

‘Tamam’ ve ‘Devam’ Etiketleriyle Atılan Tweetler ile Cumhuriyet ve Millet İttifakı Grupları Arasındaki İlişkinin İncelenmesi

Araştırma Makalesi/Research Article

 Ertuğrul GÜMÜŞSU^{1*},  Naci MURAT²

¹Akıllı Sistemler Mühendisliği, Ondokuz Mayıs Üniversitesi, Samsun, Türkiye

²Endüstri Mühendisliği, Ondokuz Mayıs Üniversitesi, Samsun, Türkiye

16210309@stu.omu.edu.tr, nacimurat@omu.edu.tr

(Geliş/Received:13.08.2018; Kabul/Accepted:07.08.2019)

DOI: 10.17671/gazibtd.453272

Özet— Teknoloji kullanımının her geçen gün artmasıyla birlikte sosyal medya ve internet kullanımı hızla yaygınlaşmaktadır. Özellikle sosyal medya kullananların sayısının artmasıyla insanlar duygu ve düşüncelerini, gündemle ilgili görüş ve fikirlerini sosyal medya araçlarını kullanarak paylaşmaktadır. En çok kullanılan sosyal medya araçlarından biri olan Twitter ve Facebook gibi sosyal ağ uygulamalarında paylaşılan içerikler yapısal olmayan veri türleridir. Bunun gibi yapısal olmayan metin içerikli verilerin yapısal hale dönüştürülmesinde ve analiz edilmesinde metin madenciliği yöntemleri kullanılmaktadır. Bu çalışmada, 24 Haziran Cumhurbaşkanlık ve Milletvekili Seçimlerinde ‘Tamam’ ve ‘Devam’ etiketleriyle atılan tweetler metin madenciliği yöntemleri kullanılarak analiz edilmiştir. Metin ön işlem aşamasında kelimelerin köklerini bulmak için Zemberek kök bulma yöntemi ve ağırlıklandırma yöntemi olarak ise terim frekansı (TF) ağırlıklandırma yöntemi uygulanmıştır. Oluşturulan sınıflar arasındaki ilişkiyi ölçmek amacıyla korelasyon analizi yapılmıştır. Kruskal Wallis Testi ile frekans terim ağırlıklarının sınıflara göre farklılığı test edilmiştir. Elde edilen özelliklerin terim frekansı (TF) değerlerine göre kelime bulutları oluşturulmuştur.

Anahtar Kelimeler— metin madenciliği, kelime bulutu, siyaset analiz, sosyal medya

To Examine of the Relationship Between Shared Tweets of the ‘Tamam’ and ‘Devam’ Tags and the People's Alliance and National Alliance

Abstract— As the use of technology increases day by day, the use of social media and internet is rapidly spreading. Especially with the increasing number of people who use social media, people share their feelings and thoughts, opinions and ideas about the agenda by using social media tools. Contents that is shared in social network applications like Twitter and Facebook, one of the most used social media tools, are non-structural data types. Text mining methods are used to transform non-structured textual data such as these into structured and analyze data. In this study, the tweets shared with the tags 'Tamam' and 'Devam' in the 24 June Presidential and Parliamentary Elections were analyzed using text mining methods. The Zemberek stemming method were used to find words root and the term frequency (TF) weighting method were used as the weighting method at text pre-processing stage. Correlation analysis was performed to measure the relationship between the created classes. With Kruskal Wallis Test, the frequency term weights were tested according to the classes. Word clouds was formed according to the term frequency (TF) values of the obtained features.

Keywords— text mining, word cloud, political analysis, social media analysis

1. GİRİŞ (INTRODUCTION)

İnternet kullanımının ve teknolojinin gelişmesiyle sosyal medya kullanımı her geçen gün artmaktadır. Ayrıca gelişen sosyal medya platformları sayesinde insanlar duygu ve düşüncelerini, görüşlerini, günlük aktivitelerini çeşitli sosyal medya araçlarını kullanarak paylaşmaktadır. Özellikle Twitter, Facebook gibi sosyal ağ uygulamaları dünya üzerinde milyonlarca kişi tarafından kullanılmaktadır. Bu platformlar sayesinde insanlar kolaylıkla örgütlenebilmekte, bir fikir üzerinde gruplaşabilmektedir. Bu duruma 27 Mayıs 2013 tarihinde ülkemizde yaşanan Gezi Parkı olayları örnek verilebilir. Çok az bir grup insanın çadır eylemleri sonrasında başlayan eylemler 5-6 gün içerisinde ülkenin birçok bölgesine yayılmıştır. Bu durumun ülkenin birçok bölgesine yayılmasının en büyük sebeplerinden biri insanların Twitter üzerinde örgütlenmesi olmuştur. Twitter üzerinde insanlar birbirlerini eyleme destek vermeye ve sokağa çıkmaya çağırıştır. Türkiye o dönemlerde günlük ortalamasının 2,5 katı tweet atıldığı ve #direngeziparkı etiketiyle yaklaşık olarak 2 milyondan fazla tweet atıldığı belirtilmektedir. Bu sayılar ve durumlar dikkate alındığında insanların olaylar ve fikirler etrafında örgütlenmesi üzerinde sosyal medya platformlarının büyük etkisi olduğu gözükmektedir.

Twitter gibi sosyal ağ sitelerinde paylaşılan bu tür içerikler yapısal olmayan nitelikteki veri türleridir. Bu verilerin analiz edilebilmesi için yapısal formlara dönüştürülmesi gerekir. Metin içerikli yapısal olmayan bu verileri analiz etmek için metin madenciliği teknikleri kullanılır. Metin madenciliği yöntemleri kullanılarak yapısal olmayan metin verileri yapısal hale dönüştürülür ve analiz edilir. Yapısal hale dönüştürülmeden yapılan analizlerde hatalı sonuçlar elde edilir.

Metin madenciliği çok büyük metin verileri içerisinde bilgiye ulaşma işlemi olarak tanımlanabilir. Metin madenciliğinde, doğal dil işleme ve veri madenciliği birlikte kullanılmaktadır [1]. Doğal dil işleme, bilgisayar bilimi ve dil biliminin bir alt dalı olarak doğal dildeki metinlerin veya seslerin bilgisayar ile tanımlanması üzerine çalışmalar yapan bilim dalıdır. Veri madenciliği, büyük miktardaki veri yığınları içerisinde değerli bilgilere ulaşarak gelecekle ilgili tahminlerde bulunabilecek bağıntı ve kuralların bilgisayar programları kullanarak aranmasıdır [2].

Metin madenciliği birçok alanda uygulanmaktadır. Metin madenciliği güvenlik uygulamalarında [3], istihbarat alanında [4], internet-yazılım alanında [1], akademik uygulamalarda ve bilimsel araştırmalarda [5], finans alanında [6], pazarlama alanında [7], müşteri ilişkileri yönetiminde [8], hile denetiminde [9], biyoloji alanında [10], basın yayın sektöründe [11] yazar tanıma sistemlerinde [12], soru cevap sistemlerinde [13], sağlık ve medikal araştırmalarda kullanılmaktadır.

