.citlak@gmail.com adresi üzerinden oluşturulmuştur. Link analizi modelinde, VirusTotal veri kümesi havuzunun kullanılabilmesi için bir API Key'i yaratılmıştır [EK-8]. Kısa link analizi yönteminde kullanılmak için geliştirilen uygulamanın yazılım script leri VirusTotal web sitesinin sayfasından alınmaktadır [81].

Twitter üzerinden iletilen bağlantılar da dâhil olmak üzere direk ileti mesajları otomatik olarak işlenir ve http://t.co/xXxXxXxXxX şeklinde kısaltılmaktadır [82]. Uzun bağlantıların iletilerde paylaşabilmesi sağlamaktadır. İletilen mesaj için daha önceden belirtilen maksimum karakter sayısının geçilmemesine faydası bulunmaktadır ve kısaltılan bağlantılara destek sağlamaktadır. Şekil 3.6'da @panteraoguz kullanıcısı tarafından atılan bir ileti içerisindeki örnek linkler gözükmektedir [83].

Şekil 3.6. Twitter'da bir ileti içerisindeki linkler [83]

Şekil 3.6'da iletilen linklerin listesi gözükmektedir. Bu iletide linkler masun bir paylaşım şeklindedir. Şekil 3.7'de ise iletilen bu linklerin Twitter sistemindeki kısa linke dönüşmüş halleri gözükmektedir.



Şekil 3.7. Twitter formatındaki linkler

Şekil 3.7'de, Python yazılımı aracılığı ile iletilen linklerin ham veri olarak nasıl işlendiği gözükmektedir. Şekil 3.7'deki verinin elde edilmesi için bir kod parçacığı çalıştırılmıştır. Çalıştırılan *tez.py* dosyasının içeriği EK-3 içerisinde paylaşılmıştır. Bu ham veri içerisinden, iletilen linklerin hepsinde *t.co* formatlı uzantıları gözükmektedir [82]. Yapılan incelemeye göre, Twitter kullanıcılarının iletilerinde link paylaşmaları durumunda, link ya da linkler *t.co* formatına dönüştürülüp işlenmektedir.

Twitter, *t.co* uzantılı linkler hakkında çok fazla ayrıntıya girmektedir ancak yukarıdaki bilgi şuan için yeterli olmaktadır [82]. Aşağıdaki Çizelge 3.2'de Twitter üzerinden iletilen uzun linklerin işlenerek kısa linklere dönüştürülmüş halleri listelenmektedir.

Çizelge 3.2. Uzun linklerin Twitter‘daki kısa link halleri

Uzun Link	Twitter Link
http://www.gazi.edu.tr	https://t.co/UK1VovPhS1
http://fbe.gazi.edu.tr	https://t.co/L1TGjxOOpq
http://tf.gazi.edu.tr	https://t.co/7sokJpEeqF

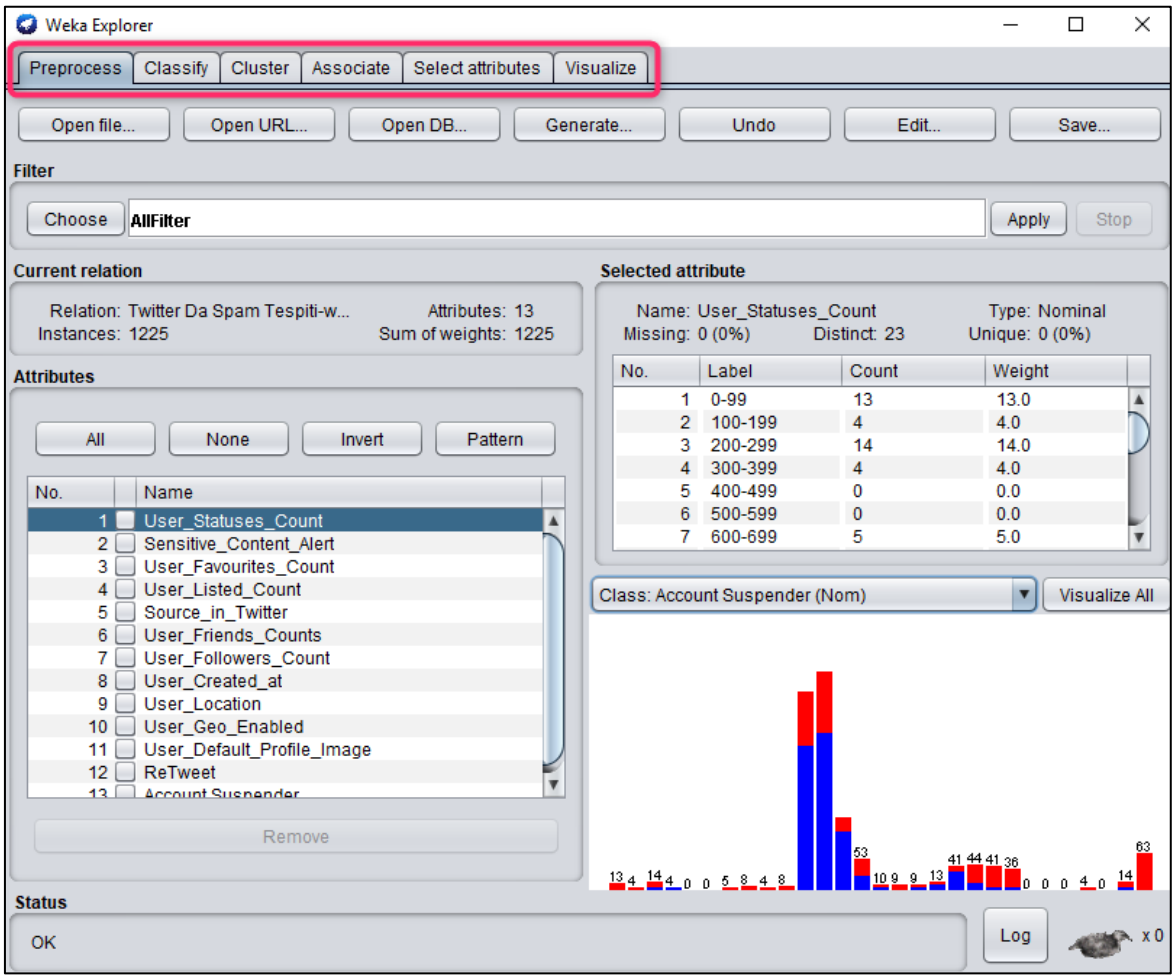
Çizelge 3.2’de bir ileti içerisindeki uzun linklerin Twitter sisteminde işlenerek dönüştükleri kısa linkler gösterilmektedir. Twitter, iletilerdeki uzun linkleri kendi sistemine özel bir şifreleme algoritması ile yeniden oluşturmaktadır. Bu şifreleme algoritmasında 10 karakter bulunmaktadır ve içerisinde rakam, büyük ve küçük harfler haricinde başka karakter bulunmaktadır [82].

3.2.2. Makine öğrenmesi yöntemi

Makine öğrenimi elinde bulundurduğu verilerde benzer motifleri açığa çıkartmak için benzer dizilere sahip algoritmaların uygulamasıdır [49]. Ancak makine öğrenimi öğrendiği bilgiler ile ilgili programın davranışını değiştirebilmektedir. Makine öğrenmesi, sunulan parametreler ve veriler ile benzerlikler kurmaktadır. İyi tespitlerde bulunabilmekte ve kendi kendini eğitebilen gelişmiş sistemlere sahip olmaktadır. Çözülme metotlarının makine tarafından analiz edildiği bir sınıflandırma yöntemidir. İstatiksel ve matematiksel yöntemler kullanılarak elde bulunan verilerden çıkarımlar yapmaktadır ve bu çıkarımlardan bilinmeyeni ön gören tahminlerde bulunmaya yarayan bir yöntem paradigması olarak da adlandırılabilmektedir [4].

Makine öğrenmesi yapay zekânın çok geniş çapta bir alt alanı olaraktan tanımlanabilmektedir. İnsan müdahalesi olmadan girilen verilerden deneyimler geliştirilmektedir. Hafızada saklanan eğitim kümesi ile yeni kümeyi karşılaştırarak ona bir sınıf etiket verilmesidir. Makine öğrenmesi, çok büyük veri kümelerini çok sayıdaki öznitelikleri ile birlikte yorumlayabilmektedir. Büyük veri kümelerinin yorumlanabilmesi için uygun denklemler ve fonksiyonların bulunmadığı zaman makine öğrenmesi yöntemi bu işi yapmaktadır [84]. Çalışma denklemlerinin oluşturulabilmesi için makine öğrenmesi yöntemi seçilmektedir. Weka, açık kaynak kodludur ve çok sayıda öğrenme algoritmasını

içerisinde bulundurmaktadır. Sonuçlarının detaylı analiz edilmesine imkân vermektedir. Java tabanlı bir veri madenciliği aracı olaraktan tanımlanabilmektedir [26]. Weka verileri basit bir dosya üzerinden okumaktadır ve bu okunan verilerin üzerinde bulunan değişkenlerin sayısal ya da yazılı değerlerden oluştuğunu kabul etmektedir [85]. Kullanılan verilerin bir dosya kümesi olması durumunda database üzerinden de veri çekebilmektedir. İstatistik ve makine öğrenmesi ile ilgili birçok kütüphane Weka üzerinde kütüphanede hazır olarak gelmektedir. Aşağıda göstrilen Şekil 3.8'de Weka üzerindeki makine öğrenmesi kütüphaneleri gösterilmektedir.



Şekil 3.8. Weka üzerindeki makine öğrenmesi kütüphaneleri

Şekil 3.8'de, Ön işleme paneli (Preprocess), Sınıflandırıcı paneli (Classify), Klasör paneli (Cluster), Birleştirme paneli (Associate), Seçim özellikleri paneli (Select Attributes) ve Görsel (Visualize) panelleri görülmektedir. Veri kümesinde 1 225 tane örnek (Instance) ve 13 tane öznitelikler (Attributes) bulunmaktadır. Kullanılan eğitim kümesinin sahip olduğu parametreler ve ölçüleme aralığı Çizelge 3.3'de detaylı olarak verilmektedir.

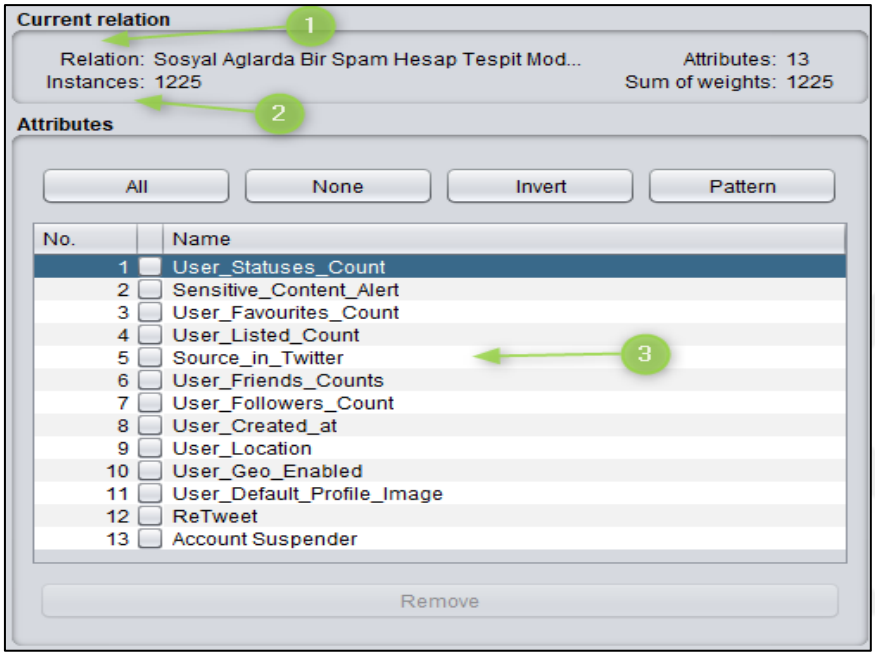
Çizelge 3.3. Weka eğitim dosyamızın parametreleri ve örnek spam hesap detayı

	Kullanılan Özellikler	Ölçütleme Aralığı	Örnek Hesap
1	Kullanıcı Kodu (Id)	-	"852624385819131904"
2	Kullanıcı Görüntü Adı (screen_name)	-	"donnelllebla41"
3	İleti (text)	-	"#bedava https://t.co/8WapCjKocF"
4	Kullanıcı Adı (name)	-	"Alison Scott"
5	İleti Sayısı (User_Statuses_Count)	0-99,100-199,...,1000000-1999999	"2356"
6	Hassas İçerik Uyarısı (Sensitive_Content_Alert)	DOĞRU / YANLIŞ (TRUE/FALSE)	"YANLIŞ"
7	Favori Eklenme Sayısı (User_Favourites_Count)	0-9,10-19,20-29,…,100000-1999999	"0"
8	Listelenme Sayısı (User_Listed_Count)	0-9,10-19,20-29,…,900-999	"0"
9	Hesap Kaynağı (Source_in_Twitter)	EVET / HAYIR (YES / NO)	"twitter.com"
10	Arkadaş Sayısı (User_Friends_Counts)	0-9,10-19,20-29,…,1000-99999	"18"
11	Takipçi Sayısı (User_Followers_Count)	0-9,10-19,20-29,…,100000-1999999	"18"
12	Oluşturulma Tarihi (User_Created_at)	2006-2008,…,2015-2017	"Salı Ağustos 25 16:36:39 +0000 2011"
13	Konum (User_Location)	EVET / HAYIR (YES / NO)	"HAYIR"
14	Konum Erişim (User_Geo_Enabled)	DOĞRU / YANLIŞ (TRUE/FALSE)	"YANLIŞ"
15	Standart Profil (User_Default_Profile_Image)	DOĞRU / YANLIŞ (TRUE/FALSE)	"YANLIŞ"
16	RT (ReTweet)	DOĞRU / YANLIŞ (TRUE/FALSE)	"YANLIŞ"
17	Hesap Askıya Alındı (Account Suspender)	DOĞRU / YANLIŞ (TRUE/FALSE)	"DOĞRU"

Çizelge 3.3'de Weka'da kullanılan eğitim kümesinin özelliklerini ve bu özelliklere karşılık gelen spam bir hesabın özellikleri gösterilmiştir. Kullanılan eğitim kümesinin sahip olduğu parametreler ve ölçüleme aralığı da detaylı olarak verilmektedir.

Öğrenmeye dayalı spam hesap tespit modelinin makine öğrenmesi bölümünde Weka

yazılımı kullanılmaktadır. Weka'ya öğrenme yapılabilmesi için bir veri girişinin olması gerekmektedir [86]. Makine öğrenmesinin gerçekleşebilmesi için öğrenmeyi sağlayan bir veri kümesine ihtiyaç duyulmaktadır. Weka yazılımında kullanılan veri kümesinin dosya uzantı formatı Arff (attribute relation file format) şeklindedir. Weka'da, Arff formatındaki dosya üzerinden öğrenme işlemi yapılmaktadır.

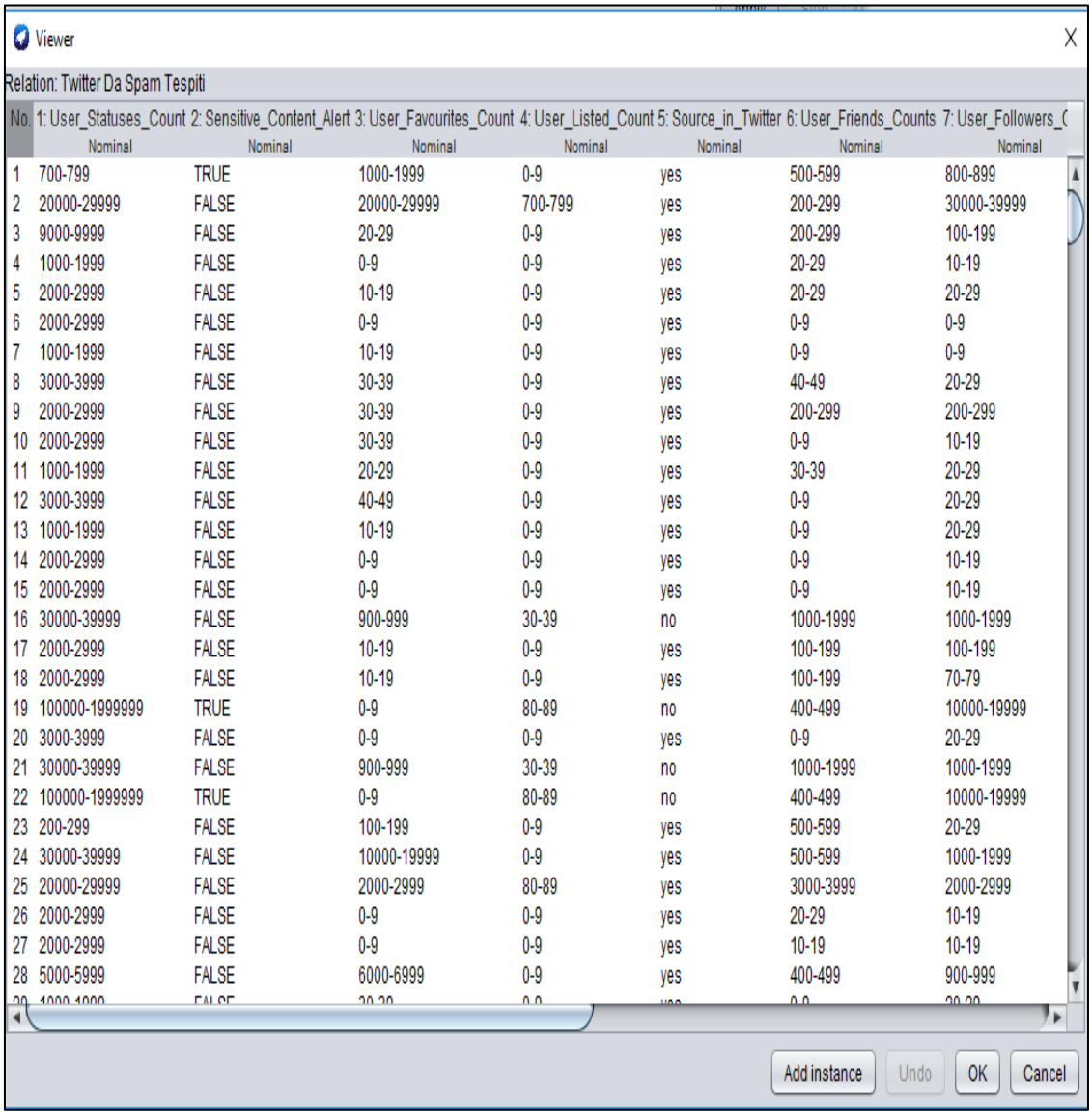


Şekil 3.9. Kullanılan Arff dosyasının özellikleri

Şekilde 3.9'da, Arff dosyasının Weka'da çalıştırdıktan sonra alınan değerleri gözükmektedir. Arff uzantılı veri kümesinin bir takım öznitelikleri bulunmaktadır. Bunlar; @relation, @attribute, ve @data özellikleridir. Bu özellikler, veri kümesini içerisinde bulunduran Arff uzantılı dosyada gösterilmektedir. Weka öğrenmesini bu dosya aracılığı ile yapmaktadır. Arff uzantılı dosyalar, Weka yazılımı için özel olarak waikato üniversitesi tarafından oluşturulmuştur [26, 87]. Arff uzantılı veri kümesinde bulunan özellikler şu şekildedir; @relation dosyanın ilk satırında dosyadaki ilişki tipini göstermektedir. @attribute ikinci satırdan başlayarak veri kümesinin özelliklerini göstermektedir. @data ile veri kümesi ve veri kümesindeki her satır bir örneği gösterecek şekilde sıralanmaktadır. Virgül ayıracı veri kümesinde bulunan herbir örneğinin her özelliği arasında kullanılmaktadır.

Kullanılan veri kümesi dosyasının formatı Arff'dir ve içeriği EK-9'da gösterilmektedir. Şekil 3.9'da belirtilen 1 numaralı kısmında relation karşılığı, 2 numaralı kısımda data

içerisindeki her bir özellik Instances sayısını ve 3 ile işaretlenen bölümde ise kullanılan attribute ları görülebilmektedir.



No.	1: User_Statuses_Count	2: Sensitive_Content_Alert	3: User_Favourites_Count	4: User_Listed_Count	5: Source_in_Twitter	6: User_Friends_Counts	7: User_Followers_(
	Nominal	Nominal	Nominal	Nominal	Nominal	Nominal	Nominal
1	700-799	TRUE	1000-1999	0-9	yes	500-599	800-899
2	20000-29999	FALSE	20000-29999	700-799	yes	200-299	30000-39999
3	9000-9999	FALSE	20-29	0-9	yes	200-299	100-199
4	1000-1999	FALSE	0-9	0-9	yes	20-29	10-19
5	2000-2999	FALSE	10-19	0-9	yes	20-29	20-29
6	2000-2999	FALSE	0-9	0-9	yes	0-9	0-9
7	1000-1999	FALSE	10-19	0-9	yes	0-9	0-9
8	3000-3999	FALSE	30-39	0-9	yes	40-49	20-29
9	2000-2999	FALSE	30-39	0-9	yes	200-299	200-299
10	2000-2999	FALSE	30-39	0-9	yes	0-9	10-19
11	1000-1999	FALSE	20-29	0-9	yes	30-39	20-29
12	3000-3999	FALSE	40-49	0-9	yes	0-9	20-29
13	1000-1999	FALSE	10-19	0-9	yes	0-9	20-29
14	2000-2999	FALSE	0-9	0-9	yes	0-9	10-19
15	2000-2999	FALSE	0-9	0-9	yes	0-9	10-19
16	30000-39999	FALSE	900-999	30-39	no	1000-1999	1000-1999
17	2000-2999	FALSE	10-19	0-9	yes	100-199	100-199
18	2000-2999	FALSE	10-19	0-9	yes	100-199	70-79
19	100000-1999999	TRUE	0-9	80-89	no	400-499	10000-19999
20	3000-3999	FALSE	0-9	0-9	yes	0-9	20-29
21	30000-39999	FALSE	900-999	30-39	no	1000-1999	1000-1999
22	100000-1999999	TRUE	0-9	80-89	no	400-499	10000-19999
23	200-299	FALSE	100-199	0-9	yes	500-599	20-29
24	30000-39999	FALSE	10000-19999	0-9	yes	500-599	1000-1999
25	20000-29999	FALSE	2000-2999	80-89	yes	3000-3999	2000-2999
26	2000-2999	FALSE	0-9	0-9	yes	20-29	10-19
27	2000-2999	FALSE	0-9	0-9	yes	10-19	10-19
28	5000-5999	FALSE	6000-6999	0-9	yes	400-499	900-999

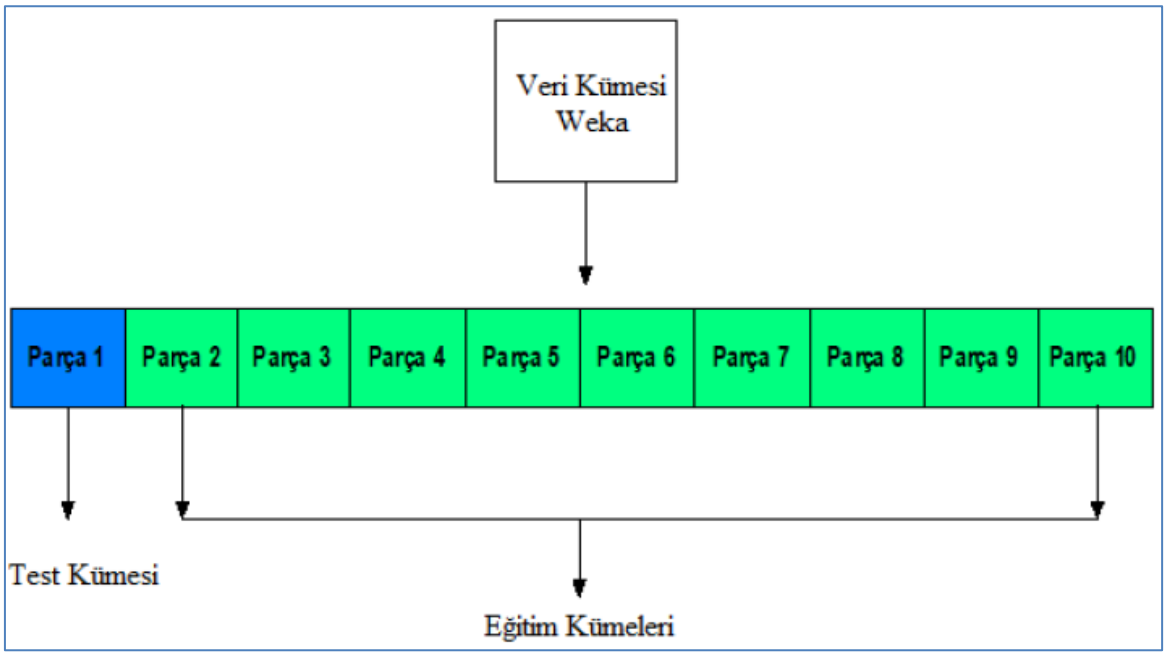
Şekil 3.10 Arff'de tanımlı değişkenlerin görüntüsü

Kullanılan Arff uzantılı veri kümesinin içerisindeki değerlerin bir kısmını şekil içerisinde gösterilmektedir. Şekilde veri kümesinde kullanılan attribute lerin metrik aralıkları bulunmaktadır. Örneğin "User_Status_Count" attributesi ilk satırında 700-799 aralığındadır. Önerilen yöntemde kullanılan Twitter kullanıcısın iletilerindeki toplam Tweet sayısını bu metrik aralığı göstermektedir. Başka bir örnekte ise "User_Listed_Count" attributesi aralığı 0-9 olarak verilmektedir. Bu aralık ise yine analiz edilen Twitter kullanıcısının sahip olduğu arkadaş listesini göstermektedir. Weka makine öğrenmesi yönteminde kullanılmak için

oluşturulan Arff dosyasında 23 650 tane kelime bulunmaktadır. Bu dosya toplam 1 276 satıra sahiptir ve boşlukları düştükten sonra 129 132 tane karakter bulunmaktadır. Dosyanın büyüklüğü ise 130 410 byte'tır.Aşağıda, Weka makine öğrenmesi yönteminde kullanılmış olan parametreler, algoritmalar ve ölçüm değerlerinden bahsedilmektedir.

