



HANGİ TWEETLER DAHA FAZLA ETKİLEŞİM ALIYOR?

İbrahim SABUNCU¹, Açıya YILDIRIM²

^{1,2}Endüstri Mühendisliği, Mühendislik Fakültesi, Yalova Üniversitesi, Yalova Türkiye

ÖZET

İşletmeler sosyal medyayı etkili bir tanıtım mecrası olarak kullanabilmektedirler. Bu tanıtım çalışmalarının etkisi, şirketin yaptığı sosyal medya paylaşımlarının etkileşim miktarıyla ölçülür. Etkileşimi arttırabilecek nitelikleri bilmek işletmeler için önemlidir ve bu çalışmada ele alınmıştır. Önemli bir sosyal medya paylaşım sitesi olan Twitter'daki paylaşımların etkileşim oranını etkileyen faktörler araştırılmış, analiz edilmiş ve gelecek paylaşımlarda değer yaratmak için kullanılmıştır. Bu amaçla, Türkiye'de faaliyet gösteren ve en fazla takipçi sayılı 25 içecek şirketinin hesapları incelenmiştir. Bu hesaplardan 1000 tweet çekilerek RapidMiner veri madenciliği yazılımı ile analiz edilmiştir. İçeriklerine göre 3.436 tweet, metin madenciliği yöntemiyle kategorilere ayrılmıştır. Sonrasında tweetlerin etkileşim değerleri ile paylaşıldığı saat periyotları, mevsimler, günler, markalar, kategoriler arasındaki ilişkiler analiz edilmiştir. Testler sonucunda, bu niteliklerin etkileşim oranı üzerinde etkili olduğu görülmüştür. Yeni oluşturulacak bir tweetin belirtilen nitelikleri bilindiğinde sahip olacağı etkileşim oranını öngörebilmek için tahmin modeli oluşturulmuştur. Model test edilmiş ve etkileşim oranının %68,53 doğruluk oranıyla doğru tahmin edilebildiği görülmüştür. Sonuç olarak, sosyal medya analizinin etkileşim oranı öngörebilmeye önemli bir yöntem olabileceği anlaşılmıştır.

Anahtar Kelimeler: Sosyal Medya Analizi, Etkileşim Oranı, Metin Madenciliği, Sınıflandırma Modeli, Twitter

WHICH TWEETS GET MORE ENGAGEMENT?

ABSTRACT

Businesses can use social media as an effective promotion platform to engage with their customers. The amount of engagement measures the impact of this promotion. Knowing the factors that can increase engagement is essential for businesses and is discussed in this study. For this purpose, Twitter accounts of 25 beverage companies operating in Turkey and has the largest number of followers have been examined. 1000 tweets were gathered from all these accounts and were analyzed with RapidMiner data mining software. 3.436 tweets were divided into categories according to their content. Afterwards, the relationships between the engagement rates and the periods, seasons, days, brands, categories of tweets were analyzed. As a result of the tests, it was observed these factors are influential on the engagement rate. A prediction model was created to predict a new tweet's engagement rate when its specified factors are known. The model was tested, and it was found that the interaction rate could be estimated correctly. As a result, it was understood that social media analysis could be an essential method in predicting engagement rates.

Keywords: Social Media Analysis, Engagement Rate, Text Mining, Classification Model, Twitter

GİRİŞ

Ekonomi, siyasal katılım, eğitim, eğlence gibi oldukça geniş bir yelpazede hayatımızın her alanına ulaşan bilgisayarlarla aracılanmış iletişim dijital kültürü de beraberinde getirmiştir. İçinde bulunduğumuz bu dijital kültürün benzersiz özellikler ve araştırma alanları ortaya koyduğu yadsınmaz bir gerçektir (Paslanmaz & Narmanlıoğlu, 2020). Bu araştırma alanlarından biri olarak sosyal medya uygulamaları da insan yaşamlarına birçok yönden aşamalı olarak giren etkili araçlardan

biri olarak gözlenmektedir (Alalwan et al., 2017). Sosyal medya ile müşteriler de organizasyonların markalaşma süreçlerine destek olmaktadır. Dünya çapındaki çoğu firma yeni müşterilere ulaşmak, mevcut müşteri deneyimini geliştirmek ve onlarla bağ kurmak için sosyal medya uygulamalarını kullanmaktadır. Şirketler sosyal medya içerikleri geliştirip paylaşarak müşterileri ile etkileşime girmeye çalışır. Etkileşim, içeriklerin hedef kitle tarafından ne derece ilgi gördüğünün ölçüsüdür. Etkileşim miktarını arttırmak marka bilinirliğini ve sosyal medya pazarlama çalışmalarının başarısını arttıracaktır. Bu çalışmada işletmelerin sosyal medya paylaşımlarının etkileşim miktarlarının nasıl artırılacağı konusu araştırılmıştır.

Hangi sosyal medya içeriklerinin daha çok etkileşim aldığı tespit etmek içerik geliştirme stratejisi belirlemede önemli bir kavramdır. Yayınlanan içeriklerin sosyal ağlarda ne miktarda etkileşim aldığı ölçüp etkileyen faktörler analiz edilerek önemli bilgiler elde edilir. Bu bilgiler ile gelecek içeriklerin ne zaman, nasıl yayılması gerektiğinden hedef kitlenin hangi tarz içerikleri beğendiğine kadar çıkarımlarda bulunulabilir. Sonrasında sosyal medya stratejisi bu doğrultuda optimize edilebilmektedir. Yaş, cinsiyet, konum gibi belirli demografikler altında konumlandırılan kullanıcılara özel geliştirmeler yapılarak dönüşüm oranı artırılabilir. Böylelikle işletmelerin tanımlanan hedeflerine ulaşması çok daha kolay hale gelmiş olur.

Dünya çapında milyonlarca insan tarafından kullanılan sosyal medya platformlarından biri de Twitter'dır. Twitter kullanıcıları bu platformda kişi, grup ve işletmelerle bağlantı kurabilir, onları takip edebilir ve kendi fikirlerini 280 karakter sınırlı paylaşımlar (Abanoz, 2012). Kısa içerik özelliğiyle geniş bir kitleye sahip olan Twitter, hızlı bir şekilde rakipleri ve sektörün müşteri trendlerini anlamaya da yardımcı olabilmektedir. Twitter'da etkileşim miktarı, etkileşim oranı ile ölçülür. Etkileşim oranı, gönderilen içeriğin hedef kitle tarafından ne derece ilgi gördüğüdür. Tweetlerin, retweet ve beğeni sayılarının toplam takipçi sayılarına oranlarının hesaplanmasıyla belirlenebilir (Kürkçü Dumanlı, 2017). Yüksek etkileşim oranı; daha fazla görünürlük, marka çekiciliği, insanların kendi aralarında markayı konuşması, daha fazla gelir gibi faydalar sağlayacaktır.

Bu çalışmada, işletmeler tarafından Twitter sosyal medya platformunda paylaşılan içeriklerin (tweetler) etkileşim oranını nasıl artırılacağı konusunu araştırmak için metin madenciliği yapılmıştır. Metin madenciliği, yapısal olmayan metin belgelerini sayısallaştırıp daha sonra veri madenciliği araç ve tekniklerini kullanarak onlardan anlamlı örüntüler çıkarmaya dayanır (Çelik et al., 2017). Yapılan çalışmada metin madenciliği toplanan tweetlerin içeriğini analiz edip kategorilere ayırmak için kullanılmıştır. Ardından bu kategorilerin etkileşim oranı üzerindeki etkisi analiz edilmiştir.

Bu çalışmayı diğer çalışmalardan ayıran en önemli fark metin madenciliği çalışmasıyla elde ettiği paylaşım özelliklerini, paylaşımın diğer özellikleriyle birleştirerek bütüncül bir değerlendirme sunmasıdır. Buna ek olarak tahmin modelinin de sunulmuş olması gelecek paylaşımlarının etkileşim değeri belirleyici olacaktır.

LİTERATÜR ARAŞTIRMASI

Sosyal medya paylaşımlarının içeriklerine göre, etkileşim miktarını belirleyen faktörler konusunda dünyada yapılmış çok sayıda çalışma bulunmaktadır (A. J. Kim & Ko, 2012) (Gray et al., 2013) (Oliveira & Welch, 2013) (Sabate et al., 2014) (Hollebeek et al., 2014) (Warren et al., 2014) (Dijkmans et al., 2015) (Alt, 2015) (Guesalaga, 2016) (Khan, 2017) (Oh et al., 2017). Bildiğimiz kadarıyla Türkiye'de bu konuyla ilgili sadece 4 çalışma tespit edilebilmiştir (Kocabaş, 2016) (Kürkçü Dumanlı, 2017) (Eralp Üstündağ, 2018) (Ayan et al., 2019). Türkiye'de yapılmış olan bu çalışmalar incelenerek, bu konuda literatürdeki boşluğun tespit edilmesi hedeflenmiştir.

