

---

## SOSYAL MEDYA MADENCİLİĞİ İLE FİRMALARIN TWITTER VERİLERİNİN İNCELENMESİ<sup>1</sup>

---

Büşra AYAN<sup>2</sup>, Mustafa CAN<sup>3</sup>, Umman Tuğba GÜRSOY<sup>4</sup>

### Öz

Bu çalışma, farklı sektörlerde faaliyet gösteren rakip firmaların Twitter verilerini analiz ederek, firmaların Twitter verilerinin firmalara göre anlamlı bir uyum gösterip göstermediğinin tespit edilmesini, firmaların Twitter'da paylaştıkları içeriklerin kümelenmesini ve hangi içerik kümesinin en fazla etkileşime yol açtığının belirlenmesini amaçlamaktadır. Bu kapsamda, 2017 yılı boyunca kozmetik, elektronik ve pazaryeri sektörlerinde faaliyet gösteren rakip firmalar tarafından paylaşılan Twitter verileri, Sosyal Medya Madenciliği süreci izlenerek analiz edilmiştir. Firmaların Twitter verilerinin firmalara göre anlamlı bir uyum gösterip göstermediği Uygunluk Analizi ile tespit edilmiştir. Firmaların Twitter paylaşımları ise Metin Madenciliği ön işleme metotlarından faydalanılarak "Özel Teklif", "Yarışma & Etkinlik", "Ürün", "Sosyal", "Destek & Geri Bildirim" ve "Özel Etkileşim" kategori başlıklarıyla kümelenmiştir. Firmaların elde ettikleri etkileşimlerin büyük bir çoğunluğunun azınlıktaki paylaşımlardan gelmesi sebebi ile hangi içerik kümesinin en fazla etkileşime yol açtığı Pareto İlkesi yardımı ile belirlenmiştir.

**Anahtar Kelimeler:** Sosyal Medya, Sosyal Medya Madenciliği, Metin Madenciliği, Pareto İlkesi, Uygunluk Analizi.

**JEL Sınıflandırması:** M10, M19.

---

## ANALYZING TWITTER DATA OF FIRMS WITH SOCIAL MEDIA MINING

---

### Abstract

This study aims to determine whether Twitter data of the firms has a significant correspondence with respect to the firms, to cluster Twitter feeds of the firms and to find out which cluster has the maximum interaction through analyzing the Twitter data of the rival firms operating in different sectors. In this context, Twitter data shared by competitors operating in the cosmetics, electronics and marketplace sectors during 2017 were analyzed by following the process of Social Media Mining. The significant correspondence of Twitter variables of the firms was determined by the Correspondence Analysis. Twitter feeds of the firms were clustered with categories "Special Offer", "Competition & Event", "Product", "Social", "Support & Feedback" and "Special Interaction" by using a number of Text Mining pre-processing methods. Since the majority of the interactions obtained by the firms came from the minority of the feeds, which cluster received more interaction was analyzed with the help of the Pareto Principle.

**Keywords:** Social Media, Social Media Mining, Text Mining, Pareto Principle, Correspondence Analysis.

**JEL Classification:** M10, M19.

---

<sup>1</sup> Bu çalışma, Dr. Öğr. Üyesi Mustafa CAN danışmanlığında Büşra AYAN tarafından hazırlanan ve İstanbul Üniversitesi Sosyal Bilimler Enstitüsü'nde kabul edilen "Sosyal Medya Madenciliği ve Bir Uygulama" başlıklı yüksek lisans tezinden türetilmiştir.

<sup>2</sup> Ar.Gör., MEF Üniversitesi, İktisadi, İdari ve Sosyal Bilimler Fakültesi, İşletme Bölümü, ayanbu@mef.edu.tr, ORCID: 0000-0002-5212-2144

<sup>3</sup> Dr. Öğr. Üyesi, İstanbul Üniversitesi, İşletme Fakültesi, Sayısal Yöntemler Anabilim Dalı, cmustafa@istanbul.edu.tr, ORCID: 0000-0002-7786-5198

<sup>4</sup> Prof.Dr., İstanbul Üniversitesi, İşletme Fakültesi, Sayısal Yöntemler Anabilim Dalı, tugbasim@istanbul.edu.tr, ORCID: 0000-0002-5143-4058

## 1. Giriş

Web 2.0 teknolojisi üzerine kurulu olan sosyal medya, bireylere özgür ve etkileşimli bir ortam sunmaktadır. Sosyal medya kullanıcıları çeşitli platformlar aracılığı ile duygu ve düşüncelerini aktarmakla beraber, başka kullanıcılar tarafından destek görebilmektedir. Bilgi ve fikirler daha önce hiçbir medya ortamında olmadığı kadar hızlı yayılmakta ve hatta sanal ortamdan gerçek dünyaya taşınarak eylemlere dönüşebilmektedir. Büyük yankı uyandırabilen bu medya ortamında üretilen verilerden günümüzde birçok alanda faydalanılmaktadır.

Bireyler, kullandıkları sosyal medya platformuna göre farklılık göstermekle birlikte, fotoğraf, video, konum bilgisi gibi çeşitli türlerde veriler üreterek, büyük verinin oluşumunda rol oynamaktadır. Kullanıcılar tarafından üretilen bu verilerin işlenmesi, son dönemlerde farklı disiplinlerdeki araştırmacıların dikkatini çekmeye başlamıştır. Sosyal Medya Madenciliği olarak adlandırılan yeni bir alan gelişmekte ve müşteri ilişkileri yönetimi, halkla ilişkiler, ürün yönetimi ve reklamcılık gibi çeşitli alanlarda uygulamaları görülmektedir.

Sosyal medya platformlarından çeşitli araçlar ve arayüzler yardımı ile verinin elde edilerek analize hazırlanması ve disiplinler arası metotlar yardımı ile analiz edilerek analiz bulgularının değerlendirilmesi süreçlerinden oluşan Sosyal Medya Madenciliği, birçok araştırmacı tarafından farklı alanlarda uygulamaların gerçekleştirilmesine olanak tanıyan bir tekniktir. Sosyal medyanın gelişmesi ve sosyal medya platformlarının çeşitlenmesi ile birlikte bu yeni medya ortamında üretilen ve paylaşılan verilerinin analizi gün geçtikçe daha da önem kazanmaktadır.

Bu çalışmada Sosyal Medya Madenciliği ve süreci açıklandıktan sonra bu alanda gerçekleştirilen çalışmalara literatür taraması bölümünde değinilmiştir. Çalışmanın uygulama bölümünde Sosyal Medya Madenciliği adımları izlenmiş, sırası ile Uygunluk Analizi, metin madenciliği ön işleme metotlarından faydalanılarak Tweet Analizi ve Pareto İlkesi ele alınarak Rt Analizi gerçekleştirilmiştir. Çalışmanın sonuç bölümünde ise elde edilen bulgular tartışılmıştır.

## 2. Sosyal Medya Madenciliği Kavramı ve Süreci

Kaplan ve Haenlein (2010:61) tarafından "Web 2.0'ın ideolojik ve teknolojik temelleri üzerine kurulmuş ve kullanıcıların oluşturduğu içeriğin üretilmesine ve paylaşılmasına izin veren internet tabanlı uygulamalar grubu" olarak tanımlanan sosyal medya, coğrafik duvarları dağıtmakta ve yeni çevrimiçi topluluklar ortaya çıkararak büyümeye devam etmektedir (Weinberg, 2009:1).

Sosyal medyanın hem bireysel hem de kurumsal amaçlarla kullanılması ile birlikte ortaya çıkan büyük veri yığını içerisinde anlamlı bilgilerin elde edilmesi için sosyal medya verilerinin analiz edilmesi gerekmektedir. Bu çalışmanın konusu olan Sosyal Medya Madenciliği, sosyal medya verilerinden anlamlı örüntülerin analiz edilerek ortaya çıkarılma süreci olarak tanımlanmaktadır. Sosyal Medya Madenciliği, bilgisayar bilimi, veri madenciliği, makine öğrenimi, sosyal ağ analizi, ağ bilimi, sosyoloji, etnografya, istatistik, optimizasyon ve matematik gibi alanları kapsayan disiplinler arası bir alandır (Zafarani vd., 2014:21).

Bir diğer tanımlamaya göre ise Sosyal Medya Madenciliği, sosyal medyadan elde edilen verilerin sistematik bir şekilde analiz edilmesi temelleri üzerine kuruludur. Verilerden çeşitli örüntülerin ve bilgilerin keşfi için kullanılan teknikler veri madenciliği olarak adlandırılırken, bu kavram sosyal medya verileri bağlamında değerlendirildiğinde Sosyal Medya Madenciliği adını almaktadır (Ravindran ve Garg, 2015:4).