K katmanlı çapraz doğrulama

Makine öğrenmesi yöntemlerinden birisi olan Weka'da K Katmanlı Çapraz Doğrulama (Cross-validation) seçeneğinin değeri birçok çalışmada 10 olarak alınarak kullanılmaktadır [88]. Weka'da yapılan çalışmalarda uygulanan yöntemin başarısını sınamak için veri kümesi, eğitim ve test kümeleri olarak ikiye ayrılmaktadır. Veri kümesi başarısının sınanması için %66'lık eğitim, %33'lık kısmı ise test kümesi olarak ayırmak, sistem eğitildikten sonra test kümesi ile başarının sınanması başka bir yöntem olaraktan kullanılabilmektedir [89]. Ancak, yaptığımız literatür araştırmalarında K katmanlı çapraz doğrulama yöntemini baz alıp K değerini 10 olarak kullanmak veri kümemiz için daha uygun olmaktadır [88]. K 10 için veri kümemiz öncelikli olarak 10 eşit parçaya bölünmektedir. Veri kümesinin tamamı K değerimiz kadar parçaya bölündükten sonra K katmanlı çapraz doğrulama sistemi çalışmaya başlamaktadır. Öncelikle 10 parçamızdan bir tanesi rastgele seçilip geri kalanı eğitim için kullanılmıştır. Hangi parçadan başlanıldığının burada hiçbir önemi yoktur.



Şekil 3.11. K katmanlı çapraz doğrulama [89]

Şekil 3.11'de Veri kümesinin K 10 için 10 eşit parçaya bölümlenmesi gözükmektedir. Rastgele olarak, Parça 1 Test kümesi olarak seçilmiştir ve kalan parçalar ise eğitim kümesi olarak yer almaktadır. K katmanlı çapraz doğrulama işleminde Parça 1 test kümesi olduğunu varsayıldığında bir sınıflandırma algoritması çalıştırıp sonucumuz S1 olarak bulunmaktadır. Parça 2 Test kümesi olduğu zaman ise bulunan sonuç S2 ve en son Parça 10 test kümesi seçildiği zaman bulunan sonucun ise S10 olduğunu söylenebilmektedir. Sonuç olarak 10 kere aynı yöntemi 10 adet farklı eğitim ve test kümelerinde uygulamış bulunmaktadır. Önerilen spam tespit modelinin başarımı ya da genel hata oranı ise elde edilen bu 10 sonucun ortalaması olarak hesaplanmaktadır. Bütün bu test ve eğitim kümelerinin hesaplamalarının sonucunda ortalama alındığı için hangi parçadan başlanıldığının hiçbir önemi bulunmamaktadır. K katmanlı çapraz doğrulama yönteminde S1,S2,…,S10 sonuçlarının hesaplamasında Eşitlik 3.1'deki formül kullanılmaktadır.

$$t_i \in VK \text{ olmak üzere, } \text{Sonuç} = \frac{\sum_{i=0}^{k} SF\,(ti,\ VK - ti)}{k} \tag{3.1}$$

Bu formülde, "t" veri kümesi içerisinden seçilen her bir test kümesini, "k" kaç parça katlama kullanılmakta olduğunu, "VK" veri kümesini ve "SF" sınıflandırma fonksiyonunu göstermektedir. K katmanlı çapraz doğrulama yöntemi için K – kere yöntem çalıştırılmaktadır. Her bir adımda veri kümesinin 1/K kadar test kümesi olarak kullanılmakta iken geri kalan kısım ise eğitim kümesi olarak kullanılmaktadır. Bu yöntemde bütün sınıflandırma fonksiyon performanslarının toplamının K sayısına bölünerek ortalamasının alınması ile bulunmaktadır.

NaiveBayes algoritması

NaiveBayes sınıflandırıcı algoritmasında, eğitim veri kümelerinin ve sınıflandırılması gereken diğer kümelerin hangi sınıflara ait oldukları bellidir [90]. Mevcutta bulunan en kısıtlayıcı sınıflandırma tekniklerinden bir tanesidir [91]. Metin sınıflandırılmasında çok başarılı olduğu kanıtlanmış bir algoritmadır [91]. NaiveBayes algoritması bir sonuç değişkeninden oluşmaktadır. Değerlendirilen sınıflandırma veri kümesi ise birçok özellikten oluşmaktadır. NaiveBayes algoritması Eşitlik 3.2'deki gibi formüle edilmektedir.

$$p(C|F_1,...F_n) = \frac{P(C)p(F_1,...F_n|C)}{p(F_1,...F_n)} \quad (3.2)$$

Verilen formülde, F eğitim kümesindeki özellikleri göstermektedir. C verilen hedefi temsil etmektedir. NaiveBayes sınıflandırıcısı, bütün koşullu olasılıkların çarpımı olarak kabul edilmektedir [90].

Random Forest algoritması

Random Forest algoritmasında birden fazla karar ağacı kullanılmaktadır [92]. Bu öğrenme algoritması, çok değişkenli karar ağaçlarının her birisinin farklı eğitim kümeleri ile eğitilmesi sonucunda ortaya çıkan sonuçların bir araya getirilmesinin önermektedir [93]. Random Forest öğrenme algoritması birden çok sınıflandırıcıyı üretmektedir. Sonrasında her bir sınıflandırıcıların tahminlerinden gelen dönütler ile yeni veri kümesini sınıflandırabilmektedir [94]. Hata dengeleme yöntemine sahip olmasından dolayı büyük veri kümelerinde sorunsuz olarak çalışmaktadır. Random Forest öğrenme algoritması topluluk yöntemlerindeki özelliklerin iyileştirilmiş halini içermektedir. Rastgele örnekleme yapması nedeniyle daha iyi genellemeler sunabilmektedir ve geçerli tahminlerde bulunabilmektedir [93]. İsabetli tahminlerde bulunabilmek için daha uyumları üretebilmektedir.

IBk algoritması

IBk algoritmasının diğer adı (K nearest neighnorhood-KNN) algoritmasıdır [94]. Veri kümesi sınıflandırılmasında çok yaygın olarak kullanılan bir algoritmadır. Sınıflandırma esnasında çıkartılan özelliklere göre, yeni sınıflandırılacak bireyin eski sınıflandırılan bireylere K tanesi yakınlığı olarak açıklanabilmektedir [94]. Örneğin, K = 5 seçildiğinde, yeni sınıflandırılacak eleman, eski sınıflandırılan elemanlardan kendisine en yakın beş tanesinin sınıfını almaktadır. Yeni sınıflandırılacak eleman, daha öncesinden sınıflandırılmış beş eleman hangi sınıfa dâhil ise o sınıfa dahil olmaktadır [95].

J48 karar ağacı algoritması

J48 karar ağacı algoritması, makine öğrenmesi yöntemlerinde kullanılan ve çok başarılı sonuçlar veren popüler bir algoritmasıdır [90]. Her karar ağacı satırında bir düğüm noktası

kullanmaktadır. Alt satırlar, üst satırların düğüm noktalarından oluşmaktadır [96]. Üzerinde çalışılan veri kümesindeki olaylar, her karar ağacı satırındaki düğümler olarak düşünülebilmektedir. Sınıflandırma algoritmaları içerisinde Random Forest algoritması gibi çok hızlı çalışmaktadır ve yüksek seviyede doğru sonuçlar vermektedir [97].

J48 algoritması, veri kümesindeki mevcut örnekler ile başlamaktadır. Yeni karar ağacı veri yapısını, yeni durumları sınıflandırabilmek için yaratmaktadır. Karar ağacının her satırındaki her bir düğüm test içermektedir, bu test düğümden sonra hangi dalın takip edileceğini belirlemek için kullanmaktadır [96, 97].

JRip algoritması

JRip, makine öğrenmesindeki en temel ve popüler algoritmalardan bir tanesidir [98]. Veri kümesinin bütün üyelerini kapsayan bir dizi kural belirleyerek ilerleyen bir algoritmadır. Belirlediği bu kuralda, veri kümesindeki eğitim verilerinin kararlarını ve tüm örneklerini bir sınıf olarak ele almaktadır. Ardında bir sonraki sınıfa geçmektedir ve aynı işlemi burada da yapmaktadır, tüm sınıfları bulana kadar bu işlemi tekrar etmektedir [99].

Doğruluk hata oranı

Doğruluk hata oranı (Accuracy-Error Rate), kullanılan modelin başarımının ölçülmesinde önemli bir yer almaktadır. Bu hata oranı işleme alınan veri kümesinin, en basit ve aynı zamanda en popüler olarak modele ait doğruluk oranını vermektedir [100]. Doğruluk hata oranını Eşitlik 3.3'deki formüle göre hesaplanmaktadır.

$$\text{Doğruluk} = \frac{(TP+TN)}{(TP+FP+FN+TN)} \tag{3.3}$$

Doğruluk hata oranının hesaplanmasında, Doğru Onaylanmış (True Pozitif-TP) ile Doğru Onaylanmamış (True Negatif-TN)'in toplamının, Doğru Onaylanmış (True Pozitif-TP),Yanlış Onaylanmış (False Pozitif-FP), Yanlış Onaylanmamış (False Negatif-FN) ve Doğru Onaylanmamış (True Negatif-TN) oranlarının toplamına oranıdır şeklinde hesaplanmaktadır [100]. Doğruluk, doğru olarak sınıflandırılmış örnek sayısının diğer bütün örnek sayılarına oranını göstermektedir.

Hata oranının formüle edilmiş gösterimi Eşitlik 3.4'de gösterilmiştir.

$$\text{Hata Oranı} = \frac{(FP+FN)}{(TP+FP+FN+TN)} \tag{3.4}$$

Hata oranı ise doğruluk oranının 1'e tamamlayanıdır. Yanlış ifadeyle sınıflandırılmış örnek sayısının, diğer bütün örnek sayılarına oranı olarak hesaplanmaktadır [100].

Kesinlik

Kesinlik (Precision), sınıfı 1 olarak tahmin edilen Doğru Onaylanmış (True Pozitif-TP) örnek sayısının, sınıfı 1 olarak tahmin edilen bütün örnek sayılarına Doğru Onaylanmış (True Pozitif-TP) ve Yanlış Onaylanmış (False Pozitif-FP) in toplamına oranıdır şeklinde hesaplanmaktadır [85]. Kesinlik hesaplamasının formüle edilmiş hali Eşitlil 3.5'te gösterilmektedir.

$$\text{Kesinlik} = \frac{TP}{(TP+FP)} \tag{3.5}$$

Kesinlik, doğru olarak tahmin edilenlerin, toplam doğru olarak tahmin edilenlere oranıdır. Kesinlik ile getirilen verinin ne kadarı istenilen veri ile alakalı olduğu anlaşılmaktadır [85].

Hassasiyet

Hassasiyet (Recall), getirilmesi gereken bilginin ne kadarının getirildiği hakkında bilgi vermektedir. Yapılan Hassasiyet hesaplamasının sonucu ile bu anlaşılmaktadır [101]. Hassasiyet hesaplamasında Eşitlik 3.6'daki formül kullanılmaktadır.

$$\text{Hassasiyet} = \frac{TP}{(TP+FN)} \tag{3.6}$$

Hassasiyet, doğru olarak sınıflandırılmış Doğru Onaylanmış (True Pozitif-TP) sayısının, toplam pozitif örnek sayısına oranıdır.

F- Ölçütü

F-Ölçütü (F-Measure), Kesinlik ve Hassasiyet değerlerinin harmoni ortalamasıdır [84]. Bu ölçütler tek başlarına anlamlı bir karşılaştırma sonucu çıkartılmasında yeterli olamamaktalardır. Daha doğru sonucu her iki ölçütü değerlendirmek ile alınabilmektedir. F-Ölçütü hesaplamasında Eşitlik 3.7'deki formül kullanılmaktadır.

$$F = \frac{2DK}{(D+K)} \tag{3.7}$$

Formülde, K ile Kesinlik, D ile de Hassasiyet gösterilmektedir.

Kappa istatistiği

Kappa istatistiği iki değerlendirici arasındaki uyuşmanın güvenilirliğini ölçen bir istatistik yöntemidir [84]. K sayısı -1 ile + 1 arasında değişmektedir. Kullanıldığı veri kümelerindeki tam uyum K'nin 1'e eşit olduğu zaman gerçekleşmektedir. Gözlemlenen uyumun şansa bağlı olan uyuma eşit olduğunda ya da şansa bağlı olan uyumdan daha büyük olması durumunda $K \geq 0$ olmaktadır. Gözlenen uyumun, şansa bağlı uyumundan daha küçük olması durumunda ise $K > 0$ olmaktadır [84]. Güvenilirlik açısında $K < 0$ değerlerinin hiçbir anlamı bulunmamaktadır. 0.4 üzerinde bulunan bir kappa skoru makul bir anlaşmayı ifade etmektedir [102]. Kappa istatistiği Eşitlik 3.8'deki formüle göre hesaplanmaktadır.

$$K = \frac{(P_0 - P_c)}{(1 - P_0)} \tag{3.8}$$

Formülde; P_0, kabul edilen oran ; P_c , ise kabul edilmesi beklenen oranı göstermektedir.

Sınıflandırma algoritmalarının karşılaştırılma kriterleri

Twitter üzerinden Twitter API aracılığı ile @oguzhancitlak hesabının ayarlarını tanımlanmıştır. Python yazılımını kullanılarak 1 Gigabayta yakın bir Twitter veri kümesi Nisan 2017 de çekilmiştir. Çekilen veri kümesinden 1 225 tane spam hesap ayıklanmıştır. Ayıklama esnasında bu spam kullanıcı hesapları etiketlenmiştir. Veri kümesinden ayıklanan

spam kullanıcı hesapların özellikleri kullanılarak Weka yazılımında kullanılacak eğitim ve test kümeleri oluşturulmuştur. Oluşturulan veri kümesinin değerlendirilmesinde bir takım algoritmalar kullanılmaktadır. Makine öğrenmesi yazılımlarında kullanılan ve literatürde sıklıkla bahsedilen beş algoritmayı ve bu algoritmaların ölçüm metriklerinde kullanılan özniteliklerinden yukarıda bahsedilmiştir. NaiveBayes, Random Forest, IBk, J48 ve JRip Algoritmaları ile veri kümesi dosyasının analiz işlemlerinde ihtiyaç duyulan metrikler Çizelge 3.4'te verilmektedir.

Çizelge 3.4. Weka'da değerlendirilen metriklerin listesi [26, 27, 84, 91, 92, 94-97, 104-106]

	Weka Özellik	Açıklama
1	Doğru Olarak Sınıflandırılan (Correctly Classified Instance)	Doğru Sınıflandırılan miktar
2	Yanlış Olarak Sınıflandırılan (Incorrectly Classified Instance)	Yanlış Sınıflandırılan miktar
3	Kappa İstatistiği	İki değerleyici arasındaki uyuşmanın güvenilirliğini ölçen bir istatistik yöntemidir. K sayısı -1 ile + 1 arasında değişir. Tam uyum K=1 olduğunda olur.
4	Ortalama Mutlak Hata (Mean Absolute Error)	Tahmindeki hataların ortalama büyüklüğünü ölçer.
5	Ortalama Hata Karekök (Root Mean Squared Error)	Sıfıra yakın değerler daha iyidir ve her zaman negatif değildir. Rastsallık sebebiyle daha doğru bir tahminde bulunabilecek bilgileri hesaba katmadığından kaynaklanır.
6	Göreli Mutlak Hata (Relative Absolute Error)	Deneysel değer ile gerçek değer arasındaki farktır. Bir ölçümdeki gerçek hata olaraktan bilinmektedir.
7	Göreli Hata Karekök (Root Relative Squared Error)	Göreli mutlak hata ile ölçülen değer arasındaki orandır.
8	Doğru Onaylanmış (True Positive-TP) Oran	Doğruya doğru demektir
9	Hatalı Onaylanmış (False Positive-FT) Oran	Yanlışa doğru demektir
10	Kesinlik (Precision)	Doğru var olarak tahmin edilenlerin, toplam var tahminlere oranıdır TP/(FP+TP). Getirilen verinin ne kadarı istenilen veri ile alakalıdır.
11	Hassasiyet (Recall)	Getirilmesi gereken bilginin ne kadarı getirilmiştir.
12	F-Ölçütü (F-Measure)	Precision ve Recall değerlerinin Harmoni ortalamasıdır.
13	Matthews Korelasyon Katsayısı (Matthews correlation coefficient-MCC)	Matthews Korelasyon Katsayısı (MCC) -1 ile 1 aralığına sahiptir, burada -1 tamamen yanlış bir ikili sınıflandırıcıyı belirtirken 1 tamamen doğru bir ikili sınıflandırıcıyı gösterir.
14	Alıcı İşletim Karakteristiği (Receiver Operating Characteristic-ROC) Alanı	ROC eğrilerinin yalnızca sınıflandırıcıların genel olarak nasıl performans gösterdiğine dair bir fikir verir
15	Hassas Hassasiyet (Precision Recall-PRC) Alanı	PRC, yalnızca sınıflandırıcının bir sınıfta nasıl davrandığıyla ilgilidir. Örneğin hastaları hastalıklı veya sağlıklı olarak sınıflandırma işlemini yaparken başarılıdır.
16	Sınıf (Class)	Sınıf Etiketi
17	Karışıklık Matrisi (Confusion Matrix)	Sınıflandırma algoritmasının performansını özetlemek için kullanılan bir tekniktir.

Çizelge 3.4'de belirtilen özellikler incelendiğinde bir veri kümesinin değerlendirilmesinde "Doğru Olarak Sınıflandırılan"," Kappa İstatistiği" ve "Doğru Onaylanmış (True Positive-TP) Oran" oranları önemli bir yer tutmaktadır. Belirtilen bu üç metriğin ölçümleri kullanıldığı modeldeki başarım sonuçlarını vermektedir. Kappa istatistiği -1 ve +1 arasında değişmektedir. Kappa istatistiği metriği 1'e eşit olduğu zaman kesin sonuç verilmektedir [84]. "Ortalama Hata Karekök" ve "Göreli Mutlak Hata" oranları bir değerlendirmede yüksek ise, bu metriklerin kullanıldığı modelin başarısı çok kötü denilebilmektedir [96, 103, 107]. Karışıklık Matrisi kullanıldığı modeldeki sınıflandırma algoritmasının performansını özetlemekte kullanılmaktadır [108].

3.2.3. Metin analizi yöntemi

Metin analizi, bir metnin veri kaynağı olarak kabul edilmesi mantığı üstüne kurulmuş veri madenciliği çalışmasıdır [108]. Yapısallaştırılmış veriyi metin üzerinden elde etmeyi amaçlamaktadır. Varlık ilişki modellemesi, metin özetleme, duygusal analiz, metinlerden konu çıkartılması, sınıf taneciklerinin üretilmesi, sınıflandırılma ve bölümleme gibi çalışmaları hedef almaktadır [109]. Metin analizinde, metin üzerinden istatistikel sonuçlara ulaşmak hedeflenmektedir. Python aracı ile Twitter havuzundan toplanan veri kümesinde *bet* hassas içerik kelimesi kullanılmaktadır. Metin analizi yönteminde, oluşturulan veri kümesi içerisindeki spam hesapları bulabilmek için hassas içerikli kelimeler kullanılmıştır. Kullandığımız hassas içerikli kelimelerin listesi EK-5'te belirtilmektedir. Bu hassas içerikli kelimelerin, kullanıcı iletilerinin içerisinde en az ikinci kez ya da daha fazla geçip geçmediklerine bakılmaktadır. Makine öğrenmesi sonucunda tespit edilen spam hesapların iletilerinde geçen hassas içerikli kelimeler, yapılan tahminin doğru olduğuna destek vermektedir. Aşağıda hassas içerik ile hesap arama komutu Eşitlik 3.9'de gösterilmekdir.

…\TEZ\DATASET\bet>python icerisinde_bet_gecen_datacek.py -q bet -d data (3.9)

Yukarıdaki komut ile yapılan aramanın sonucu EK-6 içerisinde paylaşılmaktadır. Analiz işleminde kullanılan sosyal ağ veri kümesi bu komut ile oluşturulmaktadır. Metin analizi işlemi, metin sınıflandırma veya analiz yöntemleri elde edilen metindeki aranan metnin istatistiksel analizini yapmaktır [108, 109].

Hassas içerikli kelimelerin belirlenmesinde, veri kümesindeki hesaplarda hassas içerikli

kelimelerin geçme sıklıkları hesaplanmaktadır. Twitter'ın askıya aldığı birkaç hesabın iletisinden geçen birkaç hassas içerikli ileti ve kelimeler Şekil 3.12'de gözükmektedir.

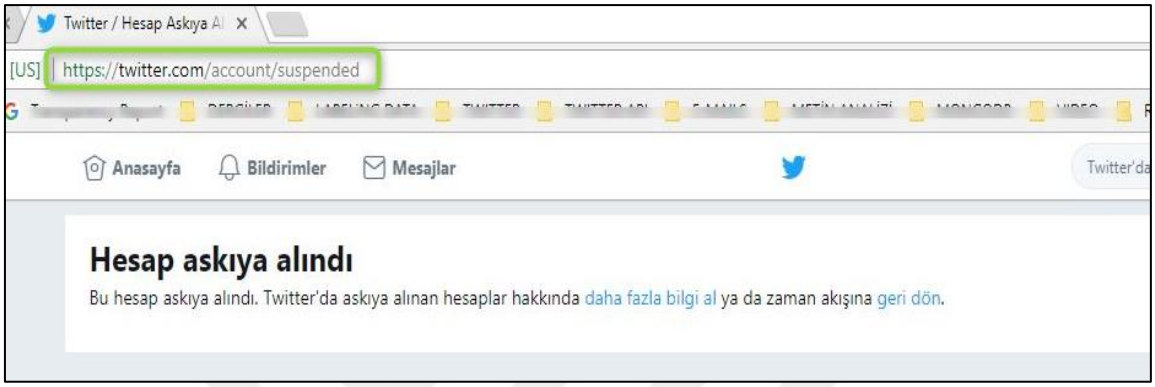
	A	B	C	D	E	F
1	İleti	#free	#naked	#teen	#bet	#sex
2	text:"sean michaels bet https:\/\/t.co\/GmyELvGCMM"	0	0	0	0	0
3	text:"#free bet 4u rocker bet https:\/\/t.co\/zDdaWZbbWA"	1	0	0	1	0
4	text:"simpsons animation bet https:\/\/t.co\/fZwEfajO4Q"	0	0	0	1	0
5	text:"#free hentei bet videos sexy fuck free video https:\/\/t.co\/umUzIVQ7yW"	1	0	0	1	1
6	text:"#home granny bet download full bet movie free https:\/\/t.co\/T7D41nUi9j"	1	0	0	1	0
7	text:"#adults with dementia oslo lesbian bet https:\/\/t.co\/MO8Ms8xFyL"	0	0	0	1	0
8	text:"#milk sexy skin head girl bet https:\/\/t.co\/fhXQWsbt75"	0	0	0	1	0
9	text:"#bet stars models naked torrents https:\/\/t.co\/UC3QSmhK79"	0	1	0	1	0
10	text:"#free nude webcha free onepiece bet https:\/\/t.co\/P6kfmlRwGw"	1	0	1	1	0
11	text:"#free mature sex online tube bet hairy https:\/\/t.co\/AwNxqgo37u"	1	0	0	1	1
12	text:"neuseeland teen bet https:\/\/t.co\/vSdEUzI9jO"	0	0	1	1	0
13	text:"full length mobile bet movies https:\/\/t.co\/jlJPuaDfnQ"	0	0	0	1	0
14	text:"#thong girls gallery message boards bet https:\/\/t.co\/a51eym5rIP"	0	0	0	1	0
15	text:"RT @betHardd: #Follow & #Retweet \"the new #erotic #bet\" betHardd	0	0	0	1	0

Şekil 3.12. İşlenen iletiler ve hassas içeriklerin bir bölümü

Şekil 3.12'de hassas içerikli birkaç kelimenin *text* içerisindeki görüntüsünü verilmektedir. EK-7 içerisinde ise bu listenin tamamı paylaşılmaktadır. Hassas içerik listesindeki kelimelerden bir tanesi *free* kelimesidir. Bu kelime, şekilde 3. ,5. ,6. ,10. ve 11. satırlardaki iletilerde geçtikleri görülmektedir. Sosyal ağ kullanıcısının iletisinde belirtilen hassas içerikli kelimelerden en az bir tane geçmesi bu hesabın spam özellikli olduğunu göstermektedir. Atılan iletiler içerisinde hassas içerik olmaması durumunda ise ilgili hesabın spam özellikte olmadığı metin analizi yöntemine göre söylenebilmektedir.