Öncelikle Kocabaş'ın beyaz eşya sektöründe önemli markalardan ikisi olan Arçelik ve Bosch'un sosyal medya platformları üzerindeki çalışması incelenmiştir (Kocabaş, 2016). Kocabaş bu iki markanın kurumsal Facebook ve Twitter hesaplarını kullanma biçimlerine dayanarak kullanıcılarıyla olan iletişimlerinde etkileşimin ne şekilde sağlandığını tespit etmeye çalışmıştır. Veriler Temmuz, Ağustos, Eylül aylarını kapsayacak şekilde 2014 tarihinde çekilmiş ve veri kaynağı olarak bahsi geçen markaların Facebook ve Twitter hesapları kullanılmıştır. Veri temin etme aşamasında içerik analizi uygulanmıştır. İçerik analizi yöntemini uygulayabilmek amacıyla kodlama

çetveli oluşturulmuş ve bu içerik analizi kodlama çetvelinde de Facebook ve Twitter'daki her bir paylaşımın içeriği yorumlanarak kaynak olarak kullanılmıştır. Bu çetvelde “bilgi verme içerikli mesaj, tanıtım içerikli mesaj, kutlama içerikli mesaj, karşılıklı etkileşim içerikli mesaj, öneri içerikli mesaj” tanımlamaları kullanılmıştır. Sonuç olarak Twitter kaynağında birinci sırada “bilgi verme” amaçlı ikinci sırada ise “tanıtım” amaçlı kullanımın yer aldığı görülmüştür. Bu neticeye ek olarak Arçelik markasının Twitter’ı Bosch markasına göre daha dinamik kullandığına ulaşılmıştır. Yaptığı çalışmalar doğrultusunda Kocabaş, markaların sosyal medya platformlarını etkileşimi daha yüksek tutacak şekilde kullanmaları ve bu amaç doğrultusunda stratejik planlar oluşturmaları gerektiğini ileri sürmüştür.

Kürkçü ise etkileşim seviyelerini belirlemeyi hedeflediği çalışmasında veri kaynağı olarak haber ajansları ve gazetelerden faydalanmıştır (Kürkçü Dumanlı, 2017). Verilerini 2016 yılı Mayıs, Haziran ve Temmuz aylarını kapsayacak şekilde beş farklı hesaptan almıştır. Kürkçü beş farklı hesaptan gönderilen 81.174 tweet ile çalışmasını gerçekleştirmiştir. Kullandığı beş hesapla etkileşim kuran kullanıcıların bu etkileşim seviyelerini ölçmek amacıyla analizlerini gerçekleştirmiştir. Bu analizinde kullanıcıların retweet etme ve beğeni sayılarını dikkate almıştır. Analizinde bu değerlerin toplamını, toplam takipçi sayısına bölerek elde ettiği oranları kullanmıştır. Çalışmasında test etmek üzere iki hipotez geliştirmiştir. İlkinde ileri sürdüğü takipçi sayısı ile retweet, beğeni sayısı arasında ilişki vardır ifadesi reddedilmiştir. İkinci hipotezinde belirttiği retweet ve beğeni sayısı ile etkileşim oranının değişimi arasında ilişki vardır ifadesi ise kısmen de olsa kabul edilmiştir. Analiz sonucunda üstüne düşülmesi gereken ve dikkat çeken bir nokta olarak Anadolu Ajansı’nın takipçi sayısının az olmasına karşın paylaşımlarının en yüksek retweet ve beğeni sayılarına ulaşması olmuştur. Dolayısıyla en yüksek etkileşim oranına sahip hesap olmuştur. Buna karşılık olarak Hürriyet Gazetesi’nin retweet ve beğeni sayılarının çok düşük olmamasına rağmen takipçi sayısının çok fazla olması etkileşim oranı formülündeki paydayı çok yükselttiği için Hürriyet Gazetesi’nin hesabını etkileşim oranı en düşük olan hesap yapmıştır. Bu neticeye bakılarak etkileşim oranını değerlendirirken takipçi sayılarının da odaklanılacak ayrı bir konu olduğu tespit edilmiştir. Bu hesapların belirli bir zaman içindeki retweet edilme ve beğeni alma durumları göz önüne alındığında on iki saat içindeki retweet sayısının yirmi dört saat içindeki beğeni sayısına eşit olduğu gözlenmiştir. Tweetlerin retweet dolaşımına sokulması ve daha fazla kişi gördükçe beğeni sayısının retweet sayısına ulaşması durumu düşünüldüğünde bu oran normal karşılanabilmektedir. Çalışmaları sonucu Kürkçü, gazete haberlerinin daha dinamik ve yüksek etkileşimli bir şekilde yayılmasında sosyal medyanın önemli bir araç olduğunu vurgulamıştır.

Eralp de çalışmasında yüksek etkileşim oranı seviyesini yakalayabilmek için içeriklerin sahip olması gereken nitelikler üzerine çalışmıştır. Bu amaç doğrultusunda etkileşim oranı yüksek seviyede olan içerikleri, içerik analizi yöntemiyle incelemiştir (Eralp Üstündağ, 2018). Başarılı etkileşim oranına ulaşmada kullanılacak içerikleri analiz etmek ve yorumlamak için bir kod tablosu oluşturulmuştur. Eralp çalışmasında hızlı tüketim sektörüne ait elli hesabı bizim çalışmamızda da kullanılacak olan “Boomsocial” sitesinden almıştır. 2017 yılının tüm verilerini kullanarak çalışmış ve başarılı oranları yakalayabilmek için 50 beğeni ve üzeri olan içerikleri araştırmıştır. Araştırma neticesinde öncelikle ürün veya hizmet çekiciliklerinin, sonrasında ise mizah çekiciliğinin kullanıldığı görülmüştür. Bunun yanı sıra değerli günler için hazırlanan içeriklerin de sıklıkla tercih edildiği gözlenmiştir. Neticede çalışmada, içeriklerin bu tür özellikleriyle kullanıcı eylemlerindeki değişimlerin önemi vurgulanmak istenmiştir.

Ayan ve arkadaşları da çalışmalarında, sektörel anlamda çeşitli hesapların içeriklerini kümeleyerek hangi kümede etkileşim düzeyinin fazla olduğunu ve buna sebebiyet veren durumları incelemiştir (Ayan et al., 2019). Çalışmalarında sosyal medya madenciliğinin süreçlerini takip ederek rakip firma verilerini analiz etmişlerdir. Bu süreçte uygunluk analizi, tweet analizi, rt analizi ve metin madenciliği yöntemleri kullanılmıştır. Metin madenciliği çalışmasına hazırlık olarak tweetler “Özel Teklif”, “Yarışma & Etkinlik”, “Ürün”, “Sosyal”, “Destek & Geri Bildirim” ve “Özel Etkileşim” etiketleri kullanılarak kümelenebilir. Gelen etkileşimlerin incelenmesi pareto analizi ile yapılmıştır. Neticede yüksek etkileşimin kozmetik sektöründe “Özel Teklif”, elektronik sektöründe “Yarışma & Etkinlik” etiketinden geldiği tespit edilmiştir. Bu tespit sonucu bu sektörlerdeki firmalara bu etiket kümesini içeren paylaşımlara yönelim önerisinde bulunulmuştur. Sektörde var olan firmaların yanı sıra sektöre yeni girecek firmalar için de bu etiketler kapsamındaki paylaşımlara

yoğunluk gösterilmesinin rekabet edilebilirlik seviyelerini hızla arttıracığı uyarısı verilmek istenilmiştir.

İncelenen çalışmalarda, sosyal medya paylaşımlarının içeriği dışındaki faktörlerin (paylaşım tarihi, saati vb.) etkileşim oranına etkisi bütüncül olarak ele alınmamıştır. Ayrıca araştırmalarımız kadarıyla etkileşim oranının tahmin edilmesi ile ilgili bir çalışma bulunmamıştır. Bu makalede eksik görülen bu konularda çalışmalar yapılmıştır.

METODOLOJİ

Metodoloji kısmında uygulanacak adımları kısaca özetleyecek olursak; öncelikle veri kümesi oluşturulmuş ve ön işleme yapılmıştır. Ön işleme aşamalarından en önemlisi metin madenciliği kullanılarak oluşturulan kategori sütunu olmuştur. Sonrasında paylaşımların etkileşim oranını belirleyecek formül sunulmuştur. Formül sonucu oluşacak değer ile paylaşım niteliklerinin ilişkilerini incelemek adına hipotezler oluşturulmuştur. Hipotezleri inceleme amacıyla değişken tiplerine göre testler sırasıyla uygulanmıştır. Testler sonucunda etkileşim oranı üzerinde önemli yere sahip parametreler ifade edilmiştir. Son olarak yeni oluşturulacak bir paylaşımın sahip olacağı parametreler doğrultusunda hangi aralıkta bir etkileşim değerine ulaşacağını görmek için tahmin modelleri oluşturulmuş ve doğruluk oranları incelenmiştir. En yüksek doğruluk değerine Karar Ağacı modeliyle ulaşılmıştır ve güncel veriler test için alınarak bu model üzerinde test edilmiştir.

İlk adımda çalışmanın amacına uygun olarak içecek sektörü Twitter hesaplarının paylaşımlarının etkileşim oranına (bağımlı değişken) etki eden faktörleri (bağımsız değişkenler) araştırılmıştır. Bu etkileşim oranına; gün, ay, mevsim, zaman periyodu, marka, kategori, beğeni sayısı, retweet sayısı değişkenlerinin etkisini belirlemek ve bu faktörlere bağlı olarak etkileşim oranını tahmin etmek hedeflenmiştir. Bu hedefle oluşturulan hipotezler ve değişkenler aşağıda ifade edilmiştir:

Tablo 1:Değişkenler Tablosu

Değişken İsmi	Değişken Ölçeği	Değişken Tipi	Değişkenin Ölçülme Yöntemi
Gün	Nominal	Bağımsız	Tweetin paylaşılma tarihinden haftanın hangi günü paylaşıldığı bilgisi alındı.
Ay	Nominal	Bağımsız	Tweetin paylaşılma tarihinden paylaşıldığı ay bilgisi alındı.
Mevsim	Nominal	Bağımsız	Tweetin paylaşılma ay bilgileri mevsimlerin kapsamına göre filtrelendi.
Zaman Periyodu	Aralık	Bağımsız	Tweetin paylaşıldığı saat 24 dilimlik periyotta ifade edildi.
Twitter Hesabı	Nominal	Bağımsız	Tweeti paylaşan şirketin kullandığı Twitter hesabı belirtildi.
Kategori	Nominal	Bağımsız	Tweet metinlerinin, metin madenciliği ile analiz edilmesi sonucu kategoriler oluşturuldu.
Beğeni Sayısı	Ordinal	Bağımsız	Tweetin beğenilme sayısı bilgisi kaydedildi.
Retweet Sayısı	Ordinal	Bağımsız	Tweetin tekrar paylaşılma bilgisi kaydedildi.
Etkileşim Oranı	Oran	Bağımlı	Beğeni ve retweet sayılarının toplamının takipçi sayısına bölünmesiyle elde edildi.