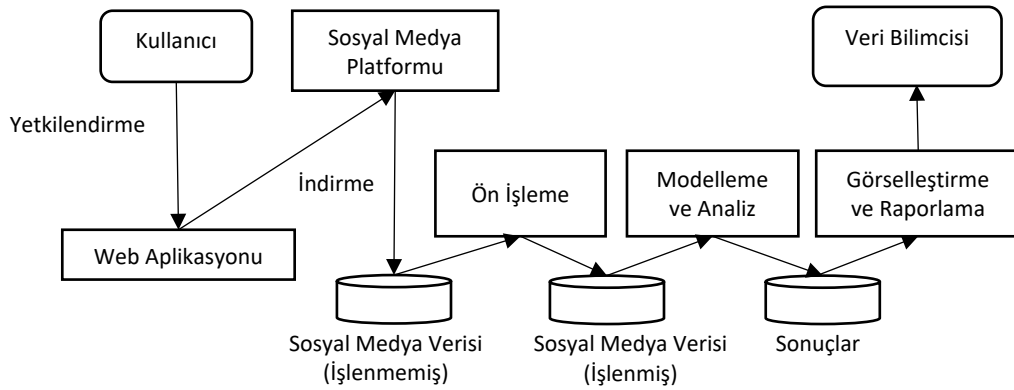
Sosyal medya verilerinin üretildiği platformlar, akademik araştırmalarda büyük miktarda veri toplanmasına olanak tanıyan bir çeşit canlı laboratuvarıdır. Sadece araştırmalar için değil, pratik çözümler için de kullanılan sosyal medya verileri, kurumlar tarafından reklamcılık, halkla ilişkiler, müşteri ilişkileri yönetimi (CRM) ve iş zekâsı (BI) hedeflemede fırsat olarak görülmektedir. Özellikle pazarlama için ek bir kanal olarak etkin bir şekilde kullanılabilmesi, birçok kurumsal firma adına önem arz etmektedir. Son zamanlarda ise siyasi kurumlar, kamuoyunu izlemek, güncel siyasi konuları tespit etmek ve sosyal medyada kendi itibarlarını yönetmek için de sosyal medyaya eğilim

göstermektedirler. Ayrıca, sağlık kurumları, sosyal medya verilerine dayanarak, zamanında müdahaleye yardımcı olacak hastalık salgınlarının teşhisi için erken uyarı sistemi oluşturabilmektedirler (Stieglitz vd., 2014:90).

Sosyal medyanın veri kaynağı olarak kullanılması eşsiz ve yeni özellikleri beraberinde getirmektedir. Anket gibi yöntemlerle veri toplama sırasında katılımcılara, davranışları veya duyguları ile ilgili geriye dönük olarak sınırlı kapsamda sorular sorulurken veya hayali bir senaryo üzerinden ileriye yönelik davranışları sorgulanırken, sosyal medya kişisel ifadeyi ve bireyler arasındaki etkileşimi gerçek zamanlı ve geniş ölçekte gözlemleme fırsatını sunmaktadır (McCormick vd., 2017:391).

Sosyal Medya Madenciliği sosyal medya verilerinin elde edilmesinden, verilerden anlamlı sonuçların çıkarımına ve değerlendirilmesine kadar olan geniş bir süreci kapsamaktadır. Süreçte izlenen adımlar Şekil 1'de yer almaktadır.

Şekil 1: Sosyal Medya Madenciliği Süreci



Kaynak: Bonzanini, 2016:17.

**Yetkilendirme:** Sosyal Medya Madenciliği sürecinin ilk adımı olan yetkilendirme, genellikle OAuth (Open Authorization) adı verilen protokol tarafından gerçekleştirilir (Bonzanini, 2016:17). Açık bir protokol olan OAuth, Web, mobil ve masaüstü uygulamalarına basit ve standart bir yöntemle güvenli yetkilendirme imkânı sunmaktadır (Ravindran ve Garg, 2015:8). Bu protokol, kullanıcının şifre gibi hassas bilgilerini paylaşmadan, izin verilen uygulamanın kullanıcının hesap verisine erişmesine olanak tanıyan bir araçtır (Russell, 2013:13).

**Veri Toplama:** Veri toplama adımı, yalnızca erişim izni verilen veri indirilebildiği için, Sosyal Medya Madenciliği sürecinin ilk adımı olan yetkilendirme aşamasına bağlıdır (Bonzanini, 2016:19). Birçok sosyal medya platformu verilerine erişim API (Uygulama Programlama Arayüzü) ile sağlanmaktadır (Ravindran ve Garg, 2015:8). Bir arayüz olan API, bir uygulamanın diğer uygulama ile iletişim halinde olmasını sağlar. Bu sayede API, örneğin bir şirketin Twitter'ı kullanmak için kendi yazılım istemcilerini geliştirmesine ya da Facebook'u diğer sosyal medya servisleriyle entegre etmesine izin vererek temel sosyal medya hizmetlerinin geliştirilmesini sağlar. Araştırmacılar için ise API, sosyal medya platformlarından veri toplamak için kullanılmaktadır. API tabanlı araştırmalarda verilerin otomatik bir şekilde toplanabilmesi, araştırmacılar için büyük kolaylık sağlamaktadır. Bu sayede, araştırmalarda sosyal medya verilerini toplamak, sosyal medya kullanıcılarının dijital izlerini elde etmek, anket ve mülakat gibi tekniklerden çok daha fazlasını verebilmektedir (Lomborg ve Bechmann, 2014:256).

**Veri Ön İşleme:** Veri ön işleme sürecinde, veri temizleme (data cleaning), veri birleştirme (data integration), veri indirgeme (data reduction) ve veri dönüştürme (data transformations) teknikleri uygulanmaktadır. Bu tekniklerden veri setinin ihtiyacı ve araştırmanın amacı doğrultusunda faydalanılabilmektedir. Veri temizleme genel olarak, veri setindeki eksik verileri doldurmak, aykırı

verileri saptamak, gürültüyü temizlemek ve veri setindeki tutarsızlıkları düzeltmek gibi adımları içermektedir. Veri birleştirme aşamasında ise farklı kaynaklardan gelen veriler, uygun bir veri tabanında birleştirilir. Veri indirgeme, fazla olan değişkenlerin çıkarılması, birleştirilmesi gibi farklı yöntemlerle veri büyüklüğünün azaltılması amaçlanır. Çeşitli sebeplerden dolayı, gerek duyulduğunda normalleştirme gibi veri dönüşüm teknikleri de ön işleme aşamasında kullanılabilir (Han vd., 2011:83).

Veri ön işleme aşaması bir önceki adımda elde edilen ve depolanan ham verilerin analiz aşamasına hazırlanmasını temsil etmektedir. Veri seti içerisindeki gürültü verileri tanımlamak ve temizlemek oldukça zor bir süreçtir. Bu süreç ne kadar başarılı gerçekleştirilirse, analiz sonuçlarındaki doğruluk oranı o denli yüksek olur (Ravindran ve Garg, 2015:14).

**Modelleme ve Analiz:** Ön işlemeden geçirilen verilere, araştırmanın amacı ve kurulan model doğrultusunda çeşitli analizler gerçekleştirilmektedir. Bu aşamada öncelikle bir araştırma modeli kurulur. Kurulan modelde hangi analiz yöntemlerinin uygulanacağına karar verildikten sonra, gerekli adımlar izlenerek sosyal medya verileri analiz edilir. Sosyal medya madenciliğinde sıklıkla kullanılan analizlere örnek olarak metin madenciliği ve sosyal ağ analizi verilebilir (Bonzanini, 2016:20).

Sosyal medya madenciliğinde uygulanan analizler, kullanıcıların belli bir ürün ve hizmet ile ilgili duygularını, düşüncelerini ve şikâyetlerini analiz eden *içerik analizi temelli*, içerikler arası, kullanıcılar arası ilişkileri ve her ikisi (kullanıcılar ve içerikler) arasındaki ilişkileri analiz eden *yapı analizi temelli* ve içeriğin/kullanıcıların dinamik analizine dayalı *değişkenlik analizi temelli* çalışmalar olmak üzere üç başlık altında toplanabilir (Ishikawa, 2015:68-69).

**Değerlendirme:** Sonuç sunumu olarak adlandırılan bu aşama sonuçların değerlendirmesi üzerine kuruludur. Son aşama olan görselleştirme, eğilimler, anomaliler, ilişkiler gibi modelleme ve analiz aşamasının sonucunda elde edilen bulguların, soyut bir boyuttan somut bir hale dönüştürülerek kolayca anlaşılmasını sağlamaktadır (Ravindran ve Garg, 2015:19). Bir veri analizi sürecinin sonuçlarını sunmanın ve aktarmanın etkili bir yolu olan görselleştirme aşaması, farklı disiplinlerce benimsenmiştir (Bonzanini, 2016:16). Kutu ve serpilme diyagramları, kelime bulutları, karar ağaçları ve sosyal ağ analizi sürecinde kullanılan araçlar gibi çeşitli yöntemlerle veri görselleştirme yapılabilmektedir (Ravindran ve Garg, 2015:14).

### 3. Literatür Araştırması

Bu bölümde Sosyal Medya Madenciliği ile ilgili literatürde yer alan çalışmalar kısaca ele alınmıştır.

Sosyal medya verilerinin finans verileriyle arasındaki ilişkiyi inceleyen çalışmaların olduğu görülmektedir. af Rosenborg vd. (2017), hazır giyim sektöründe faaliyet gösteren bir markanın sosyal medya katılımı ve finansal performansı arasındaki ilişkiyi incelemiştir. Çalışmada markanın stil ikonları ve tasarımcılarının yer aldığı yedi kampanyasının Facebook verileri ele alınmış, trend ve mevsimsellik içeren üstel düzgünleştirme yöntemi ile markanın satışları tahmin edilerek, stil ikon ve tasarımcı kampanyalarının etkisi ölçülmüştür. Lassen vd. (2014) ise, bir firma ile ilgili tweetleri firmanın satışlarını tahmin etmede kullanmak üzere ele almıştır. Doğrusal ve çoklu regresyon tahmin modeli geliştirilmiş, Morgan Stanley ve IDC'nin tahmin modellerine yakın bir ortalama hata ile üç aylık olmak üzere satışlar başarılı bir şekilde tahmin edilmiştir. Bollen vd. (2011), kullanıcıların duygu durumları ile ilgili olan tweetlerin Dow Jones Borsası Endüstri Endeksi'nde (DJIA) meydana gelen değişimlerde tahmin amaçlı kullanılıp kullanılmayacağını incelemiştir. %87,6'lık bir başarı ile DJIA'da gerçekleşen günlük iniş ve çıkışlar tahmin edilebilmiştir. Zimbra vd. (2009), şirketler hakkında kamuoyu görüşlerini değerlendirmek için otomatik konu çıkarımı ve duygu analizi yöntemlerini uygulayan bütünleşmiş bir çerçevenin, şirketlerin hisse senedi değerleri ile güçlü bir ilişki içinde olduğunu keşfetmişlerdir.