Veri kümesindeki hassas içeriklerin ve Twitter tarafından askıya alınan hesapların kontrolü için Kitle Kaynak Yöntemi kullanılmıştır. Kitle Kaynak Yönetimi, belirli bir projenin hayata geçirilebilmesi için internet üzerinde iletişim ağı içerisinde bulunan insanların bir araya gelmesidir [3, 47, 48]. Bir araya gelen insanlar, destek talebinde bulunulan projeyi bağımsız olarak incelemektelerdir ve bir sonuç çıkartmaktalardır [47, 48]. Varılan ortak sonuç, hayata geçirilmek istenen projenin etiketlenmiş bir parçası olarak kullanılmaktadır [3, 48]. Bu yöntem, ağ içerisindeki kişilerin bireysel çalışmaları sonucunda bir sorunun çözülmesini

sağlamaya yarayan ve ortak ürünler ortaya koyabilme işi olaraktan düşünülebilmektedir [3, 47]. Makine öğrenmesi yönteminde kullanılmak üzere, Twitter tarafından askıya alınmış hesapların kontrolünün yapılması gerekmektedir. Twitter tarafından askıya alınan ve kullanılan veri kümesinde bulunan hesapların belirlenmesinde kitle kaynak yönetimi ne başvurulmuştur.

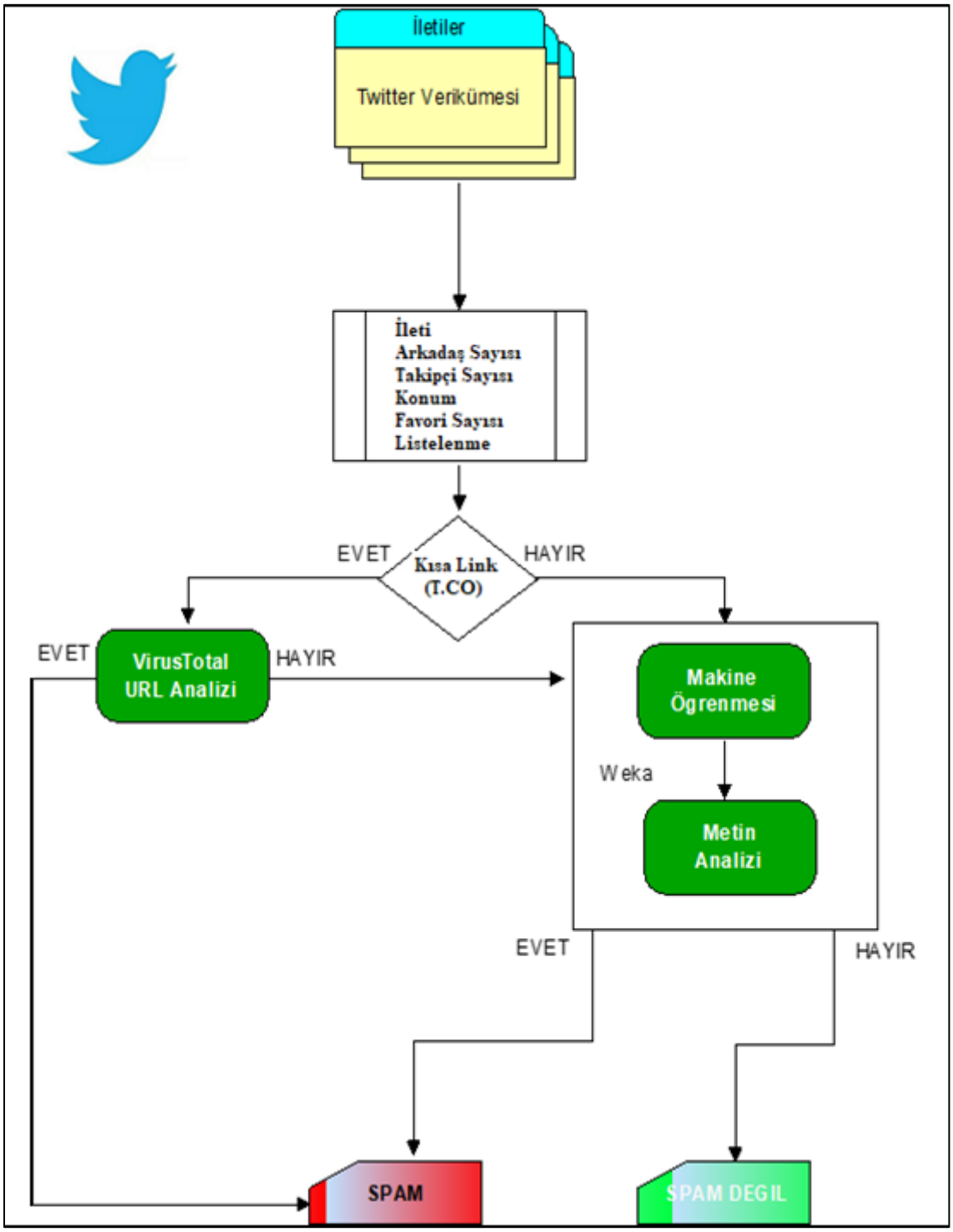


Şekil 3.13. Twitter'ın hesap askıya alındı uyarısı linki

Şekil 3.13'de, Twitter tarafından askıya alınmış bir hesabın ekran görüntüsü görülebilmektedir. Burada Twitter tarafından askıya alınma bağlantısı gözükmektedir. Twitter askıya aldığı her hesap için aynı linki paylaşmaktadır. Twitter üzerinden oluşturulan veri kümesinde 81 317 adet kullanıcı hesabının bilgisi bulunmaktadır. Veri kümesindeki bu hesapların 1 225 tanesinin özelliklerini oluşturulan Arff dosyası içerisinde kullanılmaktadır. Kullanılan Arff dosyası içerisinde spam ve spam olmayan hesaplar bulunmaktadır. Bu her iki hesabın kontrolünü yaparken kitle kaynak yöntemine başvurulmuştur. Oluşturulan veri kümesinin oluşturulma tarihi üzerinden bir yıl geçtikten sonra, aynı veri kümesi üzerinde kitle kaynak yöntemi kullanılmıştır. Veri kümesinden yeterli sayıda kullanıcı hesabının Twitter tarafından askıya alınıp alınmadıklarına bakılmıştır. Askıya alınan hesaplar için spam ve askıya alınmadan halen aktif olan hesaplar için ise spam değil şeklinde etiketlenme yapılmıştır. Makine öğrenmesi yönteminde eğitim veri kümesi olarak kullanılan Arff uzantılı dosyada spam ve spam olmayan hesapların özellikleri "Account Suspender" özniteliği altında gösterilmektedir. Weka'da kullanılan veri kümesindeki 1 225 hesabın tamamı bu yönteme göre etiketlenmektedir.

4. SOSYAL AĞLARA YÖNELİK ÖĞRENMEYE DAYALI BİR SPAM HESAP TESPİT MODELİ VE UYGULAMASI

Sosyal ağlara yönelik öğrenmeye dayalı bir spam hesap tespit modelinin uygulama algoritmasında oluşturulan modelde üç bileşen bulunmaktadır. Bu bileşenler; link analizi, makine öğrenmesi ve metin analizi.



Şekil 4.1. Uygulama çalışma algoritması

Şekil 4.1'de belirtilen Weka havuzunda, incelenecek olan hesap daha önce belirtilen Arff dosyasındaki özelliklere göre eğitilmiş bir makine öğrenmesi yöntemine maruz kalmaktadır. Makine öğrenmesinden çıkan hesaplar iletilerindeki hassas içeriklerin tespit edilmesi için metin analizi yöntemine girmektelerdir. Bu havuzda, Weka'da eğitilmiş veri kümesine göre spam ya da spam olmayan Twitter hesabı tespit edilmeye çalışılmıştır. Buna ek olarak incelenen hesabın hassas içerikli olup olmadığınada bakılmaktadır. Çizelge 4.1'de analiz edilen hesabın ihtiyaç duyulan öznitelik bilgileri bulunmaktadır.

Çizelge 4.1. Twitter Veri kümesinden ayıklanan hesap özellikleri

Sıra	Ayıklanan Özellik	Sıra	Ayıklanan Özellik
1	Kullanıcı Kodu (Id)	9	Hesap Kaynağı (Source_in_Twitter)
2	Kullanıcı Görüntü Adı (screen_name)	10	Arkadaş Sayısı (User_Friends_Counts)
3	İleti (text)	11	Takipçi Sayısı (User_Followers_Count)
4	Kullanıcı Adı (User_Name)	12	Oluşturulma Tarihi (User_Created_at)
5	İleti Sayısı (User_Statuses_Count)	13	Konum (User_Location)
6	Hassas İçerik Uyarısı (Sensitive_Content_Alert)	14	Konum Erişim (User_Geo_Enabled)
7	Favori Eklenme Sayısı (User_Favourites_Count)	15	Standart Profil (User_Default_Profile_Image)
8	Listelenme Sayısı (User_Listed_Count)	16	RT (ReTweet)

Çizelge 4.1'de kullanılan veri kümesindeki ihtiyaç duyulan ilgili özellikleri sıralanarak gösterilmektedir. Veri kümesi oluşturulmasında @oguzhancitlak Twitter hesabı kullanılmıştır. Twitter veri kümesinin oluşturulma tarihi Nisan 2017'dir. Veri kümesinde 221 756 adet Twitter kullanıcı hesabı detayları bulunmaktadır. Oluşturulan dosyanın uzunluğu 1 019 742 837 bayttır ve toplam kelime sayısı 107 307 574'tür. Python yazılımı aracılığı ile Twitter üzerinden çekilen veri kümesi Nisan 2018'de önerilen modelde kullanılmak üzere işleme alınmıştır. Toplam çekilen veri kümesindeki kullanıcılar rastgele bölünmüştür. Bu bölünme sonucunda içerisinde 81 317 tane Twitter kullanıcısının bulunduğu veri kümesinden Şekil 4.1'deki akış diyagramında belirtilen Twitter veri kümesi oluşturulmuştur. Geride kalan 140 439 tane kullanıcılı veri kümesinden rastgele 1 225 tane kullanıcı hesabı seçilmiştir.

Bir Twitter hesabından atılan ileti Json formatındaki dosyalarda *text* şeklinde

gözükmektedir. Önerilen uygulamada ilk önce kullanıcıların iletilerindeki *text* link ya da linklerin olup olmadığına bakılmaktadır.

Text içerisinde link var ise kısa link analizi yöntemi devreye girmektedir. Text içerisinde link bulunmaması durumunda, incelenen hesap spam analizi için makine öğrenmesi ve metin analizi yöntemine yönlendirilmektedir.

Önerilen yöntem ileti içerisinde kötücül bir link tespiti yaptığında analiz edilen hesabın spam olduğuna twitter spam politikasına [2, 110] göre karar verilmektedir. Analiz edilen hesabın iletisi içerisinde kötücül link bazen bulunmayabilmektedir. Atılan ileti içerisinde bazende hiç link olmayabilmektedir. Bu durumda, kötücül link bulunmayan hesaplar ile iletisinde hiç link bulunmayan hesaplar birlikte Weka havuzuna gitmektedir. Bu aşamada, kötücül link paylaşan ancak Weka havuzuna girmesi durumunda da masum olarak tespit edilebilecek hesaplar göz ardı edilmektedir. Önerdiğimiz modelin çalışma mantığında bir hesap kötücül link paylaşıyor ise bu hesap spam hesap olarak değerlendirilmektedir. Kötücül link paylaşımı yapan bir hesap başka hiçbir değerlendirilmeye sokulmamaktadır. İletilen bir ileti veri kümesi içerisindeki *text* bölümünde bulunmaktadır.

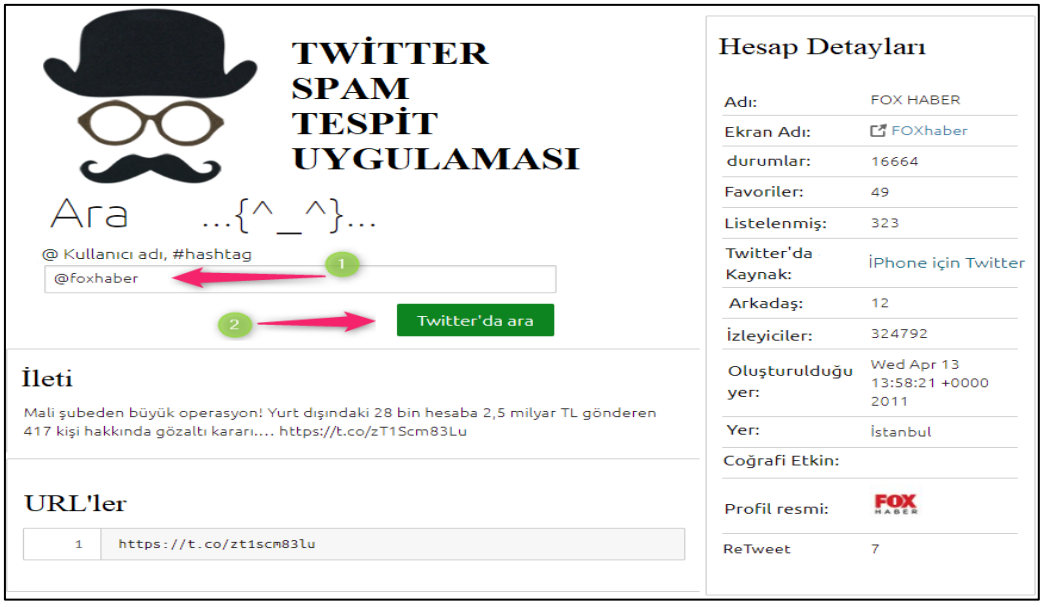
{"created_at":"Thu Aug 03 04:43:11 +0000 2017","id":892969070068367361,"id_str":"892969070068367361"
"text":"RT @Bitcoin_Friend: Can #Bitcoin Be Traded Like #Gold? https:\/\/t.co\/17uWia2Obv","source"
:"\u003ca href=\"http:\/\/Timmyzophrene.com\" rel=\"nofollow\"\u003eTimmyzophrene
favbtcf\u003c\/a\u003e","truncated":false,"in_reply_to_status_id":null,"in_reply_to_status_id_str":
null,"in_reply_to_user_id":null,"in_reply_to_user_id_str":null,"in_reply_to_screen_name":null,"user"
:{"id":852516321476661248,"id_str":"852516321476661248","name":"Timmyzophrene","screen_name":
"Timmy_zophrene","location":"Marseille, France","url":
"https:\/\/www.youtube.com\/channel\/UCxk837pMsjFaGu5MIjV4Jag","description":null,"protected":false,
"verified":false,"followers_count":351,"friends_count":323,"listed_count":1,"favourites_count":3480,
"statuses_count":3567,"created_at":"Thu Apr 13 13:38:24 +0000 2017","utc_offset":-25200,"time_zone":
"Pacific Time (US & Canada)","geo_enabled":false,"lang":"en","contributors_enabled":false,
"is_translator":false,"profile_background_color":"000000","profile_background_image_url":
"http:\/\/abs.twimg.com\/images\/themes\/theme1\/bg.png","profile_background_image_url_https":
"https:\/\/abs.twimg.com\/images\/themes\/theme1\/bg.png","profile_background_tile":false,
"profile_link_color":"19CF86","profile_sidebar_border_color":"000000","profile_sidebar_fill_color":
"000000","profile_text_color":"000000","profile_use_background_image":false,"profile_image_url":
"http:\/\/pbs.twimg.com\/profile_images\/852516989067309056\/4Pqzat9-_normal.jpg",

Şekil 4.2. Twiter'da iletilen örnek bir ileti

Şekil 4.2'de analiz edilen bir Twitter kullanıcı hesabındaki tweet(text) alanı işaretlenmiştir. Şekilde incelenen hesap @Timmy_zophene kullanıcı adlı bir Twitter hesabıdır. İşaretlenen kısımda kullanıcının iletisi içerisinde paylaştığı metin gözükmektedir. Atılan ileti içerisinde bir adet link, bir adet ileti ve iki adet bahset olduğu anlaşılmaktadır. Bu hesabın oluşturulma tarihinde Ağustos 2017 olduğu, bu ham veri analiz edildiğinde anlaşılmaktadır.

4.1. Kısa Link Analizi Bileşeni

Kısa link analizi yönteminde kullanıcı hesabından atılan iletiler içerisindeki link önemlidir. İletilen metin içerisinde T.CO formatlı link ya da linkler var ise, bu hesap kısa link analizine gönderilmektedir [111]. Gönderilen link VirusTotal tarafından değerlendirilmektedir. Kısa link analizi yönteminde kullanılan kod parçacığı EK-10'un içerisinde gösterilmektedir. Kötücül link tespiti yapılmış ise bu hesap Twitter'ın spam hesap politikası gereği spam dır şeklinde sonuçlanmaktadır. Bir ileti içerisindeki link ya da linklerden en az bir tanesinin kötücül olması bu hesabın spam olarak işaretlenmesi için yetmektedir [2, 110]. Analiz edilen hesabın iletisinde link yoksa ya da VirusTotal tarafından analiz edilen link kötücül olarak belirtilmemiş ise, önerilen model analiz edilen hesabı makine öğrenmesi ve metin analizi aşamalarının uygulandığı Weka havuzuna göndermektedir. Kısa link analizi yönteminde VirusTotal veri kümesi kullanılmaktadır. Atılan bir iletinin içerisinde bulunan link analiz edilmektedir. Önerilen model uygulamasının ara yüzü hazırlanmıştır ve bu arayüz üzerinden hesap sorglaması yapılabilmektedir. Twitter hesabına ait bir ileti aşağıda paylaşılmaktadır.



Şekil 4.3. Örnek analiz edilen bir hesap detayı

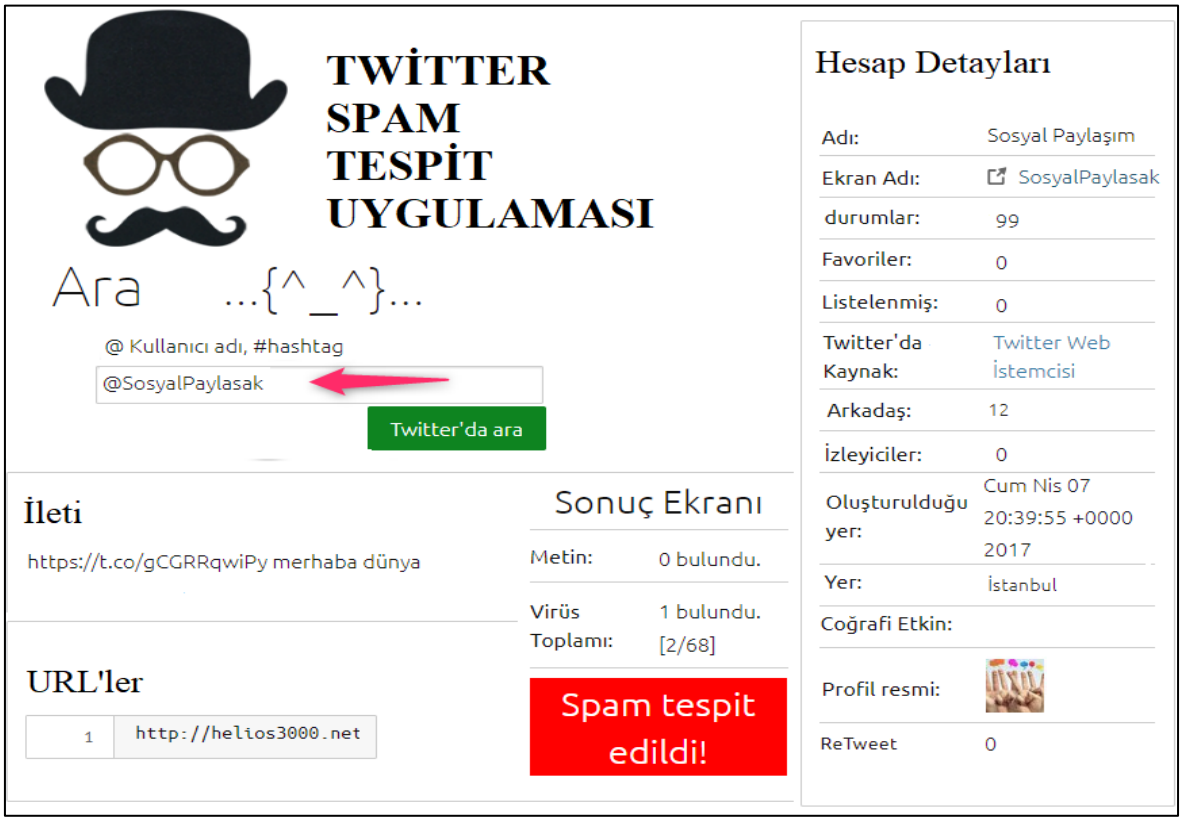
Şekil 4.3'te paylaşılan bir tweet içerisinde T.CO formatında link bulunmaktadır [111]. VirusTotal sistemi bu hesabın kötücül olup olmadığını EK-10'da belirtilen kod parçacığı ile analiz etmektedir. Link analizinde işlem göre veri kümesi detayı Çizelge 4.2'de verilmektedir. Analiz edilen Twitter veri kümesinde 81 317 adet kullanıcı hesabı bulunmaktadır. Analiz edilen hesapların 69 932 tanesinde link bulunmaktadır. Link paylaşımı yapmayan hesap sayımız 11 385'tir. İletilen linkler arasından 1 870 adet kötücül link belirlenmiştir.

Çizelge 4.2. Link analizinde işlem gören veri tablosu

No	Özellik	Adet
1	Toplam çekilen veri kümesindeki kullanıcı sayısı	221 756
2	Kullanılan veri kümesindeki hesap sayısı	81 317
3	Etiketlemenin arasından seçildiği hesap sayısı	140 439
4	Etiketlenen hesap sayısı	1 225
5	Veri Kümesinin oluşturulma tarihi	Nisan 2017
6	Veri Kümesinin işleme alınma tarihi	Nisan 2018
7	Link paylaşan hesap sayısı	69 932
8	Link paylaşımı yapmayan hesap sayısı	11 385
9	Paylaşılan kötücül link sayısı	1 870
10	Paylaşılan masum link sayısı	68 062

Link analizinde incelenen veri kümesi içerisinde paylaşılan masum link sayısı 68 062'tir. Link analizinin başarısı sürekli güncellenen VirusTotal veri kümesi kötücül link havuzuna göre zaman içerisinde değişiklik gösterebilmektedir. Analiz edilen veri kümesindeki hesapların spam olup olmadıklarına yalnızca bu kötücül link havuzundaki güncellemeler etki edebilmektedir. Bu durumda veri kümesinin içerisinde 69 932 adet link paylaşan kullanıcıdan 1 870 tanesinin kötücül link paylaşımı yaptığını kabul etmek durumundayız. Bir kötücül linki birden fazla hesap paylaşabilmektedir. Böyle bir durumla karşılaşıldığı zaman, kötücül linki paylaşan bütün hesaplar spam olarak işaretlenmektedir. Link analizi yöntemimizin başarısı %100 olarak değerlendirilmektedir. Bu kanıya, yapılan incelemelerde 1 870 adet kötücül link paylaşan hesapların hepsinin Twitter tarafından askıya alınmış olması vardırmaktadır. Aynı veri kümesi üzerinde, belirli bir zaman geçtikten sonra tekrarlanabilecek link analizi yöntemi değişik sonuçlar verebilmektedir. Bu çalışmanın link

analizi yapıldığında bulunan masum link sayısı 68 062 iken kötücül link sayısı 1 870 adettir. Ancak bir kötücül linki birden fazla hesabın paylaşmış olduğu durumlarla da karşılaşılmıştır. Aynı veri kümesine daha sonrasında tekrarlana bilecek ikinci kısa link analizi yönteminde masum link sayımız azalabilir, kötücül link sayımız buna bağlı olarak artabilmektedir. Tam tersi bir durumda mümkün olabilmektedir. Kısa link analizi yönteminde kullanılan kötücül link analiz sisteminin dinamik veri kümesi böyle bir sonuçla karşılaşılmasına neden olmaktadır. VirusTotal kötücül veri seti havuzunun zaman içerisindeki güncellemelerinin göz ardı edilmesi gerekmektedir. VirusTotal yönteminde analiz edilen Twitter veri kümesi içerisinde 1 870 adet kötücül link bulunmaktadır. İletilen bütün linkler Nisan 2017 tarihinde Twitter kullanıcıları tarafından paylaşılmıştır. Analiz işleminin yapıldığı Nisan 2018 tarihinde ise bu kötücül link paylaşım yapan hesapların tamamının Twitter tarafından askıya alındığı fark edilmiştir. İletinin atıldığı tarih ile analiz edildiği tarih arasında geçen bir yıllık süre içerisinde Twitter bu hesapları ne zaman ya da ne kadar süre geçtikten sonra askıya almıştır bilinmemektedir. Ancak VirusTotal sistemi bu iletiler atıldığı zaman aynı anda bu linkleri analiz edebilme durumunda olsa idi bu hesapların hepsinin spam olarak belirtecektir. Ancak bu işlemi iletilerin atıldığı zamandan bir yıl sonrasındaki analizinde yapmış bulunmaktadır. Şekilde 4.4'de bir sosyal hesap uygulama üzerinde aratılmıştır.



Şekil 4.4. Uygulama üzerinde kötücül bir bağlantı tespit ekranı

Şekil 4.4'de analiz edilen sosyal hesabın iletisi içerisinde kötücül bir link paylaşımı yaptığı anlaşılmaktadır. URL kısmındaki kötücül linki uygulama tespit etmiştir. VirusTotal sistemi, ileti içerisinde paylaşılan bu kötücül linki kötücül olarak tespit etmiştir. İleti içerisinde kötücül bir metin bulunmamaktadır. İleti bölümü incelendiği zaman sadece bağlantı ve iki kelimeden oluşan bir iletinin paylaşıldığı gözükmektedir. Hesap detayları bölümündeki öznitelikleri Weka makine öğrenmesi bileşeninde analiz edilmesine gerek kalmamıştır. Sistem kötücül bağlantı paylaşan bu hesabı spam olarak etiketlemiştir. Hesap öznitelikleri üzerinden makine öğrenmesinde analiz yapılması durumunda hesap yine spam olarak ortaya çıkmaktadır.