- Hipotez 1: Şirket hesabının takipçi sayısı ile tweetlerinin beğeni sayısı ve retweet sayısı değerleri arasında anlamlı bir ilişki vardır.
- Hipotez 2: Tweetlerin paylaşım saatleri (1 ve 24 arasındaki periyotlar) ile etkileşim oranı arasında anlamlı bir ilişki vardır.
- Hipotez 3: Paylaşılan tweetlerin etkileşim oranları mevsimlere göre farklılık göstermektedir. Dolayısıyla mevsimler ile etkileşim oranı arasında anlamlı bir ilişki vardır.

- Hipotez 4: Hafta sonu paylaşılan tweetler daha yüksek etkileşim oranına sahiptir (Semiz & Berger, 2017). Dolayısıyla paylaşımın günü ile etkileşim oranı arasında anlamlı bir ilişki vardır.
- Hipotez 5: Markaların paylaştığı tweetlerin etkileşim oranları arasında anlamlı bir ilişki vardır.
- Hipotez 6: Tweetin içeriği (kategorisi) ile etkileşim oranı arasında anlamlı bir ilişki vardır.

Bu hipotezleri test etmek ve etkileşim oranı tahmin modelini oluşturmak amacıyla Twitter'dan veriler toplanmış bu veriler analiz edilip yorumlanmıştır.

Veri Toplama

İncecek sektöründe faaliyet gösteren ve en fazla takipçisi olan 25 şirketin Twitter hesabında veriler toplanmıştır. Takipçi sayıları "BoomSocial"¹ sitesinden alınmıştır. Ek 1'de listelenen şirket Twitter hesaplarında paylaşılmış olan tweetleri çekebilmek amacıyla, Sabuncu ve Atmış tarafından açık kaynak kodlu R programlama dilinde yazılmış olan yazılım modifiye edilerek kullanılmıştır (Sabuncu ve Atmış 2018). Modifiye edilmiş yazılımın pseudo kodları şu şekildedir:

I) Başla

II) Gerekli kütüphane dosyasını yükle

III) Veri çekmek için, API anahtarı ve belirtecini kullanarak Twitter API'sını ayarla

IV) API kullanarak kullanıcı adı yazılmış hesabın son 1000 tweetini indir.

i) Twitter'da aranacak ifade için "account" fonksiyonu ile arama yap.

ii) Bulunan tweetleri veri çerçevesi formatına dönüştür.

iii) Bulunan tüm tweetleri, hesabın isminin yazılı olduğu dosyaya ekle.

V) Tweetleri gereksiz noktalama işaretler ve boşluklardan temizle

XII) Bitir

Bu kodlar 12 Mart 2020 tarihinde 25 hesap için de kullanılmıştır ve toplam 22.931 tweet nitelikleriyle birlikte listelenmiştir.

Veri Ön İşleme

Açık kaynak kodlu R stüdyo programı ile çekilen hesap verileriyle birlikte gelen değişkenlerden text, favorite count, created at date, created at time, screen name, retweet count kullanılmıştır. Bu değişkenlerin ifade ettiği değerler de Tablo 2'de gösterilmiştir.

Verileri hipotezlerde kullanılabilecek duruma getirmek için öncelikle ihtiyaç duyulmayan değişkenler çıkarılmış, sonrasında ilk adım olarak etkileşim oranını hesaplamak için hesapların yanına Ek 1'de belirtilen takipçi sayıları eklenmiştir.

Tweetlerin etkileşim oranlarını hesaplamak için aşağıdaki formül kullanılarak hesaplanmıştır (Sterne, 2010):

$$\text{Tweet Etkileşim Oranı} = \frac{\text{Toplam etkileşim (Tweetin beğenilme ve retweet miktarı toplamı)}}{\text{Tweeti Paylaşan Hesabın Toplam Takipçi Sayısı}} \times 100$$

Erişim sık sık dalgalanıyorsa ya da dalgalanabileceği düşünülüyorsa etkileşimi daha doğru bir şekilde ölçmek için bu yöntem kullanılabilir. Bu yöntemin dezavantajı da takipçi sayısı arttıkça katılım oranı biraz düşebilir ve bu durumun tersi de geçerlidir. Bu sebeple bu durumu, takipçi büyüme analizleriyle birlikte incelemek daha avantajlı sonuçlar verebilir (Sterne, 2010).

Bir gönderinin toplam etkileşimini hesaplarken gönderiyle ilgili gerçekleştirdiği tüm eylemleri (beğeniler, yorumlar, paylaşımlar, görüntülemeler, tıklamalar) toplanır. Fakat bu çalışmada toplanılan tweetlerin sadece beğeni ve paylaşım (retweet) sayıları elde edilmiştir. Bu nedenle, etkileşim oranı

¹ <https://www.boomsocial.com/>

hesabında sadece bu değerler dikkate alınmıştır. Tablo 2’de örnek etkileşim oranları hesabı gösterilmiştir.

Tablo 2 :Etkileşim Oranı Hesaplanmış Tweetler

Text	Favorite Count	Created at Date	Created at Time	Screen Name	Retweet Count	Follower Count	Engagement Rate
Afili Aşk'ta beklediğin haber sonunda	29	12.03.2020	09:41:00	CocaCola_TR	11	85.437	0,0468%
Mucize Doktor' da ateş bacayı sardığı	16	12.03.2020	09:36:10	CocaCola_TR	4	85.437	0,0234%
İkrakk Merhaba böyle bir problem ya	1	10.03.2020	15:42:03	CocaCola_TR	0	85.437	0,0012%
gulbbeyyaaz Keşke her gün Çukur iz	38	10.03.2020	13:55:51	CocaCola_TR	13	85.437	0,0597%
xspringdaisy Keşke her gün Çukur iz	6	10.03.2020	13:55:38	CocaCola_TR	2	85.437	0,0094%
pelinbrls Keşke her gün Çukur izleyi	3	10.03.2020	13:55:23	CocaCola_TR	2	85.437	0,0059%
cereniko Keşke her gün Çukur izleyi	6	10.03.2020	13:55:01	CocaCola_TR	1	85.437	0,0082%
azkarqueen Haftaya da Coca Cola'la	11	10.03.2020	13:53:32	CocaCola_TR	2	85.437	0,0152%
monoduyar Haftaya da Çukur' da en	5	10.03.2020	13:53:06	CocaCola_TR	0	85.437	0,0059%
sadecekendimben Keşke her gün Çu	1	10.03.2020	13:52:40	CocaCola_TR	0	85.437	0,0012%

Bir sonraki veri ön işleme adımında ise tweetlerin paylaşıldığı aylar, günler ve saatler “MS Excel” programının metneçevir formülüyle istenilen formata dönüştürülmüştür. Dönüştürülen verilerin görüntüsü Tablo 3’teki gibidir.

Tablo 3 : Gün, Ay ve Zaman Dilimlerine Ayrılmış Tweetler

Tweet Id	Günler	Aylar	Zaman Periyotları	Twitter Hesabı	Etkileşim Oranı
2	Perşembe	Mart	9	CocaCola_TR	0,0234
3	Salı	Mart	15	CocaCola_TR	0,0012
23	Pazartesi	Mart	19	CocaCola_TR	0,0094
32	Pazartesi	Mart	18	CocaCola_TR	0,0023
35	Pazartesi	Mart	18	CocaCola_TR	0,0059
40	Pazartesi	Mart	14	CocaCola_TR	0,0012
45	Pazartesi	Mart	14	CocaCola_TR	0,0117
61	Pazartesi	Mart	13	CocaCola_TR	0,0164
79	Pazartesi	Mart	12	CocaCola_TR	1,0815
83	Perşembe	Mart	14	CocaCola_TR	0,0047
89	Perşembe	Mart	14	CocaCola_TR	0,0199
97	Perşembe	Mart	14	CocaCola_TR	0,0082
100	Perşembe	Mart	14	CocaCola_TR	0,0012
103	Perşembe	Mart	14	CocaCola_TR	0

Metin Madenciliği

Bu çalışmada hipotez 6’yı test edebilmek ve tweetlere yeni bir değişken ekleyebilmek için “RapidMiner” programının “MeaningCloud” eklentisiyle aktif olan “Topic Extraction” operatörü kullanılmıştır. Tweet içeriğindeki metinlerinden kelimeler çıkarılmıştır. “MeaningCloud” programının ücretsiz versiyonu kullanıldığından, işlenebilecek tweet sayısı için 20.000 adet limiti bulunmaktadır. Bu nedenle hipotez 6’yı test etmek için sadece 13 şirketin Twitter hesabından paylaşılan veri setindeki toplam 12.560 tweet “Topic Extraction” operatörü için girdi olarak kullanılmıştır. “Topic Extraction” sonucunda 5593 kelime tweet metinleriyle birlikte matris halinde gelmiştir. Fakat bu 5593 kelimenin içinde büyük bir çoğunluk hiçbir anlam ifade etmeyen ya da hiçbir kategoriye dahil edilemeyen, yorumlanamayan terimler bulundurmaktadır. Anlamli veriler çıkarabilmek için 5593 kelime manuel olarak taranmış ve gerekli ayıklamalar sonucunda sadece 539 kelimeyle çalışılmaya devam edilmiştir.