Sosyal medya verilerinden gişe tahmini çalışmalarında faydalandığı görülmektedir. Ding vd. (2017), Facebook'un "beğen" özelliğinin filmlerin gişe tarihinden önce etkin bir sosyal pazarlama aracı olarak kullanılabileceğini önermişlerdir. Beğenilerin gişe hasılatına etkisini incelemek amacı ile kesitsel regresyon modeli kullanılmıştır. Gişeden önceki en son beğenilerin gişe hasılatında daha fazla etkili olduğu sonucuna ulaşılmıştır. Oh vd. (2017), Facebook platformu ile birlikte Twitter ve Youtube platformlarını da dahil ederek tüketici katılım davranışları perspektifinden, sosyal medyanın filmlerin gişe performansına etkisinin olup olmadığını araştırmışlardır. Bulgular sosyal medya platformlarının filmlerin gelecek performansını olumlu yönde etkilediğini ortaya koymuştur. Asur ve Huberman (2010), Twitter verilerini kullanarak gişe gelirlerinin tahmini için bir regresyon modeli geliştirmişlerdir. Bulgular Hollywood Stock Exchange verileriyle karşılaştırılmıştır ve daha başarılı sonuçlar elde edildiği görülmüştür.

Sosyal medya paylaşımlarının duygu durumlarını saptamaya yönelik çalışmalar bulunmaktadır. Kuflik vd. (2017), ulaşım ile ilgili tweetlerin manuel olarak kategori ve duygu durumlarını etiketleyerek, sınıflandırıcıları eğitmek ve test etmek için kullanmışlardır. Destek Vektör Makine ve Naive Bayes algoritması ikili n-gram tekniğiyle birlikte başarılı sonuçlar vermiştir. Çoban vd. (2015), eğitim verisi için duygu durumlarının etiketlenmesi aşamasında, his simgeleri olarak ifade ettikleri emoji göz önünde bulundurularak tweetleri pozitif ve negatif olarak etiketlenmiştir. Thelwall vd. (2011) Twitter'daki en çok konuşulan etkinliklerin aynı konuyla ilgili diğer tweetlere kıyasla pozitif ve negatif duyguların ifade gücünden dolayı oluşup oluşmadığı araştırma sorusu çerçevesinde otuz etkinlik ile ilgili elde edilen tweetleri analiz etmiştir. Tweetler hazır bir program tarafından olumlu-olumsuz olarak kodlanmış, popüler olayların olumsuz duygu artışlarıyla alakalı olduğuna ve olumlu olarak nitelendirilen olayların, olumsuz yorumları doğurduğuna ulaşılmıştır.

Sosyal medya paylaşımlarının farklı amaçlar doğrultusunda kümelendiği çalışmalar mevcuttur. Pournarakis vd. (2017), bir araç paylaşım servisi ile ilgili tweetleri kümelere ayırmış ve her kümedeki duygu durumlarını belirleyerek, son aşamada tüketicilerin algılarına ulaşmışlardır. Chae (2015) tedarik zinciri ile ilgili çeşitli anahtar kelimeler aratılarak toplanan tweetleri, tanımlayıcı istatistikler, içerik analizi ve ağ analizinden oluşan bir analitik çerçeve önererek analiz etmiştir. İçerik analizi bölümünde en popüler kelimeler ve hashtagler frekans sayılarına göre belirlenmiş ve kümelenecek, tweet duygu durumları hazır bir program kullanarak gerçekleştirilmiştir. Xiang vd. (2017), ABD'de bir bölgede bulunan otellerle ilgili üç sosyal medya platformundan (TripAdvisor, Expedia ve Yelp) veri toplamışlardır. Yorumlardaki konular analiz edilmiş, duygu durumları ise belirlenen pozitif ve negatif sözcüklerden oluşturulan bir sözlük ile sınıflandırılmıştır. Bulgularda platform temelli birçok farklılık gözlemlenmiştir.

Sosyal medya platformlarında bir konu ya da kavram ile ilgili neler konuşulduğunun saptanmasını amaçlayan çalışmalar mevcuttur. Bian vd. (2016), bireylerin "Nesnelerin İnterneti (IoT)" kavramıyla ilgili düşüncelerini araştırmak adına Twitter'dan elde edilen verileri analiz etmişlerdir. Analiz sonuçlarına göre bireylerin IoT ile ilgili algısının çoğunlukla olumlu olduğu, büyük veri, akıllı teknoloji ve sektör liderleri gibi genellikle işletme ve teknolojiye odaklandığı belirlenmiştir. Sert vd. (2014), Twitter'da akademik zam konusunda kişilerin görüşlerinin değişimini ve gelişimini sosyal ağ analizi ile analiz etmişlerdir. Bulgular ağ değerlerinin, yaşanan gelişmeler ışığında azalıp arttığını göstermiştir. Twitter aracılığıyla gündemdeki gelişmeler ile ilgili halkın tepkisini ölçmenin mümkün olabileceği gözlemlenmiştir.

Sosyal medyada doğan ya da sosyal medyaya taşınan kurumsal krizlerle ilgili yapılan araştırmalar mevcuttur. Ye ve Ki (2017), bir kurumun yaşamış olduğu krizi, Facebook üzerindeki kriz yanıtlarına ve kamunun krize tepki stratejileri konusundaki algılamalarına katkıda bulunan faktörlere içerik analizi uygulayarak araştırmışlardır. Kurumun bilgi verici stratejilerine Facebook kullanıcıları tarafından olumlu yorumların yapıldığı saptanmıştır. Guidry vd. (2015), Forbes 2013 verilerine göre en büyük on fast-food zinciri hakkında Instagram'da üretilen ve yayılan kriz bilgileri ile ilgili içeriklerini analiz etmişlerdir. Bulgular negatif yorumların müşteriler dışında şirket çalışanları tarafından da gönderildiğini ve müşterilerin servis sorunları, çalışanların ise çalışma

ortamı koşulları ile ilgili yorumlar gönderdiklerini göstermiştir. Ki ve Nekmat (2014) Fortune 500 şirketlerinin ne ölçüde Facebook'u kriz iletişim aracı olarak kullandığını, şirketlerin ve bireylerin göstermiş olduğu etkileşim seviyesini, kurum tarafından uygulanan kriz müdahale stratejilerini, bunlara verilen yanıtlardaki kamu duygu durumlarını, kriz türleri ile müdahale stratejilerinin ne ölçüde ilişkili olduklarını incelemişlerdir. van der Meer ve Verhoeven (2013), bir tesiste gerçekleşen patlama krizini ele alınarak kamunun krizler hakkındaki algısının medya içeriklerinden önce neye dayandığını, medya içerikleri ile uygunluk gösterip göstermediğini ve kriz çerçevesini kişiselleştirip kişiselleştirmedeğini araştırmışlardır.

Kurumsal krizlerden farklı olarak sosyal medya verilerinden hareketle doğal afetlerle ilgili araştırmalara odaklanan çalışmalar bulunmaktadır. Odlum ve Yoon (2015) Twitter'ı, Ebola salgınının izlenmesinde gerçek zamanlı bir bilgi yayılım gözlemeleme, erken tespit, kamusal bilgileri ve tutumları saptama yöntemi olarak ele almıştır. Kamuoyu bilgilerini ve tutumları incelemek amacı ile doğal dil işleme, bilgi yayılımını gözlemek için ise jeolojik görselleştirme ile zaman serisi analizi kullanılmıştır. Chatfield ve Brajawidagda (2012), Twitter platformunun bir afet uyarı sistemi olarak çalışıp çalışmadığını incelemek adına Endonezya'da gerçekleşen bir doğal afeti ele almışlardır. Twitter'ın uyarı sistemi olarak başarılı bir şekilde kullanabileceği bilgisinin yanı sıra, jeolojik afet uyarılarından sorumlu yerel ajans takipçilerinin incelenmesi ile birlikte, bu kullanıcıların çok kısa bir sürede mesajların yayımında etkin rol oynadığı, ağda kanaat önderliği rolü üstlendiği sonucuna ulaşılmıştır.

Literatürdeki bu uygulama konuları dışında da sosyal medya verilerinden faydalanılan çalışmalar bulunmaktadır. Poell ve Borra (2012), sosyal medyanın alternatif gazetecilik alanında kullanımını incelemiştir. Bir zirve ile ilgili içerikler Twitter, Youtube ve Flickr'dan elde edilerek analiz edilmiştir. Bulgulara göre alternatif gazetecilik adına en umut verici platform Twitter olarak değerlendirilmiştir. Kumar vd. (2011), göç hareketlerini incelemek ve göç kalıplarının neler olduğunu öğrenmek amacıyla birden fazla sosyal medya platformunda faaliyet gösteren kullanıcıları ele alarak bir çalışma gerçekleştirmişlerdir. Kullanıcıların aynı kategorideki rakip siteler arasında dolaşmaya eğilimli oldukları gözlemlenmiştir.

#### 4. Uygulama

##### 4.1. Araştırmanın Amacı

Bu çalışmada farklı sektörlerde faaliyet gösteren firmaların sosyal medya istatistiklerinin ve sosyal medyayı kullanım amaçlarının belirlenmesi hedeflenmiştir. Firmaların sosyal medya istatistikleri ele alınarak bu istatistiklerin firmalara göre uyum gösterip göstermediği ve firmaların sosyal medyada ne tür içerikler paylaştıkları, hangi içerik türlerinin daha fazla etkileşime yol açtığı saptanmaya çalışılmıştır.

##### 4.2. Araştırmanın Kapsamı ve Sınırları

Sosyal medya platformları arasında firma ve kullanıcıların katılım düzeylerinin yüksek olması ve kullanıcıların görüşlerinin paylaşılmasını teşvik eden yapısı ile mikroblog siteleri arasında en çok kullanılan sosyal medya platformu Twitter, bu uygulamanın örnekleme olarak belirlenmiştir.