VirusTotal'ın bulduğu 1 870 adet kötücül linkin sahibi hesapların tamamı aynı zamanda Twitter tarafından askıya alındıklarından dolayı, Link analizi yöntemimizin başarısını %100 olarak belirtilebilmektedir. Bunu söylerken göz ardı edilen tek bir özellik vardır. VirusTotal veri kümesi dinamik yapılı olmasıdır. Örneğin, bugün yayımlanan bir kötücül web sitesi VirusTotal database sini destekleyen servisler tarafında hemen bulunmayabilmektedir. Google Safebrowsing [72-74], Kaspersky [75], OpenPhish [73], Comodo Site Inspector [76], Forcepoint ThreatSeeker [71], Opera ve Yandex Safebrowsing [74, 78] servisleri VirusTotal'ın güçlü kötücül link analiz etme servisleridir. Sosyal medya üzerinden şuan paylaşılan yeni bir kötücül link VirusTotal dinamik yapısında daha önce tanımlanmamış ise maalesef tespit edilemeyecektir. Ancak önerilen öğrenmeye dayanan spam hesap tespit modelinde, aslında kötücül link ileten hesaplar tespit edilememektedir. Ancak makine öğrenmesi ve metin analizi yöntemlerine maruz kalacaklardır. Önerilen bu modelin başarımının hesaplanmasında, Link analizi yöntemi analiz edildiği Twitter veri kümesi üzerinde çalıştırıldığında 1 870 adet kötücül linklerin tamamını bulmaktadır. Kötücül link paylaşımını yapan hesaplar günümüzde artık yaşamamaktalardır. VirusTotal'ın dinamik yapısı göz ardı edilerek önerilen modelin link analizi bileşiminin başarımı %100'dür.

4.2. Makine Öğrenmesi Bileşeni

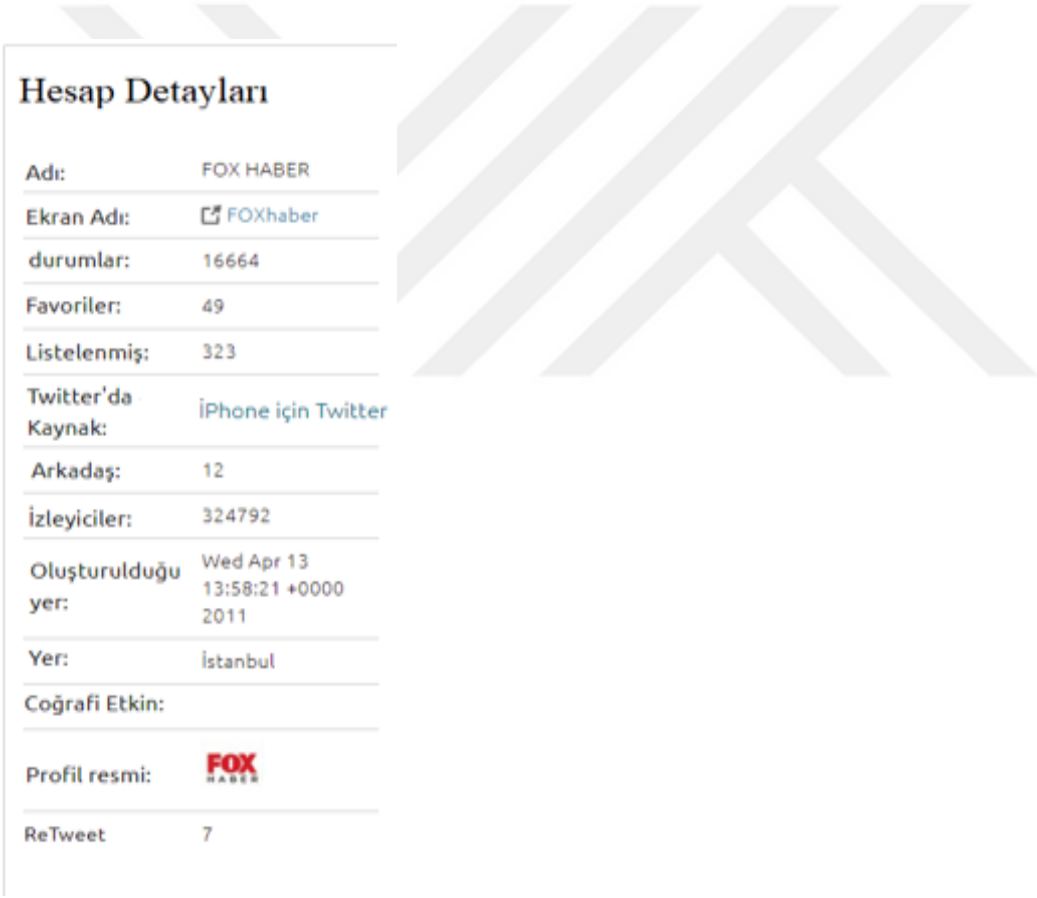
Makine öğrenmesi, önerilen model algoritması içerisinde kendisini Weka havuzu içerisinde göstermektedir. Veri kümesinde link paylaşımı yapmayan hesaplar ile masum link paylaşan hesapların incelendiği önerilen modelin ikinci bileşenidir. Başka bir deyişle kötücül olan bir hesap masum link paylaşmış olabilir ya da link analizi yöntemi bu hesabı tespit edememiş olabilir. Bu tarz hesapların gözden kaçırılmaması için makine öğrenmesi ve metin analizi

yöntemlerine başvurulmuştur. Masum görünüp kötücül özelliklere sahip hesapların tespit edilmesinde Weka havuzu kullanılmaktadır. Makine öğrenmesi yönteminde geçen hesaplar, Weka havuzunun diğer parçası olan metin analizi yönteminde analiz edilmektedir.

Makina öğrenmesi yöntemindeki veri kümesinde 1 225 adet kitle kaynak yöntemine göre etiketlenmiş Twitter kullanıcı hesaplarının özellikleri tanımlanmıştır. Makine öğrenmesindeki bu veri kümesini, Weka'da literatürde çok kabul gören oranlarda K = 10 ve percentege split = 66 olarak iki farklı şekilde değerlendirilmiştir [94, 95, 123]. K ve percentege split'in aldığı değerler Weka içerisinde eğitim kümesi ve test kümesi olarak kullanılmaktadır. Önerilen modelde literatürde sıkça kullanılan beş adet makine öğrenmesi algoritması kullanılmaktadır. NaiveBayes, RandomForest, IBk, J48 ve JRip bu algoritmalardır. Link paylaşımı yapmayan hesap sayısı 11 385'tir. Masum link paylaşan hesap sayısı ise 68 062 adettir. Önerilen modelin uygulama algortimasının işleyişine göre Makine öğrenmesi ve Metin analizi yöntemleri için toplamda 79 447 adet hesap aktarılmış olmaktadır. Şekil 4.5'te Twitter tarafından askıya alınmış bir hesabın, önerilen modeldeki arama sonucunu göstermektedir.

Şekil 4.5 Uygulama üzerinde spam hesap sorgulanması

Şekil 4.5'de geliştirilen uygulama modelinde askıya alınmış bir hesabın detayının alınamadığı ekran görüntüsü verilmektedir. Askıya alınmış bir hesabı Twitter üzerinden geliştirilen uygulama üzerinden sorgulandığı zaman bu hesabın askıya alındığı bilgisi dönmektedir. Twitter askıya alınan hesaplar için hesap detayı, arkadaş sayısı, takipçi sayısı gibi özellikleri, ayrıca normal hesaplardan uygulama aracılığı ile elde edilebilen özelliklerin hiç birisini vermemektedir. Önerilen modelin makine öğrenmesi bileşeninde elde edilen sonuçlar ve başarımın gösterimi için bir takım algoritmalara ve bu algoritmaların metriklerine ihtiyaç duyulmaktadır. Makine öğrenmesinde analiz edilen öznitelikleri ile etiketlenmiş veri kümesinin öznitelikleri kıyaslanmaktadır. Şekil 4.6'da spam olarak bulunan bir hesabın öznitelikleri paylaşılmıştır.



Şekil 4.6. Öznitelikleri analiz edilen bir hesap

Şekil 4.6'da analiz edilen hesap, daha öncesinde kötücül link paylaşımı yaptığı için spam olarak etiketlenmiştir. Aynı hesap dikkate alındığı zaman değerlendirilecek öznitelikleri etiketlenmiş veri kümesi ile karşılaştırması Çizelge 4.3'te de yapılmıştır. İlgili hesap makine öğrenmesi yöntemine göre de spam olarak ortaya çıkmaktadır.

Çizelge 4.3. Bir sosyal medya hesabının makine öğrenmesi ile analizi

Özellik	NaiveBayes	IBk	J48	Sonuç
Doğru Olarak Sınıflandırılan	%91,9951	%99,0295	%97,9863	%96,3369
Yanlış Olarak Sınıflandırılan	%8,0049	%0,9705	%2,0137	%3,6631
Kappa İstatistiği	0,8373	0,98	0,9581	0,9251
Ortalama Mutlak Hata	0,4408	0,0001	0,0241	0,155
Hesap	Spam	Spam Değil		
@FOXhaber	%3,6631	%96,3369		
Sonuç	Spam Değildir			

Çizelge 4.3'de analiz edilen ve öznitelikleri Şekil 4.6'da verile sosyal hesap (@FOXhaber)'dir, yapılan analizler sonucunda spam değil olarak bulunmaktadır. Öncelikle hesabın özniteliklerinden bir Arff dosyası oluşturulmuştur. Oluşturulan bu Arff dosyası, spam ve spam değil şeklinde iki olasılıklı değerlendirilmiştir. Twitter veri kümesinde daha önce yapılan analizler sonucunda en yüksek başarıyı veren IBk algoritması (%99,0385), en düşük başarıyı veren NaïveBayes algoritması (%91,8269) ve ortada bir değer veren J48 algoritmalari (%97,1154) ile analiz işlemi yapılmıştır. Çizelge 4.3'de incelenen sonuçlara göre hesabın spam hesap olmadığı bu üç algortimanın sonucuna göre belirlenmektedir. Bu üç algoritma için kappa istatistiği 0,9251 oranını vermektedir. Kappa istatistiği 1'e eşit olduğu zaman kesin sonuç alınmaktadır [84, 102]. Bu spam analiz işleminde 1'e çok yakın değer bulunmaktadır. Ortak mutlak hata oranı 0,155 ise oldukça düşük gözükmektedir ve başarılı bir sonuç elde edilmesi için bu oran geçerlidir [96, 103]. Analiz edilen hesabın spam olmadığını bu metrik sonuçlarına bakılarak karar verilebilmektedir. Kappa istatistiği 1'e yakın olduğu zaman doğru sonuça erişildiği ortaya çıkmaktadır [102]. Doğru olarak sınıflandırılan ve yanlış olarak sınıflandırılan metrikleri değerlendirildiğinde, analiz edilen hesabın spam hesap olmadığı %96,3369 oranla ortaya çıkmaktadır. Ortalama mutlak hata oranları ise 1'den oldukça uzaktadır. Bu değer 0'a çok yakın değerdir ve makine öğrenmesi işlemlerinde elde edilen sonucu yorumlamakta çok işe yaramaktadır.

Makine öğrenmesinde yönteminin veri kümesi üzerinde işlenilmesi sonucund elde edilen değerler aşağıda gösterilmektedir. Çizelge 4.4'te percentage split'in 66 olduğu duruma göre doğruluk oranları gösterilmektedir.

Çizelge 4.4. Makine öğrenmesinin percentage split = 66‘a göre başarım oranları

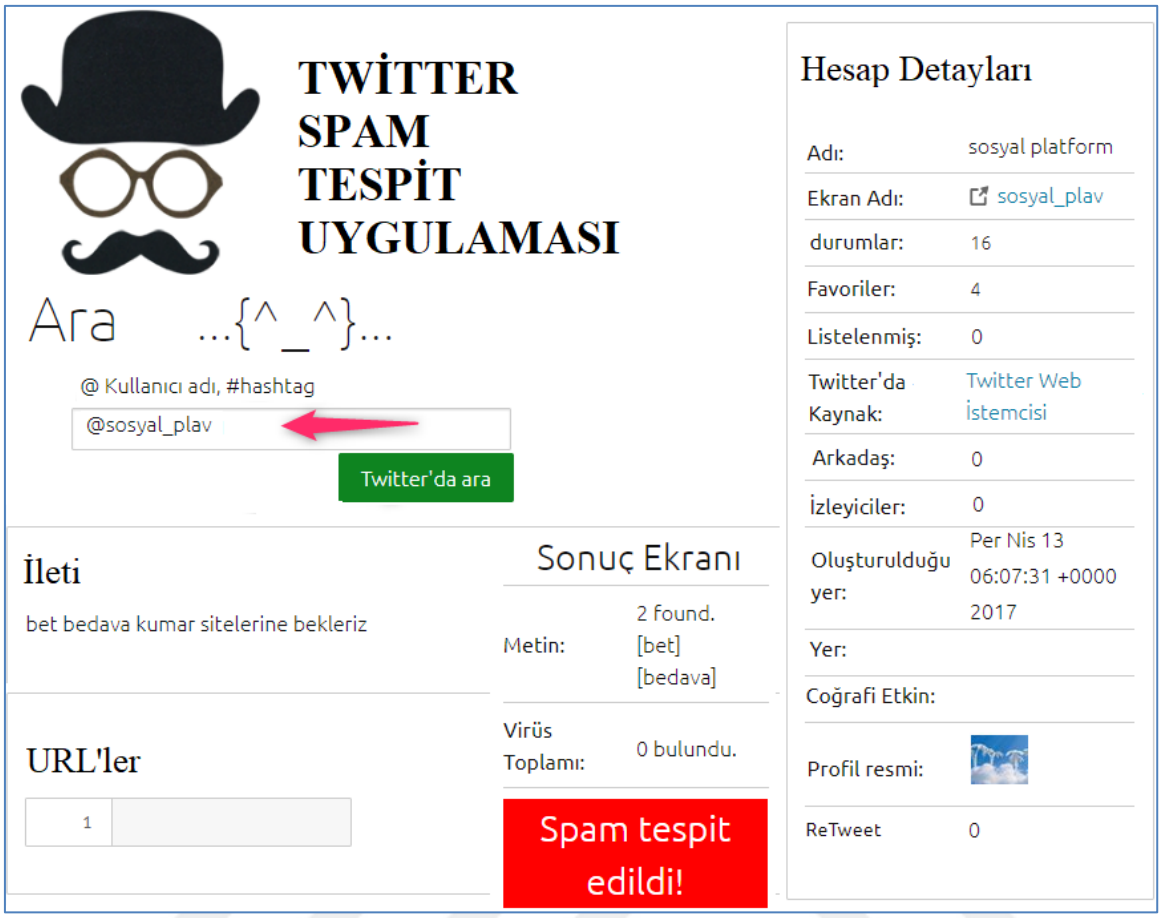
Özellik	NaiveBayes	Random Forest	IBk	J48	JRip
Doğru Olarak Sınıflandırılan	%91,8269	%97,8365	%99,0385	%97,1154	%96,875
Yanlış Olarak Sınıflandırılan	%8,1731	%2,1635	%0,9615	%2,8846	%3,125
Kappa İstatistiği	0,8341	0,9556	0,9802	0,9399	0,9355
Ortalama Mutlak Hata	0,1062	0,0325	0,0091	0,0394	0,0391
Ortalama Hata Karekök	0,2591	0,1057	0,0747	0,1567	0,1736
Göreli Mutlak Hata	%21,9015	%6,7007	%1,8766	%8,1383	%8,065
Göreli Hata Karekök	%52,6095	%21,4596	%15,164	%31,8204	%35,2452
Doğru Onaylanmış Oran (Ağırlıklı Ortalama)	0,918	0,978	0,990	0,971	0,969
Yanlış Onaylanmış Oran (Ağırlıklı Ortalama)	0,071	0,019	0,010	0,041	0,034
Kesinlik (Ağırlıklı Ortalama)	0,923	0,979	0,990	0,973	0,969
Hassasiyet (Ağırlıklı Ortalama)	0,918	0,978	0,990	0,971	0,969
F-Ölçütü (Ağırlıklı Ortalama)	0,919	0,978	0,990	0,971	0,969
Matthews Korelasyon Katsayısı (Ağırlıklı Ortalama)	0,937	0,956	0,980	0,942	0,936
Alıcı İşletim Karakteristiği Alanı (Ağırlıklı Ortalama)	0,960	1	1	0,985	0,969
Hassas Hassasiyet Alanı (Ağırlıklı Ortalama)	0,961	1	1	0,980	0,957

Çizelge 4.4’de, Makine öğrenmesi yönteminde kullanılan algoritmalarda percentage split = 66 olarak alındığı zaman IBk algoritması en başarılı sonucu verdiği görülmektedir. Bu kanıya kappa istatistiği, Doğru onaylanmış oran ve doğru sınıflandırma yüzdelerinin diğer kalan dört algoritmaya kıyasla en yüksekte olması vardırmıştır. Ortalama mutlak hata oranının da yok denilecek kadar az olması bir başka destekleyici veridir. Bu çalışmada, NaiveBayes sınıflandırma algoritması ise diğer dört algoritma içerisinde en düşük sonucu verdiği görülmektedir.

4.3. Metin Analizi Yöntemi

Metin analizi yönteminde analiz edilen hesapların iletileri göz önüne alınmaktadır. Literatürde spam hesapların sıklıkla kullandıkları kelimeler belirlenmiştir. Bu kelimelerden hassas içerikli kelimeler, iletilerdeki geçme sıklıklarına göre belirlenmiştir. Belirlenen bu kelimeler metin analizi yönteminde kullanılmıştır. Belirlenen kelimelere EK-5 içerisinde listelenmektedir. Önerilen modelin metin analizi bileşeninde de kullanılacak olan veri kümesi *bet* hassas içerik kelimesi kullanılarak oluşturulmuştur. Oluşturulan veri kümesindeki hesapların tamamında bu hassas içerikli kelime yer almaktadır. Metin analizi yönteminde incelenen iletilerde, ikinci, üçüncü ya da daha fazla sıklıkla geçen bir hassas içerik kelimesi bu hesapların spam hesap olma olasılığını arttırmaktadır. Makine öğrenmesi yöntemi çalıştırıldığında, işlem gören Weka havuzundaki bütün iletilerde *bet* hassas içerik kelimesi bulunmaktadır. Buradan sonuçla, makine öğrenmesi yöntemi analizinden gelen ve spam olduğu belirtilen hesapların tamamında da hassas içerikli kelime bulunmak zorundadır. Metin analizinde işlem gören hesapların da tamamında hassas içerikli kelime bulunmaktadır. İletilerinde hassas içerik bulunduran hesapların tamamı spam davranış göstermektedir denilebilmektedir [2, 61]. Ancak, hassas içerikli bu hesapları spamdır şeklinde işaretlemek yerine, içerisinde iki hassas içerik ya da en az üç hassas içerikli kelime geçen iletiler dikkate alınabilmektedir. Başka bir deyişle, analiz edilen veri kümesine spam özellikli veri kümesi denilebilmektedir [61]. Ancak önerilen modelin üçüncü bileşeni olan metin analizi yöntemi uygulamasında, iletilerinde ikinci ya da daha fazla hassas içerikli kelime bulunduran hesaplar daha da dikkate alınmaktadır. Analiz edilen veri kümesinde de aynı işlem dikkate alınmıştır.

Metin analizi yönteminde daha önce belirlenen hassas içerikli kelimelerin iletiler içerisinden geçip geçilmediğine bakılmaktadır. Şekil 4.7’de iletisinde hassas içerik paylaşan bir sosyal hesap gösterilmektedir.



Şekil 4.7. Uygulama üzerinde hassas içerikli metin paylaşan bir hesap

Şekil 4.7'de analiz edilen hesabın içerisinde iki adet hassas içerikli kelime bulanan bir ileti paylaşmıştır. *bet* ve *bedava* kelimelerinin bulunduğu bir ileti @sosyal_plav kullanıcısı tarafından paylaşılmıştır. Diğer hassas içerikli kelimelerin listesi EK-5'te verilmiştir. İlgili hesabın iletisinde hiçbir bağlantı bulunmamaktadır. Link analizi yönteminden iletisinde bağlantı bulunmadığı için masum kullanıcı olarak geçmiştir. Twitter veri kümesi içerisinde hassas içerikli kelimelerin olduğu hesaplar bu şekilde tespit edilmiştir. Hesap detayları bölümünde gösterilen özniteliklere göre de makine öğrenmesinde işlem yapılması durumunda, bu hesabın spam olduğu ortaya çıkmaktadır.

Analiz edilen veri kümesinde, 81 317 adet Twitter hesabının yalnızca 18 553 tanesinde ikinci ya da üçüncü hassas içerikli kelime bulunmuştur. Bu 18 553 adet hesabın 4 316 tanesi link analizi sonucunda masum link paylaşmış olarak belirtilmektedir. Geriye kalan 14 237 adet hesabın tamamı makine öğrenmesi yöntemine göre spam olarak belirlenmektedir. Makine öğrenmesinde spam olarak belirlenmiş hesapların içerisinde ikinci ya da üçüncü bir hassas içeriği barındırmasından dolayı bu hesapların spam oldukları belirtilebilmektedir. Yapılan

kontrollerde bu hesapların tamamı Twitter tarafından askıya alınmış hesaplar oldukları görülmektedir [69]. Bu çalışmada, metin analizi yönteminin yalnızca katkısı bulunmaktadır. Metin analizi yönteminde iletiler içerisinde hassas içerik olup olmadığı kontrol edilmektedir. Makine öğrenmesi sonucunda spam olarak bulunan 14 237 adet hesabın içerisinde ikinci ya da üçüncü bir hassas içerikli kelime bulunmaktadır. Bu hesapların tamamının da Twitter tarafından askıya alındığı göz önünde bulundurulduğun da metin analizi tarafından spam olarak belirtilen hesapların doğru sonuç verdikleri görülmektedir.

Metin analizi işleminde 62 073 adet hesabın spam olduğu iletilerindeki hassas içerikler incelendikten sonra bulunmuştur. Bu hesapların tamamı veri kümesinden çekilmektedir. Veri kümesindeki bu belirtilen hesapların tamamı askıya alınmış durumdadır. Metin analizi yönteminde 1 822 adet hesap masum kullanıcı olarak belirlenmiştir. Bu hesaplar askıya alınmamışlardır. Metin analizi yönteminde toplam analiz edilen hesapların yaklaşık %97,15‘i spam olarak belirtilmiştir.

5. BULGULAR VE TARTIŞMA

Bu çalışmada, oluşturulan veri kümesindeki kullanıcı sayısı 221 756 dır. Analiz edilen toplamda 81 317 adet Twitter hesabının 69 932 adet tanesi iletilerinde link paylaşımı yapmaktadır. Buradan hesaplama ile yaklaşık olarak %85,99 sının kötücül ve masum link paylaşımı yaptığını görülmektedir. Çizelge 5.1'de kullanılan veri kümesinin değerlendirme sonuçları gösterilmektedir.

Çizelge 5.1. Önerilen spam hesap tespit modelinde kullanılan veri kümesi detayları

Sıra	Özellik	Adet	Yüzdelik Oranları	Kıyaslanan Özellik
1	Veri Kümesindeki Toplam Kullanıcı	221 756	%100	
2	Kullanılan Veri Kümesindeki Toplam Kullanıcı	81 317	%36,67	Sıra 1'e göre
3	Etiketlemenin Yapıldığı Kalan Veri Kümesindeki Toplam Kullanıcı	140 439	%63,33	Sıra 1'e göre
4	Etiketlenen Hesap Sayısı	1 225	%0,87	Sıra 3'e göre
5	Link Paylaşan Hesap Sayısı	69 932	%85,99	Sıra 2'ye göre
6	Link Paylaşımı Yapmayan Hesap Sayısı	11 385	% 4,01	Sıra 2'ye göre
7	Paylaşılan Kötücül Link Sayısı	1 870	%2,67	Sıra 5'e göre
8	Paylaşılan Masum Link Sayısı	68 062	%97,32	Sıra 5'e göre
9	Weka Havuzunda Analiz Edilen	79 447	%97,70	Sıra 2'ye göre
10	Makina Öğrenmesinde Bulunan Spam	14 237	%17,92	Sıra 9'a göre
11	Metin Analizinde spam kelimeli hesap	73 337	%90,19	Sıra 2'ye göre
12	Metin Analizinde iki hassas kelimeli hesap	14 124	%19,26	Sıra 11'e göre
13	Metin Analizinde üç ve fazla hassas kelimeli hesap	4 429	%6,04	Sıra 11'e göre

Çizelge 5.1 incelendiğinde, 11 385 adet hesap link paylaşımı yapmamıştır ve doğrudan Weka havuzunda değerlendirilmiştir. Paylaşılan linklerin 1 870 tanesi kötücül olup 68 062 tanesi masum davranış sergilemektedir. 1 870 adet kötücül link, toplam link paylaşan hesaplar içerisinde yaklaşık %2,67 oranına sahip olmaktadır, aynı zamanda bu hesaplar spam olarak tanımlanmaktadır. Link paylaşımı yapmayan 11 385 adet hesap ile birlikte 68 062 masum link paylaşımı yapan hesapların tamamı makine öğrenmesi ve metin analizi yöntemlerine geçmişlerdir. Toplamda 79 447 adet hesap bu yöntemlerden geçmiştir.

