

## Kalite Ölçekleme Kriterleri ile Sosyal Ağ Hesaplarının Etkinliğinin Belirlenmesi

Hafzullah İŞ<sup>1</sup>, Taner TUNCER<sup>2\*</sup>

<sup>1</sup> Batman Üniversitesi Bilgi İşlem Dairesi Batman Türkiye

<sup>2</sup> Fırat Üniversitesi, Mühendislik Fakültesi Bilgisayar Mühendisliği Elazığ, Türkiye

\*<sup>1</sup> hafzullah.is@batman.edu.tr, <sup>2</sup> ttuncer@firat.edu.tr

(Geliş/Received: 11/04/2018;

Kabul/Accepted: 19/02/2019)

**Özet:** Sosyal ağlar; bilgi edinme, eğitim alma, içerik paylaşma, fikir beyan etme, eğlenme, haberleşme, arkadaşlık ve ticari girişimlerde bulunma gibi bireysel ve kurumsal iletişim ve etkileşime olanak sağlamaktadır. Bu ağlardaki ölçülebilir veriler üzerinde veri madenciliği teknikleri uygulanarak örüntüler tanımlanabilmekte akademik, ticari ve sosyolojik alanlarda kıymetli bilgiler elde edilebilmektedir. Ağdaki eğilimler ve ilişkiler; iş zekası, bilgi keşfi, reklam ve hizmetlerin özelleştirilmesi, suç tespitleri ve toplumsal hareketlerin izlenmesinde kullanılmaktadır. Sosyal ağlarda kullanıcıların trafik istatistikleri ve özellikleri yanı sıra etkinliklerine göre topluluklar tespit edilebilmektedir. Ancak bunların doğru bir şekilde yapılabilmesi için kullanıcı profillerinin etkinlik değerleri ortaya çıkarılmalıdır. Bu çalışmada; sosyal ağlardan çekilen kullanıcı verileri anonimleştirilip, filtrelenmiş ve sonrasında da çap (diameter), yoğunluk (density), karşılıklılık (reciprocity), merkezlilik (centralization), modülerlik (modularity) ve twit aktivitesi gibi nitelikler tespit edilmiştir. Böylece "Profil Kalite Ölçekleme" metodu ile kullanıcıların etkinlik değerleri derecelendirilmiş, gruplandırılmış ve etkin profilleri tanımlanmıştır. Analiz sonucunda algoritmanın başarı oranı %92 olarak elde edilmiştir.

**Anahtar kelimeler:** Sosyal ağ, Kalite ölçekleme kriteri, Profil kalite ölçümü.

### Determining the Effectiveness of Social Network Accounts with Quality Scaling Criteria

**Abstract:** Social networks provide individual and institutional communication and interaction, such as acquiring information, receiving training, sharing content, expressing ideas, having fun, communicating, friendship, and being in business ventures. When Data mining techniques are applied on the scalable data in these networks, the patterns are defined and valuable information is obtained in academic, commercial and sociological fields. Trends and relationships in the network are used in business intelligence, discovery of information, privatization of advertisements and services, monitoring of crime detections and social movements. In social networks, as well as traffic statistics and features of users communities can be determined according to their activities. However, the efficiency values of user profiles must be revealed so that they can be done correctly. In this study; the user data extracted from the social networks are anonymized, filtered and then the diameter, density, reciprocity, centralization, modularity and twit activity have been determined. Thus, with the "Profile Quality Scaling" method, the activity values of the users have been graded, grouped and the active profile has been defined. As a result of the analysis, the performance ratio of the algorithm has been obtained 92%.

**Key words:** Social network, quality Scaling criteria, Profile quality measurement.

#### 1. Giriş

Günümüzde teknolojinin her geçen gün daha çok gelişmesi ile birlikte birey, firma, kurum ve devlet bazında iletişim şeklimiz ve platformlarımız, sosyalleşme alanlarımız, beklenti ve duyarlılıklarımız şeklini değiştiriyor. İnternete kesintisiz erişim olanaklarının artması, sosyal ağların üye sayılarının astronomik şekilde yükselmesine ve hayatın birçok alanında kullanılmasına imkân sağlamaktadır. Sosyal ağlar bireylere arkadaşlık kurmak, haberleşmek, medya takibi yapmak ve sosyal çevrelerini genişletmek amacıyla iletişim ve etkileşimde kalma imkânları sağlamanın yanı sıra devlet hizmetlerinin çevrimiçi olarak vatandaşa ulaştırılması ve firmalara reklam yapma, müşterileri ile iletişimde kalma, ürün ve hizmetler konusunda bildirim alma gibi olanaklar sağlamaktadır. Ayrıca, sosyal ağlar üzerinden elde edilen üyelik, paylaşımlar, yorumlar, beğeni, arkadaşlık ilişki ve etkileşimleri birer veri kaynağı olarak kullanılabilir. Ölçeklenebilir bu veriler üzerinde sosyal ağ analizi ile bireyler arasındaki bağıntılar bulunabilir, profil kalite ölçümleri yapılabilir, profiller kategori bazında ayrıştırılıp incelenebilir. Bu veriler iş zekası uygulamaları, organize suç tespitleri, toplumsal nabız ölçümleri, reklam ve pazarlama ile tanıtım ve halkla ilişkilerde kullanılabilir. Bu yılın heyecan verici en önemli kilometre taşı, dünya nüfusunun yarısından fazlasının, İnternete erişilebilir ve kullanılabilir olmasıdır[1]. İnternet dünya

\* Sorumlu yazar: [ttuncer@firat.edu.tr](mailto:ttuncer@firat.edu.tr) Yazarların ORCID Numarası: <sup>1</sup> ORCID, <sup>2</sup> ORCID

genelinde ortalama %8 oranında gelişirken sosyal ağların kullanım oranı son bir yılda %20 oranında artış gösterdi. Aktif sosyal medya kullanıcı sayısı %30 artış ile 2.5 Milyar kişiye ulaştı. Mobil kullanıcıların sayısı hızlı ve pratik erişimden dolayı artmaya devam etmektedir. İnternet kullanıcılarının 2/3'ü mobil cihazlarından, bunların da %55'i akıllı telefonlarından internete erişim sağlamaktadır.