Bu çalışmada, 24 Haziran Cumhurbaşkanı ve Milletvekili Seçimi döneminde Sayın Cumhurbaşkanı Recep Tayyip Erdoğan'ın "Şayet bir gün milletimiz 'Tamam' derse ancak o zaman biz kenara çekiliriz" sözlerinden sonra Twitter'da 'Tamam' ve 'Devam' etiketiyle atılan tweetler metin madenciliği yöntemleri kullanılarak analiz edilmiş ve anahtar kelimeler baz alınarak kelime bulutları oluşturulmuştur. Daha sonra Millet ve Cumhuriyet İttifakı gruplarının miting konuşmaları incelenmiştir. 'Tamam' ve 'Devam' tweetleriyle ilişkisini analiz etmek için korelasyon analizi yapılmıştır. Kruskal Wallis Testi ile terim frekans ağırlıklarının sınıflara göre farklı olup olmadığına bakılmıştır. Çalışma sonunda Millet ve Cumhuriyet İttifakının, Recep Tayyip Erdoğan'ın, Devlet Bahçeli'nin, Muharrem İnce'nin, Meral Akşener'in ve Temel Karamollaoğlu'nun konuşmalarını özetlemek amacıyla kelime bulutları oluşturulmuştur. Bu çalışma ile kullanıcıların neden tamam ve dedikleri tespit edilmeye çalışılmıştır. Ayrıca Millet ve Cumhuriyet İttifakı grupları ve parti liderlerinin söylemleri ile kullanıcıların tweetleri arasındaki ilişki ortaya konmaya çalışılmıştır.

2. METİN MADENCİLİĞİ (TEXT MINING)

Günümüzde dijital ortamların ve internetin kullanılmasının artmasıyla birlikte dijital ortamlarda kayıt altına alınan veriler artmaktadır [14]. Genellikle bu ortamlardaki çeşitli metin belgeleri ve dokümanları, e-posta mesajları, sosyal ağ uygulamalarında paylaşılan içerikler, forum ve blog verileri yapısal olmayan nitelikteki veri türleridir. Yapısal olmayan bu verilerin analiz edilmesi ve bilgi çıkarımında bulunmak için bu verilerin yapısal bir formata dönüştürülmesi gerekir. Metin madenciliği teknikleri ile yapısal olmayan bu veriler, yapısal hale dönüştürülür ve analiz edilebilir.

Metin veri madenciliği (Text Data Mining) ve metin veritabanlarından bilgi keşfi (Knowledge Discovery from Textual Databases) olarak adlandırılan metin madenciliği, 1960'lı yıllarda bilgi erişim sistemleri ve doğal dilin işlenmesine yönelik yapılan çalışmalar sonucu ortaya çıkmıştır. Metinlere erişim, metinlerden bilgi çıkarımı, metin kategorizasyonu, metinlerin yapısal hale dönüştürülmesine yönelik çalışmalar, 1990'lı yıllarda oldukça yaygınlaşmıştır [15]. 1990'lı yıllardan sonra internet ve kişisel bilgisayar kullanımının yaygın hale gelmesiyle bilgiye erişim hızı artmaya başlamıştır. Bu durum beraberinden artan bilgi patlamasını meydana getirmiş fakat istenilen bilgiye ulaşmayı zorlaştırmıştır. Bu nedenle istenilen bilgiye erişim sorununu çözmek için yeni yöntemler geliştirilmeye başlanmıştır. Özellikle 2000'li yıllardan sonra metin madenciliğine olan ilgi artmaya başlamıştır [15].

Metin madenciliği, büyük boyuttaki belgelerin analizi ve metin tabanlı verilerden bilgi çıkarmak amacıyla metinlerin analiz edilmesi süreci olarak tanımlanabilir [16, 17]. Başka bir tanımda metin madenciliği, önceden bilinmeyen ve önemli olan bilgilerin çıkarılması amacıyla çok sayıda dokümanı analiz eden bir teknoloji olarak tanımlanmaktadır [14]. Metin madenciliği işlemlerinde

doğal dil işleme (Natural Language Processing-NLP) ve veri madenciliği bir arada kullanılmaktadır [1, 18]. Veri madenciliği (Data Mining), bilgisayar programlarını kullanarak büyük miktardaki veri yığınları içerisinde faydalı bilgilerin elde edilmesi amacıyla verinin işlenip ve modellenmesi işlemi olarak tanımlanır [19]. Doğal dil işleme insanların kullandığı dilin bilgisayara tanıtılması ve dilin doğru bir şekilde analiz edilmesi amacıyla ortaya çıkmış bir çalışma alanıdır [20].

2.1. Metin Ön İşlem Aşaması (Text Preprocessing Stage)

Yapısal olmayan nitelikteki metin dokümanları doğal dil kullanılarak oluşturulur. Metin dokümanları üzerinde veri madenciliği tekniklerinin uygulanabilmesi için yapısal olmayan veriyi yapısal formata dönüştüren veri ön işlem adımına ihtiyaç duyulur [21]. Veri ön işlem aşamaları, veriyi uygun formata getirmek ve veri seti üzerinde çeşitli analizler yapabilmek için yapılan işlemler dizisidir [14, 22]. Ön işlem; veri üzerinde bulunan problemleri çözmek, verinin doğal yapısını öğrenerek daha anlamlı ve kaliteli analiz yapabilmek ve veriden daha anlamlı bilgi üretebilmek gibi nedenlerden dolayı yapılır [1]. Veri ön işlem aşamasında metni kelimelere ayırma, kelimelerin anlamsal değerlerini bulma (isim, sıfat, zamir, fiil, zarf), kelimeleri köklerine ayırma, kelime köklerinin belirleme, dokümanları fazlalıklardan arındırma, gereksiz bilgileri çıkarma, yazım kurallarına uygunluğu tespit etme, ekleri ve genel kelimeleri çıkarma, önek ve sonekleri ayırma, kelime anlamlarını belirleme, kelimelerin doküman içerisinde sayısal olarak ifade edilmesi gibi kelimelerle ilgili işlemler yapılır [15]. Böylelikle hatalı sonuçlara neden olacak faktörler ön işlem aşaması ile ortadan kaldırılmış olur. Her veri seti için ön işlem aşaması aynı formatta değildir, verinin formatına göre farklılık gösterebilir [23]. Örneğin web ortamından alınan verilerden HTML etiketi temizlenirken, Twitter üzerinden alınan bir veriden “@” karakteriyle başlayan kullanıcı adları veya “#” karakteriyle başlayan hashtagler temizlenebilir.

2.1.1. Dönüştürme (Transformation)

İnternette elde edilen dokümanlar genellikle HTML (Hyper Text Markup Language) ve XML (Extensible Markup Language) gibi çeşitli etiketler ile tutulduğundan dolayı bunları düz metin hale dönüştürmek için dokümanlar HTML ve XML etiketlerinden temizlenir [1]. Sosyal ağlar üzerinden elde edilen verilerde veri içerisinde yer alan mentionlar (@), hashtagler (#) ve linkler silinir. Ayrıca metin içerisinde yer alan metin olmayan çift tırnak (“), nokta (.), ünlem (!), virgül (,), tek tırnak (‘) gibi noktalama işaretleri, *, +, /, # gibi çeşitli karakterler, his simgeleri ve rakamlar metin içerisinde çıkarılır [20, 24].

2.1.2. Tarama ve İşaretleme (Tokenization)

Bir doküman veya metin içerisindeki kelimelerin tamamının elde edilmesi için işaretleme işlemi yapılır [25]. Yapısal olmayan veriler ilk olarak cümlelere, daha sonra da kelimelere parçalanır [26]. Metin belgeleri bölüm,

paragraf, cümle, kelime ve hecelere ayrılabilir. En çok kullanılan işlem kelimelere ayırma işlemidir [22]. Ayrıca bütün kelimeler küçük harfe dönüştürülür.

2.1.3. Kök Bulma (Stemming)

Kök bulma işlemi, parçalanmış kelimelerin doküman içerisindeki sıklıklarına bakılırken aynı kökten gelen kelimelerin farklı ek almış hallerinin aynı kelime olarak algılanması için yapılır [1]. Bu sayede kelimeler köke indirgenerek aynı kelime kökleri ile temsil edilmesi sağlanmış olur [22]. Ayrıca hem özellik sayısı azalmış olacak hem de aynı kelime köküne sahip kelimelerin frekansı daha doğru bir şekilde hesaplanmış olacaktır [22].