Tablo 1: Araştırmanın Örneklemini Oluşturan Sektör ve Firma Bilgileri

Sektör Adı ve Açıklaması	Firma Adı	Twitter Takipçi Sayısı (01.01.2018)	Twitter'a Katılma Tarihi
<b>Kozmetik:</b> Güzellik ve kişisel bakım ürünleri perakende zincirlerinden oluşan sektördür.	Firma 1	374.654	2010
	Firma 2	128.784	2010

<b>Pazaryeri:</b> Çok kategorili ürün yelpazesine sahip olan ve yalnızca online alışverişin gerçekleştirilebildiği e-ticaret firmalarından oluşan sektördür.	Firma 3	430.313	2009
	Firma 4	308.045	2012
<b>Elektronik:</b> Elektronik ve teknoloji ürünlerin satıldığı, çok kanallı (online ve mağaza) perakende zincirlerinden oluşan sektördür.	Firma 5	547.594	2013
	Firma 6	601.926	2008

Firmaların Twitter istatistikleri ve paylaşım kategorileri ile ilgili analizleri gerçekleştirmek için örneklem dâhilinde seçilen firmaların 2017 yılı Twitter paylaşımları ele alınmıştır. Tablo 1’de örnekleme yer alan sektörler ve firmalar ile ilgili gerekli bilgiler verilmiştir.

### 4.3. Araştırmanın Yöntemi ve Sosyal Medya Madenciliği Uygulaması

Bu çalışmada yöntem olarak Sosyal Medya Madenciliği adımları izlenerek, farklı sektörlerde faaliyet gösteren firmaların Twitter kullanımları incelenmiştir. Firmaların 2017 yılına ait Twitter paylaşımlarını elde edebilmek amacı ile önce yetkilendirme işlemi gerçekleştirilmiş, veri çekim aşaması gerçekleştirilmiş ve veri ön işleme aşamasında analize hazır hale getirilmiştir. Firmaların Twitter kullanımları ile ilgili istatistikler ele alınarak rakip firmalar ve/veya sektörler arası uyumlar ise Uygunluk Analizi ile saptanmıştır. Firmaların Twitter’da ne tür içerikler paylaştıkları bir dizi Metin Madenciliği ön işleme metotlarından faydalanılarak gerçekleştirilen Tweet Analizi ve hangi içerik türü sınıfının en yüksek etkileşimi elde ettiği Pareto İlkesi kapsamında gerçekleştirilen Rt Analizi ile elde edilmiştir.

#### 4.3.1. Yetkilendirme Aşaması

Sosyal medya madenciliği sürecinin ilk adımı olan yetkilendirme aşamasını gerçekleştirmek için öncelikle bir Twitter uygulaması oluşturularak veri çekimi için gerekli olan “access token” değeri elde edilmiştir. Daha sonra veri çekiminde kullanılan RapidMiner Studio 8.1 aracında bu değer ile yetkilendirme işlemi gerçekleştirilmiştir.

#### 4.3.2. Veri Toplama Aşaması

Bu çalışmanın örneklemini oluşturan firmaların 2017 yılı Twitter paylaşımları Twitter API desteği ile RapidMiner Studio 8.1 aracı kullanılarak elde edilmiştir. Twitter API desteği ile bir kullanıcıya ait en yeni gönderilerden eskilere doğru olmak üzere, 3.200 adet tweet çekilebilmektedir. Bu şekilde firmaların paylaşmış oldukları son 3.200 adet tweet elde edilmiştir.

#### 4.3.3. Veri Ön İşleme Aşaması

RapidMiner aracında elde edilen veri setindeki 1 Ocak 2017 ile 31 Aralık 2017 tarihleri dışındaki veriler ve analizde kullanılmayacak olan değişkenler çıkarılarak veri seti indirgenmiştir. Nihai veri setinde yer alan değişkenlerin isimleri ve açıklamaları Tablo 2’de verilmiştir.

Tablo 2: Veri Setindeki Değişken Bilgileri

Değişken	Tanım
Zaman Dilimi	Firma paylaşımlarının yapıldığı saat bilgisi göz önünde bulundurularak bir zaman dilimi sınıflandırılması yapılmıştır. Belirlenen üç zaman dilimi kapsamında ( $\geq 05.00$ - $< 12.00$ =öğleden önce, $\geq 12.00$ - $< 17.00$ =öğleden sonra, $\geq 17.00$ - $< 05.00$ =akşam) zaman dilimi değişkeni elde edilmiştir.
Gün	Firma paylaşımlarının yapıldığı gün bilgisinin verildiği değişkendir.

Tweet	Tweet olarak adlandırılan firma paylaşımlarını içeren değişkendir.
Rt Sayısı	Her bir firma paylaşımının almış olduğu retweet miktarını gösteren değişkendir.
Özgün Tweet	Firmaların paylaşmış oldukları özgün içerikler, "rt" ve "mention" içermeyen tweetler olarak belirlenmiştir. Bu değişken, eğer tweet özgün ise 1, değil ise 0 olarak hesaplanmıştır.
Rt Tweet	Firmaların paylaşmış oldukları "retweet" içeren tweetlerdir. Bu değişken, eğer tweet retweet edilmiş bir paylaşımından oluşuyor ise 1, değil ise 0 olarak hesaplanmıştır.
Mention Tweet	Firmaların paylaşmış oldukları "mention" içeren (@kullanıcıadı) tweetlerdir. Bu değişken, eğer tweet mention içeriyor ise 1, değil ise 0 olarak hesaplanmıştır.

#### 4.3.4. Modelleme ve Analiz Aşaması

Uygulamada gerçekleştirilen analizlerin yer aldığı açıklama, Şekil 2'de ayrıntılı olarak verilmiştir.

Şekil 2: Uygulamada İzlenen Adımlar



## 5. Bulgular

### 5.1. Uygunluk Analizi

Firmaların paylaşmış oldukları tweet türleri (özgün, mention ve rt tweet) ile firmalar arasındaki uyum incelenmiştir. IBM SPSS Statistics (24.0) programında gerçekleştirilen Uygunluk Analizi bulgularına yer verilmiştir.

Tablo 3'te görüldüğü üzere, inertia'nın 0'dan farklı olup olmadığının test edildiği Ki-Kare istatistiği değeri ( $\alpha=0.00<0.05$ ) anlamlı çıkmıştır. Buna göre özgün, mention ve rt tweetlerinin dağılımı firmalara göre farklılaşmaktadır. Boyut sayısına göre açıklanan inertia değerleri incelendiğinde, ilk boyutun toplam değişkenliğin %95,9'unu, ikinci boyutun ise %4,1'ini açıkladığı görülmektedir. Böylece iki boyutun toplamı, değişkenliğin tümünü açıklamıştır.

Tablo 3: Firmalar Tarafından Paylaşılan Tweetlerin Türlerinin Firmalar ile Uyumundaki Boyut Sayısı ve Açıklanan Inertia Değerleri

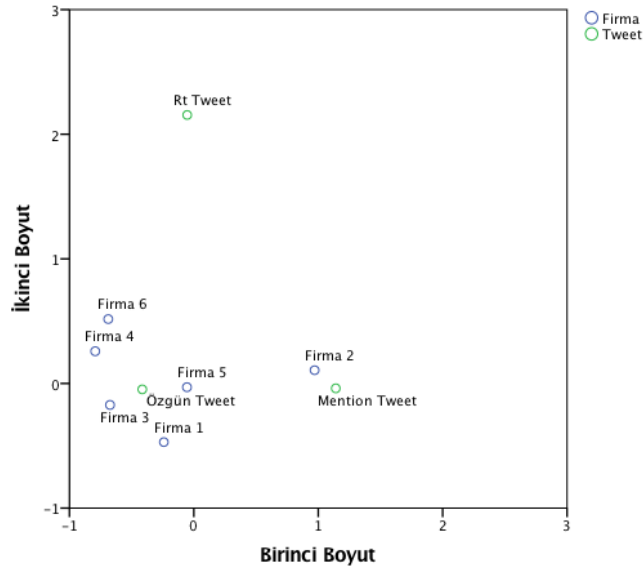
Boyut	Tekil Değer	Inertia	Ki-Kare	$\alpha$	Açıklanan Inertia	
					Açıklanan	Kümülatif
1	0,464	0,215			0,959	0,959
2	0,097	0,009			0,041	1
<b>Toplam</b>		0,225	993,148	0	1	1

Şekil 3'deki uzaklıklar incelendiğinde, tweet kategorileri arasında yer alan mention tweetin bulunduğu konumun uzaklık olarak (Ki-Kare uzaklığı) en yakın olduğu firmanın Firma 2 olduğu görülmektedir. Rt tweet kategorisi ise tüm firmalara uzak konumlanması sebebi ile bu kategorinin firmalar tarafından çok fazla tercih edilmeyen bir tweet türü olduğu söylenebilir. Özgün tweetin ise özellikle Firma 5, Firma 3 ve Firma 1'e oldukça yakın konumlandığı görülmektedir. Tüm bu



değerlendirmeler ele alındığında kozmetik sektöründe faaliyet gösteren Firma 2 dışında, diğer tüm firmaların Twitter paylaşımlarının çoğunluğunun özgün tweetlerden oluştuğu söylenebilir.