İngiltere merkezli yapılan ve 2100 kişi üzerinde denenen bir sosyal deneyde telefon kullanıcılarının %50'si telefondan kopmadığını belirtmiş ve "Nomofobi" yani cep telefonu yoluyla iletişimden kopmaktan korkma durumu yaşadığını belirtmiştir. Amazon Alexa[2] verilerine göre dünya çapında en çok ziyaret edilen ilk 500 sitenin yaklaşık 2/3'ü sosyal medya sitelerinden oluşmaktadır. GlobalWebIndex[3] verilerine göre Dünya genelinde kullanıcılar günlük ortalama 4.5 saatlerini internette geçirmekte iken, geçirdikleri vakitlerinin de ortalama 1/3'ünü sosyal ağlarda harcamaktadırlar.

İnternetin hızla yaygınlaşması, sosyal ağların gelişmesine ve dolayısıyla da astronomik oranlarda kullanıcı profiline ortaya çıkmasına sebep teşkil etmektedir. Bu kullanıcıların sosyal medyadaki aktiviteleri bireysel, kurumsal, ticari ve toplumsal çalışma ve eğilimlere yön vermektedir. Firmalar buralardan edindikleri izlenimlerle ürün ve pazar politikaları belirlerken, kurumlar itibar ve halkla iletişim kanalı yönetimleri yapmakta, devletler ise toplumsal nabız ölçümleri ile diplomatik, kültürel ve yatırımsal kararlar alabilmekte ve vizyon çalışmalarına yön verebilmektedir.

Sosyal ağlarda sahte hesaplar rahatsız edici boyutlarda olabilmektedir. Bu hesaplardan yalan haberler, yanıltıcı anketler, spam içerikler yayınlanabilmektedir. Sahte hesaplar fikirleri manipüle ederek yanlış yönlendirebilmekte, hatalı yatırım-değerlendirme kararları almada etkili olabilmektedir. Bunun yanı sıra zararlı link paylaşımları, manipüle edilmiş takipçi-takip sayıları, yasaklı spam içerik paylaşımları da yapabilmektedirler.

Bu derece etkili olabilen sosyal medya kullanıcılarının profillerinin incelenmesi, gerçekliklerinin analiz edilmesi ve bunların topluluk tespitleri yoluyla etkinliklerinin ortaya çıkarılması akademik, bilimsel ve toplumsal açıdan sağlayacağı faydalar dolayısıyla önem arz etmektedir. "Status People Fake Account" uygulaması "sahte", "iyi" ve "pasif" olarak twitter hesaplarının sınıflandırmasını yapabilmektedir. Ancak, bu sınıflandırmayı sadece twit analiz yaparak gerçekleştirdiğinden çok yönlü bir uygulama değildir.

Bu çalışmada, Twitter'dan halka açık kullanıcı profilleri ve ham verileri elde edilmiş veriler anonimleştirilerek kimlikleri gizlenmiştir. Veriler filtrelenmiş gereksiz bilgilerden arındırılarak işe yarar veri elde edilmiştir. Elde edilen bu verideki tüm kullanıcıların profil bilgileri üzerinde Anma ve Cevaplama kategorisinde olmak üzere sosyal ağ için önemli çap, yoğunluk, karşılıklılık, merkezileşme, modülerlik ve twit aktivitesi parametreleri bazında ölçümler yapılmış "Profil Ölçekleme Metodu" ile kalite dereceleri ortaya çıkarılmıştır. Ölçülen bu veriler kullanılarak tüm profiller; popüler, aktif, normal, pasif ve sahte hesap olarak sınıflandırılmış ve kümülatif bazda bir hesabın takipçilerinin genel kalite seviyeleri-nin ortaya çıkarılmasına çalışılmıştır. Böylece bir kullanıcının takipçilerinin profillerinden hesabının kalitesi ölçülenebilmektedir. Sosyal hesaplarda yoğun takipler, paylaşımlar, etkileşimler birçok konuda manipülasyona sebebiyet verebildiğinden bu çalışma profillerin gerçek değerlerinin ortaya çıkarılmasında akademik, ticari ve siyasi girişimler açısından önemlidir.

Makalenin geri kalan kısmı aşağıdaki gibi organize edilmiştir. İkinci bölümde literatür taraması ile akademik çalışmalardan bahsedilmiştir. Üçüncü bölümde sosyal ağlardan veri toplama ve toplanan verinin anonimleştirilip filtrelenmesinden bahsedilmiştir. Daha sonra Profil Ölçekleme Metodu açıklanarak Sosyal medya hesaplarının sınıflandırma işlemi sunulmuştur. Son bölümde elde edilen sonuçlar tartışılmıştır.

## 2. Literatür Taraması

Sosyal ağların popüler hale gelmesiyle birlikte bu alandaki spam trafik ve sahte hesap sayılarında ciddi problemler oluşturacak, ortamı manipüle edecek şekilde artışlar meydana gelmiştir. İçerik filtreleme bir yere kadar etkili olabilirken kullanıcı bazında sahte hesap tespiti ve engellenmesi çalışmaları önem kazanmıştır. Sosyal ağ kullanıcılarının profil bilgilerini incelemek suretiyle gerçek hesap olup olmadıkları, spam trafik oluşturup oluşturmadıkları ile ilgili yapılmış akademik çalışmalar mevcuttur[4-12]. Bu çalışmalar ağırlıklı olarak kullanıcıların oluşturduğu trafiği incelemeye odaklanmışken bir kısmı da kullanıcının sahte olup olmadığını mesaj tra-fiğinin şekline ve kullanıcının ilişki ağına göre değerlendirmektedir[13]. Bu çalışmaların tamamı zararlı trafiğin ve kullanıcıların tespitine yöneliktir. Hiçbiri bireysel, kurumsal veya ticari hesabın takipçilerinin profil kalite değerlerini ortaya çıkaran ve dolayısıyla da bir sosyal ağ hesabının etkisini ölçebilecek çalışmalar değildir.

Sosyal ağ hesapları birçok nitelikten oluşmaktadır. Bu niteliklerin her biri kullanıcının profiline ortaya çıkarılmasında önem arz etmektedir. Somut olarak ortaya çıkarılabilecek nitelikler çap, yoğunluk, karşılıklılık,

merkezileşme, modülerlik ve twit aktivitesidir. Soyut olarak değerlendirilebilecek nitelikler ise davranış analizi, hesaplar arası psikolojik geri beslemeler ve anlamsal analiz olabilmektedir.