Kök işlemlerinin uygulanabilmesi için metinde kullanılan dilin özelliklerinin çok iyi bilinmesi gereklidir. Bu bağlamda, Türkçe bitişken bir dildir. Her bir kelime, bir kökten ve kökün sonuna eklenmiş eklerden oluşur. Kelimelerin aldığı ekler, hem kelimenin anlamını hem de türünü değiştiren yapım ekleri ve kelimenin anlamını değiştirmeyen ancak türünü değiştirebilen çekim ekleri olmak üzere ikiye ayrılır. Türkçede kelime türleri isim, sıfat, zamir, zarf, edat, bağlaç, ünlem ve eylem olmak üzere 8 kısma ayrılır [27]. Yapılan çalışmanın amacına göre metin içerisindeki bu türler belirlenmeye çalışılır. Örneğin; Tablo 1’de ‘yap’ kelime kökünün farklı ek almış halleri verilmiştir.

Tablo 1. Örnek kelime kökü
(Example word root)

Kelime	Kök
Yapıyorum	Yap
Yapacaktım	Yap
Yaptım	Yap
Yaparım	Yap

Türkçe için kullanılan kök bulma yöntemleri 3 başlık altında toplanabilir. Birincisi sözlük tabanlı kök bulma yöntemidir [28]. Türkçe metinlerin kök ve eklerini bulmak için Zemberek kütüphanesi kullanılır. Zemberek, Türk dilleri ve Türkçe için geliştirilmiştir açık kaynaklı, bağımsız bir Doğal Dil İşleme kütüphanesidir [21]. Zemberek kelimelerin köklerini bulmak için kök ve ek sözlüğü kullanan sözlük tabanlı bir kök bulma yöntemidir [21].

İkinci kök bulma yöntemi, ek çıkaran kök bulucu (affix stripping stemmer) olarak adlandırılır ve sözlük tabanlı kök bulma yöntemlerinden farklı olarak Türkçe’nin kural tabanlı yapısından yararlanır [21]. Bu kök bulma yöntemi sözcüklerin sonlarındaki ekleri sondan başa doğru çıkararak kökleri bulur [28].

Kullanılan son kök bulma yöntemi ise, sabit önek (fixed prefix stemming) yöntemi olup temel olarak sözcüklerin ilk n karakterini sözcüğün kökü olarak kabul eder. Hızlı ve basit bir yöntemdir. n karakter ve daha kısa olan sözcükler olduğu gibi kabul edilir [21].

2.1.4. Durak ve Gereksiz Kelimeleri Çıkarma (Removal Of Stopwords and Redundant Words)

Edat (gibi, için, ancak, vb.), bağlaç (ve, ile, fakat vb.) ve zamir (ben, sen, o vb.) gibi metin içerisinde çok sık geçen ve tek başına bir anlam ifade etmeyen gereksiz kelimeler (stop words) metinden çıkarılması işlemidir [22]. Durak kelimeler, bir dilde sıklıkla kullanılan ve tek başına bir anlam ifade etmeyen kelime listesidir [24].

Her dilin kendine özgü durak kelimeleri var olmakla birlikte, standart bir liste yoktur [24]. Genellikle, durak kelimeler baz alınarak durak kelime listesi oluşturulur ve bu durak kelime listesine göre gereksiz kelimeler çıkarılır [17].

Etkisiz kelimelerin metinden atılması, sınıflandırmada kullanılacak olan özniteliklerin boyutunu azaltacağından öznitelik seçme yöntemi olarak da düşünülebilir [29]. Ayrıca metin içerisindeki bir kelimenin diğer kelimelere kıyasla çok veya az defa kullanılması kelimeyi ayrıştırıcı yapmadığından durak kelimesi olarak görülebilir [25].

2.1.5. Terim Ağırlıklandırma (Term Weighting)

Terim ağırlıklandırma ile sözcüklerin kendileri değil onları temsil eden sayısal değerleri kullanılır. Ağırlıklandırma işlemine kelimelerin doküman üzerindeki etkisi olarak da ifade edilebilir. Ayrıca yapısal olmayan veriler ağırlıklandırma ile tam yapısal hale dönüşmüş olur [23]. Ağırlıklandırma işlemi yapılırken çeşitli ağırlıklandırma yöntemlerinden faydalanılır. Bunlar; binary ağırlıklandırma, frekansa göre ağırlıklandırma (TF) ve terim frekansı x ters doküman frekansı (TFxIDF) ağırlıklandırma yöntemleri en çok kullanılan ağırlıklandırma yöntemlerindedir [30, 31].

2.1.5.1. Binary Ağırlıklandırma (Binary Weighting)

Bir kelimenin doküman içerisinde geçmesi ya da geçmemesini temel alan bir ağırlıklandırma yöntemidir [25]. Doküman içerisinde geçmesi durumunda 1, geçmemesi durumunda ise 0 değerini alır. Denklem 1'deki gibi ifade edilir.

$$w_{i,j} = \begin{cases} 1 & \text{eğer } i \text{ terimi} \\ j \text{ dokümanda geçiyorsa} \\ 0 & \text{diğer durumlarda} \end{cases} \quad (1)$$

2.1.5.2. Frekansa Göre Ağırlıklandırma (TF) (Term Frequency Weighting)

Kelimelerin metin içerisinde kaç defa geçtiğini gösteren basit bir yöntemdir [31]. Terim frekansı $T_i = (W_{i1}, W_{i2}, W_{i3}, \dots, W_{in})$ şeklinde ifade edilir. Burada T vektörü, i doküman sayısını, n terimlerin toplam sayısını ve W kelime frekansını ifade etmektedir [22].

2.1.5.3. Terim Frekansı x Ters Doküman Frekansı (TFxIDF) (Term Frequency x Inverse Document Frequency)

TFxIDF ağırlıklandırma, eğer bir kelime az sayıda dokümanda kullanılıyorsa, kelimenin o dokümanın sınıfının belirlenmesinde önemli olduğu, eğer bir kelime çok sayıda dokümanda kullanıyor ise kelimenin ayırt edici özelliğinin az olduğu düşüncesine dayanarak oluşturulmuştur [31]. TF, terim frekansını yani terimlerin bulunduğu metin dokümanı içerisinde kaç kez kullanıldığını, IDF ise aynı terimin tüm dokümanlar arasında kaç kez kullanıldığını gösteren ters doküman frekansıdır [32]. Bu iki değer çarpımıyla TFxIDF ağırlıklandırma değerleri elde edilmiş olur [32]. i .doküman, j .terim için oluşturulan TFxIDF formülü Denklem 3'te verilmiştir [17].

$$TF_{ij} = \frac{F_{ij}}{\sum_i F_j} \quad (2)$$

$$w_{i,d} = IDF_j \times TF_{ij} \quad (3)$$

$$IDF_j = \log\left(\frac{D}{df_j}\right) \quad (4)$$

2.1.6. Terim Ayıklama (Term Elimination)

Terim ayıklama işlemi, frekansı eşik değerinin altında olan kelimelerin kaldırılmasıdır. Düşük frekanslı kelimelerin, metin içeriğinde daha az öneme sahip olduğu varsayılarak bu kelimeler çıkarılır. Basit boyut azaltma yöntemlerinden biridir [22]. Dokümanlarda çok sık kullanılan kelimelerin ayırım güçleri çok az olduğu için maksimum geçiş sayısına göre de filtre uygulanabilir [30]. Kelime içerisinde tekrar eden harflerin bire indirgenmesi amacıyla metin normalizasyonu yapılabilir [24]. Örneğin; kelime "selammmmm" ise bu kelime "selam" şekline dönüştürülebilir. Minimum kelime uzunluğu belirlenerek de terim ayıklama yapılabilir [33]. Örneğin; kelime uzunluğu minimum 3 olarak belirlenmişse metin içerisinde bir ve iki harf uzunluğuna sahip kelimeler elenecektir.