Şekil 3: Firmalar Tarafından Paylaşılan Tweetlerin Türlerinin ve Firmaların İki Boyutlu Düzlemde Birlikte Gösterimi



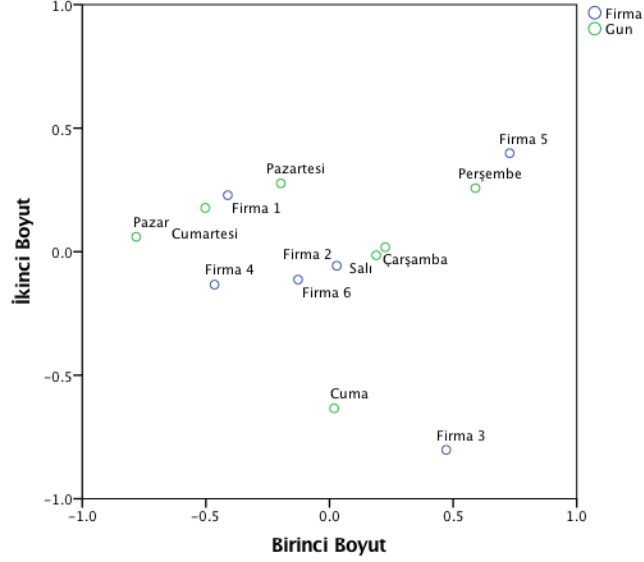
Bir diğer uyum, firma tweetlerinin paylaşım günleri ile firmalar arasında incelenmiştir. Tablo 4 incelendiğinde ilk boyutun toplam değişkenliğinin %66,1'ini, ikinci boyutun ise %21,7'sini açıkladığı görülmektedir. İki boyut, toplam değişkenliğin %87,8'ini açıklamaktadır.

Tablo 4: Firmalar Tarafından Paylaşılan Tweetlerin Paylaşım Günlerinin Firmalar ile Uyumundaki Boyut Sayısı ve Açıklanan Inertia Değerleri

Boyut	Tekil Değer	Inertia	Ki-Kare	$\alpha$	Açıklanan Inertia	
					Açıklanan	Kümülatif
1	0,165	0,027			0,661	0,661
2	0,095	0,009			0,217	0,878
3	0,059	0,004			0,086	0,964
4	0,034	0,001			0,028	0,993
5	0,017	0			0,007	1
<b>Toplam</b>		0,041	181,736	0	1	1

Şekil 4 incelendiğinde, kozmetik sektörü firmaları Firma 1'in Cumartesi, Firma 2'nin Salı gününe, elektronik sektöründe faaliyet gösteren Firma 5'in Perşembe, rakibi Firma 6'nın Salı gününe ve pazaryeri firmaları Firma 3'ün Cuma, Firma 4'ün Cumartesi gününe en yakın konumda olduğu görülmektedir. Tüm bu değerlendirmeler ele alındığında aynı sektörde yer alan rakip firmaların birbirlerinden farklı günlerde Twitter'da paylaşım yaptığı söylenebilir.

Şekil 4: Firma Tweetlerinin Paylaşım Günleri ve Firmaların İki Boyutlu Düzlemde Birlikte Gösterimi



Son olarak firma tweetlerinin paylaşıldığı zaman dilimleri ile firmalar arasındaki uyum incelenmiştir.

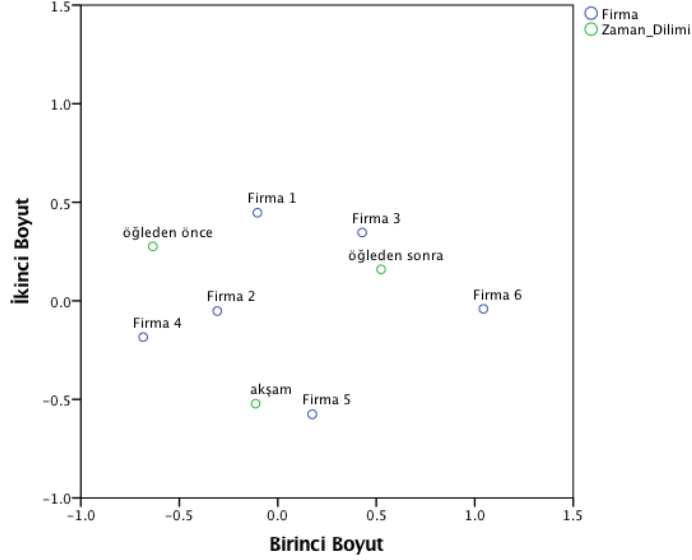
Tablo 5’de görüldüğü üzere ilk boyut %82,2’sini, ikinci boyut %17,8’ini olmak üzere iki boyutun tüm değişkenliği açıkladığı görülmüştür.

Tablo 5: Firmalar Tarafından Paylaşılan Tweetlerin Zaman Dilimlerinin Firmalar ile Uyumundaki Boyut Sayısı ve Açıklanan Inertia Değerleri

Boyut	Tekil Değer	Inertia	Ki-Kare	$\alpha$	Açıklanan Inertia	
					Açıklanan	Kümülatif
1	0,238	0,057			0,822	0,822
2	0,111	0,012			0,178	1
<b>Toplam</b>		0,069	304,676	0	1	1

Şekil 5’de görüldüğü üzere, pazaryeri sektöründe faaliyet gösteren Firma 3 ve elektronik sektöründe faaliyet gösteren Firma 6 öğleden sonra zaman dilimine yakın uzaklıkta konumlanırken, Firma 3’ün rakibi Firma 4 öğleden önce, Firma 6’nın rakibi Firma 5 ise akşam zaman dilimine yakın konumlanmıştır. Kozmetik sektörü firmaları Firma 1 ve Firma 2 ise öğleden önce zaman dilimine yakın uzaklıkta konumlanmıştır. Tüm bu değerlendirmeler ele alındığında kozmetik sektörü dışında aynı sektörde yer alan rakip firmaların birbirlerinden farklı zaman dilimlerinde Twitter’da paylaşım yaptığı söylenebilir.

Şekil 5: Firmalar Tarafından Paylaşılan Tweetlerin Zaman Dilimlerinin ve Firmaların İki Boyutlu Düzlemde Birlikte Gösterimi



## 5.2. Tweet Analizi

Firmaların 2017 yılı Twitter paylaşımları bir dizi Metin Madenciliği ön işleme metotları uygulanarak kümelendirilmiştir. RapidMiner programında gerçekleştirilen ön işleme aşamalarının açıklamaları aşağıda verilmiştir:

- URL Temizleme: Tweet içerisinde yer alan linkler metin içerisinden temizlenmiştir.
- Mention (@kullanıcıadı) Temizleme: Tweet içerisinde yer alan @kullanıcıadı ifadeleri temizlenmiştir.
- Küçük Karaktere Dönüştürme: Tweet içerisindeki tüm kelimeler küçük karaktere dönüştürülmüştür.
- Parçalara Ayırma (Tokenize): Bu operatör parçalara (kelimelere) ayırma işlemi için kullanılmıştır. Metin içerisinde yer alan her kelime boşluk ve noktalama işaretleri tanınarak ayrılmıştır.
- Durak Sözcüklerini Temizleme (Filter Stopwords): Tek başına anlam ifade etmeyen çeşitli edatlardan (diye, göre vb.), sayılardan (beş, yirmi vb.), zamirlerden (bizim, size vb.) ve bazı isim ve fiillerden (olduğunda, şey vb.) oluşan sözlük ile metin içerisinde yer alan durak sözcükleri temizlenmiştir.
- Parçaları Değiştirme (Replace Tokens): Parçaları değiştirme operatörü ile metin içerisinde yer alan aynı anlama gelen kelimelerin aynı şekilde ele alınması için çeşitli dönüştürmeler yapılmıştır. Örneğin mağaza ve mağazalar kelimelerini birlikte ele almak adına mağazalar kelimesi mağaza kelimesi ile değiştirilmiştir.
- Kelime Filtreleme (Filter Tokens-by Length): Karakter temelinde sınıflandırma yapılarak belirlenen sınırlar dışında bulunan kelimeler analizden çıkarılmıştır. Bu operatör kullanılarak en az 3 karakter bulunma şartı konularak bunun altındaki karakterler elenmiştir.
- Tamlamalar Yöntemi (Generate N-Grams-Terms): Bazı kelimeler tek başına değil beraberinde gelen kelime ile tamlama oluşturularak anlam kazandığı için tamlamalar

yöntemine başvurulmuştur. İkili tamlamalar yöntemine örnek olarak veri setinden “mağaza açmak”, “sosyal medya”, “teşekkür etmek” ve “indirim beklemek” verilebilir.

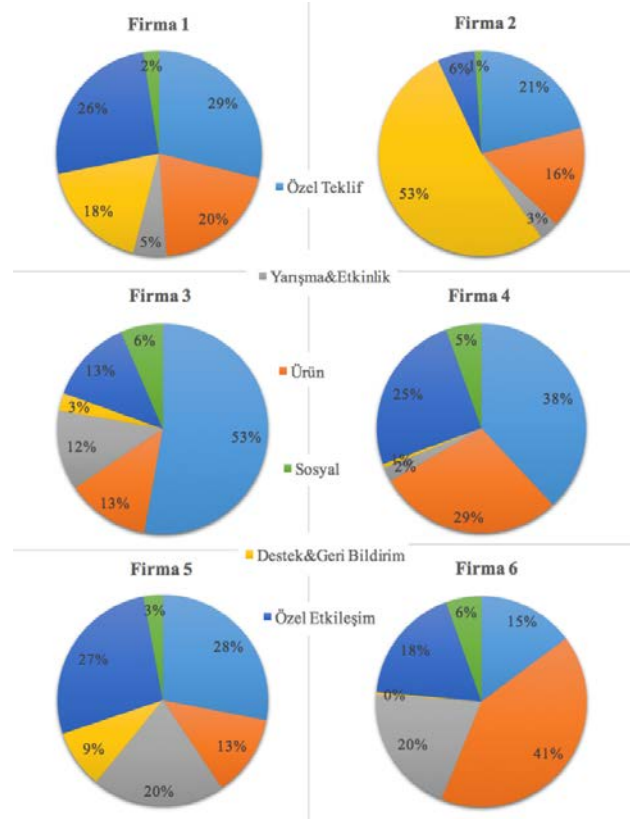
RapidMiner programında gerçekleştirilen metin ön işleme aşamaları sonucunda kelime listesine ve kelimelerin frekanslarına ulaşılmıştır. Kümelere ayırmak adına liste gözden geçirilmiş ve küme başlıkları belirlenmiştir. Anahtar kelimeler aratılarak küme isimleri veri setinde kodlanmıştır. Kümeler belirlenirken özele inmeden geniş başlıklar seçilerek her firma için ortak olabilecek küme isimleri belirlenmiştir. Küme açıklamaları ve örnek firma tweetleri Tablo 6’da verilmiştir.