Kategorileme

Konu modelleme sonucu temin edilen 539 kelime, tweetlere anlamlı içerik kategorisi oluşturabilmek için tek tek incelenmiştir. Bu kelimelerin içinde aynı başlık altında toplanabileceği düşünülen ve aynı kelimeyi ifade eden fakat içindeki farklı karakterler (büyük/küçük harf kullanımları

vb.) yüzünden farklı kelimelermiş gibi algılanan kelimeler gruplandırılmıştır. Bu işlem sonrasında geriye 184 kelime kalmıştır.

Sonraki adımda 184 kelime, tweetleri birbirinden ayırmamıza yardımcı olacak 9 kategoriye ayrılmıştır. Bu kategoriler; promosyon, reklam, magazin, spor, eğlence, tebrik, tedarik, ulusal (milli) ve toplumsal (Keskin & Sönmez, 2015) içeriklidir. Belirlenen bu kategoriler doğrultusunda Tablo 4'te görüldüğü gibi 184 kelimeyi içeren bir kategori sözlüğü oluşturulmuştur (Y. Kim et al., 2016).

Tablo 4: Kategori Sözlüğü

Kategori No	Kategori İsmi	Kelime Sayısı	1	2	3	4	5	6
1	Promosyon	6	Hediye	Edin	Son	İnternet	Kazan	Ödül
2	Reklam	17	Dizi İsmi	McDonalds	Formula1	NBA	FIFA	Çizgi Karakter
3	Magazin	24	Hadise	Murat Boz	Tv8	Al Pacino	Sertab Erener	Yüzyüzeyken Konuşuruz
4	Spor	16	Taraf	Takım	Taraftar	Gol	Paintball	Basket
5	Eğlence	50	Festival	Salla	Normal Film	Sanat	Dans	Roman
6	Tebrik	11	Özel Gün	Milli Bayram	Bayram	Tebrikler	Dini Bayram	Bayram
7	Tedarik	10	Temin	Market	Süpermarket	Tüketici	Evet	Özel
8	Ulusal	9	Türk	Atatürk	Türkiye	Demokrasi/Park	Cumhuriyet	Vatan
9	Toplumsal	41	Sosyal Uygulama	Mola	Bitti/Sosyal	Şehir	Geleneksel	Muhabet
		184						

Tablo 4'te sadece kategori sözlüğünün yapısı hakkında fikir verme amaçlanmıştır. Tabloda sadece "Promosyon" kategorisinin kelimelerinin tamamı görülmektedir, diğer kategoriler için belirlenmiş anahtar kelimelerin tamamı Ek 2'de verilmiştir.

Kelimeleri belirtilen 9 kategoriye dahil etme sürecinde öncelikle çok belirgin bir şekilde 9 kategoriden birisine dahil olabilecek kelimeler ayrılmıştır geri kalan kelimeler (55%) Excel'in filtreleme imkanıyla içlerinde geçtikleri tweetlerin incelenmesi sonucu ilgili kategorilere dahil edilmiştir.

Sözlüğün oluşturulması sonucu ilk aşamada tespit edilen ve içlerinde geçtiği tweet sütunlarıyla birlikte matris halinde mevcut olan 539 kelime, oluşturulan sözlük çerçevesinde incelenmek için Tablo 5'te özetle görüldüğü gibi 10. satıra eklenmiştir. Üstündeki 9 satır ise sırasıyla satırlarda kategori kelimeleri olacak şekilde doldurulmuştur.

Tablo 5: Kategori Kelimleri ve Topic Extraction Sonucu Gelen Düzenlenmiş Kelimeler

Hediye	Edin	Son	İnternet	Kazan	Ödül				
Dizi İsmi	McDonalds	Formula1	NBA	FIFA	Çizgi Karakter	Otomobil Markası	Karayıpler	Hürriyet	Karam
Hadise	Murat Boz	Tv8	Al Pacino	Sertab Erener	Yüzyüzeyken Konuşuruz	Duman	Demet Akalın	Ronaldinho	Mabel Matiz
Taraf	Takım	Taraftar	Gol	Paintball	Basket	Pilates	Motorspor	Derbi	Saha
Festival	Salla	Normal Film	Sanat	Dans	Roman	Kamp	Enerji	Tatil	Plaj
Özel Gün	Milli Bayram	Bayram	Tebrikler	Dini Bayram	Bayram	Yirmi Bir Aralık	Mutlu	En Uzun Gece	Emek
Temin	Market	Süpermarket	Tüketici	Evet	Özel	Bey	Merhaba	Rekabet	İkram
Türk	Atatürk	Türkiye	Demokrasi/Park	Cumhuriyet	Vatan	Millet	Zafer	Saray	
Sosyal Uygulama	Mola	Bitti/Sosyal	Şehir	Geleneksel	Muhabet	Genç	Dünya	Erkekler	Aile
Dizi İsmi	Zamir	Zamir	Dizi İsmi	Selam, Merhaba	Dizi İsmi	Zaman	Dünya	Sosyal Uygulama	Empati

Sonraki adımda Topic Extracion çıktısı 539 düzenlenmiş kelimenin alt satırına kategori numaralarını yazmak için Şekil 1'de gösterilen Excel VBA kodları kullanılmıştır. Şekil 1'de 6 kelimelik ilk kategori olan "Promosyon" kategorisi için oluşturulan adımlar her yeni kategori için dizi sayılarının sınırı ve kategori numarası değiştirilerek 9 kategori için de tekrarlanmıştır.

Şekil 1: Kelimelerin Kategori Numaralarını Belirlemek İçin Kullanılan Excel VBA Kodları

```

Sub kategorinumaraekleme ()
    Dim i (1 To 6) As String
    Dim a As Integer
    Dim j As Integer

    m = 1
    For k = 1 To 6
        i(k) = Cells(1, m)
        m = m + 1
    Next k

    For a = 1 To 539

        For k = 1 To 6

            If Cells(10, a) = i(k) Then
                Cells(11, a) = 1
            End If

        Next k
    Next a
End Sub

```

Expression	Value	Type
Module2		Module2/Module2
i(1)	"Hediye"	String(1 to 6)
i(2)	"Edin"	String
i(3)	"Son"	String
i(4)	"Internet"	String
i(5)	"Kazan"	String
i(6)	"Odul"	String

Kelimelerin altına kategori numaraları eklendikten sonra tweetlerle kategori numaralarının kesişimlerini işaretleyerek bir matris oluşturmak için Tablo 6'da özetle gösterilen Excel çalışma sayfasına Şekil 2'de gösterilen Excel VBA kodları yazılmıştır.

Tablo 6: Topic Extraction Çıktısı 539 Kelimeli Matrisin Kategori Numaraları Verilmiş Hali

	Kelimenin Tweet İçinde Geçme Şekli	ent_Devam_Mucize Doktor_Mucize	ent_Bizi	ent_Sana	ent_ZAMAN	ent_Dunya	ent_Netflix	ent_rempatiLaz
	Tweet Toplamı	1	154	624	329	43	3	1
	Kelimeleri Anlamlandırma	Kelime Sayısı	1	2	3	4	5	6
	Kategori Sayısı	2	9	9	5	9	9	9
	Tweet	ent_Devam_Mucize Doktor_Mucize	ent_Bizi	ent_Sana	ent_ZAMAN	ent_Dunya	ent_Netflix	ent_rempatiLaz
1	Afili Aşk' ta beklediğin haber sonund	0	0	0	0	0	0	0
2	Mucize Doktor' da ateş bacadı sardığ	1	0	0	0	0	0	0
3	lkrakk Merhaba böyle bir problem ya	0	1	0	0	0	0	0
4	gulbbeyyaaz Keşke her gün Çukur iz	0	0	0	0	0	0	0
5	xspringdaisy Keşke her gün Çukur iz	0	0	0	0	0	0	0
6	pelinbrls Keşke her gün Çukur izleyi	0	0	0	0	0	0	0
7	cereniko Keşke her gün Çukur izleyi	0	0	0	0	0	0	0
8	azkarqueen Haftaya da Coca Cola'la	0	0	0	0	0	0	0
9	monoduyar Haftaya da Çukur' da en	0	0	0	0	0	0	0
10	sadecendimben Keşke her gün Çu	0	0	0	0	0	0	0
11	Mila13567966 Hislerimizin karşılıkl	0	0	0	0	0	0	0
12	laarasn Hislerimiz karşılıklı laarasn	0	0	0	0	0	0	0
13	EmirKoz01 azerkaracaaa Kesinlikle	0	0	0	0	0	0	0
14	misafirta <U+2764><U+FE0F><U+	0	0	0	0	0	0	0
15	nisa23nur Hislerimiz karşılıklı nisa2	0	0	0	0	0	0	0

Şekil 2: Kategorilerin Kelimelerinin Dahil Olduğu Tweetlerin Frekanslarının Toplamını Görebilmek İçin Oluşturulan Excel VBA Kodu

```

Private Sub CommandButton1_Click()

For i = 1 To 12560 'the number of tweets
  For j = 1 To 539 'the number of words
    For k = 1 To 9 'the number of category
      If Cells(4, j + 9) = k Then
        Cells(5 + i, 548 + k) = Cells(5 + i, 548 + k) + Cells(5 + i, j + 9)
      End If
    Next k
  Next j
Next i

End Sub

```

Kodun çalıştırılmasıyla tweetlerde o kategoriye ait kelimelerin ne sıklıkla bulunduğu ve bulunup bulunmadığı Tablo 7’deki gibi gözlenmiştir.