Makine öğrenmesi ve metin analizinde analiz edilen 79 447 adet kullanıcı hesaplarının yaklaşık %17,97'lik oranı buda yaklaşık olarak 14 237 adet hesap makine öğrenmesine göre spam olarak bulunmaktadır. Bu hesapların tamamının iletilerinde *bet* hassas içerik kelimesi bulunmaktadır. Metin analizi yönteminde, makine öğrenmesi sonucunda bulunan sonucu destekleyici nitelikte olmaktadır. Metin analizi yönteminde spamlı kelime içeren 73 337 adet hesabın %19,26 tanesinde yaklaşık olarak 14 124 tane hesabın iletisinde içerisinde ikinci bir hassas içerikli kelime bulunmaktadır. Spam olarak belirtilen hesapların %6,04 tanesinde yaklaşık 4 429 adet hesabın iletisi içerisinde üçüncü ya da daha fazla hassas içerikli kelime geçtiği gözlemlenmektedir. Aşağıda kullanılan algoritmalar, analizler ve elde edilen sonuçlar sırasıyla paylaşılmaktadır.

NaiveBayes'e göre önerilen model başarısı

NaiveBayes sınıflandırıcı algoritmasın, literatürde çok kabul gören percentage split = 66 oranında işaretlenerek test ve eğitim veri kümeleri oluşturulmuştur ve literatürde kullanıldığı gibi sonuç bulunmuştur [9, 112].

NaiveBayes algoritması Eşitlik 4.1'deki gibi formüle edilmektedir.

$$p(C|F_1,...F_n) = \frac{P(C)p(F_1,...F_n|C)}{p(F_1,...F_n)} \tag{4.1}$$

NaiveBayes algoritmasının başarım sonucu ve metriklerin değerleri aşağıdaki gibidir. NaiveBayes algoritması, percentege split = 66 olduğu zamanki kappa istatistiği 0,8341 olarak bulunmuştur. Doğru olarak sınıflandırılan yüzdesi %91,8269 oranında bulunmaktadır. Doğru onaylanmış oran ise 0,918 seviyesindedir.

NaiveBayes sınıflandırıcı algoritmasını literatürde sıklıkla kullanılan K = 10 olarak etiketlenmiş veri kümesinde işlendiğinde alınan doğrulama sonuçları Çizelge 5.2'de gösterilmektedir [27, 94, 113]. Aşağıda kullanılan veri kümesinin NaiveBayes algoritması literatürde sıklıkla kullanılan K değeri 10 alınarak hesaplanmıştır.

Çizelge 5.2. NaiveBayes algoritması ile veri kümesi analiz sonuçları (K=10)

Naïve Bayes									
Doğru Olarak Sınıflandırılan	1 129				%92,1633				
Yanlış Olarak Sınıflandırılan	96				%7,8367				
Kappa İstatistiği	0,8405								
Ortalama Mutlak Hata	0,094				Karışıklık Matrisi				
Ortalama Hata Karekök	0,294				a	b			
Göreli Mutlak Hata	%19,3916				648	72	a = DOĞRU		
Göreli Hata Karekök	%50,3731				24	481	b = YANLIŞ		
Toplam Örnek Sayısı	1 225								
	Doğru Onaylanmış Oran	Hatalı Onaylanmış Oran	Kesinlik	Hassasiyet	F-Ölçütü	Matthews Korelasyon Katsayısı	Alıcı İşletim Karakteristiği Alanı	Hassas Hassasiyet Alanı	Sınıf
	0,900	0,048	0,964	0,900	0,931	0,843	0,970	0,978	DOĞRU
	0,952	0,100	0,870	0,952	0,909	0,843	0,970	0,961	YANLIŞ
Ağırlıklı Ortalama	0,922	0,069	0,925	0,922	0,922	0,843	0,970	0,971	

Çizelge 5.3 incelendiği zaman NaiveBayes algoritması, K=10 olduğu zamanki kappa istatistiği 0,8405 olarak bulunmuştur. Doğru olarak sınıflandırılan yüzdesi %92,1633 oranında bulunmaktadır. Doğru Onaylanmış Oran ise 0,922 seviyesindedir.

Percentege split ve K modellerinin sonucuna göre NaiveBayes algoritmasının başarısı %92,1633 ile %91,8269 aralığında olacağı hesaplanmaktadır. NaiveBayes algoritmasının başarısı her iki yöntemde de (K=10 ve percentege split = 66) aynı veri kümesini işlediği için iki başarı yüzdesinin ortalaması olarak %91,9951 olarak bulunmaktadır.

RandomForest'e göre önerilen model başarısı

Random Forest sınıflandırıcı algoritmasını da literatürde percentage split = 66 oranında

sıklıkla kullanılmaktadır [92, 93]. Random Forest algoritmasında kullanılan etiketlenmiş veri kümesi işlendiğinde, alınan doğrulama sonuçları şu şekildedir. RandomForest algoritması, percentege split = 66 olduğu zamanki kappa istatistiği 0,9556 olarak bulunmuştur. Doğru olarak sınıflandırılan yüzdesi %97,8365 oranında bulunmaktadır. Doğru Onaylanmış Oran ise 0,978 seviyesindedir.

Random Forest sınıflandırıcı algoritmasını literatürde sıklıkla kullanılan K = 10 değerinde hesaplama yapılmıştır [94]. Makine öğrenmesinde bu algoritmada etiketlenmiş veri kümesinde işlendiğinde alınan doğrulama sonuçları Çizelge 5.3'de gösterilmektedir. Aşağıda kullanılan veri kümesinin K değeri 10 alınarak Random Forest değerleri hesaplanmıştır.

Çizelge 5.3. Random Forest algoritması ile veri kümesi analiz sonuçları (K=10)

<table>
<tr><td>Random Forest</td><td colspan="8"></td></tr>
<tr><td>Doğru Olarak Sınıflandırılan</td><td colspan="4">1 215</td><td colspan="4">%99,1837</td></tr>
<tr><td>Yanlış Olarak Sınıflandırılan</td><td colspan="4">10</td><td colspan="4">%0,8163</td></tr>
<tr><td>Kappa İstatistiği</td><td colspan="4">0,9831</td><td colspan="4" rowspan="6"></td></tr>
<tr><td>Ortalama Mutlak Hata</td><td colspan="4">0,0171</td></tr>
<tr><td>Ortalama Hata Karekök</td><td colspan="4">0,0727</td></tr>
<tr><td>Göreli Mutlak Hata</td><td colspan="4">%3,53</td></tr>
<tr><td>Göreli Hata Karekök</td><td colspan="4">%14,7696</td></tr>
<tr><td>Toplam Örnek Sayısı</td><td colspan="4">1 225</td></tr>
<tr><td></td><td colspan="4"></td><td colspan="4">Karışıklık Matrisi</td></tr>
<tr><td></td><td colspan="4"></td><td>a</td><td>b</td><td colspan="2" rowspan="3">a = DOĞRU
b = YANLIŞ</td></tr>
<tr><td></td><td colspan="4"></td><td>716</td><td>4</td></tr>
<tr><td></td><td colspan="4"></td><td>6</td><td>499</td></tr>
<tr><td></td><td>Doğru Onaylanmış Oran</td><td>Hatalı Onaylanmış Oran</td><td>Kesinlik</td><td>Hassasiyet</td><td>F-Ölçütü</td><td>Matthews Korelasyon Katsayısı</td><td>Alıcı İşletim Karakteristiği Alanı</td><td>Hassas Hassasiyet Alanı</td><td>Sınıf</td></tr>
<tr><td></td><td>0,994</td><td>0,012</td><td>0,992</td><td>0,994</td><td>0,993</td><td>0,983</td><td>1</td><td>1</td><td>DOĞRU</td></tr>
<tr><td></td><td>0,988</td><td>0,006</td><td>0,992</td><td>0,988</td><td>0,990</td><td>0,983</td><td>1</td><td>1</td><td>YANLIŞ</td></tr>
<tr><td>Ağırlıklı Ortalama</td><td>0,992</td><td>0,009</td><td>0,992</td><td>0,992</td><td>0,992</td><td>0,983</td><td>1</td><td>1</td><td></td></tr>
</table>

Çizelge 5.3 incelendiği zaman RandomForest algoritması, K=10 olduğu zamanki kappa

istatistiği 0,9831 olarak bulunmaktadır. Doğru olarak sınıflandırılan yüzdesi %99,1837 oranında bulunmaktadır. Doğru onaylanmış oran ise 0,992 seviyesindedir. RandomForest algoritmasının başarısı %99,1837 ile %97,8365 aralığında olacağı hesaplanmaktadır. Random Forest algoritmasının başarısı her iki yöntemde de (K=10 ve percentege split = 66) aynı veri kümesini işlediği için iki başarı yüzdesinin ortalaması olarak %98,5101 olarak bulunmaktadır.

IBk'e göre önerilen model başarısı

IBk sınıflandırıcı algoritmasını literatüre göre percentage split = 66 olarak çalıştırılmıştır [94]. Bu algoritmada, makine öğrenmesinde kullanılan etiketlenmiş veri kümesi işlendiğinde alınan doğrulama sonuçları şu şekildedir. IBk algoritması, percentege split = 66 olduğu zamanki kappa istatistiği 0,9802 olarak bulunmuştur. Doğru olarak sınıflandırılan yüzdesi %99,0385 oranında bulunmaktadır. Doğru Onaylanmış Oran ise 0,990 seviyesindedir.

IBk sınıflandırıcı algoritmasını literatürde çok kullanılan K = 10 olarak alınarak analiz edilmiştir [95]. Bu algoritma tercih edildiğinde, makine öğrenmesinde kullanılan etiketlenmiş veri kümesinin analizinden alınan sonuçlar Çizelge 5.4'de gösterilmektedir. Aşağıda kullanılan veri kümesinin K değeri 10 alınarak hesaplanmış IBk değerleri gösterilmektedir.

Çizelge 5.4. IBk algoritması ile veri kümesi analiz sonuçları (K=10)

<table>
<tr><td>IBk</td><td colspan="4"></td></tr>
<tr><td>Doğru Olarak Sınıflandırılan</td><td>1 213</td><td colspan="3">% 99,0204</td></tr>
<tr><td>Yanlış Olarak Sınıflandırılan</td><td>12</td><td colspan="3">% 0,9796</td></tr>
<tr><td>Kappa İstatistiği</td><td>0,9798</td><td colspan="3" rowspan="7">Karışıklık Matrisi
<table><tr><td>a</td><td>b</td><td></td></tr><tr><td>716</td><td>4</td><td>a = DOĞRU</td></tr><tr><td>8</td><td>497</td><td>b = YANLIŞ</td></tr></table></td></tr>
<tr><td>Ortalama Mutlak Hata</td><td>0.0089</td></tr>
<tr><td>Ortalama Hata Karekök</td><td>0,0688</td></tr>
<tr><td>Göreli Mutlak Hata</td><td>1,8299</td></tr>
<tr><td>Göreli Hata Karekök</td><td>13,9855</td></tr>
<tr><td>Toplam Örnek Sayısı</td><td>1 225</td></tr>
</table>

Çizelge 5.4.(devam) IBk algoritması ile veri kümesi analiz sonuçları (K=10)

	Doğru Onaylanmış Oran	Hatalı Onaylanmış Oran	Kesinlik	Duyarlılık	F-Ölçütü	Matthews Korelasyon	Alıcı İşletim Karakteristiği Alanı	Hassas Duyarlılık Alanı	Sınıf
	0,994	0,016	0,989	0,994	0,992	0,980	1,000	1,000	DOĞRU
	0,984	0,006	0,992	0,984	0,988	0,980	1,000	1,000	YANLIŞ
Ortalaması Ağırlığı	0,990	0,012	0,990	0,990	0,990	0,980	1,000	1,000	

Çizelge 5.4 incelendiği zaman IBk algoritması, K=10 olduğu zamanki kappa istatistiği 0,9798 olarak bulunmuştur. Doğru olarak sınıflandırılan yüzdesi %99,0204 oranında bulunmaktadır. Doğru Onaylanmış Oran ise 0,990 seviyesindedir. Buradan sonuçla IBk algoritmasının başarısı %99,0204 ile %99,0385 aralığında olacağı hesaplanmaktadır. IBk algoritmasının başarısı her iki yöntemde de (K=10 ve percentege split = 66) aynı veri kümesini işlediği için iki başarı yüzdesinin ortalaması olarak %99,0295 olarak bulunmaktadır.

J48'e göre önerilen model başarısı

J48 sınıflandırıcı algoritmasını literatürde tercih edilen percentage split = 66 olarak makine öğrenmesinde kullandığımız etiketlenmiş veri kümesi işlenmiştir [79, 80]. J48 algoritması, percentege split = 66 olduğu zamanki kappa istatistiği 0,9399 olarak bulunmuştur. Doğru olarak sınıflandırılan yüzdesi %97,1154 oranında bulunmaktadır. Doğru onaylanmış oran ise 0,971 seviyesindedir.J48 sınıflandırıcı algoritmasını literatürde kullanılan K = 10 olarak makine öğrenmesinde kullanılan etiketlenmiş veri kümesinde işlenmiştir [90].
Aşağıda kullandığımız veri kümesinin K değeri 10 alınarak hesaplanmış J48 değerleri gösterilmektedir.

Çizelge 5.5. J48 algoritması ile veri kümesi analiz sonuçları (K=10)

J 48		
Doğru Olarak Sınıflandırılan	1 211	% 98,8571
Yanlış Olarak Sınıflandırılan	14	% 1,1429

Çizelge 5.5. (devam) J48 algoritması ile veri kümesi analiz sonuçları (K=10)

Kappa İstatistiği	0,9763								
Ortalama Mutlak Hata	0,0223								
Ortalama Hata Karekök	0,1057								
Göreli Mutlak Hata	% 4,598								
Göreli Hata Karekök	%21,4747								
Toplam Örnek Sayısı	1 225								

Karışıklık Matrisi

a	b	
720	0	a = DOĞRU
14	491	b = YANLIŞ

	Doğru Onaylanmış Oran	Hatalı Onaylanmış Oran	Kesinlik	Duyarlılık	F-Ölçütü	Matthews Korelasyon	Alıcı İşletim Karakteristiği	Hassas Duyarlılık Alanı	Sınıf
	1	0,028	0,981	1	0,990	0,977	0,985	0,984	DOĞRU
	0,972	0	1	0,972	0,986	0,977	0,985	0,988	YANLIŞ
Ortalaması Ağırlığı	0,989	0,016	0,989	0,989	0,989	0,977	0,985	0,986	

Çizelge 5.5 incelendiği zaman J48 algoritması, K=10 olduğu zamanki kappa istatistiği 0,9763 olarak bulunmuştur. Doğru olarak sınıflandırılan yüzdesi %98,8571 oranında bulunmaktadır. Doğru onaylanmış oran ise 0,989 seviyesindedir. Buradan sonuçla J48 algoritmasının başarısı %98,8571 ile %97,1154 aralığında olacağı hesaplanmaktadır. J48 algoritmasının başarısı her iki yöntemde de (K=10 ve percentege split = 66) aynı veri kümesini işlediği için iki başarı yüzdesinin ortalaması olarak %97,9863 olarak bulunmaktadır.

JRip'e göre önerilen model başarısı

JRip sınıflandırıcı algoritmasını literatürde kullanılan percentage split = 66 olarak makine öğrenmesinde kullanılan etiketlenmiş veri kümesinde işlenmiştir [99]. JRip algoritması, percentege split = 66 olduğu zamanki kappa istatistiği 0,9355 olarak bulunmuştur. Doğru olarak sınıflandırılan yüzdesi %96,875 oranında bulunmaktadır. Doğru Onaylanmış Oran ise 0,969 seviyesindedir. JRip sınıflandırıcı algoritmasını literatüre göre K = 10 olarak alınmasıyla makine öğrenmesinde kullanılan etiketlenmiş veri kümesinde işlenmiştir [98]. Alınan doğrulama sonuçları Çizelge 5.6'de gösterilmektedir.

Çizelge 5.6'da kullandığımız veri kümesinin K değeri 10 alınarak hesaplanmış JRip değerleri gösterilmektedir.

Çizelge 5.6. JRip algoritması ile veri kümesi analiz sonuçları (K=10)

JRip									
Doğru Olarak Sınıflandırılan	1 209				%98,6939				
Yanlış Olarak Sınıflandırılan	16				%1,3061				
Kappa İstatistiği	0,973								
Ortalama Mutlak Hata	0,019				Karışıklık Matrisi				
Ortalama Hata Karekök	0,1044				a	b			
Göreli Mutlak Hata	%3,9274				715	5	a = DOĞRU		
Göreli Hata Karekök	%21,2142				11	494	b = YANLIŞ		
Toplam Örnek Sayısı	1 225								
	Doğru Onaylanmış Oran	Hatalı Onaylanmış Oran	Kesinlik	Hassasiyet	F-Ölçütü	Matthews Korelasyon Katsayısı	Alıcı İşletim Karakteristiği Alanı	Hassas Hassasiyet Alanı	Sınıf
	0,993	0,022	0,985	0,993	0,989	0,973	0,991	0,988	DOĞRU
	0,978	0,007	0,990	0,978	0,984	0,973	0,991	0,992	YANLIŞ
Ağırlıklı Ortalama	0,987	0,016	0,987	0,987	0,987	0,973	0,991	0,990	

Çizelge 5.6 incelendiği zaman JRip algoritması, K=10 olduğu zamanki kappa istatistiği 0,973 olarak bulunmuştur. Doğru olarak sınıflandırılan yüzdesi %98,6939 oranında bulunmaktadır. Doğru onaylanmış oran ise 0,987 seviyesindedir. Buradan sonuçla JRip algoritmasının başarısı %98,6939 ile %96,875 aralığında olacağı hesaplanmaktadır. JRip algoritmasının başarısı her iki yöntemde de (K=10 ve percentege split = 66) aynı veri kümesini işlediği için iki başarı yüzdesinin ortalaması olarak %97,7845 olarak bulunmaktadır.

Makine öğrenmesi yönteminde veri kümesinin analiz edildiği diğer mertiğe göre alınan

sonuçlar aşağıdaki gibi gözükmektedir. Çizelge 5.7, literatüre göre sıklıkla kullanılan K = 10 parametresine göre alınan sonuçları göstermektedir [88, 89].

Çizelge 5.7. Makine öğrenmesinin K = 10'a göre başarım oranları

Özellik	NaiveBayes	Random Forest	IBk	J48	JRip
Doğru Olarak Sınıflandırılan	%92,1633	%99,1837	%99,0204	%98,8571	%98,6939
Yanlış Olarak Sınıflandırılan	%7,8367	%0,8163	%0,9796	%1,1429	%1,3061
Kappa İstatistiği	0,8405	0,9831	0,9798	0,9763	0,973
Ortalama Mutlak Hata	0,094	0,0171	0,0089	0,0223	0,019
Ortalama Hata Karekök	0,248	0,0727	0,0688	0,1057	0,1044
Göreli Mutlak Hata	%9,3916	%3,53	%1,8289	%4,598	%3,9278
Göreli Hata Karekök	%50,3731	%14,7696	%13,9855	%21,4747	%21,2142
Doğru Onaylanmış Oran (Ağırlıklı Ortalama)	0,922	0,992	0,990	0,989	0,987
Yanlış Onaylanmış Oran (Ağırlıklı Ortalama)	0,069	0,009	0,012	0,016	0,016
Kesinlik (Ağırlıklı Ortalama)	0,925	0,992	0,990	0,989	0,987
Hassasiyet (Ağırlıklı Ortalama)	0,922	0,992	0,990	0,989	0,987
F-Ölçütü (Ağırlıklı Ortalama)	0,922	0,992	0,990	0,989	0,987
Matthews Korelasyon Katsayısı (Ağırlıklı Ortalama)	0,943	0,983	0,980	0,977	0,973
Alıcı İşletim Karakteristiği Alanı (Ağırlıklı Ortalama)	0,970	1	1	0,985	0,991
Hassas Hassasiyet Alanı (Ağırlıklı Ortalama)	0,971	1	1	0,986	0,990

Bu çizelgede Makine öğrenmesi yönteminde kullandığımız algoritmalarda K = 10 olarak alındığı zaman RandomForest algoritması en başarılı sonucu verdiği görülmektedir. Bu kanıya kappa istatistiği, Doğru onaylanmış oran ve doğru sınıflandırma yüzdelerinin diğer kalan dört algoritmaya kıyasla en yüksek değerde olması vardırmaktadır. NaiveBayes sınıflandırma algoritması ise diğer dört algoritma içerisinde en düşük sonucu vermektedir.

Makine öğrenmesi yönteminin başarımının ölçülmesinde kullanılan algoritmalar ve elde edilen metrik değerleri Çizelge 5.8'de gösterilmektedir. Bu çizelgedeki değerlendirmeler sonucunda Makine öğrenmesi başarımı %97,06 olarak bulunmaktadır. Önerilen modelin başarısı için her bir yöntemin toplam analiz ettiği hesaplar içerisindeki spam hesap oranlarını dikkate alınmaktadır. Makine öğrenmesinin çalışmasında kullanılan algoritmaların

başarımının yüksek olması, tespit edilen hesapların ihmal edilebilecek az bir yanılma ile doğru olduğunun bir göstergesidir. Çizelgede elde ettiğimiz öğrenme algoritma sonuçları gösterilmektedir.

Çizelge 5.8. Kullanılan algoritmaların detayları ve başarı oranları

Algoritma	Özellik	K = 10	Percentege Split = 66	Başarı Aralığı Ortalaması	Toplam Başarı Ortalaması
NaiveBayes	Doğru Olarak Sınıflandırılan	%92,1633	%91,8269	%91,9951	%97,0611
	Doğru Onaylanmış Oran	0,922	0,918		
	Kappa İstatistiği	0,8405	0,8341		
Random Forest	Doğru Olarak Sınıflandırılan	%99,1837	%97,8365	%98,5101	
	Doğru Onaylanmış Oran	0,992	0,978		
	Kappa İstatistiği	0,9831	0,9556		
IBk	Doğru Olarak Sınıflandırılan	%99,0204	%99,0385	%99,0295	
	Doğru Onaylanmış Oran	0,990	0,990		
	Kappa İstatistiği	0,9798	0,9802		
J48	Doğru Olarak Sınıflandırılan	%98,8571	%97,1154	%97,9863	
	Doğru Onaylanmış Oran	0,989	0,971		
	Kappa İstatistiği	0,9763	0,9399		
JRip	Doğru Olarak Sınıflandırılan	%98,6939	%96,875	%97,7845	
	Doğru Onaylanmış Oran	0,987	0,969		
	Kappa İstatistiği	0,973	0,9355		

Çizelge 5.8'de belirtilen algoritmalar aynı Twitter veri kümesindeki hesapların analizinde kullanılmaktadır. Önerilen modelin her yönteminde aynı veri kümesi işleme alınmaktadır bu yüzden K = 10 ve percentege split = 66 yöntemlerinde birbirine yakın elde edilmektedir. Her bir algoritmanın çalışma mantıkları da farklıdır. Örneğin Random Forest ve J48 Karar ağacı algoritmalarıdır. Kullanılan beş algoritma içerisinde kıyaslama yapıldığı zaman, IBk algoritması %99,0295 oranında ortalama sonuç vererek en başarılı sonuç veren algoritma olduğu anlaşılmaktadır. Buna karşın NaiveBayes algoritması, kullanılan beş algoritma içerisinde en düşük ortalamaya %91,9951 sahip olmaktadır. Bu değerlendirilen beş algoritma %90'nın üzerinde başarı göstermektedir. Makine öğrenmesi uygulamalarında bu oran ve üstü oranlar gayet başarılı sonuçlar vermektedir [4, 26, 49, 94]. Makine öğrenmesi

sonucunda spam olarak bulunan Twitter hesaplarına, metin analizi destekleyici nitelikte çalışmaktadır [114]. Makine öğrenmesi yöntemimizin başarısını Çizelge 5.8'de belirtildiği gibi %97,06 olarak bulunmaktadır. Metin analizi yöntemimizin başarısı da aynı orandadır. Bu çalışmada analiz edilen Twitter veri kümesindeki hesapların tamamında hassas içerik geçmektedir ve spam olarak işaretlenen hesapların tamamı en az bir metin analizi koşulunu yerine getirmektedir. Bu nedenden dolayı, metin analizi yöntemimizin amacı makine öğrenmesinin başarısını arttırmak olarak kabul edilmektedir. Spam olarak bulunacak bütün hesaplarda hassas içerik bulunması bize makine öğrenmesinin başarısı ile aynı başarıyı metin analizinde yakaladığımızı göstermektedir. Makine öğrenmesi sonucunda belirtilen spam hesapların iletileri içerisindeki hassas içeriklere bakılmaktadır. Link analizi, makine öğrenmesi ve metin analizi yöntemlerinin değerlendirilmesi sonucunda geliştirdiğimiz sosyal ağlara yönelik öğrenmeye dayalı bir spam hesap tespit modelimizin başarısı %96,23 olarak karşımıza çıkmaktadır. Çizelge 5.9'te önerilen model sonucunda bulunan başarı detaylarını gösterilmektedir.