Erşahin ve ark.[14] Twitter'daki sahte he-sapların bulunması üzerine çalıştılar. Sahte sosyal hesapların yanlış bilgilerin yayılmasında kullanıldığında ciddi problemler ortaya çıktığını gösterdiler. Çalışmalarında Entropi Minimizasyon Ayırıklaştırma methodunu kullandılar. Sayısal özellikleri ve analiz sonuçlarını Naïve Bayes algoritmaları yardımıyla elde ederek sahte hesap sınıflandırmayı gerçekleştirdiler. Bu yöntem, kullanıcıların twit ve arkadaş sayılarını temel almaktadır. Bunun yanı sıra niteliklerin bağımsız olarak değerlendirildiği danişmalı öğrenmeyi kullanan Naïve Bayes algoritması yardımıyla ağırlıklı olarak anlamsal analiz ve spam filtreleme için yapılmış çalışmalar da vardır [15,16]. H.H. John ve arkadaşlarının[15] çalışmasında; kullanıcıların profil adı, hesap tanımı, profil fotoğrafı, takipçiler ve arkadaş sayıları, atılan twit sayısı ve twitlerin içeriklerine göre sınıflandırma yapılmıştır. 501 sahte, 499 gerçek hesaptan veri çekilmiştir. Naive Bayes doğruluk sonuçlarında ayırıklaştırmadan önce başarı oranı %86.1 iken ayırıklaştırmadan sonra %90.9 elde edilmiştir. Fabricio ve ark. [16] yaptığı çalışmada 23 nitelik göz önünde bulundurulmuş, yapılan spam hesap filtrelemesinde %84.5 doğruluk oranı elde edilmiştir.

Gurajala ve ark. [17] sahte hesapların ortaya çıkarılmasında kullanılan niteliklerin sayısını 10 olarak belirlemiştir. Daha fazla nitelik kullanılmasına rağmen sonuçlar istenen derecede başarılı olamamıştır. Buna rağmen sahte twitleri bulmada oldukça başarılı sonuçlar elde etmişlerdir. Stringhini ve ark. [18] spam profil tespiti için sadece 6 nitelik kullandılar. Bu niteliklerle sadece sisteme tanımlanmış spam ve içeriklerin filtrelenmesinde başarılı sonuçlar elde edilmiştir. Social Bakers web uygulamasında Random Forest Algoritması kullanılmak suretiyle Twitter hesaplarından başarılı sonuçlar elde edilmiştir. Online olarak sahte hesap taraması yapan uygulamalardan “Fake Follower Check”[19] nitelik olarak 8 kategoride tarama yapmakta olup taramanın detay verileri ve sınıflandırma teknikleri ile ilgili bir bilgi sitede paylaşılmamaktadır. Tarama neticesinde isim bazında bir sonuç verilmemektedir. Sadece “Sahte”, “Pasif”, “İyi” kategorileri bazında yüzdeler dilimlere ulaşılabilmektedir.

2012 yılında yapılan Amerika başkanlık seçimlerinde Obama ve Romney arasındaki rekabeti sosyal medyadaki sahte hesapların tespitini Camisani ve ark.[20] tarafından yapılmıştır. Bu çalışmada, 22 nitelik göz önünde bulundularak başarı elde etmişlerdir. Kullandıkları algoritmada takipçilerin hesapları nitelik değerlendir-melerine göre bir skor puanı alıp doğru veya sahte hesap olarak iki kategoride değerlendirilmiştir.

### 3. Profil Kalite Ölçekleme Metodu

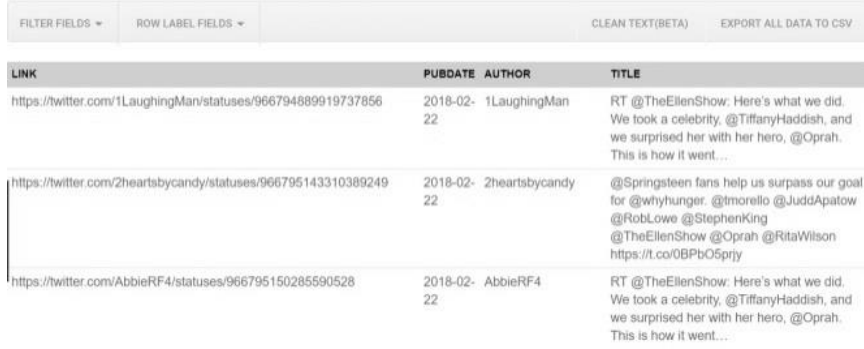
Bu makalede, kullanıcıların etkinlik değerlerini belirlemek ve sınıflandırmasını yapmak 3 aşamalı süreç uygulanmaktadır. Bu aşamalar sosyal ağdan veri toplama ve filtreleme, anonimleştirme ve analiz, işlemleridir.

#### 3.1. Verinin Toplanması Depolanması ve Filtrelenmesi

Sosyal ağlardan verilerin teknik açıdan seri şekilde çekilmesi ve işlenmesi için uygun formatlarda toplanmasına olanak sağlayan birçok web tabanlı uygulama mevcuttur. Toplanan verinin bozulmamış yani bütünselliğini korumuş, filtrelenmemiş ve yeni veri olması uygulamalara değer katmaktadır. Import.io, Netvizz, Netlytic, NodeXL, ile Facebook, Twitter, Instagram ve RSS gibi güncel sosyal ağ platformlarından veri çekilebilmektedir. Facebook, Twitter ve Instagram gibi sosyal ağların veri çekiminde kullanılan kendi API'leri mevcuttur. Bu çalışmada Netlytic, sosyal ağlardan kullanıcı hesap bilgilerini çekme için kullanılmıştır. Twitter Rest API V1.1'den veri çekiminde istifade edilmiştir. Netlytic [21], çevrimiçi olarak tarayıcı üzerinden kullanılabilen veri toplama ve görselleştirme aracıdır. Daha yüksek hacimli verilerin çekilmesi gereksinimi durumunda NodeXL Twitter verisi çekmede başarılı bir araç olarak çalışabilmektedir.

Netlytic ile çekilen veri; Bağıntı, Yayın Tarihi, Yazar ve Başlık gibi kategorilerde elde edilebilmektedir. Bu veriler üzerinden gereksiz görülmesi durumunda herhangi bir kolon silinebilmektedir. Ya da veri üzerinde sadece yazar, başlık, yayın tarihi ve bağıntı bazında filtreleme yapılabilmektedir. Temizlenen veri csv formatında dışarı alınabilmekte ve WEKA, R, UCINET, SPSS ve Statnet gibi platformlar üzerinde çalışmalar yapılabilmektedir[22]. Bu makalede veriler üzerindeki çalışmalar Netlytic kullanılarak yapılmıştır. Şekil.1 Netlytic ile çekilen verinin filtrelenmesi göstermektedir. Şekil.1 Etik problemlerden dolayı içerikler bulanıklaştırılarak verilmiştir.