Tablo 7: Tweet-Kategori Matrisi

Tweet Id	Tweet	1	2	3	4	5	6	7	8	9		
2	Mucize Doktor' da ateş bac	0	1	0	0	0	0	0	0	0	1	Kategori 2
3	İkrak Merhaba böyle bir p	0	0	0	0	0	0	0	0	1	1	Kategori 9
23	NKFC9 Çukur' da en güzel	0	0	0	0	0	0	0	0	1	1	Kategori 9
32	nomatter Özetten dizi sonun	0	1	0	0	0	0	0	0	0	1	Kategori 2
35	kubican t Özetten dizi donu	0	1	0	0	0	0	0	0	0	1	Kategori 2
40	yoruwonukeru Merhaba bö	0	0	0	0	0	0	0	0	1	1	Kategori 9
45	booom71956926 Selamla	0	0	0	0	0	0	1	0	0	1	Kategori 7
61	tuuceess Çukur' un en heye	0	0	0	0	0	0	0	0	1	1	Kategori 9
79	AzKar aşkında hararet arttı	0	1	0	0	0	0	0	0	0	1	Kategori 2
83	detokscum232 Perşembe a	0	0	0	0	0	0	0	0	1	1	Kategori 9
89	JrMahshid Perşembe akşam	0	0	0	0	0	0	0	0	1	1	Kategori 9
97	DisasterQueenE Perşembe	0	0	0	0	0	0	0	0	1	1	Kategori 9
100	JustSofia1009 MucizeDokt	0	0	0	0	1	0	0	0	0	1	Kategori 5
103	mucizecftalnaz MucizeDok	0	0	0	0	0	0	0	0	1	1	Kategori 9
108	teacherralla Geri sayım başl	0	0	0	0	0	0	0	0	1	1	Kategori 9
116	Gke77728820 MucizeDokt	0	0	0	0	1	0	0	0	0	1	Kategori 5
122	Gençler kendi geleceğine k	0	0	0	0	0	0	0	0	1	1	Kategori 9

Sonrasında satırda bulunan en yüksek değer frekans değeri olarak belirlenmiş ve bu en yüksek değer hangi kategori sütunundan geldiği Excel çoklu “eğer” formülü ile not edilmiştir. Bu aşama sonrası hiçbir kategoriye dahil olmayan yani kategori frekansı sıfır olan tweetler çıkarılmıştır ve 3.436 tweet kalmıştır. Tweetlerin içeriklerine göre hangi kategoride olacaklarını da belirledikten sonra 3436 verinin önceki düzenlenmiş değişkenlerle birleştirilmiş yeni halinin özeti Tablo 8’deki gibi olmuştur.

Tablo 8:R Programı ile Çekilen Tweet Verilerinin Son Hali

Tweet Id	Günler	Aylar	Zaman Periyotları	Twitter Hesabı	Kategori	Etkileşim Oranı
2	Perşembe	Mart	9	CocaCola_TR	2	0,0234
3	Salı	Mart	15	CocaCola_TR	9	0,0012
23	Pazartesi	Mart	19	CocaCola_TR	9	0,0094
32	Pazartesi	Mart	18	CocaCola_TR	2	0,0023
35	Pazartesi	Mart	18	CocaCola_TR	2	0,0059
40	Pazartesi	Mart	14	CocaCola_TR	9	0,0012
45	Pazartesi	Mart	14	CocaCola_TR	7	0,0117
61	Pazartesi	Mart	13	CocaCola_TR	9	0,0164
79	Pazartesi	Mart	12	CocaCola_TR	2	1,0815
83	Perşembe	Mart	14	CocaCola_TR	9	0,0047
89	Perşembe	Mart	14	CocaCola_TR	9	0,0199
97	Perşembe	Mart	14	CocaCola_TR	9	0,0082
100	Perşembe	Mart	14	CocaCola_TR	5	0,0012
103	Perşembe	Mart	14	CocaCola_TR	9	0

2.5 Hipotez Testleri

Hipotez 1, 2, 3, 4 ve 5'te 22.931, hipotez 6'da ise 3.436 tweet verisi kullanılmıştır. Hipotezlerde kullanılacak testlere karar verilmesi için veri setlerinin normal dağılıp dağılmadığı incelenmiştir.

Normallik testleri

Tablo 9:Hipotezlerde Kullanılacak Veri Setleri İçin Normallik Değerlendirmeleri

	Değerlendirme	Beğeni Sayısı	Retweet Sayısı	Takipçi Sayısı	Etkileşim Oranı
1	Histogram Grafikleri	-	-	+	-
2	Varyasyon Katsayısı	-	-	-	-
3	Skewness/Kurtosis	-	-	-	-
4	Detrended	-	-	+	-
5	Normallik Testleri	-	-	-	-

Hipotezlerde kullanılacak veri setlerinin hiçbirinde 3/5 skoru elde edilemediği için veri setlerinin hiçbiri normal dağılıma uymaz. Bu durumda hipotez 1'de Korelasyon testinin **Spearman Korelasyon Analizi** ile yapılmasına karar verilmiştir.

Hipotez 2'de tweetlerin saatleri (1 ve 24 arasındaki periyotlar) etkileşim oranını etkilemektedir ifadesi kullanılmıştır. Bu durumda saat periyodu grupları normal dağılıma uymayan ana kitleden geldikleri için grupların ortalamaları arasındaki farklılığı analiz etmek amacıyla **Kruskal Wallis H Testi** uygulanmıştır.

Hipotez 3'te Paylaşılan tweetlerin etkileşim oranları mevsimlere göre farklılık göstermektedir ifadesi yer almaktadır. Grupların ortalamaları arasındaki farklılığı analiz etmek amacıyla **Kruskal Wallis H Testi** uygulanmıştır.

Hipotez 4'te hafta sonu paylaşılan tweetlerin daha yüksek etkileşim oranına sahip olduğu ileri sürülmüştür. Hipotezi test etmek için T-testin nonparametrik karşılığı olan **Mann Whitney U Testi** uygulanmıştır.

Hipotez 5'te paylaşılan tweetlerin etkileşim oranları arasında markalar çerçevesinde anlamlı bir fark vardır ifadesi kullanılmıştır. Grupların ortalamaları arasındaki farklılığı analiz etmek amacıyla **Kruskal Wallis H Testi** uygulanmıştır.

Son olarak hipotez 6'da ise metin analizi sonucu oluşturulan kategorilerin etkileşim oranları ortalamalarının değerleri arasında anlamlı bir farklılık vardır ifadesi yer almaktadır. Bu durumda

kategori grupları normal dağılıma uymayan ana kitleden geldikleri için grupların ortalamaları arasındaki farklılığı analiz etmek amacıyla **Kruskal Wallis H Testi** uygulanmıştır.

Sınıflandırma Modeli

Düzenlenmiş ve üzerinde analizler yapılmış 3436 veri ile yeni oluşturulacak bir tweetin etkileşim aralığını tahmin edebilmek için etkileşim oranı sürekli değişkeni kategorik değişkene dönüştürülmüştür.

Oran ölçeği olan Etkileşim Oranı değişkeni aralık ölçeğe dönüştürülürken veriler normal dağılmadığı için aralıkları içeren değerler literatürde (Wadhwa et al., 2017) yer alan ifadelerle göre incelenmiş bizim etkileşim oranı formülümüzde yorum sayısının yer almadığı düşünülerek tahmini olarak Tablo 10'daki gibi belirlenmiştir.

Tablo 10 :Etkileşim Oranı Değerlerinin Kategorileştirilmesi

Aralık		Tanım
3	5,4753	Çok Başarılı
2	2,999	Başarılı
1	1,9999	Kısmen Başarılı
0,0012	0,9999	Kısmen Başarısız
0	0,0011	Başarısız

Sonrasında araştırma hedefi doğrultusunda "RapidMiner" programı ile analiz edilecek model Tablo 11'deki gibi özetlenmiştir.

Tablo 11 : Tahmin Edilecek Değişkenin Tahmin Edici Değişkenlerle Birlikte Modellenmesi

Araştırma Hedefi	Paylaşılacak bir tweetin "Etkileşim Oranı" nın tahmin edilmesi				
	Değişken	Değişken Ölçeği	Değişkenin Ölçülme Yöntemi	Veri Kaynağı	
Hedef Değişken	Etkileşim Oranı	Aralık	(0,0000-0,0011) Başarısız, (0,0012-0,9999) Kısmen Başarısız, (1,0000-1,9999) Kısmen Başarılı, (2,0000-2,9999) Başarılı, (3,0000-5,4753) Çok Başarılı	Twitter	
Öznitelikler	1	Gün	Nominal	Tweetin paylaşılma tarihinden haftanın hangi günü paylaşıldığı bilgisi alındı.	Twitter
	2	Ay	Nominal	Tweetin paylaşılma tarihinden paylaşıldığı ay bilgisi alındı.	Twitter
	3	Zaman Periyodu	Aralık	Tweetin paylaşıldığı saat 24 dilimlik periyotta ifade edildi.	Twitter
	4	Twitter Hesabı	Nominal	Tweeti paylaşan şirketin kullandığı Twitter hesabı belirtildi.	Twitter
	5	Kategori	Nominal	Tweet metnlerinin, metin madenciliği ile analiz edilmesi sonucu kategoriler oluşturuldu.	Twitter

Veri seti Şekil 3'te verildiği gibi RapidMiner'a tanıtılmıştır.

Şekil 3: Veri Setini Rapidminer Programına Tanıtma

Import Data - Format your columns.

Format your columns.