2.1.7. Vektör Uzay Modeli (Vector Space Model)

Gerard Salton tarafından bilim dünyasına kazandırılan vektör uzay modelinde her bir kelime veri uzayında bir nokta olarak ifade edilmektedir [34]. Verinin sahip olduğu farklı özellikler noktanın koordinatları olarak gösterilmektedir [35]. Dokümanların sınıflandırılmasında sıklıkla kullanılan özellik vektör uzayı kelimelerin dokümanlardaki geçme sıklıklarına dayanmaktadır. Kısaca her bir doküman içindeki kelimelerin dokümanlardaki frekanslarının hesaplanması ile oluşturulur [17]. Vektör uzay modelinde terim ağırlıklandırma yöntemleri kullanılır. Sıklıkla binary, terim frekansı ve TFxIDF ağırlıklandırma yöntemleri kullanılır [36]. Ağırlıklar hesaplandıktan sonra Doküman Terim Matrisi (Document Term Matrix-DTM) oluşturulur [5]. Matris n adet kelime ve d adet doküman olmak üzere $n \times d$ boyutunda bir matris olarak ifade edilir. Şekil 1'de doküman matrisi

gösterilmiştir. D dokümanları, t terimleri, w ise terim ağırlıklarını ifade eder. Matrisin boyutu arttıkça üzerinde hesaplama yapabilmek için daha fazla bilgi işlem gücü ve hafıza gerekir [37]. Ayrıca doğru bilgiye ulaşma ihtimali de azalacaktır [34]. Özellik çıkarımı yöntemleri ile boyut indirilmesi yapılabilir [31]. Vektör uzay modeli iki büyük kısıta sahiptir. Birincisi vektör uzayındaki boyut sayısının büyük olması doküman vektörleri arasındaki yakınlığın hesaplanmasını zorlaştırmaktadır. İkincisi olarak eşanlamlı-çokanlamlı kelimelerin, dokümanlar arasında var olan ilişkiyi yok, olmayan ilişkiyi var gibi göstermesidir [37].

$$\begin{matrix}
 & t_1 & t_2 & t_3 & \dots & t_n \\
 \begin{matrix} D_1 \\ D_2 \\ \dots \\ D_d \end{matrix} & \begin{pmatrix} w_{11} & w_{12} & w_{13} & \dots & w_{1n} \\ w_{21} & w_{22} & w_{23} & \dots & w_{2n} \\ \dots & \dots & \dots & \dots & \dots \\ w_{d1} & w_{d2} & w_{d3} & \dots & w_{dn} \end{pmatrix}
 \end{matrix}$$

Şekil 1. Doküman terim matrisi
(Document term matrix)

Doküman vektörleri hem eğitim hem de test verisi için kullanılabilir ve benzerlik hesaplama işlemleri bu vektörler üzerinden gerçekleştirilir [23].

2.2. Özellik Seçimi (Feature Selection)

Metin sınıflandırma işlemleri yapılırken bütün dokümanlarda geçen ve metni oluşturan tüm kelimelerin çalışmaya dahil edilmesi çalışma zamanı arttırır [23]. Ayrıca bütün kelimelerin dahil edilmesi sınıflandırma başarısını da etkileyecektir. Metinleri oluşturan kelimelerin tamamının alınması yerine metni temsil ettiği düşünülen kelimeler seçilmelidir. Düşük boyutlu özellik vektörleri ile daha başarılı sonuçlar elde edilebilir [23]. Özellik seçimi, çalışma zamanının düşürülmesi ve daha başarılı sınıflandırma performansı elde etmek amacıyla büyük boyutlu veri kümesinin daha küçük boyutlara indirgenerek ifade edilmesini sağlar [22]. Özellik seçiminde ilgisiz ve gereksiz veriler silinir [14].

3. KELİME BULUTU (WORD CLOUD)

Görsel erişim araçlarından biri olan kelime bulutu, kelimelerin metin içerisindeki frekanslarına göre daha büyük boyutta ve daha belirgin bir renkte sıralandığı ağırlıklı liste gösterimidir. İlk defa 2002 yılında bir sunum şekli olarak Jim Flanagan tarafından kullanılmıştır. Bazı çalışmalarda etiket bulutu olarak isimlendirilir [38].

Kelime bulutunda, metin madenciliği işlemleri sonucunda belirlenen özniteliklerin ağırlıkları dikkate alınır. Frekans değeri büyük olan yani bulunma sıklığı yüksek olan kelimeler bulutta daha büyük boyutta ve daha belirgin renkte görülür. Frekans değeri küçük olan kelimeler ise daha küçük boyutta ve daha az belirgin bir renkte görünür. Büyük boyuttaki kelimeler daha fazla öneme sahiptir.

küçük boyuttaki kelimeler daha az öneme sahiptirler. Bulutu oluşturan kelimeler alfabetik, frekans sayılarına göre veya aralarında rasgele bir hiyerarşi olmadan yerleşir.

Sayıda çok fazla olan metin dokümanlarının veya metin içerikli verilerin özetlenmesinde büyük fayda sağlamaktadır. Metni oluşturan kelimelerin frekansına bağlı olduğu için anahtar kelimelerin belirlenmesinde yardımcı olur.

4. UYGULAMA (APPLICATION)

4.1. Veri kümesi (Data set)

Cumhurbaşkanı Recep Tayyip Erdoğan'ın "Şayet bir gün milletimiz 'Tamam' derse ancak o zaman biz kenara çekiliriz" sözlerinden kısa bir süre sonra, Twitter üzerinde 'Tamam' etiketi üzerinde muhalif kesimler tarafından tweetler atılmıştır. Buna karşılık olarak Sayın Cumhurbaşkanını destekleyenler tarafından 'Devam' etiketi altında tweetler paylaşmaya başlanmıştır. 15 binden fazla 'Tamam' ve 'Devam' etiketleri ile ilgili olmak üzere 30 bin tweet toplanmıştır. Bunlar arasından benzersiz olan 4886 'Tamam' ve 7430 'Devam' tweetleri veri kümesi olarak belirlenmiştir.

Toplanan tweetler, Twitter API üzerinden KNIME programı kullanılarak çekilmiştir. Toplanan tweetler Excel dosyası olarak kaydedilmiştir.

Cumhur İttifakını ve Millet İttifakını destekleyen parti liderlerine ve Cumhurbaşkanı adaylarına ait toplanan miting konuşma sayıları Tablo 2'de gösterilmiştir.

Tablo 2. Cumhur İttifakı ve Millet İttifakını destekleyen liderlerin miting konuşma sayısı
(Number of rally speeches of leaders supporting the People's Alliance and the National Alliance)

	Cumhurbaşkanı Adayları ve Parti Liderleri	Toplanan Miting Konuşma Sayısı
Cumhur İttifakı	Recep Tayyip Erdoğan	36
	Devlet Bahçeli	10
Millet İttifakı	Muharrem İnce	86
	Meral Akşener	61
	Temel Karamollaoğlu	17

Konuşmalar, çeşitli haber sitelerinden ve parti gruplarına ait siteler üzerinden elde edilmiştir.