Tablo 6: **Firmalar Tarafından Paylaşılan Tweetlerin Kümeleri ve Açıklamaları**

Küme	Açıklama	Örnek Tweet
<b>Özel Teklif</b>	Firmaların ürün ve hizmetleri ile ilgili paylaştıkları indirim, fırsat, kampanya ve hediye tweetlerini içeren kümedir.	En güzel hafta sonu indirimi 27-29 Ocak tarihleri arasında tüm Firma 1’lerde sizleri bekliyor!
<b>Ürün</b>	İndirim gibi özel teklifler dışında kalan, içerisinde sadece ürünlerle ilgili bilgileri (yeni çıkan ürün, katalog bilgisi, ürünle ilgili çekilmiş video içerikleri vb.) barındıran ürün temelinde yapılan paylaşımlardan oluşan kümedir.	Hindistan cevizinin yenileyen ve güzelleştiren mucizesini barındıran tüm #ürünadı cilt bakım ürünleri mağazalarımızda!
<b>Yarışma &amp; Etkinlik</b>	Takipçilere yönelik çeşitli şekillerle düzenlenen (soru, katılım, başvuru, çekiliş) yarışmalar ile ilgili mesajları, söyleşi ve davet gibi etkinlik tweetlerini ve bu yarışmaların sonuçlarının duyurulduğu tweetleri içeren kümedir.	Hafta sonunun en güzel yarışması Instagram hesabımızda başladı!
<b>Destek &amp; Geri Bildirim</b>	Kullanıcılar tarafından bildirilen sorunlara, sorulara ve bildirimlere yönelik direkt olarak kullanıcılara yanıt verildiği tweetleri içeren kümedir.	@kullanıcıadı Merhabalar, şikâyet kaydınızı oluşturabilmemiz için iletişim bilgilerinizi bizimle D.M yolu ile paylaşmanızı rica ediyoruz.
<b>Özel Etkileşim</b>	Firmaların belirlediği özel içeriklerle (hashtag vb.) takipçilerini katılıma teşvik ettiği tweetleri, kullanıcılara özel olarak paylaşılan teşekkür mesajlarını, ürünler dışında paylaşılan önerileri, etkileşimi arttırmaya yönelik hazırlanan özel videoları ve firma ile ilgili özel bilgilendirme tweetlerini (mağaza açılış, kazanılan ödüller vb.) içeren kümedir.	500. mağaza mutluluğunu hep beraber yaşadık. Sizlerden gelen tüm eğlenceli yorumlar için çok teşekkürler. 💖 Siz de #Firma1veBen hashtagi ile en yaratıcı tweet ve mesajlarınızı atarak bir sonraki videomuzda yer alabilirsiniz. 😊
<b>Sosyal</b>	Bayram, maç, özel gün gibi kutlama, dilek, baş sağlığı tweetlerini içeren kümedir.	Ortaköy’de yaşanan terör saldırısında hayatını kaybedenlere Allah’tan rahmet, yakınlarına sabır ve tüm yaralılara acil şifalar diliyoruz.

Firmaların Twitter paylaşımlarının kategorilerinin yüzdelik dağılımı Şekil 6’daki gibi firmalara göre görselleştirilmiştir. Bulgulara göre pazaryeri sektöründe faaliyet gösteren rakip firmalar en çok paylaşımı aynı kümede yaparken, diğer sektörlerde farklılıklar saptanmıştır.

Şekil 6: Firmalar Tarafından Paylaşılan Tweetlerin Kümelerinin Firmalara Göre Frekans Dağılım Yüzdeleri



Aynı sektörde (kozmetik) yer almalarına rağmen, Firma 1 ve Firma 2 tweet paylaşımlarının küme yüzdeleri ele alındığında büyük farklılık gözlemlenmiştir. Firma 2 tweetlerinin yarısından fazlasını (%53) Destek & Geri Bildirim kategorisinde paylaşırken, Firma 1 ise bu alanda yalnızca %18 oranında paylaşım yapmıştır. Bu farkın sebebi, Firma 2'nin online alışveriş kanalı ile de hizmet vermesi olduğuna tweet içerikleri ayrıntılı olarak incelendiğinde ulaşılmıştır. Online alışveriş ile ilgili kullanıcıların sordukları sorulara, kullanıcı şikâyet ve sorunlarına, Firma 2 Twitter aracılığı ile yanıt vermiştir. Bu iki firma arasındaki bir diğer büyük farklılık ise Özel Etkileşim kategorisi yüzdelerinde gözlemlenmektedir. Firma 1 paylaşımlarının %26'sını bu kümede yaparken Firma 2 bu kümeye ait yalnızca %6'lık içerik üretmiştir.

Pazaryeri sektöründe faaliyet gösteren Firma 3 ve Firma 4 tweet paylaşımlarının küme yüzdeleri ele alındığında, en çok paylaşım yapılan kategori başlığının (Özel Teklif) aynı olduğu görülmektedir. Özel Teklif kategorisinde Firma 3 %53, Firma 4 ise %38 oranında paylaşım yapmıştır. Ürün ve Özel Etkileşim kategorileri, firma paylaşım yüzdeleri farklılık göstermekle birlikte, iki rakip firma tarafından en çok tercih edilen diğer iki kategoridir.

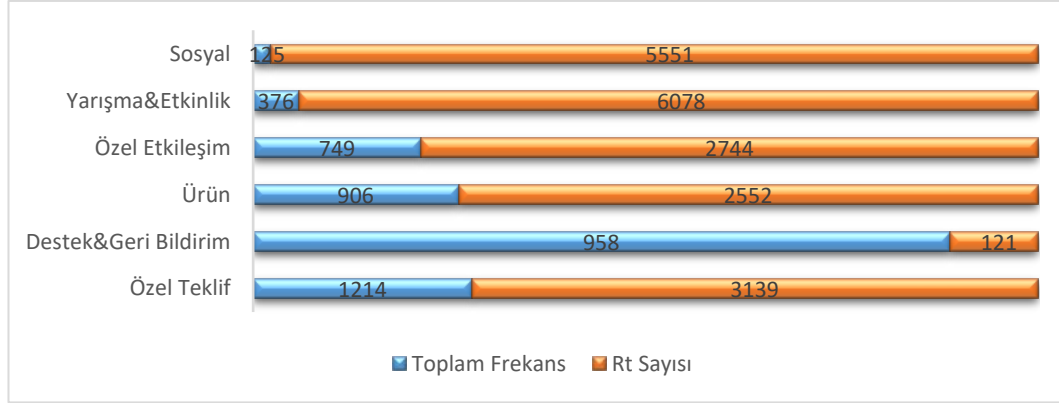
Elektronik sektörü firmaları incelendiğinde, Firma 5'in en çok Özel Teklif (%28), Firma 6'nın ise en çok Ürün (%41) kategorisinde paylaşım yaptığı görülmektedir. Firma 5'in en çok paylaşım yaptığı ikinci kategori (%27) olan Özel Etkileşim, Firma 6 tarafından daha az bir oranla (%18) paylaşılmıştır. Her iki firma da aynı oranlarda Yarışma & Etkinlik kategorisinde paylaşım yapmışlardır.

Pazaryeri ve elektronik sektöründe faaliyet gösteren firmaların Destek & Geri Bildirim kategorisine ait tweet yüzdelerinin oldukça düşük olduğu gözlemlenmiştir. Bu firmaların kurumsal hesapları dışında destek hesaplarının da mevcut olduğu ve bu hesaplarından 2017 yılı boyunca kullanıcılarına yanıt tweetler paylaştıkları görülmüştür.

### 5.3. Rt Analizi

Firmaların tweet paylaşımları birlikte değerlendirilerek, Tweet Analizi bölümünde belirlenen altı küme kapsamında kümelerin içerdiği tweet frekansları ve bu tweetlerin almış oldukları rt sayıları Şekil 7'deki gibi görselleştirilmiştir.

Şekil 7: Firmalar Tarafından Paylaşılan Tweetlerin Kümelere Göre Frekans Dağılımları ve Almış Oldukları Rt Sayıları



Şekil 7 incelendiğinde, Yarışma & Etkinlik ve Sosyal kategorilerinin en düşük tweet frekansına sahip olmasına rağmen (Sosyal; 125, Yarışma & Etkinlik; 376) en çok rt edilme frekansına sahip olduğu (Sosyal; 5.551, Yarışma & Etkinlik; 6.078) görülmektedir. Bu bulgu değerlendirildiğinde, firmaların ürün ve indirim tweetlerinden ziyade, firmalar tarafından kullanıcılara özel olarak paylaşılan yarışma ve sosyal mesaj içeren tweetlerin kullanıcılar tarafından başka kullanıcılarla paylaşılabilir görüldüğü söylenebilir.

Kategorilere göre alınan rt sayılarının dışında, firmaların bir yıl boyunca yapmış oldukları paylaşımlar sonucunda elde ettikleri toplam rt sayılarının büyük çoğunluğunun, firmaların tweetlerinin yüzde kaçından oluştuğu incelenmiştir. Bu incelemeyi gerçekleştirmek adına veri seti Pareto İlkesi kapsamında ele alınmıştır. Pareto İlkesi ilk defa, adını aldığı iktisatçı ve sosyolog Vilfredo Pareto tarafından ülke servetinin yaklaşık yüzde 80'inin popülasyonun yaklaşık yüzde 20'sinde yoğunlaştığını bulması sonucu gözlemlenmiş, daha sonra başka araştırmacılar tarafından yapılan incelemeler sonucu bu ilkenin birçok alanda sirayet ettiği keşfedilmiş ve bir kural haline gelmiştir (Sanders, 1987:37).