## Kalite Ölçekleme Kriterleri ile Sosyal Ağ Hesaplarının Etkinliğinin Belirlenmesi



LINK	PUBDATE	AUTHOR	TITLE
https://twitter.com/1LaughingMan/statuses/966794889919737856	2018-02-22	1LaughingMan	RT @TheEllenShow: Here's what we did. We took a celebrity, @TiffanyHaddish, and we surprised her with her hero, @Oprah. This is how it went...
https://twitter.com/2heartsbycandy/statuses/966795143310389249	2018-02-22	2heartsbycandy	@Springsteen fans help us surpass our goal for @whyhungry. @tmorello @JuddApatow @RobLowe @StephenKing @TheEllenShow @Oprah @RitaWilson https://t.co/0BPbO5prjy
https://twitter.com/AbbieRF4/statuses/966795150285590528	2018-02-22	AbbieRF4	RT @TheEllenShow: Here's what we did. We took a celebrity, @TiffanyHaddish, and we surprised her with her hero, @Oprah. This is how it went...

**Şekil.1** Netlytic ile çekilen verinin filtrelenmesi

100 sosyal medya kullanıcısının verileri 3 ay boyunca haftalık olarak Netlytic kullanılarak Twitter API ile periyodik olarak çekilip incelenmiştir. Her bir kullanıcının son 1000 paylaşımı filtrelemek ve temizlemek üzere dikkate alınmış ve incelenmeye tabii tutulmuştur. Paylaşımlar, kullanıcının hem kendi attığı hem de başkasından geleni ilettiği olmak üzere bir bütün olarak ele alınmıştır. Her bir kullanıcının son 1000 paylaşımı için ham verisi ortalama 700 KB kadar olduğu görülmüştür. Bu da toplu sosyal hesap incelemelerinde kümülatif anlamda ciddi bir veri miktarına tekabül etmektedir. Tablo 1 de Twitter ağından çekilen veri ve özelliklerini göstermektedir.

**Tablo 1.** Periyodik çekilen veri ve özellikleri

Kullanıcı Sayısı	Paylaşım Adedi	Süre	Boyutu
100	1000	Haftalık/3 Ay	700 KB/Ay

Toplanan ham veri miktarı  $100 \times 4 \times 3 \times 700 \text{KB} = 820 \text{ MB}$  'dır. Ham veriden bağıntılar, yazarlar, yayın tarihleri gibi parametreler temizlenerek filtrelenmiş toplam veri miktarı  $100 \times 4 \times 3 \times 620 \text{ KB} = 726 \text{ MB}$  elde edilmiştir.

### 3.2. Kategorik Kalite Ölçekleme Kriterlerinin Hesaplanması

Sosyal ağlardan Netlytic ile elde edilen verinin işlenmesi için farklı yöntemler ile veriler analiz edilmiştir [23]. Twitter'dan elde edilen veri üzerinde sosyal ağ analiz yöntemleri ile çalışılmış ve tweet etkinliği tespiti için "Kim Kimi Andi" ve "Kim Kimi Cevapladı" olmak üzere iki kategoride bağıntılı gönderiler, bağıntılar ve geçen isim verileri elde edilmesi ile ilgili çalışma yapılmıştır.

Bu çalışmada, Twitter'da politika, magazin, spor ve marka alanında tespit edilen en popüler 100 farklı hesabın oluşturduğu profiller incelenmiş ve hesapların aktiviteleri ortaya çıkarılmıştır. Sosyal hesap profillerinin incelenmesinde 9 farklı nitelik bazında inceleme yapılmıştır. Bunlar; Takip edilen ile Takipçi oranı (T/T), Karşılıklılık(K), Merkezlilik( $M_c$ ), Yoğunluk(Y), Modüler-lik( $M_o$ ), çapta anılmanın cevaplanmaya oranı ( $\hat{C}_{AC}$ ), bağıntılı göndericilerde anılmanın cevaplanmaya oranı ( $G_{AC}$ ), Bağıntılı iletilerde anılmanın cevaplanmaya oranı ( $\hat{I}_{AC}$ ), twit aktivitesinin beğeni aktivitesine oranı ( $A_{TB}$ ). The Ellen Show sosyal medya hesabı incelendiğinde en hit 10 takipçinin trafiğin %90'ından fazlasını oluşturduğu görülmüştür. Takipçiler kendi aralarında analiz edildiğinde popüler olanların ortalama aynı değerlerde aktivite değerleri taşıdıkları gözlenmiştir.

Sosyal ağ hesaplarının etkinlikleri ve profillerinin kalitesi incelendiğinde, seçilen kullanıcıların 4 farklı kategorideki ortalama değerleri Tablo.2'deki gibidir. Bu kullanıcılardan bir kısmı politik lider, bir kısmı talkshow programcısı, bir kısmı spor kulübü, bir kısmı da ticari marka kullanıcılar olmak üzere farklı alanlardan seçilmiştir. Bir hesabın etkin kabul edilmesinin temel kriteri takip edilen sayısının takipçi sayısından oldukça düşük olmasıdır. Aksi durumda, temel eşik değerleri aşılamadığından hesap popüler kabul edilmeyecek ve dolayısıyla da aktifliğinin hesaplanmasına bile gerek kalmayacaktır. Tablo 2 de verilen 4 farklı kategorideki popüler hesapların tamamında görüldüğü üzere hesapların takip edilen sayısı takipçi sayısından çok düşüktür.

Profil bilgileri incelendiğinde 4 farklı kategorideki bu popüler hesapların oldukça fazla ortak noktasının bulunduğu tespit edilmiştir. Bu ortak noktalardan en önemlisi çok az takip ve çok fazla takipçi sayılarının olmasıdır. Bir diğer önemli nokta %92.5 oranında, anma bazında çapın, cevaplama bazında çapa oranla açık ara yüksek olmasıdır. Yoğunluk tüm popüler hesaplarda oldukça düşük değerdedir. Merkezlilik popüler hesaplarda

0.5 altında iken, ticari ve politik hesaplarda modülerlikler yüksektir. Spor ve eğlence hesaplarında ise aynı gündem konuları konuşulduğundan modülerlikleri düşüktür.