4.2. Metin Ön İşlem Aşaması (Text Preprocessing Stage)

İlk olarak metin ön işlem aşamasının ilk adımı olan dönüştürme işlemi yapılmıştır. Toplanan tweetler içerisinde yer alan ve çalışma için bir anlam ifade etmeyen mention (@), hashtag (#) ve link (http//...) etiketleri

temizlenmiştir. Veriler içerisinde yer alan noktalama işaretleri ve simgeler silinmiştir.

Kök bulma işlemleri ve kelime ağırlıklandırma işlemleri, Yıldız Teknik Üniversitesi Kemik Doğal Dil İşleme Grubu tarafından geliştirilen Text2ArffV5 yazılımı kullanılarak yapılmıştır. Kelimelerin köklerini bulmak için Java kullanılarak geliştirilen açık kaynak kodlu Zemberek kullanılmıştır. Kelimeleri ağırlıklandırma yöntemi olarak terim frekansı ağırlıklandırma yöntemi kullanılmıştır.

Tablo 3. ‘Tamam’, ‘Devam’ tweetlerinde, Cumhuriyet İttifakı ve Millet İttifakı gruplarında en çok geçen 20 kelime (The 20 most frequently words in ‘Tamam’, ‘Devam’ tweets, People’s Alliance and Nation Alliance groups)

Sınıflar	En çok kullanılan 20 kelime
Millet İttifakı	iş, git yıl, ülke, genç, cumhurbaşkanı, çocuk, para, çalış, devlet, seçim, kardeş, söyle, gör, haziran, yeni, çık, dolar, zaman, gün
Cumhuriyet İttifakı	sistem, ittifak, haziran, ülke, yıl, yeni, çalış, cumhur, hareket, kardeş, milliyet, bay, hak, milli, çık, cumhurbaşkanı, son, hizmet, gelecek, yer
‘Tamam’ tweetleri	yeter, sıkıl, anla, git, gör, zaman, yaz, günaydın, iş, gün, ülke, yıl, güzel, haziran, bak, insan, genç, iyi, ders, adam
‘Devam’ tweetleri	başkan, dur, reis, çalış, ülke, genç, gün, son, yer, yıl, seçim, zaman, insan, gör, yeni, git, kal, iyi, dava, bil

Tablo 4. Parti liderlerinin ve Cumhurbaşkanı adaylarının konuşmalarında en sık geçen 20 kelime (The 20 most frequently words in the speeches of party leaders and presidential candidates)

Parti liderleri ve Cumhurbaşkanı adayları	En çok kullanılan 20 kelime
Recep Tayyip Erdoğan	bay, ülke, muharrem, proje, ak, yeni, kardeş, çalış, kemal, haziran, yatırım, bil, terör, iş, şehir, yol, vakit, çık, hizmet, cumhurbaşkanı
Devlet Bahçeli	sistem, ittifak, hareket, milliyet, cumhur, milli, haziran, hak, yeni, evet, gelecek, vatan, cumhurbaşkanlığı, karşı, mücadele, alın, çalış, ülke, in, akıl
Muharrem İnce	çocuk, cumhurbaşkanı, para, ince, lira, iş, saray, köprü, devlet, genç, dolar, bak, söz, kardeş, çık, konuş, muharrem, adam, meydan, geç
Meral Akşener	iş, para, genç, dolar, çocuk, devlet, ülke, iyi, kaynak, fabrika, sat, lira, seçim, sor, çık, kardeş, konuş, bakan, bul, vatandaş
Temel Karamollaoğlu	ülke, seçim, iktidar, zaman, iş, insan, konuş, problem, çöz, bugün, cumhurbaşkanı, adalet, politika, değiş, gör, çalış, kal, bulun, başla, fark

Kök bulma işlemi uygulandıktan sonra bulunan kelime kökleri listesinden daha önceden oluşturulan durak

kelimesi listesinde yer alan kelimeler çıkarılmıştır. Dört sınıfı temsil eden toplam 3209 tane ortak kelime kökü belirlenmiştir. Metin ayıklama işleminde bulunan özelliklerin boyutunu azaltmak için bir eliminasyon geliştirilmiştir. Ortak kelimelerin, bu dört gruba ait frekans değerlerinin ortalaması alınmıştır. Elde edilen ortalama değerlerine göre bir eşik değeri belirlenmiş ve ortalama değeri 200’ün üstündeki ve 5’in altındaki kelimeler elenmiştir. Terim ayıklama sonucunda dört gruba temsil eden kelime sayısı 1109’a indirgenmiştir. Tablo 3’te ‘Tamam’, ‘Devam’ tweetlerinde ve Cumhuriyet, Millet İttifak gruplarında en çok geçen 20 kelime kökü verilmiştir.

Aynı işlemler Cumhuriyet ve Millet İttifak gruplarının parti liderlerinin ve Cumhurbaşkanı adayları için de yapılmıştır. Metin ayıklama işleminde minimum değeri 5 olarak belirlenmiştir. Minimum değerin altındaki kelimeler elenmiştir. Tablo 4’te 5 liderin 24 Haziran seçimleri için yapmış oldukları konuşmalarda en sık kullandıkları 20 kelime verilmiştir.

4.3. Verilerin İstatistiksel Analizi (Statistical Analysis Of Data)

Metin ön işlem aşamasından sonra elde edilen verilerin analizlerinde R programı kullanılmıştır. Sınıflar arasındaki ilişkilerin belirlenmesinde Spearman sıra korelasyonu, sınıfların terim ağırlıklarını karşılaştırmak için de Kruskal Wallis Testi kullanılmıştır.

Cumhuriyet ve Millet İttifakı, ‘Tamam’ ve ‘Devam’ tweetleri arasındaki ilişkiyi ölçmek için korelasyon analizi yapılmıştır. Ayrıca Cumhuriyet ve Millet İttifakını destekleyen liderler arasındaki ilişkiyi ölçmek amacıyla da korelasyon analizi yapılmıştır. Her iki gruba ait veriler normal dağılmadığı için Spearman’s rho (Spearman Sıra Korelasyonu) testi kullanılmıştır.

Tablo 5. Sınıflara ait korelasyon katsayıları (Correlation coefficients of the classes)

Sınıflar	Millet İttifakı	Cumhuriyet İttifakı	‘Tamam’ Tweetleri
Millet İttifakı	1		
Cumhuriyet İttifakı	0,320**	1	
‘Tamam’ Tweetleri	0,393**	0,276**	1
‘Devam’ Tweetleri	0,376**	0,462**	0,593**

** %1 düzeyinde önemli

İncelenen sınıflara ilişkin korelasyon katsayıları Tablo 5 ve Tablo 6’da verilmiştir. Tablo 5 incelendiğinde, Millet İttifakı ile ‘Tamam’ tweetleri ($r=0,393$), Cumhuriyet İttifakı ile ‘Devam’ tweetleri ($r=0,376$) arasında zayıf düzeyde, pozitif ve anlamlı; ‘Tamam’ tweetleri ile Cumhuriyet İttifakı ($r=0,276$) arasında zayıf düzeyde, pozitif ve anlamlı; ‘Tamam’ tweetleri ile ‘Devam’ tweetleri ($r=0,593$) arasında orta düzeyde, pozitif ve anlamlı bir ilişki tespit edilmiştir. Benzer şekilde Cumhuriyet İttifakı ile ‘Devam’ tweetleri ($r=0,462$) arasında zayıf düzeyde, pozitif ve anlamlı bir ilişki vardır.