Çizelge 5.9. Önerilen model bileşen detayları

Link Analizi	
Link Analizi Spam	1 870
Link Analizi Spam Değil	68 062
Link Analizi Analiz Edilen Toplam Hesap	69 932
Link Başarı Oranı	100%
Makine Öğrenmesi	
Makine Öğrenmesi Spam	14 237
Makine Öğrenmesi Spam Değil	1 315
Makine Öğrenmesi Analiz Edilen Toplam Hesap	15 552
Makine Öğrenmesi Oranı	% 91,54
Metin Analizi	
Metin Analizi Spam	62 073
Metin Analizi Spam Değil	1 822
Metin Analizi Analiz Edilen Toplam Hesap	63 895
Metin Analizi Oranı	%97,15

Önerilen modelin başarım sonuçları	
Model Toplam Spam Değil	3 137
Model Toplam Spam Sayısı	78 180
Model Toplam Analiz Edilen Hesap Sayısı	81 317
Model Başarım Oranı	%96,23

Çizelge 5.9'da önerilen modelde üç bileşen bulunmaktadır. Kısa link analizi yönteminde VirusTotal tarafından yapılan kontrollerde bütün kötücül linkler tespit edilmiştir. Başarı oranımız %100'dür. Makine öğrenmesi modelimizin başarısı, uygulanan algoritmaların

başarısına yakındır. Yapılan işlemlerde farklı algoritmaların kullanılması bize yakın sonuç vermektedir. Kontroller sonrasında makine öğrenmesinin başarısı %91,54 olarak gözükmektedir. Metin analizi başarımız ise %97,15 olarak hesaplanmaktadır. Modelin toplam başarısı bu üç bileşenin ortalaması olarak alındığında yaklaşık %96,23 hesaplanmaktadır. Link analizinin dinamik yapısı model başarı oranını değiştirebilmektedir. Link analizinde spam olarak işaretlenen hesapların hepsi kötücül link paylaşmışlardır. Kalan 68 062 hesap içerisinde yine spam vardır ancak bunu link analizi tespit edememektedir. Bunu makine öğrenmesi ve metin analizi yöntemleri tespit etmek durumundalardır. Kalan 68 062 adet hesap içerisinde iletilen linklerin içerisinde VirusTotal‘e göre kötücül link bulunmamaktadır. Kötücül ve masum olan hesapların tamamı Link analizi yöntemine göre tespit edilmiştir. Makine öğrenmesinin sonucunda en düşük başarılı algoritma %91,99 oranında başarım göstermektedir. Model başarımının hesaplanmasında üç modelin başarımının ortalaması alınmaktadır. Kısa link analizi bileşeni başarı oranı %100’dür, Makine öğrenmesi bileşeni başarı oranı %91,54’tür ve metin analizi bileşeninin başarı oranı %97,15 olarak hesaplanmaktadır. Önerilen modelin başarısı ise bu sonuçlara göre %96,23 olarak ortaya çıkmaktadır.

6. SONUÇ

Bu tezde, önerilen öğrenmeye dayalı spam hesap tespit modelinin veri kümesinde ve canlı sosyal ağlar üzerindeki başarımları incelenmiştir. Daha sonrasında önerilen modelin üç bileşeni ayrı ayrı çalıştırılmıştır ve her bir yöntemin sonuçları gerçek sosyal ağ üzerinde kontrol edilmiştir. Her bir yöntemin sonucunda işaretlenen hesaplar, veri kümesinin oluşturulma ve analiz edilme zamanı arasındaki fark göz önünde bulundurularak tekrar sosyal ağ üzerinde sorgulanmıştır. Sosyal ağlara yönelik öğrenmeye dayalı bir spam hesap tespit modeli ile spam hesapların Twitter üzerinde tespit edilmesine çalışılmaktadır. Önerilen modelin oluşturulmasında, Twitter'da, spam ileti ve kötücül hesapların tespitinde kullanılan ve farklı açılardan yaklaşan literatürdeki çalışmalar göz önünde bulundurulmaktadır. Sosyal ağlara yönelik öğrenmeye dayalı yeni bir spam hesap tespit modeli, yapılan bu incelemeler sonucu şekillenmektedir.

Önerilen yöntemde; kullanılan metrikler, değerlendirilen algoritmalar, veri kümeleri ve değerlendirme kriterleri literatürde geçtikleri gibi hesaplanmaktadır. Bu tezde önerilen modelde üç adet bileşen kullanılmaktadır. Link analizi, makine öğrenmesi ve metin analizi yöntemleri bir araya getirilerek yeni bir öğrenmeye dayalı spam hesap tespit modeli geliştirilmiştir. Link analizinde, VirusTotal Apı si kullanılmıştır. Kötücül ve masum linkler sosyal hesap iletilerinden ayıklanarak analiz edilmiştir. Kötücül link paylaşan hesaplar başka hiçbir işleme tabi tutulmadan direk spamdır şeklinde etiketlenmiştir.

Link analizin başarımı, bu yöntemin oluşturulan veri kümesinden bulduğu spam hesapların analizi sonucunda %100 olarak bulunmaktadır. Makine öğrenmesi yöntemimizde beş farklı algoritma kullanılmıştır. Bu algoritmalardan IBk algoritması %99,0295 oranla en başarılı algoritma olarak bulunmuştur. Makine öğrenmesi yönteminde, oluşturulan veri kümesindeki spam hesapları tespit etmek için etiketlenmiş bir veri kümesi kullanılması önemlidir. Eğitim küme setinde kullanılan verinin hatasız olması elde edilecek sonucu ve önerilen modelin başarısını etkilemektedir. Yapılan hesaplamalar ve değerlendirmeler sonucunda makine öğrenmesi yönteminde yaklaşık %92'nin üzerinde bir başarıyı yakalandığı gözükmektedir. Metin analizi yöntemi, sosyal hesaplardan atılan iletiler içerisinde hassas içerik bulunup bulunmadığına bakmaktadır. Makine öğrenmesi sonucunda bulduğumuz sonuca ek

sağlamak için metin analizi yöntemi geliştirilmiştir. Bu da yapılan çalışmanın geçerliğini ve güvenirliğini arttırmaktadır.

Bu çalışmanın Literatürde kısmında, incelenen çalışmaların birbirlerine karşı avantaj ve dezavantajaları, elde ettikleri doğruluk değerleri, kullandıkları ölçüm metrikleri, algoritma ve teknikler ayrıntılarıyla ele alınmıştır. Geliştirilen yöntemde ölçüm metrikleri, avantaj ve dezavantajları, algoritma, teknik ayrıntılar elde edilen doğruluk değerleri kıyaslanmıştır. Link analizi, makine öğrenmesi ve metin analizi yöntemlerinin başarım ortalamasını modelin başarısı olarak dikkate alınmıştır. Tüm yapılan inceleme ve sonuçlar ışığında, önerilen modelin başarım oranının %96,23 olduğunu hesaplanabilmektedir. Ayrıca, sosyal ağ spam tespit politikasının geliştirilmesi için önceden çekilen veri kümelerinin analizinin önemi kadar, canlı sosyal ağ veri kümelerinin de analizinin önemli olduğu düşünülmektedir.

KAYNAKLAR

1. Erdoğan, G., Bahtiyar, Ş. (2014). Sosyal Ağlarda Güvenlik. *Akademik Bilişim Konferansı*, Mersin, 1-6.

2. İnternet: Twitter Unsafe Links. URL: http://www.webcitation.org/query?url=https%3A%2F%2Fhelp.twitter.com%2Fen%2Fsafety-and-security%2Fphishing-spam-and-malware-links&date=2018-10-14, Son Erişim Tarihi: 06.09.2018.

3. Buecheler, T., Sieg, J. H., Füchslin, R. M. and Pfeifer, R. (2010). Crowdsourcing, open innovation and collective intelligence in the scientific method: a research agenda and operational framework. In *The 12th International Conference on the Synthesis and Simulation of Living Systems, Odense, Denmark*, MIT Press, 679-686.

4. Wagh, B., Shinde, J. V. and Kale, P. A. (2018). A Twitter Sentiment Analysis Using NLTK and Machine Learning Techniques. *International Journal of Emerging Research in Management and Technology*, 12, 37-44.

5. Ahmed, F., Abulaish, M. (2013). A generic statistical approach for spam detection in Online Social Networks. *Computer Communications, 36*(10-11), 1120-1129.

6. Stringhini, G., Kruegel, C. and Vigna, G. (2010). Detecting spammers on social networks. *In Proceedings of the 26th annual computer security applications conference,*, USA, 1-9.

7. İnternet: Türkiye İstatistik Kurumu Araştırması. URL: http://www.webcitation.org/query?url=http%3A%2F%2Fwww.tuik.gov.tr%2FPreTablo.do%3Falt_id%3D1028&date=2018-10-14, Son Erişim Tarihi: 06.09.2018.

8. Hambrick, M. E. (2012). Six degrees of information: Using social network analysis to explore the spread of information within sport social networks. *International Journal of Sport Communication*, *5*(1), 16-34.

9. Stephen, A. T., Galak, J. (2012). The effects of traditional and social earned media on sales: A study of a microlending marketplace. *Journal of Marketing Research*, *49*(5), 624-639.

10. İnternet: Digital Reports on Global Overview. URL: http://www.webcitation.org/query?url=https%3A%2F%2Fwearesocial.com%2Fuk%2Fspecial-reports%2Fdigital-in-2017-global-overview&date=2018-10-14, Son Erişim Tarihi: 04.04.2018.

11. Yavanoğlu, U., Sağıroğlu, Ş., ve Çolak, İ (2012). Sosyal ağlarda bilgi güvenliği tehditleri ve alınması gereken önlemler. *Politeknik Dergisi*, 15(1), 8-12.

12. Watts, D. J., Dodds, P. S. (2007). Influentials, networks, and public opinion formation. *Journal of consumer research*, 34(4), 441-458.

13. Öztürk, M. F., Talas, M. (2015). Sosyal medya ve eğitim etkileşimi. *Zeitschrift für die Welt der Türken/Journal of World of Turks*, 7(1), 101-120.

14. Peleja, F., Santos, J. and Magalhães, J. (2014). Ranking linked-entities in a sentiment graph. In *Proceedings of the 2014 IEEE/WIC/ACM International Joint Conferences on Web Intelligence (WI) and Intelligent Agent Technologies (IAT)*, IEEE Computer Society, 118-125.

15. İnternet: Popular Social Networking Sites on eBizMBA. URL: http://www.webcitation.org/query?url=http%3A%2F%2Fwww.ebizmba.com%2Farticles%2Fsocial-networking-websites&date=2018-10-14, Son Erişim Tarihi: 06.06.2018.

16. İnternet: Cisco Annually Security Report. URL: http://www.webcitation.org/query?url=http%3A%2F%2Fwww.cisco.com%2Fc%2Fm%2Fen_us%2Foffers%2Fsc04%2F2016-annual-security+report%2Findex.html%3FKeyCode%3D001031952&date=2018-10-14, Son Erişim Tarihi: 11.04.2018.

17. Kabakus, A. T., Kara, R. (2017). A Survey of Spam Detection Methods on Twitter. *International Journal of Advanced Computer Science and Applications*, 8(3), 5.

18. İnternet: Twitter Listelerini Kullanma URL: http://www.webcitation.org/query?url=https%3A%2F%2Fhelp.twitter.com%2Ftr%2Fusing-twitter%2Ftwitter-lists&date=2018-10-14, Son Erişim Tarihi: 05.09.2018.1

19. Kim, D., Jo, Y., Moon, I. C. and Oh, A. (2010). Analysis of twitter lists as a potential source for discovering latent characteristics of users. In *ACM CHI workshop on microblogging*, Seul, 4-9.

20. Tardelli, S. (2017). Social media e finanza: studio dei cashtags su Twitter per l'individuazione di spam e pattern di utilizzo anomali, *Tesi di laurea*, Universita di Pisa, Pisa, 22-53.

21. Mateen, M., Iqbal, M. A., Aleem, M. and Islam, M. A.1(2017). A hybrid approach for spam detection for Twitter. *In Applied Sciences and Technology (IBCAST),14th International Bhurban Conference on*, IEEE, 446-471.

22. Yıldırım, N., Varol, A. (2013). Sosyal ağlarda güvenlik: Bitlis Eren ve Fırat Üniversitelerinde gerçekleştirilen bir alan çalışması. *Türkiye Bilişim Vakfı Bilgisayar Bilimleri ve Mühendisliği Dergisi*, 6(1), 14-17.

23. Rashidi, B., Fung, C. J. (2015). A Survey of Android Security Threats and Defenses. *Journal of Wireless Mobile Network,*, 6(4), 3-35.

24. Osborne, M., Moran, S., McCreadie, R., Von Lunen, A., Sykora, M. D., Cano, E. and Jackson, T. (2014). Real-time detection, tracking, and monitoring of automatically discovered events in social media, 50, 15-20.

25. Egele, M., Stringhini, G., Kruegel, C., and Vigna, G. (2013). Compa: Detecting compromised accounts on social networks. *In the Network and Distributed System Symposium*, San Diego, CA 1-8.

26. Holmes, G., Donkin, A., Witten, I. H. (1994, November). Weka: A machine learning workbench. In *Intelligent Information Systems, 1994. Proceedings of the 1994 Second Australian and New Zealand Conference on*, IEEE, 357-361.

27. Dener, M., Dörterler, M. ve Orman, A. (2009). Açık Kaynak Kodlu Veri Madenciliği Programları: Weka'da Örnek Uygulama. *Akademik Bilişim*, 9, 11-13.

28. Akiyama, M., Yagi, T., Yada, T., Mori, T. and Kadobayashi, Y. (2017). Analyzing the ecosystem of malicious URL redirection through longitudinal observation from honeypots. *Computers & Security*, 69, 155-173.

29. Gupta, N., Aggarwal, A. and Kumaraguru, P. (2014). *bit. ly/malicious: Deep dive into short url based e-crime detection. In Electronic Crime Research (eCrime*), APWG Symposium on, IEEE Computer Society, 14-24.

30. Hasan, M., Orgun, M. A., and Schwitter, R. (2017). A survey on real-time event detection from the twitter data stream. *Journal of Information Science*, 0165551517698564.

31. Li, Y., Sundaramurthy, S. C., Bardas, A. G., Ou, X., Caragea, D., Hu, X. and Jang, J. (2015). Experimental study of fuzzy hashing in malware clustering analysis. In *8th workshop on cyber security experimentation and test*, USENIX Association Washington, DC,5-52.

32. Benevenuto, F., Magno, G., Rodrigues, T. and Almeida, V. (2010). Detecting spammers on twitter. In *Collaboration, electronic messaging, anti-abuse and spam conference (CEAS)*,Washington, 12-15.

33. Verma, M., Sofat, S. (2014). Techniques to detect spammers in twitter-a survey. *International Journal of Computer Applications*, 85(10).

34. Edwards, C., Edwards, A., Spence, P. R. and Shelton, A. K. (2014). Is that a botrunning the social media feed? Testing the differences in perceptions of communication quality for a human agent and a bot agent on Twitter. *Computers in Human Behavior*, 33, 372-376.

35. Ntoulas, A.,Najork, M., Manasse, M. and Fetterly, D.(2006). Detecting spam web pages through content analysis. *the 15th international conference on World Wide Web*,ACM, 83-92.

36. Reeves, B., Nass, C. I. (1996). *The media equation: How people treat computers, television, and new media like real people and places*. Cambridge university press.

37. Fernandes, M. A., Patel, P. and Marwala, T. (2015). Automated detection of human users in Twitter. *Procedia Computer Science*, 53, 224-231.

38. Clark, E. M., Williams, J. R., Jones, C. A., Galbraith, R. A., Danforth, C. M. and Dodds, P. S. (2016). Sifting robotic from organic text: a natural language approach for detecting automation on Twitter. *Journal of computational science*, 16, 1-7.

39. Wu, F., Shu, J., Huang, Y. and Yuan, Z. (2016). Co-detecting social spammers and spam messages in microblogging via exploiting social contexts. *Neurocomputing*, 201, 51-65.

40. Chen, C., Wen, S., Zhang, J., Xiang, Y., Oliver, J., Alelaiwi, A. and Hassan, M. M. (2017). Investigating the deceptive information in Twitter spam. *Future Generation Computer Systems*, 72, 319-326.

41. Cao, Q., Sirivianos, M., Yang, X. and Pregueiro, T. (2012). Aiding the detection of fake accounts in large scale social online services. In *Proceedings of the 9th USENIX conference on Networked Systems Design and Implementation*, USENIX Association, 15-16.

42. Boshmaf, Y., Logothetis, D., Siganos, G., Lería, J., Lorenzo, J., Ripeanu, M. and Beznosov, K. (2015). Integro: Leveraging Victim Prediction for Robust Fake Account Detection in OSNs. In *NDSS*, 15, 8-11.

43. Romo, J., Araujo, L. (2013). Detecting malicious tweets in trending topics using a statistical analysis of language. *Expert Systems with Applications*, 8, 2992-3000.

44. Wang, A. H. (2010). Don't follow me: Spam detection in twitter. In *Security and cryptography (SECRYPT), proceedings of the 2010 international conference on*, IEEE, 1-10.

45. Hacıefendioğlu, Ş. (2011). Reklam ortamı olarak sosyal paylaşım siteleri ve bir araştırma. *Dergipark Bilgi Ekonomisi ve Yönetimi Dergisi*, 6(1).

46. Jeong, S., Noh, G., Oh, H. and Kim, C. K. (2016). Follow spam detection based on cascaded social information. *Information Sciences*, 369, 481-499.

47. Bücheler, T., Sieg, J. H. (2011). Understanding science 2.0: Crowdsourcing and open innovation in the scientific method. *Procedia Computer Science*, 7, 327-329.

48. Brabham, D. C. (2008). Crowdsourcing as a model for problem solving: An introduction and cases. *Convergence*, 14(1), 75-90.

49. Wang, A. H. (2010). Machine learning for the detection of spam in twitter networks. In *International Conference on E-Business and Telecommunications*, Springer, Berlin, Heidelberg, 319-333.

50. Liu, S., Wang, Y., Zhang, J., Chen, C. and Xiang, Y. (2017). Addressing the class imbalance problem in twitter spam detection using ensemble learning. *Computers & Security*, 69, 35-49.

51. Blanzieri, E., Bryl, A. (2008). A survey of learning-based techniques of email spam filtering. *Artificial Intelligence Review*, 29(1), 63-92.

52. Sagi, E., Dehghani, M. (2014). Moral rhetoric in Twitter: A case study of the US Federal Shutdown of 2013. In *Proceedings of the Annual Meeting of the Cognitive Science Society*, 36, 36-40.

53. Lee, S., Kim, J. (2014). Early filtering of ephemeral malicious accounts on Twitter. *Computer Communications*, 54, 48-57.

54. Chan, P. P., Yang, C., Yeung, D. S. and Ng, W. W. (2015). Spam filtering for short messages in adversarial environment. *Neurocomputing*, 155, 167-176.

55. Davidov, D., Tsur, O. and Rappoport, A. (2010). Enhanced sentiment learning using twitter hashtags and smileys. In *Proceedings of the 23rd international conference on computational linguistics: posters*, Association for Computational Linguistics, 241-249.

56. Miller, Z., Dickinson, B., Deitrick, W., Hu, W. and Wang, A. H. (2014). Twitter spammer detection using data stream clustering. *Information Sciences*, 260, 64-73.

57. Mathioudakis, M., Koudas, N. (2010). Twittermonitor: trend detection over the twitter stream. In *Proceedings of the 2010 ACM SIGMOD International Conference on Management of data*, ACM, 1155-1158.

58. Dagon, D., Qin, X., Gu, G., Lee, W., Grizzard, J., Levine, J. and Owen, H. (2004). Honeystat: Local worm detection using honeypots. In *International Workshop on Recent Advances in Intrusion Detection*, Springer, Berlin, Heidelberg, 39-58.

59. Yang, C., Zhang, J. and Gu, G. (2014). A taste of tweets: reverse engineering Twitter spammers. In *Proceedings of the 30th annual computer security applications conference*, ACM, 86-95.

60. Hansen, D. L., Shneiderman, B. and Smith, M. A. (2010). *Analyzing social media networks with NodeXL: Insights from a connected world*. Morgan Kaufmann.

61. İnternet: Twitter Communities URL: http://www.webcitation.org/query?url=https%3A%2F%2Fdeveloper.twitter.com%2Fen%2Fcommunity%23%2C&date=2018-10-14, Son Erişim Tarihi: 18.05.2018.

62. Bray, T. (2017). *The javascript object notation (json) data interchange format*, RFC, 8259.

63. İnternet: Twitter Uygulama Sayfası URL: http://www.webcitation.org/query?url=https%3A%2F%2Fapps.twitter.com%2F&date=2018-10-14, Son Erişim Tarihi: 01.09.2018.

64. İnternet: Twitter Geliştirici Platformu URL: http://www.webcitation.org/query?url=https%3A%2F%2Fdeveloper.twitter.com%2Fen%2Fdocs%2Ftweethttps%3A%2F%2Fdeveloper.twitter.com%2Fen%2Fdocs%2Fbasics%2Fgetting-starteds%2Fdata-dictionary%2Foverview%2Fintro-to-tweet-json.html%2C&date=2018-10-14, Son Erişim Tarihi: 12.09.2018.

65. İnternet: Libraries Built for the Twitter Platform URL: http://www.webcitation.org/query?url=https%3A%2F%2Fdeveloper.twitter.com%2Fen%2Fdocs%2Fdeveloper-utilities%2Ftwitter-libraries&date=2018-10-14, Son Erişim Tarihi: 01.06.2018.

66. Van Rossum, G. (2007). Python Programming Language. In *USENIX Annual Technical Conference*, 41, 36.

67. Roesslein, J. (2009). tweepy Documentation. Tweepy. *Python programming language module*, 3, 5.

68. Tapkan, P. Z., Özmen, T. (2018). Determining the spam quality by feature selection and classification in a social media. *Pamukkale University Journal of Engineering Sciences*, 4, 713-719.

69. İnternet: Twitter'da Askıya Alınmış Bir Hesap URL: http://www.webcitation.org/query?url=https%3A%2F%2Ftwitter.com%2Fsearch%3Fq%3Ddonnelllebla41%26src%3Dtypd&date=2018-10-14, Son Erişim Tarihi: 21.09.2018.

70. İnternet: Twitter'da Tweet JSON URL: http://www.webcitation.org/query?url=https%3A%2F%2Fdeveloper.twitter.com%2Fen%2Fdocs%2Ftweets%2Fdata-dictionary%2Foverview%2Fintro-to-tweet-json.html%2C&date=2018-10-14, Son Erişim Tarihi: 18.05.2018.

71. İnternet: VirusTotal, Ücretsiz Çevrimiçi Virüs, Kötü Amaçlı Yazılım ve URL Tarayıcı URL:http://www.webcitation.org/query?url=https%3A%2F%2Fwww.virustotal.com%2Ftr%2F&date=2018-10-14, Son Erişim Tarihi: 12.08.2018.

72. Thomas, K., Nicol, D. M. (2010). The Koobface botnet and the rise of social malware. In *Malicious and Unwanted Software 5th International Conference on*, IEEE, 63-70.

73. Cranor, L. F., Egelman, S., Hong, J. I. and Zhang, Y. (2007). Phinding Phish: An Evaluation of Anti-Phishing Toolbars. In *the 11th Annual Network and Distibuted System Security Symposium*, Sen Diego, CA.

74. Gerbet, T., Kumar, A. and Lauradoux, C. (2016). A privacy analysis of Google and Yandex safe browsing. In *Dependable Systems and Networks (DSN),46th Annual IEEE/IFIP International Conference on*, IEEE, 347-358.

75. Canali, D., Bilge, L. and Balzarotti, D. (2014). On the effectiveness of risk prediction based on users browsing behavior. In *Proceedings of the 9th ACM symposium on Information, computer and communications security*, ACM, 171-182.

76. Mihai, I. C., Giurea, L. (2016). Management of eLearning platforms security. In *The International Scientific Conference eLearning and Software for Education*, Carol National Defence University,1-422.

77. Devi, P., Kumar, S. and Sharma, N. (2016). Botnet Malicious Activity Detection Based on DNS Traffic Analysis, *International Journal of Research in Advent Technology*, 4(7).

78. Garera, S., Provos, N., Chew, M. and Rubin, A. D. (2007). A framework for detection and measurement of phishing attacks. In *Proceedings of the 2007 ACM workshop on Recurring malcode*, ACM., 1-8

79. İnternet: VirusTotal'da Taranmış Masum Bir Link URL: http://www.webcitation.org/query?url=https%3A%2F%2Fwww.virustotal.com%2Ftr%2Furl%2F4bb0b8e9d55bcf67ba39317cce876c7df22aa83d5285b21c768d7b1f4e2ed397%2Fanalysis%2F1539519713%2F&date=2018-10-14, Son Erişim Tarihi: 10.07.2018.

80. İnternet: VirusTotal'da Taranmış Kötücül Bir Link URL: http://www.webcitation.org/query?url=https%3A%2F%2Fwww.virustotal.com%2Ftr%2Furl%2Fec4212c4829cdafacd82dd2c49e16b7f443fc455837ebfc7df9f5fd11090b764%2Fanalysis%2F1539519935%2F&date=2018-10-14, Son Erişim Tarihi: 12.08.2018.

81. İnternet: VirusTotal Script URL: http://www.webcitation.org/query?url=https%3A%2F%2Fwww.virustotal.com%2Ftr%2Fdocumentation%2Fpublic-api%2F&date=2018-10-14, Son Erişim Tarihi: 22.03.2018.

82. İnternet: Twitter'da URL Bağlantı Gönderme URL: http://www.webcitation.org/query?url=https%3A%2F%2Fhelp.twitter.com%2Ftr%2Fusing-twitter%2Fhow-to-tweet-a-link&date=2018-10-14, Son Erişim Tarihi: 28.08.2018.

83. İnternet: Tez Hazırlanma Sırasında Twitter'da Testlerin Yapıldığı Bir Hesap URL: http://www.webcitation.org/query?url=https%3A%2F%2Ftwitter.com%2Fpanteraoguz&date=2018-10-14, Son Erişim Tarihi: 22.06.2018.

84. Nizam, H., Akın, S. S. (2014). *Sosyal medyada makine öğrenmesi ile duygu analizinde dengeli ve dengesiz veri setlerinin performanslarının karşılaştırılması.* XIX. Türkiye'de İnternet Konferansı, Bornova, İzmir.

85. Witten, I. H., Frank, E. (2000). Weka. *Machine Learning Algorithms in Java*, 265-320.

86. Alan, M. A. (2012). Veri madenciliği ve lisansüstü öğrenci verileri üzerine bir uygulama. *Dumlupınar Üniversitesi Sosyal Bilimler Dergisi*, 33, 5-25.