**Tablo 2.** Kategorik Bazda Kalite Ölçekleme Parametrelerinin Ortalamaları

	Politik	Eğlence	Sportif	Ticari
T/T	625K/100M	35.9K/77.2M	53/29.7M	141/378K
K	0.02538339	0.0009429	0.0048146	0.0621046
M <sub>e</sub>	0.17473273	0.40712826	0.4222332	0.3492379
Y	0.00114076	0.0022380	0.0002521	0.0014437
M <sub>o</sub>	0.73463678	0.21801095	0.3954690	0.6148167
Ç <sub>AC</sub>	2.873042	1.0623156	2.563120	2.0421794
G <sub>AC</sub>	3.518151	2.4591259	2.735201	5.106194
İ <sub>AC</sub>	1.996059	3.5759188	2.117342	11.205422
A <sub>T/B</sub>	15.5K/10	16.3K/863	61.2K/12	18.2K/9.6K

### 3.3. Metodoloji ve Uygulama

Bu bölümde, sosyal medya hesaplarından elde edilen veriler üzerinde kalite ölçeklenme-sinde kullanılan yöntemler, uygulanan algoritmalar ve elde edilen verilerden çıkan sonuçlar üzerinde kullanılan “Profil Kalite Ölçekleme Metodu” ile sınıflandırmaya değinilmiştir.

Sosyal medya hesaplarının etkinliğinin ölçeklenmesinde birçok yöntem uygulanabilmektedir. Ancak hızlı ve etkili bir yöntem olması önem arz etmektedir.

Heterojen sosyal ağ ilişkilerinde ilişki çeşitleri ağırlık merkezleri ile karakterize edilebilir. Her bir ilişki bir ağırlık matrisi ile ifade edilebilir. Matriste her bir eleman nesnelere arası ilişkinin gücünü ifade eder. Burada algoritmalar ağırlık matrislerinin lineer kombinasyonlarını bulmaya çalışır. Elde edilecek lineer kombinasyon, ilişkilerin gücünü ifade eder. İlişkinin değeri, hesabın gücü, takip/takipçi oranı, Direk mesaj, yeniden tweetlemek, anmak ve cevaplamak ilişkilerin değerini ölçeklemede kullanılabilir. Bu ilişkilerin değer ölçümleri ise kalite ölçüm metodları ile tespit edilebilir.

Netlytic ile elde edilen verilerden hesaplardaki bağ güçleri yani hesaptaki yeniden tweetlemeler, cevaplamalar ve anmalar, arkadaşlık gibi ilişki şekilleri üzerinden hesap yoğunluğu, merkeziliği, çapı, etkileşim gibi değerler bulunabilmektedir.

Şahsi popüler hesaplarda, politik ve eğlence gibi hesaplarda takipçi sayısı arttıkça karşılıklılık azalmaktadır. Yoğun takipçi sayılarına erişildiğinde kullanıcıların ortak ilgi alanları, beğeni ve eğilimleri oldukça değişkenlik gösterebilmektedir. Küçük şahsi ve ticari hesaplarda etkileşim oldukça fazladır. Etkileşimi artıran direk mesajlar, yeniden tweetleme ve cevaplamak değil anma sayılarıdır.

Popüler hesaplarda cevaplamak çok düşük, anma yüksektir. İçe dönük derece merkezliliğinde en çok anma alan veya cevap veren kullanıcılar öne çıkar. Dışa dönük derece merkezliliğinde ağda çevreleriyle etkileşimleri yüksek aktif Twitter kullanıcılarını ifade eder.

Yoğunluk, ağdaki mevcut bağlantıların muhtemel tüm bağlantılara oranını ifade eder. Eğer yoğunluk değeri 1 ise tüm kullanıcılar birbirleri ile bağlantılıdır denilebilir. Bu ölçü, bilgi akışının hızını da ifade ettiğinden çapa tamamlayıcıdır. Bu ölçüm 1'e yakınsa topluluğun birbirine sıkı sıkıya bağlı olduğunu gösterir. Profilin çapı ağı genişliği konusunda bilgi verir. Birbirine en uzak iki düğüm arasındaki olabilecek en uzun erişim yolunu ifade eder. Karşılıklılık profildeki takipçilerden kaç tanesinin aynı zamanda birbirleri ile ilişkili olduğunu ifade eder. Eğer tüm takipçiler birbirlerine cevap veriyorsa karşılıklılık 1'dir denebilir.

Tabana yayılmış, birbirleri ile güçlü bağları olan politik hesap takipçileri ve güçlü şekilde bir markaya bağımlılığı olan profillerin birbirleri ile etkileşimleri çok yüksektir. Politik bir hesaptaki takipçilerin karşılıklılık ilişkileri düşük ise lider hareketi olarak ifade edilebilir. Bu tabandan güç almadığına işaret eder. Merkezilik, bir profilin birkaç kullanıcıya dayalı olmasını ifade eder. Bu kullanıcılar aradan çekilirse ağ çöker yada profil alt üyeleri arasında iletişim kopar. Eğer merkezilik yüksek ise değer 1'e, eğer merkezilik düşük ise değer 0'a yakındır. Düşük bir merkezilik bilginin birçok kullanıcı arasında daha özgürce aktığını gösterir. Merkezilik yüksek ise kurumsallık, denetim, merkezilik vardır. Düşük ise rahatlık, özgürlük, esneklik ve çeşitlilik vardır.

Ticari marka ve eğlence programlarında merkezilik yüksek, politik hesaplarda düşüktür. Modülerlik, bir profilin takipçileri arasında belli ortak konularda yoğun etkileşim olması durumunda yüksektir. Alt gruplar farklı konulardan konuşuyorsa modularity düşüktür. Konu bütünlüğü yoktur denilebilir. Modülerliğin yüksek olması,

birbirlerine düşkün, birbirlerine değer veren ve iyi iletişim kuran sıkı bir taban olmasına işaret eder. Modülerlik yüksek ise, kümelenme oranı da yüksektir. Modülerlik ağdaki farklı toplulukların temsili yetini ifade eder. Daha yüksek modülerlik değerleri topluluktaki daha fazla ayrışımara işaret eder. Modülerliğin düşük değerleri genellikle 0.5 altındadır.

Bu makalede, bir hesabın kalitesini ortaya çıkaran 9 nitelik, genel anlamda 0 ya da 1 değerine tekabül ettirilmiştir. 0 değeri, bahse konu profil için kalite ölçüm kriterinin yetersiz kaldığını, 1 ise kalite ölçüm kriterinin profili yeterli olduğunu gösterir. Tüm bu ölçümler 0-1 arasında olduğundan değerlendirme daha pratik yapılabilecektir. Eşik değerlerin hesaplanması neticesinde etki, hesabın türü fark etmeksizin tümünde nitelik bazında aynı referans değerleri [0,1] aralığında olacaktır. Tablo 3 temel kriterlerin tanımlamalarını vermektedir.