Tablo 6 incelendiğinde, Recep Tayyip Erdoğan'ın miting konuşmaları ile Devlet Bahçeli'nin ($r=0,264$), Muharrem İnce'nin ($r=0,279$), Temel Karamollaoğlu'nun ($r=0,362$) konuşmaları arasında zayıf düzeyde, pozitif ve anlamlı bir ilişki varken, Meral Akşener'in ($r=0,242$) konuşmaları arasında çok zayıf düzeyde, pozitif ve anlamlı ilişki olduğu tespit edilmiştir. Devlet Bahçeli'nin konuşmaları ile Muharrem İnce'nin ($r=0,21$) konuşmaları arasında zayıf düzeyde, pozitif ve anlamsız; Meral Akşener'in ($r=0,141$) konuşmaları arasında çok zayıf düzeyde, pozitif ve anlamlı; Temel Karamollaoğlu'nun ($r=0,318$) konuşmaları arasında zayıf düzeyde, pozitif ve anlamlı bir ilişki olduğu bulunmuştur. Muharrem İnce'nin konuşmaları ile Meral Akşener'in ($r=0,430$) ve Temel Karamollaoğlu'nun ($r=0,266$) konuşmaları arasında zayıf düzeyde, pozitif ve anlamlı bir ilişki olduğu görülmektedir. Meral Akşener'in konuşmaları ile Temel Karamollaoğlu'nun ($r=0,405$) konuşmaları arasında zayıf düzeyde, pozitif ve anlamlı bir ilişki olduğu tespit edilmiştir.

Tablo 6. Liderlere ait korelasyon katsayısı
(Correlation coefficient of the leaders)

Liderler	R.T.E.	D.B.	M.İ.	M.A.	T.K.
R.T.E.	1				
D.B.	0,264**	1			
M.İ.	0,279**	0,21 ^{öd}	1		
M.A.	0,242**	0,141**	0,430**	1	
T.K.	0,362**	0,318**	0,266**	0,405**	1

** %1 düzeyinde önemli, ^{öd} önemli değil, R.T.: Recep Tayyip Erdoğan, D.B.: Devlet Bahçeli, M.İ.: Muharrem İnce, M.A.: Meral Akşener, T.K.: Temel Karamollaoğlu

Terim frekansı ağırlıklarının 'Tamam' ve 'Devam' tweetlerine, Cumhuriyet ve Millet İttifakı gruplarına göre anlamlı bir farklılık gösterip göstermediğinin belirlenmesi için non-parametrik istatistiksel yöntemlerden olan Kruskal Wallis testi uygulanmıştır. Yapılan test sonucu elde edilen bulgular Tablo 7'de verilmiştir.

H_0 : Terim ağırlıkları gruplara göre farklılık göstermemektedir.

H_1 : Terim ağırlıkları gruplara göre farklılık göstermektedir.

Kruskal Wallis Testi sonucunda p değeri 0,05'ten küçük olduğu için H_0 hipotezi reddedilir. Terim frekans değerlerinin sınıflara göre anlamlı bir farklılık gösterdiği tespit edilmiştir ($\chi^2=1137,94$; $p<0,05$). Farkın hangi sınıfın terim ağırlıkları arasında olduğu harflendirme ile gösterilmiştir. Bu sonuçlara göre Millet İttifakı ve 'Tamam' tweetleri, Millet ittifakı ve Cumhuriyet İttifakı, Millet İttifakı ve 'Devam' tweetleri, 'Tamam' tweetleri ve Cumhuriyet İttifakı, 'Tamam' tweetleri ve 'Devam' tweetleri grupları arasında istatistiksel açıdan anlamlı bir farklılık bulunmaktadır. Cumhuriyet İttifakı ve 'Devam' tweetleri arasında ise anlamlı bir fark yoktur. Bu farklılığın veya denkleğin hangi grup lehine olduğuna karar vermek için sıra ortalaması değerlerine bakılır. Millet İttifakı grubunun

sıra ortalaması değeri, 'Tamam' ve 'Devam' tweetlerinin, Cumhuriyet İttifakının sıra ortalaması değerinden yüksek olması istatistiksel açıdan anlamlı bir farklılığa yol açmıştır. Ayrıca Cumhuriyet İttifakının ve 'Devam' tweetlerinin sıra ortalaması değeri, 'Tamam' tweetlerinin sıra ortalaması değerinden yüksek olması da anlamlı bir farklılığa yol açmıştır. Cumhuriyet İttifakı grubunun sıra ortalaması değeri, 'Devam' tweetlerinin sıra ortalaması değerinden yüksek olması anlamlı bir farklılığa yol açmamıştır.

Tablo 7. Grupların terim ağırlıklarının karşılaştırılması
(Comparison of term weights of groups)

Gruplar	N	Medyan (Maksimum - Minimum)	Sıra Ortalaması	χ^2	p
Millet İttifakı	1109	18(517-0) ^c	2817,64	1137,94	<0,001
Cumhuriyet İttifakı	1109	13(221-0) ^a	2478,67		
'Tamam' Tweetleri	1109	3(164-0) ^b	1123,72		
'Devam' Tweetleri	1109	12(250-0) ^a	2453,96		

a-c; aynı harfe sahip gruplar arasında fark yoktur.

4.3. Kelime Bulutunun Oluşturulması (Creating Word Cloud)

'Tamam' ve 'Devam' tweetleriyle ilgili toplamda 4 adet kelime bulutu oluşturulmuştur. Şekil 2 ve Şekil 3'deki kelime bulutları metin ön işlem sonucunda bulunan kelime köklerin frekans değerleri esas alınarak oluşturulmuştur. Şekil 4 ve Şekil 5 ise bulunan özelliklerden yola çıkarak Twitter kullanıcılarının neden 'Devam' ve 'Tamam' dediklerini ifade eden anahtar kelimelerden oluşturulmuştur.

'Devam' tweetlerine metin ön işlem aşaması uygulanması sonucunda eşik değeri 2'nin üzerinde olan toplam 2008 tane kelime kökü bulunmuştur. Bu kelime köklerinden frekans değeri en yüksek olan 998 tane kelime, bulut oluşturmak için kullanılmıştır.

'Devam' tweetleri ile oluşturulan bulut incelediğinde ise daha çok 'devam', 'et', 'yol', 'dur', 'ol', 'yap', 'millet', 'genç', 'reis', 'parti', 'ver' gibi kelimeler sıkça geçmektedir. Örneğin;

"RT @ibrahimozden06: Hayırlı Cumalar Bizim yolumuz Recep Tayyip Erdoğan'ın yoludur. Son nefese kadar DEVAM ??? @RT_Erdogan <https://t.co/n...>"

"Bilge Başkan" Temel Karamollaoğlu şehit eşine haddini bildirirken. Millet sillesi yaman olacak #DEVAM"

"Daima İleriyi düşünen Reis ile yolumuza DEVAM #CumhuriyetDEVAMdiyo@furkiyasın @trombositkan @feritozkartal02... <https://t.co/z3AfMF5dFk>"



Şekil 10. Muharrem İnce'nin en sık kullandığı kelimeleri gösteren kelime bulutu
(Word cloud showing the most frequently used words of Muharrem İnce)



Şekil 11. Meral Akşener'in en sık kullandığı kelimeleri gösteren kelime bulutu
(Word cloud showing the most frequently used words of Meral Akşener)



Şekil 12. Temel Karamollaoğlu'nun en sık kullandığı kelimeleri gösteren kelime bulutu
(Word cloud showing the most frequently used words of Temel Karamollaoğlu)