Replace errors with missing values ⓘ

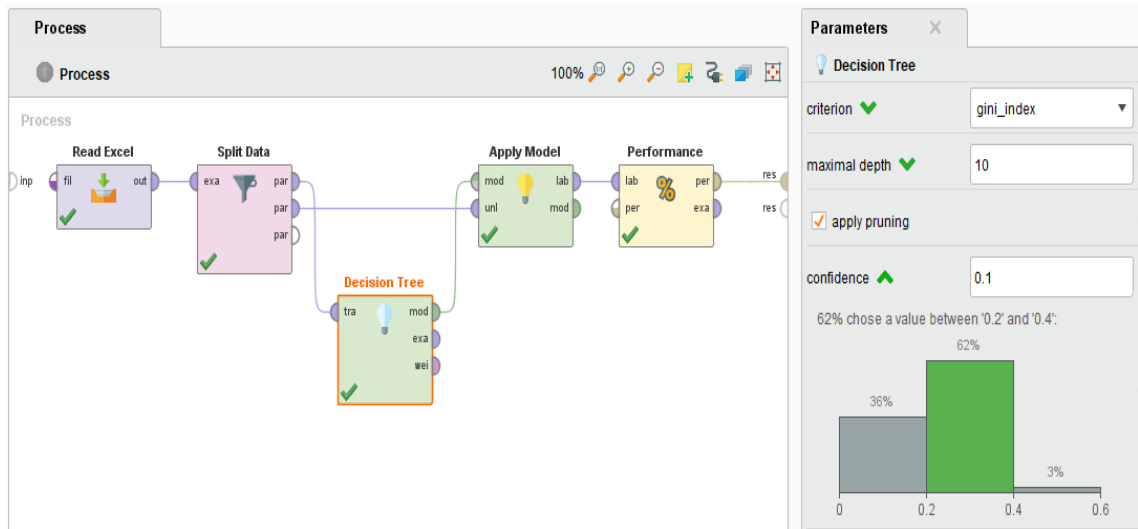
	Days <i>polynomial</i>	Months <i>polynomial</i>	Time Periods <i>integer</i>	Account <i>polynomial</i>	Category <i>polynomial</i>	Engagemen... <i>polynomial label</i>
1	Thursday	March	9	CocaCola_TR	2	Relatively Unsuc...
2	Tuesday	March	15	CocaCola_TR	9	Relatively Unsuc...
3	Monday	March	19	CocaCola_TR	9	Relatively Unsuc...
4	Monday	March	18	CocaCola_TR	2	Relatively Unsuc...
5	Monday	March	18	CocaCola_TR	2	Relatively Unsuc...
6	Monday	March	14	CocaCola_TR	9	Relatively Unsuc...
7	Monday	March	14	CocaCola_TR	7	Relatively Unsuc...
8	Monday	March	13	CocaCola_TR	9	Relatively Unsuc...
9	Monday	March	12	CocaCola_TR	2	Relatively Succe...
10	Thursday	March	14	CocaCola_TR	9	Relatively Unsuc...
11	Thursday	March	14	CocaCola_TR	9	Relatively Unsuc...
12	Thursday	March	14	CocaCola_TR	9	Relatively Unsuc...

no problems.

Previous Finish Cancel

Veri setine en uygun modeli bulma amaçlı K-NN, Naive Bayes, Decision Tree, Random Forest, Gradient Boosted Trees, Decision Stump, Random Tree sınıflandırma teknikleri tek tek denenmiş ve %74,42 (Tablo 12) ile en yüksek accuracy değerine Şekil 4'te görülen **Decision Tree** yöntemi modeliyle ulaşılmıştır.

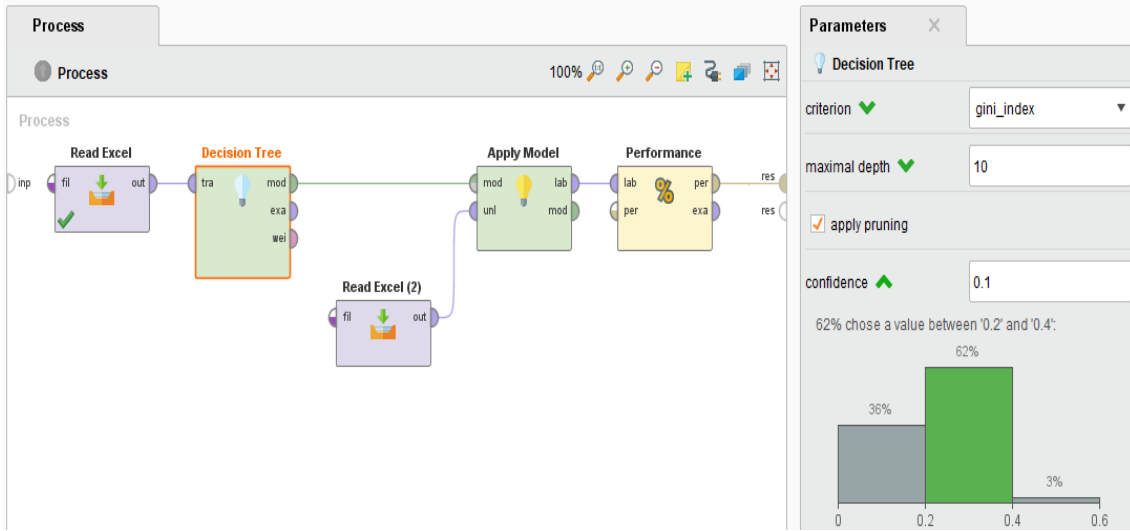
Şekil 4 : Etkileşim Oranı Aralığını Tahmin Etmek İçin Oluşturulan Decision Tree Modeli



Tablo 12 : Rapidminer Decision Tree Sınıflandırma Modeli Accuracy Çıktısı

Accuracy : 74.42 %						
	True Relatively Unsuccessful	True Relatively Successful	True Unsuccessful	True Very Successful	True Successful	Class Precision
Prediction Relatively Unsuccessful	425	9	103	3	2	78.41 %
Prediction Relatively Successful	1	0	0	0	0	0.00 %
Prediction Unsuccessful	58	0	87	0	0	60.00 %
Prediction Very Successful	0	0	0	0	0	0.00 %
Prediction Successful	0	0	0	0	0	0.00 %
Class Recall	87.81 %	0.00 %	45.79 %	0.00 %	0.00 %	

Verinin eğitimi için 3436 (80%) tweet kullanılmıştır. Aralıklara bölünen etkileşim oranı değerini test için ise 19 Mayıs 2020 tarihinde çekilen (13 hesaptan 66 tweet) 858 (20%) tweetlik yeni veri seti oluşturulmuştur. Bu veri seti Şekil 5'te görüldüğü gibi modele eklenmiştir.

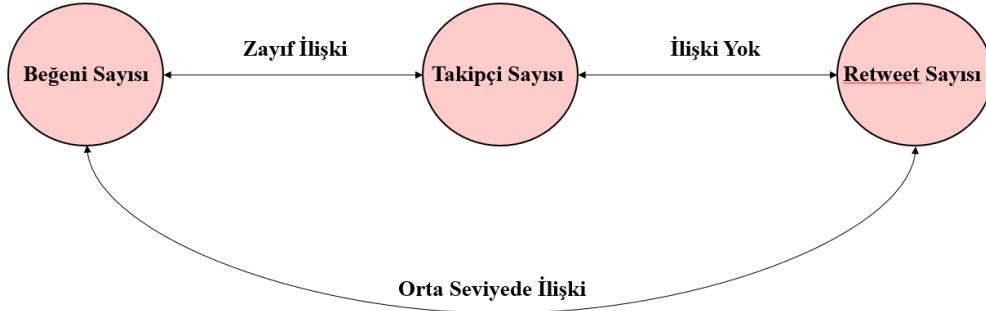
Şekil 5 :19 Mayıs 2020' de Çekilmiş Yeni Veri Setiyle Oluşturulmuş Decision Tree Modeli

BULGULAR VE SONUÇLAR

Tablo 13 : İlk Hipotezin Sonucu

Spearman Korelasyon Analizi		
Beğeni Sayısı	Zayıf İlişki	Takipçi Sayısı
Retweet Sayısı	Orta Seviyede İlişki	Beğeni Sayısı
Takipçi Sayısı	İlişki Yok	Retweet Sayısı

Şekil 6 : İlk Hipotezin Sonucu

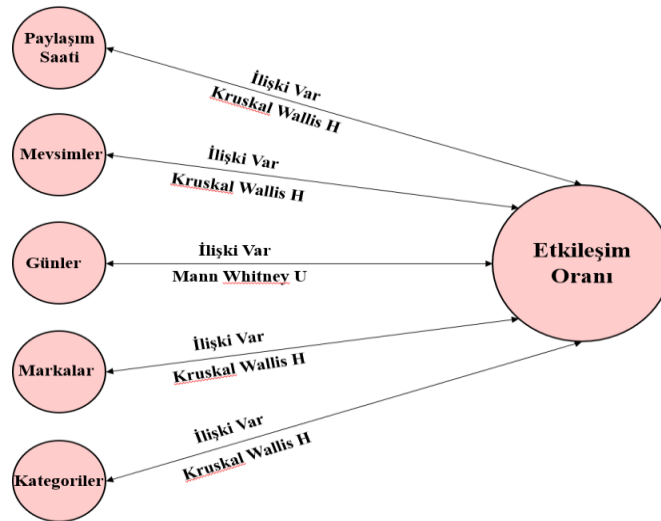


Spearman korelasyon analizi sonucunda beğeni sayısı ile takipçi sayısı arasında aynı yönde zayıf bir ilişki, beğeni sayısı ile retweet sayısı arasında aynı yönde orta seviyede bir ilişki olduğu söylenebilir. Ancak takipçi sayısı ile retweet sayısı arasında anlamlı bir ilişki tespit edilememiştir.