**Tablo 3.** Genel kalite ölçüm kriterlerinin açıklamaları

Nitelikler	Açıklamalar
Ç	İki ağ katılımcısı arasındaki en uzun mesafe.
Y	Mevcut bağlantıların bir ağdaki toplam olası bağlantı sayısına oranıdır.
M <sub>e</sub>	Tüm düğümlerin ortalama derece merkezliği.
K	Mevcut ilişkilerin toplam sayısına göre karşılıklı ilişkiye sahip bağların oranıdır.
M <sub>o</sub>	Birbirleriyle iletişim kurma olasılığı yüksek olan, birbirine bağlı alt düğüm gruplarının tespitini yapar. Kenarlar rasgele dağıtılsa, verilen gruplar içinde düşecek kenarların kesir oranıdır.

Burada ifade edilen nitelikler bulunurken kullanılan formüller aşağıdaki gibidir[23].

N düğüm sayısını, B bağıntı sayısını ve P'de olası tüm bağıntıların sayısını göstermek üzere, ağdaki olası bağlantıların sayısı değeri denklem.1' de gibi verilir. Yoğunluk (Y) ise denklem.2' deki orantı ile bulunur.

$$B = (N * (N - 1))/2 \quad (1)$$

$$Y = B/P \quad (2)$$

Dinamik ağlarda (Twitter gibi.) bir düğümün merkezliliğinin(M<sub>e</sub>) hesaplanmasında (denklem 3) dinamik merkezlilik metriği kullanılmaktadır. G:=(V,E) şeklindeki bir grafta de V kenar, her bir kenar çifti için (s,t) en kısa yol olmak üzere her bir köşe çifti için söz konusu tepe noktasından geçen en kısa yolların kesri belirlenir. Bu fraksiyon tüm çift tepe noktalarında toplanır. Burada σ<sub>st</sub> düğümünden t düğümünü en kısa yolların toplamı ve σ<sub>st</sub>(v) bu kısa yollardan V kenarından geçenler olmak üzere Merkezlilik;

$$M_e = \sum_{s:=v:=t \in V} \frac{\sigma_{st}(v)}{\sigma_{st}} \quad (3)$$

Düğümlerin bağlantı yapılarının tespit edilmesi üzerlerinden akan verinin akış yapısı hakkında önemli bilgiler sağlamaktadır. Normalleştirilmiş derece merkezliliği(M<sub>D</sub>) denklem 4'te verilmiştir.

$$M_D = d_i * (N - 1) \quad (4)$$

Karşılıklılık(K) bir düğümün etkileşim yapısının ortaya çıkarılmasında önem taşımaktadır. İki tarafla bağlantısı olan linklerin sayısı L<sup><-></sup> ve tüm aktif bağlantıların sayısı L olmak üzere,

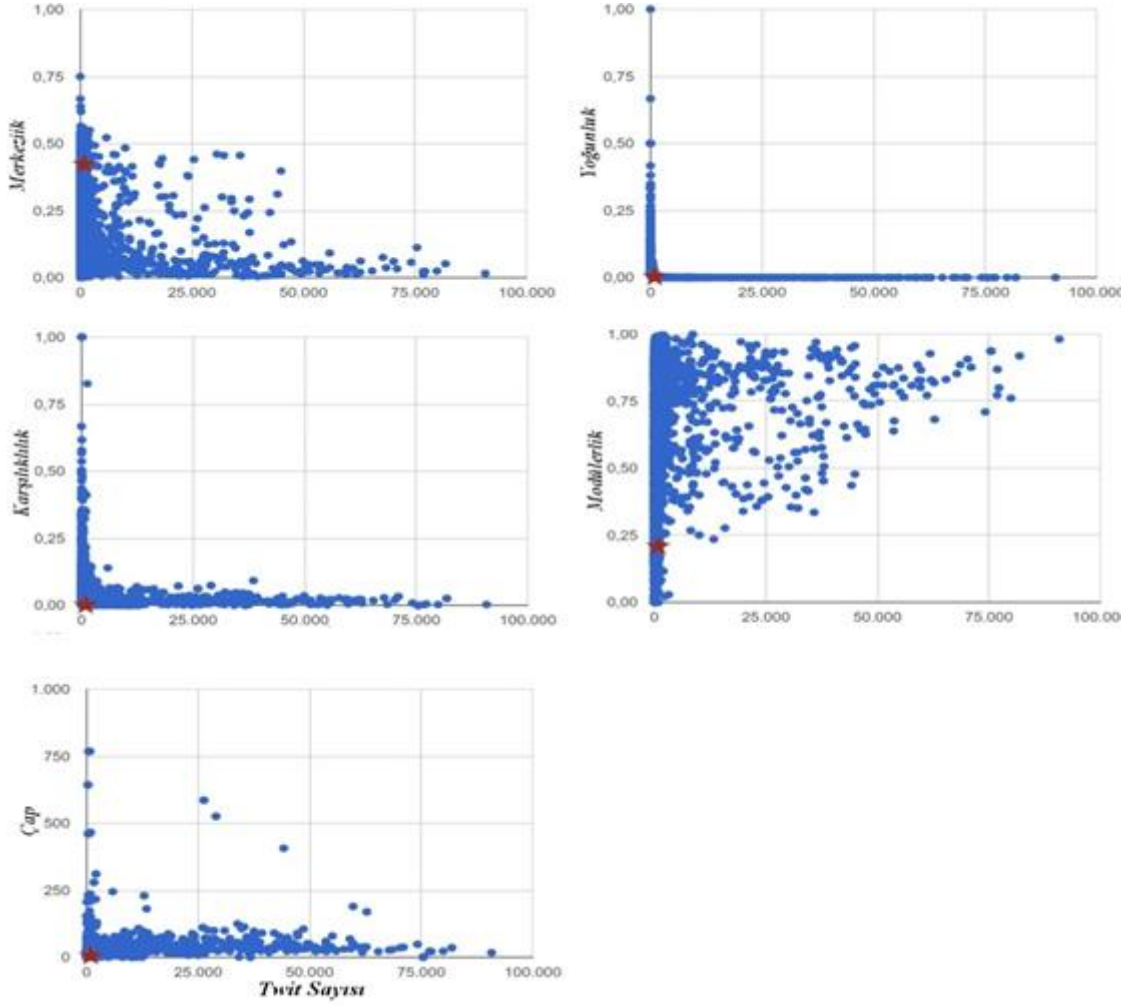
$$K = L^{<->}/L \quad (5)$$

Modülerlik(M<sub>o</sub>), Modülde yüksek kenar sayıları yüksek modülerliğe işaret eder. (u,v) bağıntı, V düğüm, E kenar ve denklem 6'da e<sub>ii</sub> = i modülündeki muhtemel kenar yüzdesi ve denklem 7'de a<sub>i</sub>, i modülünde en az bir tarafta sonlanan kenar yüzdesi olmak üzere M<sub>o</sub> denklem 8 ile elde edilmektedir.