5. SONUÇ (CONCLUSIONS)

Bu çalışmada; Sayın Cumhurbaşkanı Recep Tayyip Erdoğan'ın "Şayet bir gün milletimiz tamam derse ancak o zaman biz kenara çekiliriz" sözlerinden sonra Twitter'da atılan 'Tamam' ve 'Devam' tweetleri metin madenciliği yöntemleri kullanılarak analiz edilmiş ve kelime bulutları oluşturulmuştur. Metin ön işlem sonucunda 'Devam' tweetlerini temsil eden 2008 ve 'Tamam' tweetlerini temsil eden 787 kelime kökü bulunmuştur. İki arasındaki kelime sayılarının farklı olması 'Tamam' tweetlerinin daha kısa ve atılan tweetlerin benzer olmasından kaynaklanmaktadır. Kelimelerin ağırlıklandırılmasında ve vektörel olarak ifade edilmesinde terim frekansı (TF) kullanılmıştır. Millet ve Cumhuriyet İttifakı ile 'Devam' ve 'Tamam' tweetleri arasında anlamlı bir ilişkinin olup olmadığını ölçmek için korelasyon analizi yapılmıştır. Twitter'da 'Devam' tweeti atan kullanıcılar Cumhuriyet İttifakını, 'Tamam' tweeti atanlar ise millet ittifakını desteklemektedir. Korelasyon analizi sonucunda bu ikili gruplar arasında zayıf bir ilişki tespit edilmiştir. Bunun nedeni gruplar arasında ifade edilen ortak kelime sayısı fazladır ancak kullanılan kelimelerin sıklıkları gruplar arasında önemli derecede farklılık göstermesinden dolayıdır. Terim frekansları ağırlıklarının gruplara göre farklı olup olmadığını belirlemek amacıyla Kruskal Wallis Testi uygulanmıştır ve önemli derece terim ağırlıkları arasında farklılıklar olduğu saptanmıştır. Cumhuriyet İttifakını temsil eden Recep Tayyip Erdoğan ile Devlet Bahçeli'nin; Millet İttifakını destekleyen Muharrem İnce, Meral Akşener ve Temel Karamollaoğlu'nun miting konuşmaları arasında zayıf bir ilişki olduğu görülmektedir.

Çalışma sonunda atılan tweetleri özetlemek ve sıklıkla ifade edilen kelimeleri göstermek için 4 tane kelime bulutu oluşturulmuştur. Birinci kelime bulutu 'Devam' tweetlerinin özelliklerini göstermektedir. 'Devam' etiketi altında atılan tweetlerde genellikle devam kelimesi daha çok ön plana çıkmaktadır. Dur ve yol kelimesini içeren tweetler, 'Durmak yok yola devam' ifadeleri şeklindedir. Cumhurbaşkanı'nun lakabı olarak atılan 'Reis' kelimesini 'Devam' tweetlerinde sıkça geçmektedir. Parti kelimesinin geçtiği tweetler daha çok Ak parti ile ilişkili tweetleri oluşturmaktadır. Genç kelimesini içeren tweetler gençlere hitaben atılan tweetlerdir. Ayrıca gençlik kolları ve teşkilatıyla ilgili tweetleri içermektedir. İkinci kelime bulutu ise 'Tamam' tweetlerini temsil eden özellikleri gösterir. Genellikle atılan tweetlerin çoğu sıklık ve yeter şeklinde paylaşılmıştır. 'Sıklık yeter', 'sıklık bırak yeter' şeklinde kısa ifadelerden oluşmaktadır. İş kelimesini içeren tweetler 'bu iş tamam' şeklindeki ifadelerden oluşmakta, ayrıca iş ve işsizlik ilgi ifadeler yer almaktadır. Bıktık kelimesini içeren tweetler, sıklıkla Cumhurbaşkanı'nun konuşma tarzından, devamlı ekranlarda görmekten, iftira ve yalanlardan bıktıklarını ifade etmektedir. Görme kelimesinin geçtiği tweetlerin çoğunluğunda televizyonda ve her yerde cumhurbaşkanı görmekten sıklıkla ifade etmektedir.

Üçüncü ve dördüncü kelime bulutları ise kullanıcıların neden 'Devam' ve 'Tamam' dediklerini göstermektedir.

Analiz edilen tweetler sonucunda ‘Devam’ tweeti atan kullanıcılar genellikle 25 tane nedenden dolayı devam demektedir. ‘Tamam’ tweeti atan kullanıcılar ise 28 tane nedenden dolayı sıkılıp tamam dediklerini ifade etmektedir.

Millet ve Cumhuriyet İttifakı konuşmalarını, Recep Tayyip Erdoğan’ın, Devlet Bahçeli’nin, Muharrem İnce’nin, Meral Akşener’in ve Temel Karamollaoğlu’nun miting konuşmalarını özetlemek amacıyla kelime bulutları oluşturulmuştur. Cumhuriyet İttifakı grubunda ülke, ittifak ve sistem gibi kelimeler daha çok kullanılmaktayken, Millet İttifakı grubunda iş, genç, para ve çocuk gibi kelimeler daha sık kullanılmaktadır.

Çalışma sonunda liderlerin söylemleri ile halkın söylemlerinde geçen kelimeler büyük oranda ortak ancak halkın sıklıkla dile getirdiği kelimeler ile liderlerin sıklıkla kullandığı kelimeler arasında geniş ölçüde farklılıklar olduğu bulgular sonucunda ortaya konmuştur. Ayrıca kullanıcıların neden tamam ve devam dedikleri tespit edilmiş ve sonuçlar kelime bulutları ile gösterilmiştir. Parti liderlerinin miting konuşmalarında en çok kullandıkları kelimeler kelime bulutları ile gösterilmiş ve atılan tweetler ile aralarındaki ilişki incelenmiştir.

KAYNAKLAR (REFERENCES)