Tablo 14 : Bağımsız Değişkenler ile Bağımlı Değişken Arasındaki İlişki Durumları ve Uygulanan Analizler

Bağımsız Değişkenler	Kullanılan Test	İlişki	Bağımlı Değişken
Paylaşım Saati	Kruskal Wallis H	Var	Etkileşim Oranı
Mevsimler	Kruskal Wallis H	Var	
Günler (Hafta sonu / Hafta içi)	Mann Whitney U	Var	
Markalar	Kruskal Wallis H	Var	
Kategoriler	Kruskal Wallis H	Var	

Şekil 7 : Bağımsız Değişkenler ile Bağımlı Değişken Arasındaki İlişki Durumları ve Uygulanan Analizler



Kruskal Wallis H testi çıktılarının incelenmesi sonucunda, en fazla etkileşim oranı ortalama değerine sahip zaman periyodu grubunun "1" yani "01:00:00 – 02:00:00" arasında paylaşılan tweetler

olduğu gözlenmiştir. 3, 16 ve 17 numaralı saat periyotları en fazla etkileşim oranına sahip saat periyodu sıralamasında 1'den sonra gelerek ilk 4'ü oluşturmuştur.

Mevsimlerin incelendiği diğer bir analizde İlkbahar'da paylaşılan tweetlerin daha fazla etkileşim alarak diğer mevsimlerde paylaşılan tweetlerin etkileşim oranı değerlerine göre fark oluşturduğu görülmüştür.

Tweetlerin hafta sonu ve hafta içi paylaşımlarına göre etkileşim oranları değişimlerini inceleme amaçlı yapılan Mann Whitney U testinde ise hafta sonu paylaşılan tweetlerin etkileşim oranlarının ortalamasının daha yüksek olduğu sonucuna ulaşılmıştır.

Etkileşim oranı değerleri üzerinde farkı yaratan markalar gözlemlendiğinde en fazla etkileşim oranı değerlerine sahip markanın "Red Bull" olduğu gözlenmiştir. Doğu Çay, Hamidiye Su ve Sırma Su markaları en fazla etkileşim oranına sahip markalar sıralamasında Red Bull'dan sonra gelerek ilk 4'ü oluşturmuştur.

Farkı yaratan kategorileri yakalayabilmek için yapılan analizde ise en fazla etkileşim oranı değerlerine sahip kategorinin 6. yani "Tebrik" kategorisi olduğu gözlenmiştir. Ulusal, Promosyon ve Magazin kategorileri en fazla etkileşim oranına sahip kategoriler sıralamasında "Tebrik" kategorisinden sonra gelerek ilk 4'ü oluşturmuşlardır.

Sınıflandırma Modeli Sonuçları

Yöntem kısmında belirtilen Şekil 5'teki model çalıştırıldığında %68,53 doğruluk değeri elde edilmiştir.

Modelin 282 "Başarısız" etkileşim oranı değerlerinden 36'sını doğru sınıflandırabildiği, 246'sını ise ona çok yakın bir değer aralığına sahip "Kısmen Başarısız" kategorisine dahil ettiği görülmüştür.

Test veri setinde bulunan "Kısmen Başarısız" kategorisine dahil 571 değer 552'sini yani %96,67'lik bir kısmı doğru şekilde tahmin etmiştir.

"Başarılı" kategorisine dahil 3 değeri "Kısmen Başarısız" olarak "Kısmen Başarılı" kategorisindeki iki değeri ise birisini "Başarısız" diğerini "Kısmen Başarısız" olarak tahmin etmiştir.

Decision Tree modeli %10 güven aralığı değeriyle Şekil 5'te görüldüğü gibi çalıştırıldığında "Kısmen Başarısız" kategorisi için oldukça başarılı bir tahmin gerçekleştirmiştir. "Başarısız" kategorisi için de kısmen tahmin gerçekleştirebilmiştir. Fakat test verisinde çok az sayıda bulunan ve eğitim veri setinde de sayısı az olan diğer kategorilerin tahminini gerçekleştirememiştir. Bu model ile "Başarısız" ve "Kısmen Başarısız" etkileşim oranı değerlerini tahmin etmek ve hangi niteliklerin düşük etkileşim alacağını görüp yeni yapılacak tweet paylaşımları için bu doğrultuda iyileştirmeler yapmak mümkündür.

Literatür kısmında detayları incelenen çalışmalara bakıldığında etkileşim oranını etkileyen faktörler bütüncül olarak ele alınmamıştır. Sosyal medya paylaşımlarının hem içeriği hem de diğer faktörleri (paylaşım tarihi, saati vb.) bu çalışmada detaylıca incelenmiştir. Bunun yanında incelemelerimiz kadarıyla etkileşim oranını etkileyen faktörler üzerinde çalışılsa da tahmin edilmesi ile ilgili bir çalışma bulunmamıştır. Bu çalışmada bu konuya da odaklanılarak yeni oluşturulacak gönderilerin değerini artırma amaçlanmıştır.

GENEL SONUÇLAR VE ÖNERİLER

Sosyal medya araçlarının, müşterinin ihtiyaçlarını anlamada önemli platformlar olduğu ve etkileşim oranının, kurumun sosyal medya performansını ölçmede önemli bir metrik olduğu literatürden faydalanılarak açıklanmaya çalışılmıştır.

Bu doğrultuda, giriş kısmında önemli bir sosyal medya paylaşım sitesi olduğu vurgulanan Twitter'da işletmelerin yaptıkları paylaşımların (tweetler); beğenilme ve yeniden paylaşılma (retweet) miktarları ile hesaplanan etkileşim oranını belirleyen faktörler araştırılmıştır. Bu amaçla Türkiye'de faaliyet gösteren ve en fazla takipçi sayısına sahip, içecek sektöründeki 25 şirketin Twitter hesaplarının hepsinden ayrı ayrı 12 Mart 2020'ye kadar paylaşılmış son 1000 tweet R programıyla

çekilmiştir. Toplam 22.931 tweet ilgili nitelikleriyle birlikte listelenmiştir. Bu tweetlerin içerisinde 12.560 tweet “RapidMiner MeaningCloud” uzantısının “Topic Extraction” operatörü kullanılarak analiz edilmiştir. Analiz sonucu programdan kelimeler tek tek incelenerek anlamlandırılmış ve anlamlandırılmayan kelimeler ayrılmıştır. Geriye kalan kelimelerle tweetlere içerikleri hakkında bilgi verebilecek bir değişken eklemek için kategorileme uygulanmıştır. Kategoriler için belirlenmiş anahtar kelimeleri içeren 3.436 tweet, içeriklerine göre 9 kategoriye ayrılarak tweetlere kategori niteliğini eklemek için Excel VBA programı ile kullanılmıştır. Sonrasında tweetlerin etkileşim değerleri ile paylaşıldığı saat periyotları, mevsimler, günler, markalar, içerikler (kategoriler) arasındaki ilişkiler analiz edilmiştir. Testler sonucunda, belirtilen bu değişkenlerin etkileşim oranı üzerinde etkili olduğu görülmüştür. Örneğin İlkbahar’da, hafta sonları, gece 01-02 saatlerinde paylaşılan ve tebrik içerikli tweetler daha fazla etkileşim oranı almaktadır. Beklenilenden farklı olarak promosyon içerikli tweetlerin, tebrik ve ulusal (milli) içerikli tweetlerden sonra en fazla etkileşim alan üçüncü kategori olduğu görülmüştür. Bu analizler sonrası yeni oluşturulacak bir tweetin belirtilen nitelikleri bilindiğinde sahip olacağı etkileşim oranını öngörebilmek için “RapidMiner” programında tahmin modeli oluşturulmuştur. 19 Mayıs 2020’de toplanan yeni tweetlerle modelin geçerliliği test edilmiştir ve %68,53 doğruluk oranıyla etkileşim oranının tahmin edilebildiği görülmüştür. Bu doğruluk oranını daha yukarılara çıkarma amaçlı daha fazla anlamlı kelimedenden oluşan ya da daha spesifik bir alanı (soğuk içecekler) içeren veri kümesiyle çalışılabilir. Örnekler benzer veriler içeren bir kümeden geldiğinde etkileşim oranını tahmin etme doğruluğu da yükselecektir.

Sonuç olarak, sosyal medya analizinin etkileşim oranı öngörebilmeye önemli bir yöntem olabileceği ve belirlenen kategorilerdeki tweetlerin paylaşım maliyetlerini optimum hale getirebileceği anlaşılmıştır. Anlamlı kelime sayılarının artması ve kategorileme sürecinde daha fazla anlamlı kelimeyle çalışılabilmesinin sosyal medya analizinin etkileşim oranı öngörmedeki başarısını daha da artırabileceği düşüncesiyle metin madenciliğinde kullanılan mevcut yapay zeka araçlarının dil seçeneklerinin artırılması önerilmiştir. Dil seçeneğine ek olarak büyük harf, küçük harf, boşluk gibi karakter farklılıklarından dolayı aynı kelimeleri farklı kelimeler gibi algılama sıkıntısı çalışmada zaman kaybına sebebiyet vermektedir. Metin madenciliği araçlarında kullanılan karakter tanımlamaları da güçlendirildiğinde asıl amaca daha fazla odaklanılabileceği ve daha verimli modeller geliştirilebileceği düşünülmüştür.