$$e_{ii} = \frac{|\{(u, v): u \in V_i, v \in V_i, (u, v) \in E\}|}{|E|} \quad (6)$$

$$a_i = \frac{|\{(u, v): u \in V_i, (u, v) \in E\}|}{|E|} \quad (7)$$

$$M_o = \sum_{i=1}^c (e_{ii} - a_i^2) \quad (8)$$



**Şekil 2.** Tüm Kullanıcıların 100.000 Twitinden Elde Edilen Etkin Kalite Ölçüm Parametrelerinde Ortalamalar

Profillerin kalite değerleri ortaya çıkarılırken, başvurulan kalite ölçleme metodunda niteliklerin referans değerlerine göre aldıkları katsayı değeri Tablo 4'te verilmiştir. Burada verilen niteliklerin sırası aşağıda verilen nedenlerden dolayı çok önemlidir:

**Tablo 4.** Hesapların Kategori Bazında Nitelik Eşik Değerlerine Göre Referans Değerleri

Nitelik	Politik	Eğlence	Sportif	Ticari
$T/T > 1$	1	1	1	1
$K > 0.01$	1	1	1	1
$M_e > 0.5$	0	0	0	0
$Y > 0.01$	1	1	1	1
$M_o > 0.5$	1	1	1	1
$\hat{C}_{AC} > 1$	1	1	1	1
$\hat{G}_{AC} > 1$	1	1	1	1
$\hat{I}_{AC} > 1$	1	1	1	1
$A_{TB} > 1$	1	1	1	1

Popüler bir profilin takipçi sayısı daima takip edilen sayısından daha fazladır. Dolayısıyla bu değerin 0 olması durumunda diğer bütün parametreleri 1 olsa bile asla popüler bir hesap olduğu iddia edilemez. Bu bakımdan tabloda verilen niteliklerin sırası profil kalitesinin tespitindeki önem sıralarına tekabül etmekte olup değerlendirmeleri de bu makalede tavsiye edilen yöntemin temelini oluşturmaktadır. Baştan sona doğru olmak üzere tüm nitelikler 3 kümede toplanmaktadır. Tablo. 5 niteliklerin aldıkları katsayıya göre Profil Kalite Değerlerini göstermektedir.

Küme 1 = {T/T, K, M<sub>e</sub> }  
 Küme 2 = {Y, M<sub>o</sub>, Ç<sub>A/C</sub> }  
 Küme 3 = { G<sub>A/C</sub>, İ<sub>A/C</sub>, A<sub>T/B</sub> }

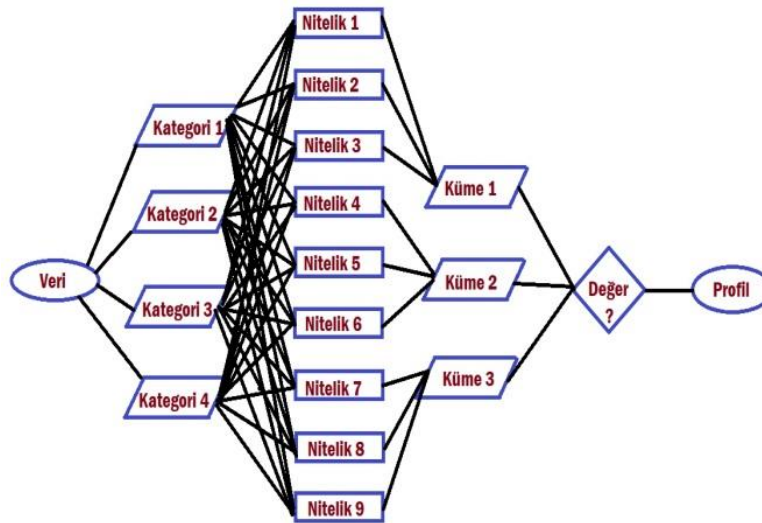
**Tablo 5.** Niteliklerin Aldıkları Katsayıya Göre Profil Kalite Değerleri

Değer	T/T	K	M <sub>e</sub>	Y	M <sub>o</sub>	Ç <sub>A/C</sub>	G <sub>A/C</sub>	İ <sub>A/C</sub>	A <sub>T/B</sub>
000	0	0	0	0	0	0	0	0	0
001	0	0	0	0	0	0	0	0	1
002	0	0	0	0	0	0	0	1	0
003	0	0	0	0	0	0	0	1	1
004	0	0	0	0	0	0	1	0	0
005	0	0	0	0	0	0	1	0	1
006	0	0	0	0	0	0	1	1	0
007	0	0	0	0	0	0	1	1	1
...	...	...	...	...	...	...	...	...	...
077	0	0	0	1	1	1	1	1	1
700	1	1	1	0	0	0	0	0	0
705	1	1	1	0	0	0	1	0	1
777	1	1	1	1	1	1	1	1	1

Tablo5'e göre her bir profilin kalite değerinin hesaplanması için aşağıda belirtilen ölçekleme kullanılarak profil sınıflandırma gerçekleştirilir. D, profilin niteliklerden elde edilen dereceyi belirtmektedir. Şekil 3 profilin sınıflandırılması için kullanılan yöntemin akış şemasını göstermektedir.

- D > 700 ise, Profil Popülerdir.
- 100 < D < 700 ise, Profil Aktiftir.
- 77 < D < 100 ise, Profil Normaldir.
- 7 < D < 77 ise, Profil Pasiftir.
- D < 7 ise, Profil Sahtedir.

şeklinde sınıflandırılmaktadır.



**Şekil 3.** Profil Kalite Ölçekleme Yöntemi Akış Diyagramı



100 kullanıcının, 3 aylık periyotlarla haftalık çekilen verilerinin ortalaması alınarak oluşturulan veri setinden profil hesaplarının kalite ölçekleme parametrelerinin alacağı değerler için temel oluşturmuştur. Bu parametrelerin algoritmik olarak tanımlanan kurallar çerçevesinde [0,1] olmak üzere aldıkları katsayılarla göre kalite dereceleri tanımlanmış ve {0,777} arasında aldıkları puanlardan profillerin; popüler, aktif, normal, pasif ve sahte olmak üzere 100 hesaptan 92'sinin doğru sınıflandırılması ile %92 başarı oranı elde edilmiştir.