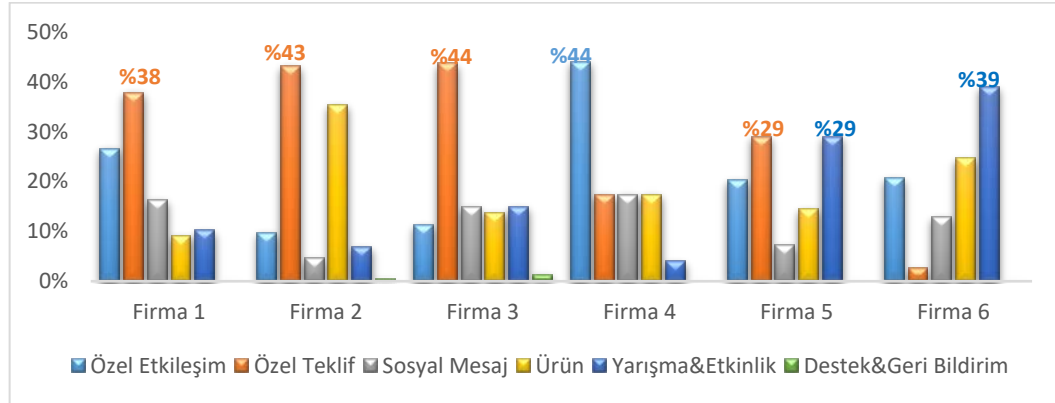
Rt içerikleri temizlenen veri setinde yer alan tweet sayıları, toplamda alınan rt miktarı ve toplam rt miktarının %80'inin tweetlerin yüzde kaçından oluştuğunun oranları Tablo 7'de verilmiştir.

Tablo 7: Firmalar Tarafından Paylaşılan Tweetlerin Almış Olduğu Rt Frekanslarının Pareto İlkesi Kapsamında Dağılım Yüzdeleri

	Firma 1	Firma 2	Firma 3	Firma 4	Firma 5	Firma 6
<b>Toplam Tweet Sayısı</b>	997	1301	364	475	667	524
<b>Toplam Rt Sayısı</b>	7352	1180	1013	710	3206	6724
<b>Rt- %80</b>	%10	%17	%22	%16	%10	%15

Bulgulara göre dağılımın Pareto ilkesine uyum sağladığı görülmektedir. Toplamda elde edilen rt miktarının büyük çoğunluğunun veri setinde yer alan tweetlerin azınlık bölümünden elde edildiğine ulaşılmıştır. Örneğin; Firma 3'ün 2017 yılı boyunca paylaşmış olduğu 364 tweetin almış olduğu toplam rt sayısının (1.013) %80'i yalnızca %22'lik orana denk gelen tweetlerden elde edilmiştir. Bu bulgular neticesinde 80/20 Kuralı ele alınarak rt miktarının büyük çoğunluğunun elde edildiği tweet kategorileri incelenmiştir. Toplam etkileşimin yüzde 80'inin elde edildiği tweet kategorileri Şekil 8'deki gibi görselleştirilmiştir.

Şekil 8: Firmalar Tarafından Paylaşılan Tweetlerin Almış Olduğu Rt Frekanslarının Pareto İlkesi Kapsamında Firma Tweet Kümelerine Göre Dağılım Yüzdeleri



Bulgulara göre kozmetik sektöründe faaliyet gösteren iki firmanın en çok etkileşim aldığı tweetlerin kategorilerinin büyük çoğunluğunun Özel Teklif kategorisinden (Firma 1; %38, Firma 2; %43) oluştuğu söylenebilir. Pazaryeri sektöründe faaliyet gösteren Firma 3 en çok Özel Teklif (%44) ve rakibi Firma 4 en çok Özel Etkileşim (%44) içeren tweetler paylaşarak, elektronik sektöründe yer alan Firma 5 ve Firma 6 ise en çok Yarışma & Etkinlik kategorisinde paylaşımda bulunarak toplam etkileşimlerinin büyük çoğunluğunu elde ettikleri görülmektedir. Firma 5 aynı oranla Özel Teklif (%29) kategorisinde de paylaşımlarda bulunarak yüksek etkileşim elde etmiştir.

## 6. Sonuç

Bu çalışmada farklı sektörlerde faaliyet gösteren firmaların, Twitter istatistiklerinin firmalara/sektörlere göre uyumu, Twitter'da ne tür içerikler paylaştıkları ve hangi içeriklerin yüksek etkileşimi sağladığı Sosyal Medya Madenciliği adımları izlenerek saptanmaya çalışılmıştır. Firmaların Twitter istatistikleri ile firmalar arasındaki uyum Uygunluk Analizi, firmaların Twitter'da ne tür içerikler paylaştıkları Tweet Analizi ve hangi tür içeriklerin en çok etkileşimi gerçekleştirdiği Rt Analizi uygulanarak elde edilmiştir ve analiz bulguları bu bölümde tartışılmıştır.

Çalışmanın Uygunluk Analizi bölümünde, firmaların paylaştıkları tweet türü için ortak bir bulgu elde edilmiştir. Kozmetik sektöründe faaliyet gösteren bir firma dışındaki tüm firmaların mention ve rt tweetlerden daha fazla oranda kendi ürettikleri içerikleri paylaştıkları saptanmıştır. Farklı sektörlerde faaliyet göstermelerine karşın firmaların benzer bir şekilde özgün içerikler üretmek Twitter'da varlıklarını sürdürdüklerine ulaşılmıştır. Bu bulgu değerlendirildiğinde firmaların Twitter'da genellikle özgün içerik paylaşmaya odaklı olmaları önerilebilir.

Tweet Analizi bölümünde ise, firmaların Twitter paylaşımlarının içerikleri altı küme başlığı altında sınıflandırılmıştır. Firmaların Twitter paylaşımlarının kategorileri; indirim/fırsat mesajlarından oluşan Özel Teklif, kullanıcılara özel düzenlenen yarışmalar/etkinlikler ile ilgili mesajlardan oluşan Yarışma & Etkinlik, ürünle ilgili bilgilendirmelerden oluşan Ürün, bayram, özel gün vb. gibi kutlama ve dilek mesajlarından oluşan Sosyal, kullanıcılara yönelik bilgilendirme/yanıt mesajlarından oluşan Destek & Geri Bildirim ve firmalara özgü hashtag vb. içerikler paylaşarak kullanıcıların içerik üretmeye teşvik edildiği Özel Etkileşim olarak belirlenmiştir. Tweet Analizi bölümünde elde edilen bulgulara göre pazaryeri sektöründeki rakip firmaların her ikisinin de Özel Teklif kategorisinde daha fazla paylaşım yaptıkları saptanmıştır. Bu sektörde faaliyet gösteren ya da sektöre yeni girecek olan firmaların Özel Teklif sınıfındaki tweetlere ağırlık vererek diğer firmalar ile sosyal medya ortamında rekabet edebileceği düşünülmektedir. Ayrıca sosyal medyada faaliyet gösteren tüm firmalar, çalışmanın sonucunda ulaşılan altı kategori ile ilgili almış oldukları rt sayısı, yorum sayısı, beğeni sayısı vb. gibi çeşitli etkileşimleri inceleyebilir ve bulgular neticesinde aksiyon olarak bir sosyal medya içerik planı oluşturup paylaşım kategorilerinin oranlarını belirleyebilir.

Çalışmadaki bir diğer önemli bulgu ise sosyal medya verilerinin disiplinler arası bir yaklaşım izlenerek Pareto İlkesi ile açıklanabilir oluşunun saptanmasıdır. Rakip firmaların en çok etkileşim aldığı tweetlerin büyük çoğunluğunun kozmetik sektörü için Özel Teklif, elektronik sektörü için ise Yarışma & Etkinlik kategorisinden olduğu sonucuna ulaşılmıştır. Bu bulgular değerlendirildiğinde firmaların tweet etkileşimlerinin büyük çoğunluğunun, paylaştıkları tweetlerin azınlıktaki bir bölümünden elde edildiği görülmüştür. Bu sebeple, firmaların etkileşimine büyük katkı sağlayan bu içerik kategorilerine yoğunlaşarak tüketicilere yönelik içerik paylaşımlarını arttırması tavsiye edilebilir. Ayrıca sektöre yeni girecek olan firmaların da, etkileşimlerinin büyük çoğunluğunu sağlayan kategorilere daha fazla odaklanıp içerikler paylaşarak, var olan rakip firmalar ile sosyal medyada rekabet edebilir düzeye ulaşabileceği düşünülmektedir.

Gelecek çalışmalarda sosyal medya platformu, değişken sayısı, veri setindeki sektör ve firma sayısı arttırılarak, veri setine firmaların finansal raporlarına ait değişkenler eklenerek, firmaların sosyal medya katılımları ile finansal başarısı arasındaki ilişki incelenebilir. Firmaların yapmış oldukları sosyal medya paylaşımları ile birlikte kullanıcıların ürettiği içerikler (user-generated content) de analize dâhil edilerek tüketici boyutunda etkileşim ölçülebilir. Tüketicilerin firmaların paylaştıkları kategorilere göre ürettikleri içeriklerin duygu durumları (olumlu, nötr, olumsuz) saptanabilir ve bulgulara göre firmalara yönelik olumsuz yorum paylaşılan kategorilerde/konularda, tüketicilerin ürettiği bu içeriklere eş zamanlı olarak, firmalar tarafından olası sosyal medya kriz aksiyon planı hazırlanabilir.