KAYNAKÇA

- Abanoz, E. (2012). Using Twitter’s “Hashtag” in Television Series to Create A Social Capital: Season Final Episodes of Turkish Television Series. In *AJIT-e: Bilişim Teknolojileri Online Dergisi* (Vol. 3, pp. 75–85). Akademik Bilişim Araştırmaları Derneği.
- Alalwan, A. A., Rana, N. P., Dwivedi, Y. K., & Algharabat, R. (2017). Social media in marketing: A review and analysis of the existing literature. *Telematics and Informatics*, 34(7), 1177–1190. <https://doi.org/10.1016/j.tele.2017.05.008>
- Alt, D. (2015). College students’ academic motivation, media engagement and fear of missing out. *Computers in Human Behavior*, 49, 111–119. <https://doi.org/https://doi.org/10.1016/j.chb.2015.02.057>
- Ayan, B., Can, M., & Gürsoy, U. T. (2019). Sosyal Medya Madenciliği İle Firmaların Twitter Verilerinin İncelenmesi. *Uluslararası İktisadi ve İdari İncelemeler Dergisi*, 0–2. <https://doi.org/10.18092/ulikidince.475092>
- Çelik, U., Akçetin, E., & Gök, M. (2017). *RapidMiner ile Uygulamalı Veri Madenciliği*.
- Dijkmans, C., Kerkhof, P., & Beukeboom, C. J. (2015). A stage to engage: Social media use and corporate reputation. *Tourism Management*, 47, 58–67. <https://doi.org/https://doi.org/10.1016/j.tourman.2014.09.005>
- Eralp Üstündağ, M. (2018). *Sosyal Medyada Gerçek Zamanlı Pazarlamanın Rolü : Hızlı Tüketim Markalarının Twitter Paylaşımları Üzerine İnceleme* (Vol. 7).
- Gray, R., Vitak, J., Easton, E. W., & Ellison, N. B. (2013). Examining social adjustment to college in the age of social media: Factors influencing successful transitions and persistence. *Computers & Education*, 67, 193–207. <https://doi.org/https://doi.org/10.1016/j.compedu.2013.02.021>
- Guesalaga, R. (2016). The use of social media in sales: Individual and organizational antecedents, and the role of customer engagement in social media. *Industrial Marketing Management*, 54, 71–79.

- <https://doi.org/https://doi.org/10.1016/j.indmarman.2015.12.002>
- Hollebeek, L. D., Glynn, M. S., & Brodie, R. J. (2014). Consumer Brand Engagement in Social Media: Conceptualization, Scale Development and Validation. *Journal of Interactive Marketing*, 28(2), 149–165. <https://doi.org/10.1016/j.intmar.2013.12.002>
- Keskin, S., & Sönmez, M. F. (2015). Liderler ve Takipçileri: Siyasi Parti Liderlerinin Twitter Performansları ve Takip İlişkisi. *Selçuk Üniversitesi İletişim Fakültesi Akademik Dergisi*, 9(1), 340–362.
- Khan, M. L. (2017). Social media engagement: What motivates user participation and consumption on YouTube? *Computers in Human Behavior*, 66, 236–247. <https://doi.org/https://doi.org/10.1016/j.chb.2016.09.024>
- Kim, A. J., & Ko, E. (2012). Do social media marketing activities enhance customer equity? An empirical study of luxury fashion brand. *Journal of Business Research*, 65(10), 1480–1486. <https://doi.org/https://doi.org/10.1016/j.jbusres.2011.10.014>
- Kim, Y., Jeong, S. R., Dwivedi, R., & Zhang, J. (2016). *Competitive Intelligence in Social Media Twitter : iPhone 6 vs. Galaxy S5*.
- Kocabaş, İ. (2016). Özel Kuruluşların Halkla İlişkiler Aracı Olarak Sosyal Medya Kullanımı: Arçelik ve Bosch Markalarının Kurumsal Facebook Ve Twitter Hesaplarının Kullanımı Üzerine Çalışma. *Selçuk Üniversitesi İletişim Fakültesi Akademik Dergisi*, 9(2). <https://doi.org/10.18094/si.66046>
- Kürkçü Dumanlı, D. (2017). Haber Tweetleri Üzerinden Kullanıcı Etkileşiminin Analizi. *Erciyes İletişim Dergisi*, 221–234.
- Oh, C., Roumani, Y., Nwankpa, J. K., & Hu, H.-F. (2017). Beyond likes and tweets: Consumer engagement behavior and movie box office in social media. *Information & Management*, 54(1), 25–37. <https://doi.org/https://doi.org/10.1016/j.im.2016.03.004>
- Oliveira, G. H. M., & Welch, E. W. (2013). Social media use in local government: Linkage of technology, task, and organizational context. *Government Information Quarterly*, 30(4), 397–405. <https://doi.org/https://doi.org/10.1016/j.giq.2013.05.019>
- Paslanmaz, İ., & Narmanlıoğlu, H. (2020). Instagram Hikayelerinde Benliğin Sunumu: Influencer'lar Üzerine Bir Araştırma. *AJIT-e: Online Academic Journal of Information Technology*, 10(39). <https://doi.org/10.5824/ajit-e.2019.04.002>
- Sabate, F., Berbegal-Mirabent, J., Cañabate, A., & Leberherz, P. R. (2014). Factors influencing popularity of branded content in Facebook fan pages. *European Management Journal*, 32(6), 1001–1011. <https://doi.org/https://doi.org/10.1016/j.emj.2014.05.001>
- Sabuncu, İ., & Atmış, M. (2018). Twitter'da Türk Hava Yolları İçin Sosyal Medya Analitiği Uygulamaları. 23. *Pazarlama Kongresi*, 853–860.
- Semiz, G., & Berger, P. D. (2017). Determining the Factors That Drive Twitter Engagement-Rates. *Archives of Business Research*, 5(2), 38–47. <https://doi.org/10.14738/abr.52.2700>
- Sterne, J. (2010). *Social media metrics: How to measure and optimize your marketing investment*. John Wiley & Sons.
- Wadhwa, V., Latimer, E., Chatterjee, K., McCarty, J., & Fitzgerald, R. T. (2017). Maximizing the Tweet Engagement Rate in Academia: Analysis of the AJNR Twitter Feed. *American Journal of Neuroradiology*, 38(10), 1866–1868. <https://doi.org/10.3174/ajnr.A5283>
- Warren, A. M., Sulaiman, A., & Jaafar, N. I. (2014). Social media effects on fostering online civic engagement and building citizen trust and trust in institutions. *Government Information Quarterly*, 31(2), 291–301. <https://doi.org/https://doi.org/10.1016/j.giq.2013.11.007>

EK-1

Tablo 15 : Analizlerde Kullanılan 25 İçecek Şirketinin Twitter Hesapları (17 Aralık 2019, <https://www.boomsocial.com/>)

Twitter Hesabı	Takipçi Sayısı
Coca-Cola Türkiye	85.437
Dimes Türkiye	66.491
Cappy Türkiye	46.211
Lipton Türkiye	41.843
Bum Türkiye	37.387
Çamlıca	36.464
Fanta Türkiye	35.744
Café Crown	31.675
Fusetea	31.577
Red Bull Türkiye	31.439
Damla	28.777
Doğuş Çay	25.244
Nescafé Xpress	25.092
Monster Energy TR	23.421
Sırma	20.119
Doğadan	20.067
Erikli	16.455
Hamidiye İstanbul	15.198
Sprite Türkiye	13.730
Obsesso	13.699
Coffee-Mate	11.311
Yedigün	10.577
Çaykur	9.458
Saka Su	8.306
Pepsi Türkiye	6.785

EK-2

Tablo 16 : Kategori Kelimelerinin Sözlüğü

Kategori No	Kategori İsmi	Kelime Sayısı			
1	Promosyon	6	Hediye	Edin	Son
			İnternet	Kazan	Ödül
2	Reklam	17	Dizi İsmi	McDonalds	Formula 1
			Otomobil Markası	Karayıpler	Hürriyet
			Yaban Mersini	Bitki	Papatya
			NBA	FIFA	Çizgi Karakter
			Karam	Devam	Yasemin
			Gazoz	Kaynak	
3	Magazin	24	Hadise	Murat Boz	Tv8
			Duman	Demet Akalın	Ronaldinho
			Schumi	Ronaldo	Zeki Müren
			Cemal Süreya	Program	Ceza
			Al Pacino	Sertab Erener	Yüzyüzyken Konuşuruz
			Mabel Matiz	David Bowie	Magazin
			Bob Marley	Marc Marquez	Messi
			Kurucu	Eve	Hafta
4	Spor	16	Taraf	Takım	Taraftar
			Pilates	Motorspor	Derbi
			Hepiniz	Yolculuk	Hikaye
			Gol	Paintball	Basket
			Saha	Final	Park
			Gurur		
5	Eğlence	50	Festival	Salla	Normal Film
			Kamp	Enerji	Tatil
			Zaman	Sence	Gün
			İnsan	İlham	Limon
			Deniz	Birlikte/Plan	Pratik
			Sabah	Portakal	Kadar
			Aralar	Herkes	Video
			Ses	Pasta	Mangal
			Sanat	Dans	Roman
			Plaj	Sahne	Tür Gerilim
			Arada	Normal Gün	Hadi
			Biraz	Elma	Kabul
			Atom	Serin	Dayan
			Gül	Meyve	Muz
			Performans	Tutku	Kim
Takip	Gülmek	Mavi			
Sadece	Bugün				
6	Tebrik	11	Özel Gün	Milli Bayram	Bayram
			Yirmi Bir Aralık	Mutlu	En Uzun Gece
			Tebrikler	Dini Bayram	Bayram
			Emek	Nice	
7	Tedarik	10	Temin	Market	Süpermarket
			Bey	Merhaba	Rekabet
			Tüketici	Evet	Özel
			İkram		
8	Ulusal	9	Türk	Atatürk	Türkiye
			Demokrasi/Park	Cumhuriyet	Vatan
			Millet	Zafer	Saray
9	Toplumsal	41	Sosyal Uygulama	Mola	Bitti/Sosyal

			Genç	Dünya	Erkekler
			Üniversite	Akademi	Yerli
			Zamir	Empati	Hayat
			Yap	Hayal	Minik
			Filiz	Kendi	Ara
			Rahat	Güzel	Gurme
			Şehir	Geleneksel	Muhabbet
			Aile	Avrupa	Birlikte
			Engel	Sapanca	Email
			Sevgi	Tad	Bitti
			Kilometreler	Belki	Anne
			Afiyet	Ev	Sevmek
			Koru	Erik	
		184			