## 5. Sonuçlar

İnternetin geniş bir kullanım alanı bulması ve erişim olanaklarının hızla artması kullanıcı sayısını artırmakta dolayısıyla da sosyal ağların kullanım trendleri yükselmektedir. Sosyal ağların kullanıcı sayılarının artması ve çok rağbet görmeleri bireyler, topluluklar, firmalar, kurumlar ve devletler tarafından yoğun şekilde iletişim ve etkileşimde kullanılmalarının önünü açmıştır. Ancak bu eğilim paralel bir şekilde manipülasyon yaratma amacıyla oluşturulmuş, oldukça problemlili olabilecek etkileşimler oluşturan sahte kullanıcı hesaplarının da hızla artmasına sebebiyet vermiştir. Bu sahte hesapların tespit edilmesi konusunda birçok akademik çalışma ve çoğu ilkel olsa da birçok web tabanlı uygulama mevcuttur. Ancak, bireyler arası ilişkilere, firmaların üretim ve pazarlamasına, kurumların hizmet portföylerini geliştirmelerine, toplumsal dinamiklerden elde edilen izlenimlerle devletlerin vizyon çalışmalarına yön veren sosyal medya kullanıcılarının etkileşimlerinin temelini oluşturan hesapların profil kalitelerinin tespiti, analizi, değerlendirilmesi bir çok açıdan fayda sağlayacaktır.

Bu makalede, sosyal medya kullanıcılarının profillerinin doğru şekilde analiz edilmesi ve doğrudan sahte veya aktif olarak tanımlanmaları yerine genel olarak sahip oldukları 9 nitelik üzerinden derecelendirilmeleri için sunulan “Profil Ölçekleme Metodu” ile değerlerinin 0-777 arasında tanımlanmaları daha geniş ve verimli bir bakış açısı kazandıracaktır. Çalışma; niteliklerin daha spesifik hale getirilmesi ile daha değerli sonuçlar üretebilecek potansiyeldedir. Çalışmada geçen 9 kalite ölçekleme kriteri sadece fiziksel ağ ve aktivite değerlerini temel almaktadır. Ancak anlamsal analizinde dâhil edileceği bir değerlendirme daha değerli sonuçlar üretecektir.

## Kaynaklar

- [1] Kemp S, Digital in 2017 Global Overview. <https://wearesocial.com/special-reports/digital-in-2017-global-overview>, son inceleme March 20, 2018.
- [2] Chaffey D, Global Social Media Research Summary 2017. <http://www.smartinsights.com/social-media-marketing/social-media-strategy/new-global-social-media-research/>, son inceleme March, 2018.
- [3] Young K. Trends 2017. <https://blog.globalwebindex.net/tag/trends-2017/>, son inceleme March, 2018.
- [4] Washha M, Qaroush A, Mezghani M, Sedes F. A Topic-Based Hidden Markov Model for Real-Time Spam Tweets Filtering. KES2017; 6-8 September; Marseille, France, 2017.
- [5] Verma M, Ivya D, Sofat S. Techniques to Detect Spammers in Twitter- A Survey. International Journal of Computer Applications 2014; 85(10).
- [6] Hirve S, Kamble S. Twitter Spam Detection. IJESC 2016; 6(10).
- [7] Azab A, Idrees A, Mahmoud M, Hefny H. Fake Account Detection in Twitter Based on Minimum Weighted Feature set. World Academy of Science, Engineering and Technology International Journal of Computer and Information Engineering 2016; 10(1): 2016.
- [8] Chavoshi N, Hamooni H, Mueen A. Identifying Correlated Bots in Twitter” in Springer International Publishing. Lecture Notes Computer Science 2016: 10047; pp. 14-21, 2016
- [9] Herzallah W, Faris H. Feature Engineering for Detecting Spammers on Twitter: Modeling and Analysis. Journal of Information Science 2016; 44(2): 230-247.
- [10] Bhattacharya P, Zafar M B, Ganguly N. Inferring User Interests in the Twitter Social Network. RecSys'14; 6-10 October; Foster City, Silicon Valley, CA, USA, 2014.
- [11] Song J, Lee S, Kim J. Spam Filtering in Twitter Using Sender-Receiver Relationship. RAID'2011; Menlo Park, California, USA: 2011, pp.301-317.
- [12] Alahmadi B, Legg P, Nurse J. Using Internet Activity Profiling for Insider-Threat Detection. International Workshop on Security in Information Systems; 2015.
- [13] İş H, Müngen A A, Tuncer T, Kaya M. Frequent Pattern Mining for Community Detection in Web Logs. International Conference on Artificial Engineering and Data Processing; 16 Ekim; Malatya, Türkiye: 2017. pp.1-7.
- [14] Erşahin B, Aktaş Ö, Kılınç D, Akyol C. Twitter Fake Account Detection. 2nd International Conference on Computer Science and Engineering (UBMK'17); 5-8 October; Antalya, Türkiye: 2017.
- [15] John G H, Langley P. Estimating Continuous Distributions in Bayesian Classifiers. 11. Conference on Uncertainty in Artificial Intelligence, San Mateo, USA: 1995. pp.338-345.
- [16] Benevenuto F, Magno G, Orozco A. Detecting Suversion of Twitter. Collaboration-Electronic-Anti Abuse and Spam Conference (CEAS); 2010.

- [17] Gurajala S, White J S, Hudson B, Matthew J N. Fake Twitter Accounts: Profile Characteristics Obtained Using an Activity-Based Pattern Detection Approach. SMSociety'15; Toronto, Canada: 2015.
- [18] Stringhini G, Kruegel C, Vigna G. Detecting Spammers on Social Networks. 26. Annual Computer Security Applications Conference; 6-10 December; Austin, Texas, USA: 2010, pp. 1-9.
- [19] Social Bakers. (Online) <http://www.socialbakers.com/products/analytics?ref=fakefollowers-top-bar> , March, 2018.
- [20] Camisani-Calzolari M. Analysis of Twitter followers of the US Presidential Election Candidates: Barack Obama and Mitt Romney. August, 2012.
- [21] Netlytic,(Online), [https://netlytic.org/home/?page\\_id=10834](https://netlytic.org/home/?page_id=10834), son inceleme March, 2018.
- [22] Weka, "Data Mining Software in Java", <https://cs.waikato.ac.nz/ml/weka/>, son inceleme March, 2018
- [23] Wikipedia, (Online), <https://wikipedia.org/wiki/Metrics/>, son inceleme March, 2018.