#### Kaynakça

- af Rosenborg, D. C. K., Buhl-Andersen, I., Nilsson, L. B., Rebild, M. P., Mukkamala, R. R., Hussain A. ve Vatrappu, R. (2017). Buzz vs. Sales: Big Social Data Analytics of Style Icon Campaigns and Fashion Designer Collaborations on H&M's Facebook Page. *50th Hawaii International Conference on System Sciences*, 1861-1870.
- Asur, S. ve Huberman, B. A. (2010). Predicting the Future with Social Media. *IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology*, 492-499.
- Bian, J., Yoshigoe, K., Hicks, A., Yuan, J., He, Z., Xie, M., Guo, Y., Prosperi, M., Salloum R. ve Modave, F. (2016). Mining Twitter to Assess the Public Perception of the "Internet of Things". *PLoS ONE*, 11(7), 1-14.
- Bollen, J., Mao, H. ve Zeng, X. (2011). Twitter Mood Predicts the Stock Market. *Journal of Computational Science*, 2(1), 1-8.
- Bonzanini, M. (2016). *Mastering Social Media Mining with Python*. Birmingham: Packt Publishing.
- Chae, B. K. (2015). Insights from Hashtag# Supplychain and Twitter Analytics: Considering Twitter and Twitter Data for Supply Chain Practice and Research. *International Journal of Production Economics*, 165, 247-259.
- Chatfield, A. ve Brajawidagda, U. (2012). Twitter Tsunami Early Warning Network: A Social Network Analysis of Twitter Information Flows. *23rd Australasian Conference on Information Systems*, 1-10.
- Çoban, Ö., Özyer, B. ve Özyer, G. T. (2015). Sentiment Analysis for Turkish Twitter Feeds. *23rd Signal Processing and Communications Applications Conference*, 2388-2391.
- Ding, C., Cheng, H. K., Duan, Y. ve Jin, Y. (2017). The Power of the "Like" Button: The Impact of Social Media on Box Office. *Decision Support Systems*, 94, 77-84.
- Guidry, J. D., Messner, M., Jin, Y. ve Medina-Messner, V. (2015). From# Mcdonaldsfail to# Dominossucks: An Analysis of Instagram Images about the 10 Largest Fast Food Companies. *Corporate Communications: An International Journal*, 20(3), 344-359.



- Han, J., Pei, J. ve Kamber, M. (2011). *Data Mining: Concepts and Techniques (3rd ed.)*. Burlington, MA: Morgan Kaufmann Publishers.
- Ishikawa, H. (2015). *Social Big Data Mining*. Boca Raton: Taylor & Francis Group, CRC Press.
- Kaplan, A. M. ve Haenlein, M. (2010). Users of The World, Unite! The Challenges and Opportunities of Social Media. *Business Horizons*, 53 (1), 59-68.
- Ki, E. J. ve Nekmat, E. (2014). Situational Crisis Communication and Interactivity: Usage and Effectiveness of Facebook for Crisis Management by Fortune 500 Companies. *Computers in Human Behavior*, 35, 140-147.
- Kuflik, T., Minkov, E., Nocera, S., Grant-Muller, S., Gal-Tzur, A. ve Shoor, I. (2017). Automating a Framework to Extract and Analyse Transport Related Social Media Content: The Potential and the Challenges. *Transportation Research Part C: Emerging Technologies*, 77, 275-291.
- Kumar, S., Zafarani, R. ve Liu, H. (2011). Understanding User Migration Patterns in Social Media. *25th AAAI Conference on Artificial Intelligence*, 1204-1209.
- Lassen, N. B., Madsen, R. ve Vatrapu, R. (2014). Predicting iPhone Sales from iPhone Tweets. *18th International Enterprise Distributed Object Computing Conference*, 81-90.
- Lomborg, S. ve Bechmann, A. (2014). Using APIs for Data Collection on Social Media. *The Information Society*, 30(4), 256-265.
- McCormick, T. H., Lee, H., Cesare, N., Shojaie, A. ve Spiro, E. S. (2017). Using Twitter for Demographic and Social Science Research: Tools for Data Collection and Processing. *Sociological Methods & Research*, 46(3), 390-421.
- Odlum, M. ve Yoon, S. (2015). What Can We Learn about the Ebola Outbreak from Tweets?. *American Journal of Infection Control*, 43(6), 563-571.
- Oh, C., Roumani, Y., Nwankpa, J. K. ve Hu, H. F. (2017). Beyond Likes and Tweets: Consumer Engagement Behavior and Movie Box Office in Social Media. *Information & Management*, 54(1), 25-37.
- Poell, T. ve Borra, E. (2012). Twitter, YouTube, and Flickr as Platforms of Alternative Journalism: The Social Media Account of the 2010 Toronto G20 Protests. *Journalism*, 13(6), 695-713.
- Pournarakis, D. E., Sotiropoulos, D. N. ve Giaglis, G. M. (2017). A Computational Model for Mining Consumer Perceptions in Social Media. *Decision Support Systems*, 93, 98-110.
- Ravindran, S. K. ve Garg, V. (2015). *Mastering Social Media Mining with R*. Birmingham: Packt Publishing.
- Russell, M. A. (2013). *Mining the Social Web: Data Mining Facebook, Twitter, LinkedIn, Google+, GitHub, and More (2nd ed.)*. Newton, MA: O'Reilly Media, Inc.
- Sanders, R. (1987). The Pareto Principle: Its Use and Abuse. *Journal of Services Marketing*, 1(2), 37-40.
- Sert, A. G. F., Tüzüntürk, S. ve Gürsakal, N. (2014). NodeXL ile Sosyal Ağ Analizi: #akademikzam Örneği. 15. *Uluslararası Ekonometri, Yöneylem Araştırmaları ve İstatistik Sempozyumu Bildiriler Kitabı*, 464-482.
- Thelwall, M., Buckley, K. ve Paltoglou, G. (2011). Sentiment in Twitter Events. *Journal of the American Society for Information Science and Technology*, 62(2), 406-418.
- van der Meer, T. G. ve Verhoeven, P. (2013). Public Framing Organizational Crisis Situations: Social Media versus News Media. *Public Relations Review*, 39(3), 229-231.

- Weinberg, T. (2009). *The New Community Rules: Marketing on the Social Web*. Newton, MA: O'Reilly Media, Inc.
- Xiang, Z., Du, Q., Ma, Y. ve Fan, W. (2017). A Comparative Analysis of Major Online Review Platforms: Implications for Social Media Analytics in Hospitality and Tourism. *Tourism Management*, 58, 51-65.
- Ye, L. ve Ki, E. J. (2017). Organizational Crisis Communication on Facebook: A Study of BP's Deepwater Horizon Oil Spill. *Corporate Communications: An International Journal*, 22(1), 80-92.
- Zafarani, R., Abbasi, M. A. ve Liu, H. (2014). *Social Media Mining: An Introduction*. New York: Cambridge University Press.
- Zimbra, D., Fu, T. ve Li, X. (2009). Assessing Public Opinions Through Web 2.0: A Case Study on Wal-Mart. *30th International Conference on Information Systems*, 1-10.

---

**ANALYZING TWITTER DATA OF FIRMS WITH SOCIAL MEDIA MINING**

---

**Extended Abstract**

**Aim:** Social media has become an indispensable tool especially for firms due to the advantages it has achieved in reaching the target audiences. In this study, the Twitter data of the rival firms operating in different sectors were collected to determine whether Twitter data of the firms has a significant correspondence with respect to the firms. This study also aims to cluster Twitter feeds of the firms to understand what firms share on Twitter and to find out which cluster has the maximum interaction.

**Method(s):** In this study, Twitter data shared by competitors operating in the cosmetics, electronics and marketplace sectors during 2017 were analyzed by following the process of Social Media Mining. In the authentication step, Twitter app was created to get tokens needed for accessing to Twitter data, and then the data was gathered. In data pre-processing step, data reduction was done to prepare the data for further analysis. Correspondence Analysis was applied to find if there was a significant correspondence between Twitter data and the firms. Twitter feeds of the firms were clustered by using a number of Text Mining pre-processing methods. Which cluster received more interaction was analyzed with the help of the Pareto Principle.

**Findings:** According to the results of Correspondence Analysis, it was found that all firms except the firm which operates in the cosmetics sector, were positioned close to the original tweet share. Considering the correspondence between the days of the tweet shares and the firms, the result showed that each firm shared on Twitter on different days. Considering the timeframes of the tweet shares, it was found that only the firms operating in the cosmetics sector made more shares in the same time period. As a result of a series of Text Mining pre-processing methods, Twitter feeds of the firms were clustered under the six categories. Categories of firms' feeds were defined as "Special Offer" consisting of discount messages, "Competition & Event" consisting of messages about events / special events organized for consumers, "Product" consisting of messages about firms' products, "Social" consisting of celebration and wish messages, "Support & Feedback" consisting of messages about support and reply messages and "Special Interaction" consisting of messages in which consumers are encouraged to produce content. The percentages of each firms' shares were calculated according to the clusters and it was determined that the most shared tweet cluster is the same among the competitors in the marketplace sector and different among the firms in the cosmetics and electronics sectors. When the tweets shared by all the firms in the data set are evaluated together, it is an interesting finding that the most retweeted feeds belong to the "Social" and "Competition & Event" categories, although less content is shared. The fact that the total number of retweet obtained by the firms can be explained by the Pareto Principle is another important finding. It was determined that the majority of the total number of retweet firms acquired during 2017 was obtained from a minority of the tweets shared by the firms. According to the findings, it was concluded that both firms operating in cosmetics sector, the majority of the tweets that received the most interaction were from the "Special Offer" cluster. It was observed that the majority of the interaction received by the marketplace sector firms differ according to the firms. Finally, it was determined that the shares of the electronic sector companies in the "Competition & Event" cluster constitute the majority of the interaction received by the firms.

**Conclusion:** It can be stated from the results of the Correspondence Analysis that mostly the Twitter variables of the rival firms differ from each other. According to the results obtained from the Text Mining pre-processing methods, firms use Twitter to inform their customers about discounts, events, products, to support them, to encourage them to produce content and to give social messages. In the scope of this research it has been noticed that the total number of retweet obtained by the firms can be explained by the Pareto Principle.