87. Garner, S. R. (1995). *Weka: The waikato environment for knowledge analysis.* In Proceedings of the New Zealand computer science research students conference, Nez Zealand, 57-64.

88. Baskin, I. I., Marcou, G., Horvath, D. and Varnek, A. (2017). Cross-Validation and the Variable Selection Bias. *Tutorials in Chemoinformatics*, 163-173.

89. İnternet: K Katmanlı Çarpraz Doğrulama URL: http://www.webcitation.org/query?url=http%3A%2F%2Fbilgisayarkavramlari.sadievrenseker.com%2F2013%2F03%2F31%2Fk-fold-cross-validation-k-katlamali-carpraz-dogrulama%2F&date=2018-10-14, Son Erişim Tarihi: 21.09.2018.

90. Patil, T. R., Sherekar, S. S. (2013). Performance analysis of Naive Bayes and J48 classification algorithm for data classification. *International journal of computer science and applications*, *6*(2), 256-261.

91. Ren, J., Lee, S. D., Chen, X., Kao, B., Cheng, R. and Cheung, D. (2009). Naive bayes classification of uncertain data. In *Data Mining, Ninth IEEE International Conference on*, IEEE, 944-949.

92. Genuer, R., Poggi, J. M., and Tuleau-Malot, C. (2010). Variable selection using random forests. *Pattern Recognition Letters*, *31*(14), 2225-2236.

93. Akar, Ö., Güngör, O. (2012). Classification of multispectral images using Random Forest algorithm. *Journal of Geodesy and Geoinformation*, 2, 105-112.

94. Breiman, L. (2001). Random forests. *Machine learning*, 45(1), 5-32.

95. Moradian, M., Baraani, A. (2009). KNNBA: K-Nearest Neighbor Based Association Algorithm. *Journal of Theoretical & Applied Information Technology*, 6(1).

96. Kaur, G., Chhabra, A. (2014). Improved J48 classification algorithm for the prediction of diabetes. *International Journal of Computer Applications*, 98(22).

97. Bhargava, N., Sharma, G., Bhargava, R. and Mathuria, M. (2013). Decision tree analysis on j48 algorithm for data mining. *Proceedings of International Journal of Advanced Research in Computer Science and Software Engineering*, 3(6), 8-12.

98. Rajput, A., Aharwal, R. P., Dubey, M., Saxena, S. P. and Raghuvanshi, M. (2011). J48 and JRIP rules for e-governance data. *International Journal of Computer Science and Security (IJCSS)*, 5(2), 201.

99. Zhang, T. (2011). Sparse recovery with orthogonal matching pursuit under RIP. *IEEE Transactions on Information Theory*, *57*(9), 6215-6221.

100. Kaur, G., Chhabra, A. (2014). Improved J48 classification algorithm for the prediction of diabetes. *International Journal of Computer Applications*, 98(22), 76-80.

101. Kaynar, O., Görmez, Y., Yıldız, M. Albayrak, A. (2016). Makine Öğrenmesi Yöntemleri ile Duygu Analizi. In *International Artificial Intelligence and Data Processing Symposium (IDAP'16)*, 17-18.

102. Landis, J. R., Koch, G. G. (1977). The measurement of observer agreement for categorical data, *biometrics*, 2(5), 159-174.

103. Karegowda, A. G., Manjunath, A. S. and Jayaram, M. A. (2010). Comparative study of attribute selection using gain ratio and correlation based feature selection. *International Journal of Information Technology and Knowledge Management*, 2(2), 271-277.

104. Platt, J. (1998). Sequential minimal optimization. *A fast algorithm for training support vector machines*, 4(3), 13-16.

105. Quinlan, J. R. (1986). Induction of decision trees. *Machine learning*, 1(1), 81-106.

106. Aha, D. W., Kibler, D. and Albert, M. K. (1991). Instance-based learning algorithms. *Machine learning*, 6(1), 37-66.

107. Willmott, C. J., Matsuura, K. (2005). Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance. *Climate research*, 30(1), 79-82.

108. Baldry, A., Thibault, P. J. (2005). *Multimodal transcription and text analysis*. London: Equinox, 26.

109. Nasukawa, T., Nagano, T. (2001). Text analysis and knowledge mining system. *IBM systems journal*, 40(4), 967-984.

110. Song, J., Lee, S. and Kim, J. (2011). Spam filtering in twitter using sender-receiver relationship. In *International workshop on recent advances in intrusion detection*, Springer, Berlin, Heidelberg, 301-317.

111. İnternet: Twitter T.CO Formatlı Kısa Link URL: http://www.webcitation.org/query?url=http%3A%2F%2F55.%09https%3A%2F%2Fhelp.twitter.com%2Ftr%2Fusing-twitter%2Furl-shortener&date=2018-10-14, Son Erişim Tarihi: 01.09.2018.

112. Priyadharshini, J. M. H., Kavitha, S. and Bharathi, B. (2017). Classification and analysis of human activities. In *Communication and Signal Processing (ICCSP), International Conference on*, IEEE, 1207-1211.

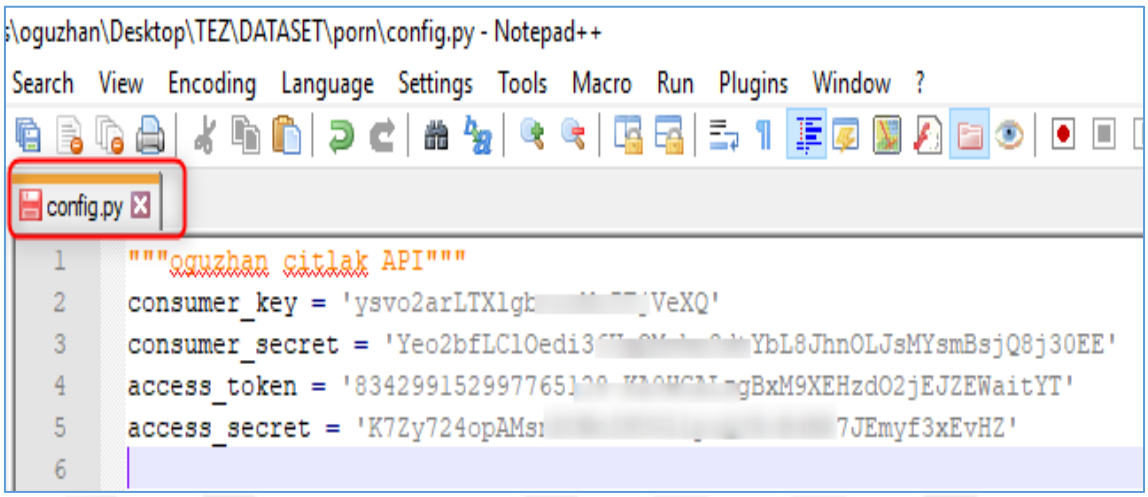
113. Han, J., Pei, J. and Kamber, M. (2011). *Data mining: concepts and techniques*, Elsevier, 5-75.

114. Luyckx, K., Daelemans, W. (2005). Shallow text analysis and machine learning for authorship attribtion. *LOT Occasional Series*, 4, 149-160.

115. Citlak, O., Doğru, İ. A., and Dorterler, M. (2017). A Spam Detection System with Short Link Analysis. *Cyber Security and Artificial Intelligence (ISC Turkey 2017)*, Ankara, 151-155.

EKLER

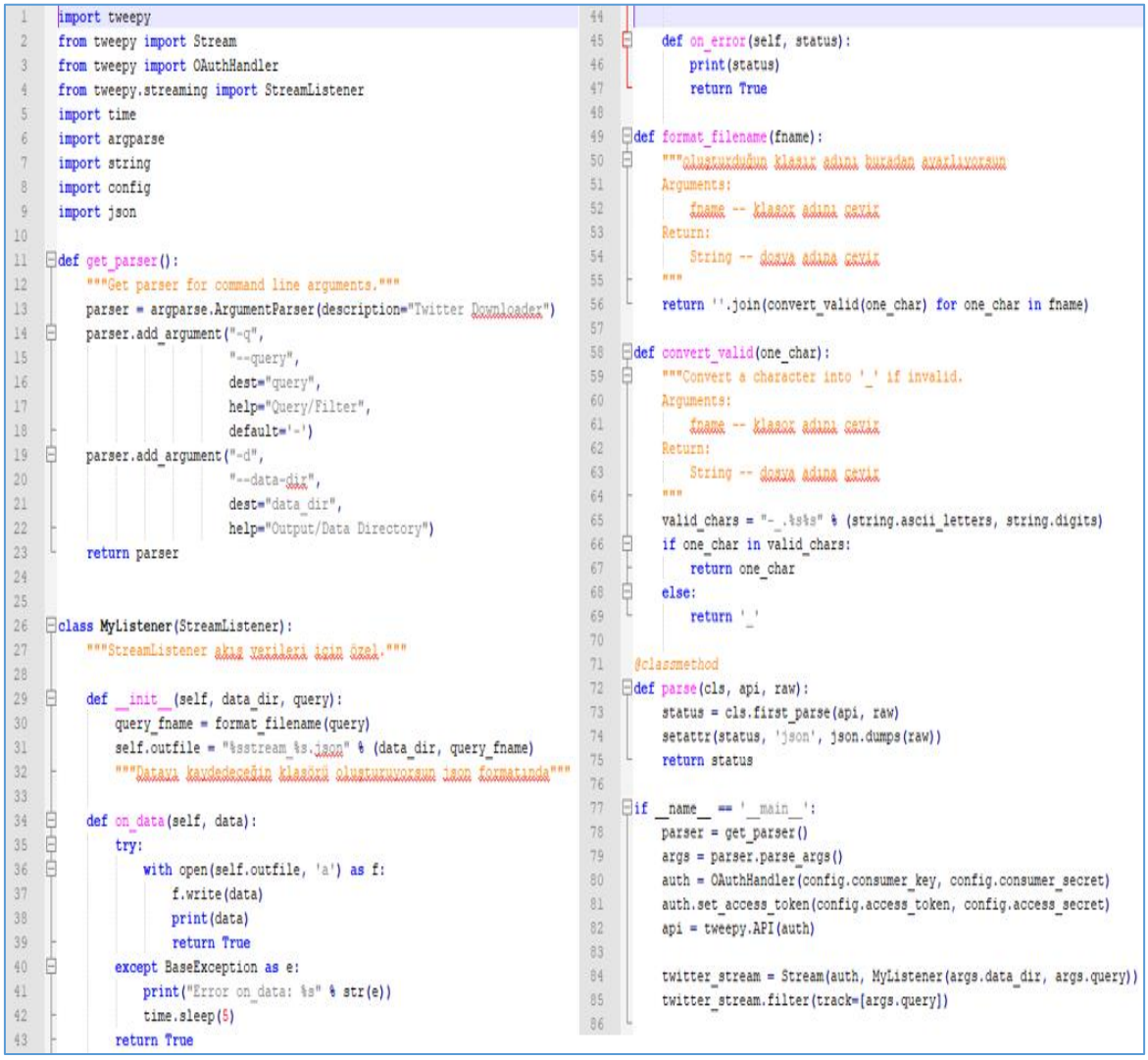
EK-1. Twitter API config.py kod parçacığı



EK-2. Çekilen veri kümesi içerisindeki kullanıcı özellikleri

No	Öznitelik	Tür	Açıklama	Örnek	Örnekteki Değer
1	attributes	Object	Contains a hash of variant information about the place.	attributes:{"street_address":"795 Folsom	Örnek Hesapta bulunamadı
2	bounding_box	Object	A bounding box of coordinates which encloses this place	"bounding_box":{"coordinates":[[[2.2241006,48.8155414], [2.4699099,48.8155414],	Örnek Hesapta bulunamadı
3	contributors	Katkıda Bulunanlar (lagency)	Tweet'e katkıda bulunanın gösterir, çok sık	contributors:[{"id":819797,"id_str ":"819797","screen_name":"epis	"contributors":null,
4	contributors_enabled	Boolean	Indicates that the user has an account with "contributor mode" enabled, allowing for Tweets issued by the user to be co-	"contributors_enabled": false	"contributors_enabled":false,
5	coordinates	coordinates	Kullanıcı yada Tweet atan uygulamanmın geographic lokasyonunu belirler.The	coordinates:{"coordinates":[-75.14310264,40.05701649],"typ e":"Point"}	"coordinates":null,
6	country	String	Name of the country containing this	"country":"France"	Örnek Hesapta bulunamadı
7	country_code	String	Shortened country code representing the country containing this place.	"country_code":"FR"	Örnek Hesapta bulunamadı
8	created_at	String	Tweet'in oluşturulduğu UTC zamanı	"created_at":"Wed Aug 27	"created_at":"Tue May 23 20:37:40 +0000 2017",
9	default_profile	Boolean	When true, indicates that the user has not altered the theme or background of	"default_profile": false	"default_profile":true,
10	default_profile_image	Boolean	When true, indicates that the user has not uploaded their own profile image	"default_profile_image": false	"default_profile_image":false,
11	description	String	The user-defined UTF-8 string describing	"description": "The Real Twitter API."	"description":null,
12	entities	entities	Entities which have been parsed out of the text of the Tweet.	entities:{"hashtags":[],"urls":[],"u ser_mentions":[]}	"entities":{"hashtags":[],"urls":[],"user_mentio ns":[]
13	entities	entities	Entities which have been parsed out of the url or description fields defined by the user.	entities:{"url": {"urls": [{"url": "http://dev.twitter.com", "expanded_url": null,"indices": [0, 22]}]},"description": {"urls":[] }}	"entities":{"hashtags":[],"urls":[],"user_mentio ns":[],"symbols":[],"media":[{"id":86591018734 9217280,"id_str":"865910187349217280","indice s":[24,47],"media_url":"http:\/\/pbs.twimg.co m\/ext_tw_video_thumb\/8659101873492172 80\/pu\/img\/MaaTVpSHdII7ZeQd.jpg","media_ url_https":"https:\/\/pbs.twimg.com\/ext_tw_ video_thumb\/865910187349217280\/pu\/img\ /MaaTVpSHdII7ZeQd.jpg","url":"https:\/\/t.co\
14	favorite_count	Integer	Indicates approximately how many times	favorite_count:1138	"favorite_count":0,
15	favorited	Boolean	Perspectival Indicates whether this Tweet	favorited:true	"favorited":false,
16	favourites_count	Int	The number of Tweets this user has liked	"favourites_count": 13	"favourites_count":0,
17	filter_level	String	Indicates the maximum value of the ilter_level parameter which may be used and still stream this Tweet. So a value	filter_level: "medium"	"filter_level":"low",
18	follow_request_sent	Type	When true, indicates that the authenticating user has issued a follow request to this	"follow_request_sent": false	"follow_request_sent":null,
19	followers_count	Int	The number of Tweets this user has liked	"followers_count": 21	"followers_count":54,
20	following	Type	Deprecated. When true, indicates that the authenticating user is following this user. Some false negatives are possible when set to "false," but these false	"following": true	"following":null,
21	friends_count	Int	The number of users this account is following (AKA their "followings"). Under certain conditions of duress, this field will temporarily indicate "0"	"friends_count": 32	"friends_count":20,
22	full_name	String	Full human-readable representation of the place's name.	"full_name":"San Francisco, CA"	Örnek Hesapta bulunamadı
23	geo	Object	Use the coordinates field instead.	Örnek Belirtilmemiş	"geo":null,
24	geo_enabled	Boolean	When true, indicates that the user has enabled the possibility of geotagging their Tweets. This field must be true for the current user to attach geographic	"geo_enabled": true	"geo_enabled":false,
25	hashtags	Array of Object	Represents hashtags which have been parsed out of the Tweet text	"hashtags":[{"indices":[32,36],"t ext":"lol"}]	"hashtags":[],
26	id	Int64	The integer representation of the unique identifier for this Tweet.This number is greater than 53 bits and some programming languages may have difficulty/silent defects in interpreting it.Using a signed 64 bit integer	"id":114749583439036416	"id":867117349282959360,
27	id	Int64	The integer representation of the unique identifier for this User. This number is greater than 53 bits and some programming languages may have difficulty/silent defects in interpreting it. Using a signed	"id": 6253282	Örnek Hesapta bulunamadı
28	id	String	ID representing this place. Note that this is represented as a string, not an	"id":"7238f93a3e899af6"	Örnek Hesapta bulunamadı

EK-3. Twitter'dan veri seti çekilirken kullanılan kod parçaçığı



```python
import tweepy
from tweepy import Stream
from tweepy import OAuthHandler
from tweepy.streaming import StreamListener
import time
import argparse
import string
import config
import json

def get_parser():
    """Get parser for command line arguments."""
    parser = argparse.ArgumentParser(description="Twitter Downloader")
    parser.add_argument("-q",
                        "--query",
                        dest="query",
                        help="Query/Filter",
                        default='-')
    parser.add_argument("-d",
                        "--data-dir",
                        dest="data_dir",
                        help="Output/Data Directory")
    return parser


class MyListener(StreamListener):
    """StreamListener akış verileri için özel."""

    def __init__(self, data_dir, query):
        query_fname = format_filename(query)
        self.outfile = "%sstream_%s.json" % (data_dir, query_fname)
        """Datayı kaydedeceğin klasörü oluşturuyoruz json formatında"""

    def on_data(self, data):
        try:
            with open(self.outfile, 'a') as f:
                f.write(data)
                print(data)
                return True
        except BaseException as e:
            print("Error on_data: %s" % str(e))
            time.sleep(5)
        return True

    def on_error(self, status):
        print(status)
        return True

def format_filename(fname):
    """oluşturduğun klasör adını buradan ayarlıyorsun
    Arguments:
        fname -- klasör adını cevir
    Return:
        String -- dosya adına cevir
    """
    return ''.join(convert_valid(one_char) for one_char in fname)

def convert_valid(one_char):
    """Convert a character into '_' if invalid.
    Arguments:
        fname -- klasör adını cevir
    Return:
        String -- dosya adına cevir
    """
    valid_chars = "-_.%s%s" % (string.ascii_letters, string.digits)
    if one_char in valid_chars:
        return one_char
    else:
        return '_'

@classmethod
def parse(cls, api, raw):
    status = cls.first_parse(api, raw)
    setattr(status, 'json', json.dumps(raw))
    return status

if __name__ == '__main__':
    parser = get_parser()
    args = parser.parse_args()
    auth = OAuthHandler(config.consumer_key, config.consumer_secret)
    auth.set_access_token(config.access_token, config.access_secret)
    api = tweepy.API(auth)

    twitter_stream = Stream(auth, MyListener(args.data_dir, args.query))
    twitter_stream.filter(track=[args.query])
```

EK-4. Veri setinde örnek bir spam hesabın Json formatında detayları

{"created_at":"Tue May 23 20:37:40 +0000 2017","id":867117349282959360,"id_str":"867117349282959360","text":"best lesbian porn sex videos https:\/\/t.co\/6rFmMqWMOI","source":"\u003ca href=\"http:\/\/twitter.com\" rel=\"nofollow\"\u003eTwitter Web Client\u003c\/a\u003e","truncated":false,"in_reply_to_status_id":null,"in_reply_to_status_id_str":null,"in_reply_to_user_id":null,"in_reply_to_user_id_str":null,"in_reply_to_screen_name":null,"user":{"id":856815625758543872,"id_str":"856815625758543872","name":"Gabrielle Russell","screen_name":"LoErA5Ngngo7EEr","location":null,"url":null,"description":null,"protected":false,"verified":false,"followers_count":54,"friends_count":20,"listed_count":1,"favourites_count":0,"statuses_count":9972,"created_at":"Tue Apr 25 10:22:18 +0000 2017","utc_offset":null,"time_zone":null,"geo_enabled":false,"lang":"ru","contributors_enabled":false,"is_translator":false,"profile_background_color":"F5F8FA","profile_background_image_url":"","profile_background_image_url_https":"","profile_background_tile":false,"profile_link_color":"1DA1F2","profile_sidebar_border_color":"C0DEED","profile_sidebar_fill_color":"DDEEF6","profile_text_color":"333333","profile_use_background_image":true,"profile_image_url":"http:\/\/pbs.twimg.com\/profile_images\/860717475897516036\/a3AzHpcQ_normal.jpg","profile_image_url_https":"https:\/\/pbs.twimg.com\/profile_images\/860717475897516036\/a3AzHpcQ_normal.jpg","default_profile":true,"default_profile_image":false,"following":null,"follow_request_sent":null,"notifications":null},"geo":null,"coordinates":null,"place":null,"contributors":null,"is_quote_status":false,"retweet_count":0,"favorite_count":0,"entities":{"hashtags":[],"urls":[],"user_mentions":[],"symbols":[],"media":[{"id":865910187349217280,"id_str":"865910187349217280","indices":[24,47],"media_url":"http:\/\/pbs.twimg.com\/ext_tw_video_thumb\/865910187349217280\/pu\/img\/MaaTVpSHdII7ZeQd.jpg","media_url_https":"https:\/\/pbs.twimg.com\/ext_tw_video_thumb\/865910187349217280\/pu\/img\/MaaTVpSHdII7ZeQd.jpg","url":"https:\/\/t.co\/6rFmMqWMOI","display_url":"pic.twitter.com\/6rFmMqWMOI","expanded_url":"https:\/\/twitter.com\/BeataPowell121\/status\/865910230122737664\/video\/1","type":"photo","sizes":{"large":{"w":640,"h":480,"resize":"fit"},"small":{"w":340,"h":255,"resize":"fit"},"medium":{"w":600,"h":450,"resize":"fit"},"thumb":{"w":150,"h":150,"resize":"crop"}},"source_status_id":865910230122737664,"source_status_id_str":"865910230122737664","source_user_id":860199653949788162,"source_user_id_str":"860199653949788162"}]},"extende

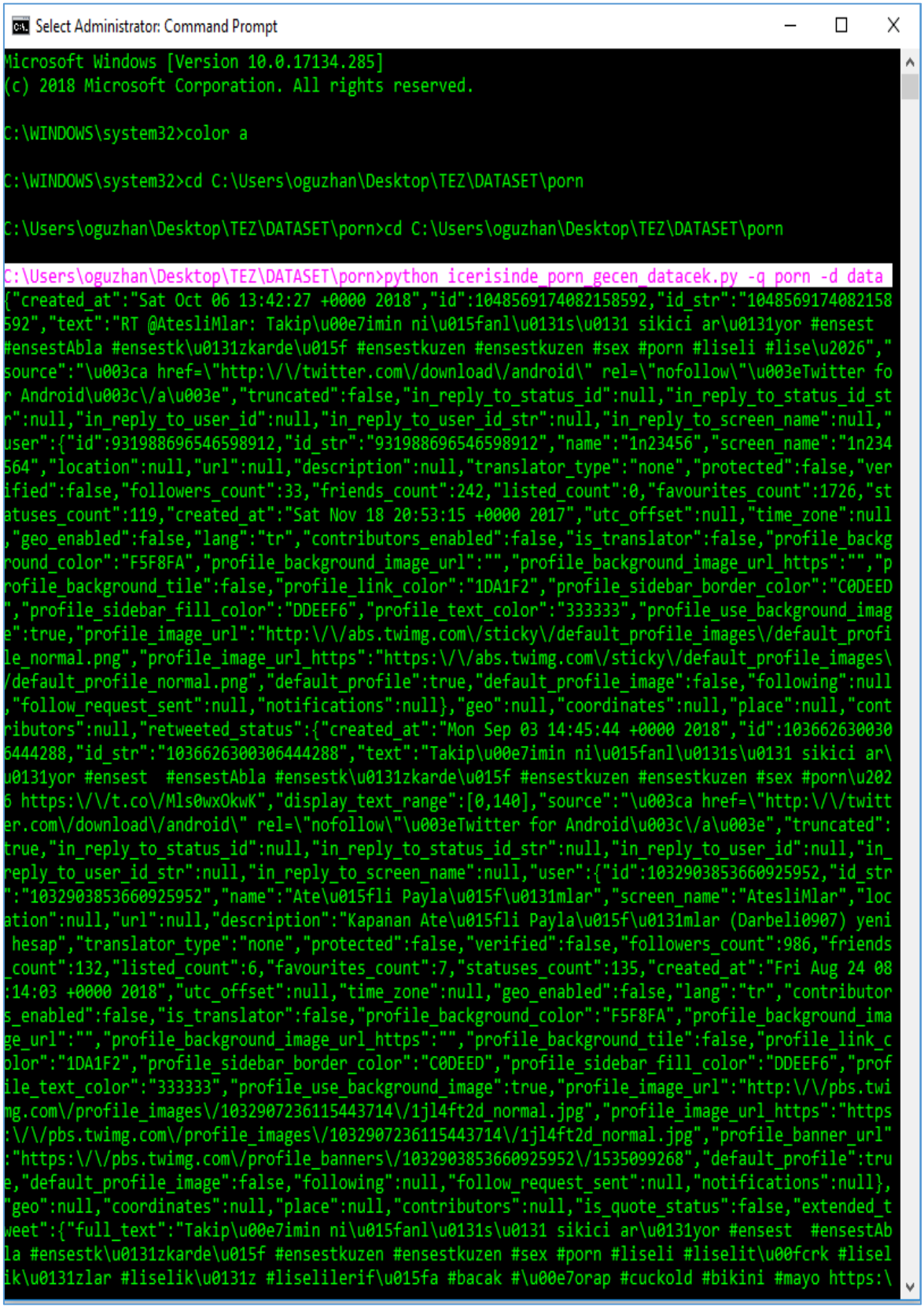
EK-4.(devam) Veri setinde örnek bir spam hesabın Json formatında detayları

d_entities":{"media":[{"id":865910187349217280,"id_str":"865910187349217280","indices":[24,47],"media_url":"http:\/\/pbs.twimg.com\/ext_tw_video_thumb\/865910187349217280\/pu\/img\/MaaTVpSHdII7ZeQd.jpg","media_url_https":"https:\/\/pbs.twimg.com\/ext_tw_video_thumb\/865910187349217280\/pu\/img\/MaaTVpSHdII7ZeQd.jpg","url":"https:\/\/t.co\/6rFmMqWMOI","display_url":"pic.twitter.com\/6rFmMqWMOI","expanded_url":"https:\/\/twitter.com\/BeataPowell121\/status\/865910230122737664\/video\/1","type":"video","sizes":{"large":{"w":640,"h":480,"resize":"fit"},"small":{"w":340,"h":255,"resize":"fit"},"medium":{"w":600,"h":450,"resize":"fit"},"thumb":{"w":150,"h":150,"resize":"crop"}},"source_status_id":865910230122737664,"source_status_id_str":"865910230122737664","source_user_id":860199653949788162,"source_user_id_str":"860199653949788162","video_info":{"aspect_ratio":[4,3],"duration_millis":10360,"variants":[{"bitrate":832000,"content_type":"video\/mp4","url":"https:\/\/video.twimg.com\/ext_tw_video\/865910187349217280\/pu\/vid\/480x360\/PcHZ_chImfSBDIp_.mp4"},{"bitrate":320000,"content_type":"video\/mp4","url":"https:\/\/video.twimg.com\/ext_tw_video\/865910187349217280\/pu\/vid\/240x180\/OQYpe7vKC1itDliN.mp4"},{"content_type":"application\/x-mpegURL","url":"https:\/\/video.twimg.com\/ext_tw_video\/865910187349217280\/pu\/pl\/-
9p2I_jPxsqdbjYA.m3u8"}]}}]},"favorited":false,"retweeted":false,"possibly_sensitive":false,"filter_level":"low","lang":"en","timestamp_ms":"1495571860987"}

EK-5. Metin analizi işleminde kullanılan hassas içerikli kelimeler

#free,#mature,#porn,#hot,#xxx,#sex,#nude,#adult,#teen,#pornvideos,#erotic,#bet,#bitcoin,#pussy,#ass,#cam,#boobs,#hooker,#tits,#massive,#um,#babe,#beat,#casino,#bonus,#naked,#gays,#erotizm,#fantasy,#panty,#tips,#blockchain.