- [1] T. Kaşıkçı, H. Gökçen, “Metin Madenciliği İle E-Ticaret Sitelerinin Belirlenmesi”, *Bilişim Teknolojileri Dergisi*, 7(1), 25-32, 2014.
- [2] Ö. Bilen, D. Hotaman, Ö. E. Aşkın, A. H. Büyüklü, “LYS Başarılarına Göre Okul Performanslarının Eğitsel Veri Madenciliği Teknikleriyle İncelenmesi: 2011 İstanbul Örneği”, *Eğitim ve Bilim Dergisi*, 39(172), 78-94, 2011.
- [3] S. Nicandro, C. M. F. Benjamin, H. K. Rashid, “Mining Known Attack Patterns From Security-Related Events”, *PeerJ Computer Science*, 1, 1-21, 2015.
- [4] Y. J. Lee, J. Y. Park, “Identification Of Future Signal Based On The Quantitative And Qualitative Text Mining: A Case Study On Ethical Issues In Artificial Intelligence”, *Quality and Quantity*, 52(2), 653-667, 2018.
- [5] D. Kılınc, E. Borandağ, F. Yücelar, V. Tunalı, M. Şimşek, A. Özçift, “KNN Algoritması Ve R Dili İle Metin Madenciliği Kullanılarak Bilimsel Makale Tasnifi”, *Marmara Fen Bilimleri Dergisi*, 28(3), 89-94, 2016.
- [6] M. F. Tsai, C. J. Wang, “On The Risk Prediction And Analysis Of Soft Information In Finance Reports”, *European Journal of Operational Research*, 257(1), 243-250, 2017.
- [7] G. João, R. Paulo, T. Duarte, “A Text Mining-Based Review Of Cause-Related Marketing Literature”, *Journal of Business Ethics*, (1) 111, 2016.
- [8] İ. O. Yigit, A. F. Ateş, M. Guvercin, H. Ferhatosmanoğlu, B. Gedik, “Call Center Text Mining Approach”, **25th Signal Processing and Communications Applications Conference (SIU)**, Antalya, 27 Haziran, 2017.
- [9] N. Akyel, K. Seçkin, “K-En Yakın Komşuluk Algoritmasının Hile Denetiminde Kullanımı”, *Muhasebe ve Vergi Uygulamaları Dergisi*, 5(1), 21-39, 2012.
- [10] J. Fluck, M. Hofmann-Apitius, “Review: Text Mining For Systems Biology”, *Drug Discovery Today*, 19, 140-144, 2014.
- [11] D. Kılınc, F. Bozyiğit, F. Yücelar, Ö. Akın, E. Borandağ, “Metin Madenciliği Kullanarak Yazılım Kullanımına Dair Bulguların Elde Edilmesi”, **Ulusal Yazılım Mühendisliği Sempozyumu (UYMS)**, İzmir, 2015.
- [12] B. Kırmacı, H. Oğul, “Author Recognition From Lyrics”, **23rd Signal Processing & Communications Applications Conference (SIU)**, Malatya, 16-19 May, 2489-2492, 2015.
- [13] S. İlhan, N. Duru, Ş. Karagöz, M. Sağır, “Metin Madenciliği İle Soru Cevaplama Sistemi”, **Elektrik-Elektronik ve Bilgisayar Mühendisliği Sempozyumu**, Bursa, 2008.
- [14] H. Göker, H. Tekedere, “Fatih Projesine Yönelik Görüşlerin Metin Madenciliği Yöntemleri İle Otomatik Değerlendirilmesi”, *Bilişim Teknolojileri Dergisi*, 10(3), 291-299, 2017.
- [15] B. Oğuz, U. Bilge, O. Saka, “Tıpta Metin Madenciliği”, **4. Tıp Bilişimi Kongresi**, Antalya, 108-112, 2007.
- [16] M. Ö. Dolgun, T. G. Özdemir, D. Oğuz, “Veri Madenciliğinde Yapısal Olmayan Verinin Analizi: Metin Ve Web Madenciliği”, *İstatistikçiler Dergisi*, 2, 48-58, 2014.
- [17] O. Yıldız, “Metin Madenciliğinde Anahtar Kelime Seçimi Bir Üniversite Örneği”, *Yönetim Bilişim Sistemleri Dergisi*, 2(3), 29-50, 2017.
- [18] M. Alaeddinoğlu, T. Aydın, D. Dal, “Biriklilik Kuralları İle Mekansal-Zamansal Veri Madenciliği”, *Fen Bilimleri Enstitüsü Dergisi*, 5(2), 191-212, 2014.
- [19] A. Yıldız, M. Demirci, “Corporate E-Mail Classification System”, **(UBMK’17) 2nd International Conference on Computer Science and Engineering**, Antalya, 559-564, 2017.
- [20] F. Başkaya, İ. Aydın, “Haber Metinlerinin Farklı Metin Madenciliği Yöntemleriyle Sınıflandırılması”, **International Artificial Intelligence and Data Processing Symposium (IDAP)**, Malatya, 1-5, 16-17 Sept., 2017.
- [21] V. Tunalı, T. T. Bilgin, “Türkçe metinlerin kümelenmesinde farklı kök bulma yöntemlerinin etkisinin araştırılması”, **Elektrik, Elektronik ve Bilgisayar Mühendisliği Sempozyumu (ELECO)**, Bursa, 598-602, 29 Kasım-1 Aralık, 2012.
- [22] A. Haltaş, A. Alkan, M. Karabulut, “Metin Sınıflandırmada Sezgisel Arama Algoritmalarının Performans Analizi”, *Gazi Üniversitesi Mühendislik Mimarlık Fakültesi Dergisi*, 30(3), 417-427, 2015.
- [23] M. F. Karaca, M. Günel, A. A. Taştan, “Metin Sınıflandırmada Benzerlik Hesaplama Tekniklerinin Değerlendirilmesi”, **17. Akademik Bilişim Konferansı**, Anadolu Üniversitesi, 4-6 Şubat, 2015.
- [24] Ö. Çoban, B. Özyer, G. T. Özyer, “Sentiment Analysis For Turkish Twitter Feeds”, **23th Signal Processing and Communications Applications Conference (SIU)**, Malatya, 2388-2391, 16-19 May, 2015.
- [25] E. Canhası, **Edebiyat eserlerinin web verilerine dayanarak sınıflandırılması**, Yüksek Lisans Tezi, Selçuk Üniversitesi, Fen Bilimleri Enstitüsü, 2007.

- [26] F. Gürbüz, E. K. Özyirmidokuz, "A Decision Support System By Analysis Of The Meeting Reports", *Sakarya Üniversitesi Fen Bilimleri Enstitüsü Dergisi*, 2018.
- [27] K. Ergün, C. Kubat, G. Çağır, R. Cesur, "Abstract Information Extraction From Consumer's Comments On Internet Media", *SAÜ Fen Bilimleri Dergisi*, 1, 33-40, 2013.
- [28] B Açıkalm, N G Bayazıt, "The Importance Of Preprocessing In Turkish Text Classification", **24th Signal Processing and Communication Application Conference (SIU)**, Zonguldak, 2053-2056, 16 – 19 May, 2016.
- [29] H. Gündüz, **Borsa İstanbul (BIST) 100 endeksi yönünün ekonomi haberleri ile tahmin edilmesi**, Yüksek Lisans Tezi, İstanbul Teknik Üniversitesi, Fen Bilimleri Enstitüsü, 2013.
- [30] M. F. Amasyalı, S. Balcı, E. Mete, E. N. Varlı, "Türkçe Metinlerin Sınıflandırılmasında Metin Temsil Yöntemlerinin Performans Karşılaştırılması" *EMO Bilimsel Dergi*, 2(4), 95-104, 2012.
- [31] K. Seçkin, **Metin madenciliğinde kullanılan yöntemlerin karşılaştırılması: Siyasi parti liderlerinin grup genel toplantı konuşmaları ile bir uygulama**, Yüksek Lisans Tezi, Sakarya Üniversitesi, Sosyal Bilimler Enstitüsü, 2011.
- [32] M. Ö. Cingiz, B. Diri, "Classification Of Microblogging Users", **IEEE 20. Sinyal İşleme ve Uygulama Kurultayı**, Muğla, 1-5, 18-20 Nisan, 2012.
- [33] E. Kılıç, M. R. Tavus, Z. Karhan, "Classification Of Breaking News Taken From The Online News Sites". **23rd Signal Processing & Communications Applications Conference (SIU)**, Malatya, 363-366, 16-19 May, 2015.
- [34] H. Bayer H., **Veri maddenciliğinde bir metin madenciliği uygulaması**, Yüksek Lisans Tezi, Beykent Üniversitesi, Fen Bilimleri Enstitüsü, 2011.
- [35] K Çaltış, O. Gazdağı, O. Yıldız, "Reklam İçerikli E-Postaların Metin Madenciliği Yöntemleri İle Otomatik Tespiti", *Bilişim Teknolojileri Dergisi*, 6(1), 2012.
- [36] M. F. Karaca, S. Görgünoğlu, "Columnready: İnternet Gazeteleri Köşe Yazılarını Hazırlama Uygulama Yazılımı", **14. Akademik Bilişim Konferansı**, Uşak Üniversitesi, 1-3 Şubat, 2012.
- [37] A. Güven, Ö. Ö. Bozkurt, O. Kalıpsız, "Veri Madenciliğinin Geleceği", **9. Akademik Bilişim Konferansı**, Dumlupınar Üniversitesi, 31 Ocak - 2 Şubat, 2007.
- [38] S. Savaşan, B. Diri, "Automatic Tag Cloud Generation From Turkish Contents", *Journal of Engineering and Natural Sciences*, 29, 156-169, 2011.