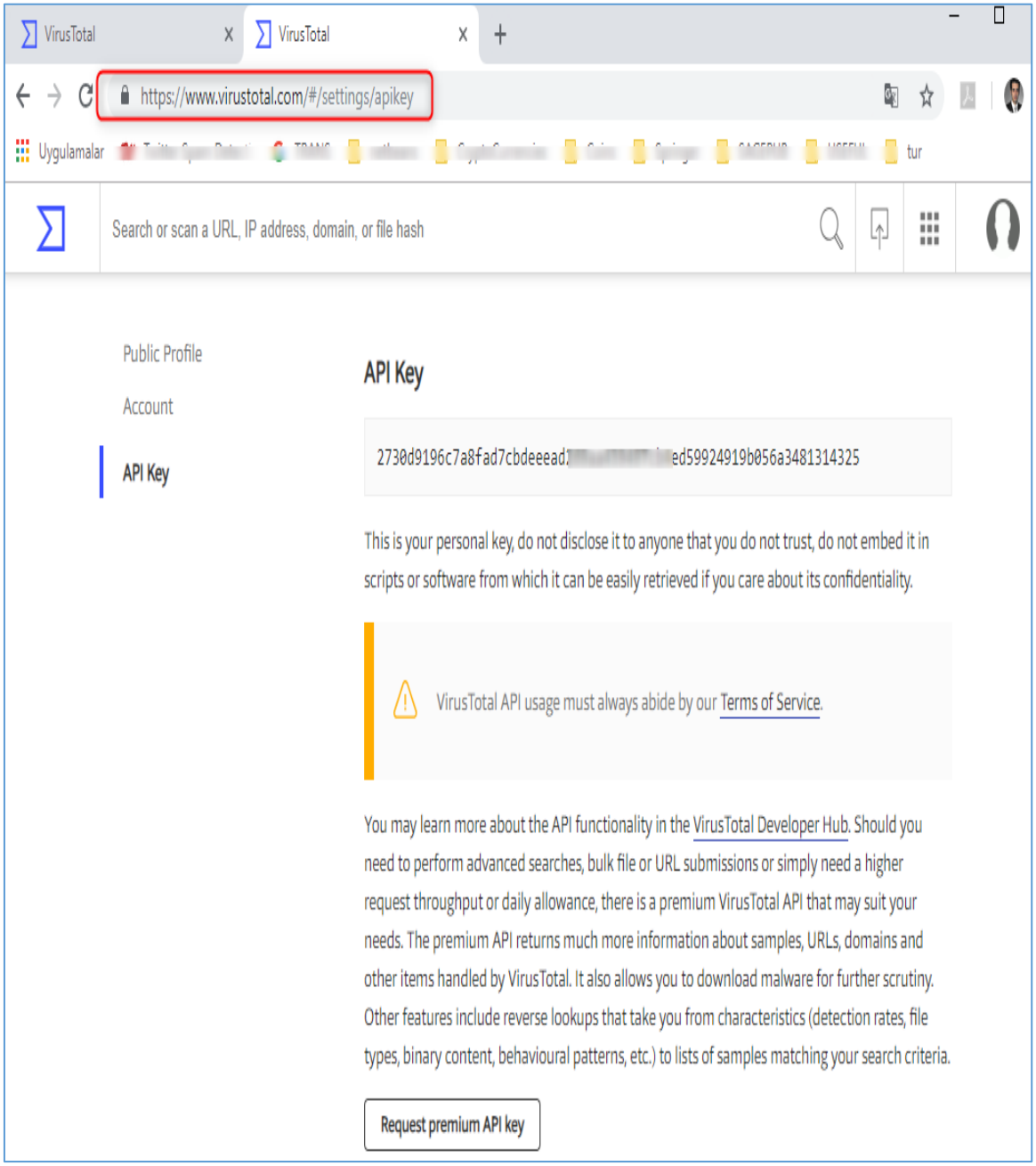
EK-6. Hassas içerikli kelime ile yapılan örnek bir arama



EK-7 Hassas içerikli kelimelerin “text” içerisindeki görüntüsü

	#free	#nsfw	#mature	#bet	#hot	#xxx	#sex	#nude	#adult	#teen	#erotic
text:"RT @VICE: Meet the woman creating lesbian Mormon bet in Utah	0	0	0	1	0	0	0	0	0	0	0
text:"#big booby bet sexuality photo https:\/\/t.co\/6t0zd4eeaE"	0	0	0	1	0	0	0	0	0	0	0
text:"sean michaels bet https:\/\/t.co\/GmyELvGCMM"	0	0	0	1	0	0	0	0	0	0	0
text:"#free bet 4u rocker bet https:\/\/t.co\/zDdaWZbbWA"	1	0	0	1	0	0	0	0	0	0	0
text:"#sailor moon bet games naked soccer game https:\/\/t.co\/MNa0EN5ZqZ"	0	0	0	1	0	0	0	0	0	0	0
text:"simpsons animation bet https:\/\/t.co\/fZwEfajO4Q"	0	0	0	1	0	0	0	0	0	0	0
text:"#mature women lesbian bet realteen sex pics https:\/\/t.co\/LMLBx4eBB5"	0	0	1	1	0	0	1	0	0	1	0
text:"Gay bet scheren schaamhaar Een Sadistische val voor Twink Scott https:\/\/t.co\/WQ205HuyBP"	0	0	0	1	0	0	0	0	0	0	0
text:"toolbar bet https:\/\/t.co\/yM8Ln56dRz"	0	0	0	1	0	0	0	0	0	0	0
text:"#free mature and young bet videos uncensored teen bet https:\/\/t.co\/iSXpXHOsDN"	1	0	1	1	0	0	0	0	0	1	0
text:"#movie bet site beto mexicanas gratis https:\/\/t.co\/HYhpjEqdH6"	0	0	0	1	0	0	0	0	0	0	0
text:"#hot pussy tgp orgasmics pussies bet movies galleries https:\/\/t.co\/c7sehijqiO"	0	0	0	1	1	0	0	0	0	0	0
text:"most viewed bet ever https:\/\/t.co\/cj5wyUSJde"	0	0	0	1	0	0	0	0	0	0	0
text:"#bet teens foto lupe free birthday sex rawsession https:\/\/t.co\/LtKHL8GrCz"	1	0	0	1	0	0	1	0	0	1	0
text:"#elve bet cartoon sexuality https:\/\/t.co\/VvFJST5ZZm"	0	0	0	1	0	0	1	0	0	0	0
text:"#free sex stores filipino bet pictures https:\/\/t.co\/LwuDjMY9b3"	1	0	0	1	0	0	1	0	0	0	0
text:"#movies shemale bet teenbrother sister fucking https:\/\/t.co\/bWvYVnPOoO"	0	0	0	1	0	0	0	0	0	1	0
text:"#sex chart free bet videos anime https:\/\/t.co\/FvlaALJcHm"	1	0	0	1	0	0	1	0	0	0	0
text:"#free non-sexual nude videos cockroach bet https:\/\/t.co\/8WapCjKocF"	1	0	0	1	0	0	1	0	0	1	0
text:"#gta iv naked bet suffolk https:\/\/t.co\/ChZcBee172"	0	0	0	1	0	0	0	0	0	0	0
text:"13 Strange Facts About The bet Industry You Probably Didn't Know https:\/\/t.co\/7RFylAKgfP"	0	0	0	1	0	0	0	0	0	0	0
text:"#soft bet shows ebony teen blowjob gifs https:\/\/t.co\/clVhxGrAo7"	0	0	0	1	0	0	0	0	0	1	0
text:"#amateur interracial video hd bet download https:\/\/t.co\/mXqIn55Iuq"	0	0	0	1	0	0	0	0	0	0	0
text:"@PSchydlowski Best bet ever"	0	0	0	1	0	0	0	0	0	0	0
text:"#ca bet #betogratis debutante #nylon #fuckass dope https:\/\/t.co\/9SFsePF6NF"	0	0	0	1	0	0	0	0	0	0	0
text:"#kendra wilkinson sex pics anime free bet games https:\/\/t.co\/xCMfkg11vQ"	1	0	0	1	0	0	1	0	0	0	0
text:"#no to bet fullscreenteen https:\/\/t.co\/LHCFb4E8iD"	0	0	0	1	0	0	0	0	0	1	0
text:"#triple play sex miss candy bet https:\/\/t.co\/0HfSIPXDGX"	0	0	0	1	0	0	1	0	0	0	0
text:"#violent anal sex brazilian bet tgp https:\/\/t.co\/pFl004ogki"	0	0	0	1	0	0	1	0	0	0	0
text:"neuseeland teen bet https:\/\/t.co\/vSdEUzI9jO"	0	0	0	1	0	0	0	0	0	1	0
text:"#our free bet black man fucking white woman images https:\/\/t.co\/mKfBSmjdnk"	1	0	0	1	0	0	0	0	0	0	0
text:"full length mobile bet movies https:\/\/t.co\/jlJPuaDfnQ"	0	0	0	1	0	0	0	0	0	0	0
text:"#thong girls gallery message boards bet https:\/\/t.co\/a51eym5rIP"	0	0	0	1	0	0	0	0	0	0	0
text:"RT @betHardd: #Follow & #Retweet \"the new #erotic #bet\" betHardd	0	0	0	1	0	0	0	0	0	0	1
text:"#free hentei bet videos sexy fuck free video https:\/\/t.co\/umUzIVQ7yW"	1	0	0	1	0	0	1	0	0	0	0
text:"#home granny bet download full bet movie free https:\/\/t.co\/T7D41nUi9j"	1	0	0	1	0	0	0	0	0	0	0
text:"#adults with dementia oslo lesbian bet https:\/\/t.co\/MO8Ms8xFyL"	0	0	0	1	0	0	0	0	1	0	0
text:"#milk sexy skin head girl bet https:\/\/t.co\/fhXQWsbt75"	0	0	0	1	0	0	0	0	0	0	0
text:"#bet stars models naked torrents https:\/\/t.co\/UC3QSmhK79"	0	0	0	1	0	0	0	0	0	0	0
text:"#free nude webchat free onepiece bet https:\/\/t.co\/P6kfmlRwGw"	1	0	0	1	0	0	0	0	0	1	0
text:"#free mature sex online tube bet hairy https:\/\/t.co\/AwNxqgo37u"	1	0	1	1	0	0	1	0	0	0	0
text:"#sexy teens making love muscle bet stars https:\/\/t.co\/fjQSnHnjtq"	0	0	0	1	0	0	1	0	0	1	0
text:"anal bet thumbs https:\/\/t.co\/YW2sQFU7NX"	0	0	0	1	0	0	0	0	0	0	0
text:"#tight pussy syndrome view bet images https:\/\/t.co\/1SA0rfoZ43"	0	0	0	1	0	0	0	0	0	0	0
text:"hot girls bet pic https:\/\/t.co\/LrcYyoMtA6"	0	0	0	1	1	0	0	0	0	0	0
text:"#4shared fuck bet japanese hot girls hypnotism sex videos https:\/\/t.co\/6pMDFp8iqI"	0	0	0	1	1	0	1	0	0	0	0
text:"#black bet big dicks top adult video sites https:\/\/t.co\/LTX34sqvND"	0	0	0	1	0	0	0	0	1	0	0

EK-8. VirusTotal'daki hesabın API detayı



EK-9. Arff dosyasının örnek içeriği

```
%     Sosyal Aglarda Bir Spam Hesap Tespit Modeli ve Uygulamasi
%     Oğuzhan Catlak
%     1225 Tane Hesap

@relation 'Sosyal Aglarda Bir Spam Hesap Tespit Modeli ve Uygulamasi'
%
@attribute User_Statuses_Count {'0-99','100-199','200-299','300-399','400-499','500-599','600-699','700-799','800-899','900-
@attribute Sensitive_Content_Alert {'FALSE','TRUE'}
@attribute User_Favourites_Count {'0-9','10-19','20-29','30-39','40-49','50-59','60-69','70-79','80-89','90-99','100-199','2
@attribute User_Listed_Count {'0-9','10-19','20-29','30-39','40-49','50-59','60-69','70-79','80-89','90-99','100-199','200-2
@attribute Source_in_Twitter {'yes','no'}
@attribute 'User_Friends_Counts' {'0-9','10-19','20-29','30-39','40-49','50-59','60-69','70-79','80-89','90-99','100-199','2
@attribute 'User_Followers_Count' {'0-9','10-19','20-29','30-39','40-49','50-59','60-69','70-79','80-89','90-99','100-199','
@attribute 'User_Created_at' {'2006-2008','2009-2011','2012-2014','2015-2017'}
@attribute 'User_Location' {'yes','no'}
@attribute 'User_Geo_Enabled' {'FALSE','TRUE'}
@attribute 'User_Default_Profile_Image' {'FALSE','TRUE'}
@attribute 'ReTweet' {'FALSE','TRUE'}
@attribute 'Account Suspender' {'TRUE','FALSE'}
@data

'700-799','TRUE','1000-1999','0-9','yes','500-599','800-899','2015-2017','no','FALSE','TRUE','TRUE','FALSE'
'20000-29999','FALSE','20000-29999','700-799','yes','200-299','30000-39999','2006-2008','yes','TRUE','FALSE','TRUE','FALSE'
'9000-9999','FALSE','20-29','0-9','yes','200-299','100-199','2009-2011','no','FALSE','FALSE','FALSE','TRUE'
'1000-1999','FALSE','0-9','0-9','yes','20-29','10-19','2009-2011','yes','FALSE','FALSE','FALSE','TRUE'
'2000-2999','FALSE','10-19','0-9','yes','20-29','20-29','2012-2014','no','FALSE','TRUE','FALSE','TRUE'
'2000-2999','FALSE','0-9','0-9','yes','0-9','0-9','2015-2017','no','FALSE','TRUE','FALSE','TRUE'
'1000-1999','FALSE','10-19','0-9','yes','0-9','0-9','2012-2014','no','FALSE','TRUE','FALSE','TRUE'
'3000-3999','FALSE','30-39','0-9','yes','40-49','20-29','2012-2014','no','FALSE','TRUE','FALSE','TRUE'
'2000-2999','FALSE','30-39','0-9','yes','200-299','200-299','2015-2017','no','FALSE','TRUE','FALSE','TRUE'
'2000-2999','FALSE','30-39','0-9','yes','0-9','10-19','2012-2014','no','FALSE','TRUE','FALSE','TRUE'
'1000-1999','FALSE','20-29','0-9','yes','30-39','20-29','2012-2014','no','FALSE','TRUE','FALSE','TRUE'
'3000-3999','FALSE','40-49','0-9','yes','0-9','20-29','2009-2011','no','FALSE','TRUE','FALSE','TRUE'
'1000-1999','FALSE','10-19','0-9','yes','0-9','20-29','2012-2014','no','FALSE','TRUE','FALSE','TRUE'
'2000-2999','FALSE','0-9','0-9','yes','0-9','10-19','2009-2011','no','FALSE','TRUE','FALSE','TRUE'
'2000-2999','FALSE','0-9','0-9','yes','0-9','10-19','2009-2011','no','FALSE','TRUE','FALSE','TRUE'
'30000-39999','FALSE','900-999','30-39','no','1000-1999','1000-1999','2012-2014','no','FALSE','TRUE','FALSE','FALSE'
'2000-2999','FALSE','10-19','0-9','yes','100-199','100-199','2012-2014','no','FALSE','TRUE','FALSE','TRUE'
'2000-2999','FALSE','10-19','0-9','yes','100-199','70-79','2012-2014','yes','FALSE','FALSE','FALSE','TRUE'
'100000-1999999','TRUE','0-9','80-89','no','400-499','10000-19999','2015-2017','no','FALSE','TRUE','FALSE','FALSE'
'3000-3999','FALSE','0-9','0-9','yes','0-9','20-29','2009-2011','no','FALSE','TRUE','FALSE','TRUE'
'30000-39999','FALSE','900-999','30-39','no','1000-1999','1000-1999','2012-2014','no','FALSE','TRUE','FALSE','FALSE'
'100000-1999999','TRUE','0-9','80-89','no','400-499','10000-19999','2015-2017','no','FALSE','TRUE','FALSE','FALSE'
'200-299','FALSE','100-199','0-9','yes','500-599','20-29','2015-2017','no','FALSE','TRUE','TRUE','FALSE'
'30000-39999','FALSE','10000-19999','0-9','yes','500-599','1000-1999','2012-2014','no','TRUE','TRUE','TRUE','FALSE'
'20000-29999','FALSE','2000-2999','80-89','yes','3000-3999','2000-2999','2012-2014','yes','TRUE','TRUE','TRUE','FALSE'
'2000-2999','FALSE','0-9','0-9','yes','20-29','10-19','2009-2011','yes','FALSE','TRUE','FALSE','TRUE'
'2000-2999','FALSE','0-9','0-9','yes','10-19','10-19','2009-2011','no','FALSE','TRUE','FALSE','TRUE'
'5000-5999','FALSE','6000-6999','0-9','yes','400-499','900-999','2012-2014','no','TRUE','FALSE','TRUE','FALSE'
'1000-1999','FALSE','30-39','0-9','yes','0-9','20-29','2012-2014','yes','TRUE','FALSE','FALSE','TRUE'
'20000-29999','FALSE','0-9','0-9','yes','2000-2999','1000-1999','2015-2017','yes','FALSE','TRUE','FALSE','TRUE'
'1000-1999','FALSE','0-9','0-9','yes','20-29','10-19','2009-2011','yes','FALSE','TRUE','FALSE','TRUE'
'1000-1999','FALSE','0-9','0-9','yes','2000-2999','1000-1999','2015-2017','no','TRUE','TRUE','FALSE','TRUE'
'0-99','FALSE','20-29','0-9','yes','80-89','0-9','2012-2014','yes','FALSE','TRUE','TRUE','FALSE'
'100000-1999999','FALSE','0-9','0-9','yes','200-299','1000-1999','2009-2011','no','FALSE','FALSE','FALSE','FALSE'
'10000-19999','FALSE','20000-29999','0-9','yes','300-399','400-499','2015-2017','no','TRUE','FALSE','TRUE','FALSE'
'4000-4999','TRUE','0-9','0-9','yes','40-49','10-19','2015-2017','yes','FALSE','TRUE','FALSE','FALSE'
'2000-2999','FALSE','0-9','0-9','yes','0-9','20-29','2012-2014','no','FALSE','TRUE','FALSE','FALSE'
'2000-2999','FALSE','0-9','0-9','yes','4000-4999','800-899','2015-2017','no','FALSE','TRUE','TRUE','TRUE'
'1000-1999','TRUE','0-9','0-9','yes','20-29','20-29','2015-2017','no','FALSE','TRUE','FALSE','FALSE'
'200-299','FALSE','100-199','0-9','yes','700-799','100-199','2015-2017','yes','FALSE','TRUE','TRUE','TRUE'
'1000-1999','TRUE','0-9','0-9','yes','10-19','0-9','2015-2017','no','FALSE','TRUE','FALSE','FALSE'
'3000-3999','TRUE','0-9','0-9','yes','0-9','10-19','2012-2014','no','FALSE','TRUE','FALSE','FALSE'
'2000-2999','TRUE','0-9','0-9','yes','0-9','10-19','2009-2011','no','FALSE','TRUE','FALSE','FALSE'
'2000-2999','TRUE','0-9','0-9','yes','0-9','10-19','2009-2011','no','FALSE','TRUE','FALSE','FALSE'
'3000-3999','FALSE','3000-3999','0-9','yes','1000-1999','200-299','2015-2017','no','FALSE','TRUE','TRUE','TRUE'
'2000-2999','TRUE','0-9','0-9','yes','0-9','20-29','2012-2014','yes','TRUE','FALSE','FALSE','FALSE'
'7000-7999','FALSE','10000-19999','10-19','yes','200-299','500-599','2015-2017','yes','FALSE','TRUE','TRUE','TRUE'
'1000-1999','TRUE','0-9','0-9','yes','30-39','10-19','2012-2014','no','FALSE','TRUE','FALSE','FALSE'
'4000-4999','TRUE','0-9','0-9','yes','2000-2999','300-399','2009-2011','yes','FALSE','TRUE','FALSE','FALSE'
'2000-2999','TRUE','0-9','0-9','yes','0-9','10-19','2009-2011','no','FALSE','TRUE','FALSE','FALSE'
'90000-99999','FALSE','70000-79999','100-199','yes','400-499','1000-1999','2009-2011','no','FALSE','TRUE','TRUE','TRUE'
```

EK-10. VirusTotal'da link analizi için kullanılan kod parçacığı

```php
<?php

class VirusTotal
{
    private $_url;
    private $_key;
    private $_result;

    function __construct($virusTotalKey)
    {
        if (!function_exists('curl_version')) {
            trigger_error('cURL is not found!', E_USER_ERROR);
            exit();
        } else {
            if (!empty($virusTotalKey)) {
                $this->_key     = $virusTotalKey;
                $this->_result  = array();
            } else {
                trigger_error('VirusTotal API key is not found!', E_USER_ERROR);
                exit();
            }
        }
    }
    public function setUrl($url = '')
    {
        !empty($url) ? $this->_url = $url : false;
    }

    public function getUrl()
    {
        return($this->_url);
    }

    public function load($url = '')
    {
        !(empty($url)) ? $this->_url = $url : false;

        $post = array('apikey' => $this->_key,'resource'=> $this->_url);
        $curl = curl_init();
        curl_setopt($curl, CURLOPT_URL, 'https://www.virustotal.com/vtapi/v2/url/report');
        curl_setopt($curl, CURLOPT_POST, True);
        curl_setopt($curl, CURLOPT_VERBOSE, 1); // remove this if your not debugging
        curl_setopt($curl, CURLOPT_ENCODING, 'gzip,deflate'); // please compress data
        curl_setopt($curl, CURLOPT_USERAGENT, "gzip, My php curl client");
        curl_setopt($curl, CURLOPT_RETURNTRANSFER ,True);
        curl_setopt($curl, CURLOPT_POSTFIELDS, $post);

        $response = curl_exec ($curl);
        $status = curl_getinfo($curl, CURLINFO_HTTP_CODE);

        curl_close ($curl);

        if ($status == 200) {
            $json = json_decode($response, true);
            return($json['response_code'] == 1 ? $json : null);
        } else {
            return(false);
        }
    }
}
```

ÖZGEÇMİŞ

Kişisel Bilgiler

Soyadı, adı	: ÇITLAK, Oğuzhan
Uyruğu	: T.C.
Doğum tarihi ve yeri	: 22.06.1985, İstanbul
Medeni hali	: Evli
Telefon	: 0 (555) 500 05 29
Faks	: 0 (312) 427 82 75
e-mail	: oguzhan.citlak@gmail.com

Eğitim

Derece	Eğitim Birimi	Mezuniyet Tarihi
Yüksek lisans	Gazi Üniversitesi / Bilgisayar Mühendisliği	Devam Ediyor
Lisans	ODTÜ / Bilgisayar ve Öğretim Tek. Eğitimi	2011
Lise	İstanbul İnönü Endüstri Meslek Lisesi	2002

İş Deneyimi

Yıl	Yer	Görev
2012-Halen	Tüprag Metal Madencilik AŞ / Ankara	Bilgi İşlem Uzmanı
2009-2011	HP Computer Datacenter / ODTÜ / Ankara	Datacenter Asistanı

Yabancı Dil

İngilizce, Almanca

Yayınlar

Çıtlak, O., Doğru İ. A, Dörterler, M. (2017). A Spam Detection System with Short Link Analysis. *10. Uluslararası Bilgi ve Güvenliği Kriptoloji Konferansı Bildiri Dergisi* /Ankara, 178-185

Hobiler

Yüzme, Gitar, Dans,

GAZİ GELECEKTİR..