

Kripto Para Birimlerinin Twitter Verileri ile Metin Madenciliği Kapsamında İncelenmesi¹

Examination of Crypto Money Currencies within the Scope of Text Mining via Twitter Data

Nur Kuban TORUN² 

Abdulkadir ŞENGÜL³ 

DOI:10.33461/uybisbbd.1089670

Öz

Makale Bilgileri

Makale Türü:
Araştırma Makalesi

Geliş Tarihi:
18.03.2022

Kabul Tarihi:
08.06.2022

©2022 UYBİSBBD
Tüm hakları saklıdır.



Günümüzde sosyal medya platformları, kullanıcıların duygu ve düşüncelerini paylaştığı bir mecra haline gelmiştir. Sosyal medyanın yaygınlaşması nedeniyle, insanlar herhangi bir kategori hakkında, bir fikir hakkında duygu ve düşüncelerini ifade etmekte zorlanmamaktadır. En popüler sosyal medya platformu, kullanıcıların tweet adı verilen ifadelerini yayınlatabildikleri Twitter'dır. Bu çalışmanın amacı, Twitter'da kripto para etiketi ile atılan tweetlerin kripto para ile ilgili yazımlarını ortaya çıkarmaya ve kripto para ile ilgili gerçek verilerle kişilerin duygularını tespit etmeye çalışmaktır. 21 gün boyunca Twitter'dan "#kriptopara" etiketi ile gönderilen 41949 Türkçe tweet toplanmıştır. Açık kaynak kodlu R programla dili ile analiz gerçekleştirilmiştir. Tweetler ile içerik analizi yapılarak en çok tekrar eden kelimeleri bulunmuş ve kelime bulutu oluşturulmuştur. Twitter kullanıcılarının konu hakkındaki duygularını tespit etmek için duygu analizi yapılmıştır. Analiz sonucunda polarite puanı 0,003639, duygu ağırlığı ise pozitif yönde bulunmuştur. Son olarak tweetlerin en çok tekrarlanan ilk 5 kelimesi: Bitcoin, Resistance, Mana, Metaverse ve Altcoin'dir.

Anahtar Kelimeler: kripto para, bitcoin, twitter, r programlama, metin madenciliği.

Abstract

Article Info

Paper Type:
Research Paper

Received:
18.03.2022

Accepted:
08.06.2022

©2022 UYBİSBBD
All rights reserved.



Nowadays, social media platforms have become a medium for the users to share their thoughts and feelings. Due to the widespread use of social media, people have no difficulty in expressing their feelings and thoughts on an idea about any categories. The most popular social media platform is Twitter where users can post their expressions called tweets. The aim of this study is to try to reveal the crypto money-related words of the tweets posted with the crypto money tag on Twitter and to try to detect the emotions of people with real data about crypto money. During the 21 days, 41949 tweets which were sent from Twitter with the "#kriptopara" hashtag were collected. Content analysis was made with these tweets, and a word cloud was created to find out the most repetitive words, and at the same time, sentiment analysis was carried out to ascertain the emotions of the Twitter users. As a result of the analysis, the polarity score was found at 0.003639 and the emotion weight was found in a positive direction. Finally, the top 5 most repeated words of the tweets were: Bitcoin, Resistance, Mana, Metaverse and Altcoin.

Keywords: : crypto money, bitcoin, twitter, r programming, text mining.

Atıf/ to Cite (APA): Torun, N. K. ve Şengül A., (2022). Kripto Para Birimlerinin Twitter Verileri ile Metin Madenciliği Kapsamında İncelenmesi. Uluslararası Yönetim Bilişim Sistemleri ve Bilgisayar Bilimleri Dergisi, 6(1), 54-65

¹ Bu çalışma, 17-22 Şubat 2022 tarihlerinde, Uluslararası Bilişim Kongresi 2022 (IIC2022) sözlü bildiri olarak sunulmuştur.

² Dr. Öğretim Üyesi, Bilecik Şeyh Edebali Üniversitesi, İktisadi ve İdari Bilimler Fakültesi, İşletme Bölümü, nurkuban.akdemir@bilecik.edu.tr.

³ Öğrenci, Bilecik Şeyh Edebali Üniversitesi, İktisadi ve İdari Bilimler Fakültesi, Yönetim Bilişim Sistemleri, kadir.sengul61@hotmail.com.

1. GİRİŞ

İnsanlık tarih boyunca veri üretmiş ve ardında veri bırakmıştır. Gelişen teknoloji ile birlikte artık daha hızlı ve çok daha çeşitli veri üretilmektedir. Özellikle internetin yaygın kullanılması ve gelişmesine bağlı olarak internetin hızında da değişiklik olmuş ve hızı 4.5 g ve ardından günümüz 5g teknolojisine ulaşmıştır. Bu hızdaki artış ise verinin hacmi ve çeşidinde büyük değişiklikler meydana getirmiştir. Verinin çeşitlendirmesinde sosyal medya büyük bir öneme sahiptir. Sosyal medyada üretilen veriler işlenerek bilgi üretmemizi sağlamaktadır. Sosyal medyadan elde edilen verilerin işlenmesiyle ortaya çıkan bilgiden, finans, pazarlama, sağlık, siyaset ve turizm gibi birçok alanda yoğun bir şekilde yararlanılmaktadır. Özellikle müşteri portföyü oluşturmak, yeni bir ürüne karşı tepkileri ölçümlemek, kampanyaların etkilerini araştırmak ya da tüketicilerin davranışlarını analiz etmek adına sosyal medyadan elde edilen verilere yoğun bir ilgi bulunmaktadır.

Teknolojideki gelişimler sadece verinin elde edildiği kaynakları değil, aynı zamanda birçok alanı da etkilemiş, değişimler ve yenilikler ortaya çıkmıştır. Bu yeniliklerden bir tanesi finans alanında da kendini göstererek, paranın merkezîyetçi yapısına karşı içerisinde kriptografiyi içeren bir yatırım aracı olarak lanse edilen, merkezîyetçilikten uzak kripto paralar ortaya çıkmıştır. Kripto paraların en bilineni olan Bitcoin, 2009 yılında Satoshi Nakamoto takma isimli kişi/kişiler tarafından ortaya atıldığından beri çok değişikliğe uğramış ve alt coinler diye tabir edilen kripto para birimleri türemiştir. Kripto paranın arz ve talebe dayalı yapısını etkileyen en büyük veri kaynağı ise yatırımcılar için beyaz sayfalar (whitepages) kadar, büyük yatırımcıların söylemlerini paylaştığı sosyal medya mecraları olmaktadır. Özellikle yakın zamanda Elon Musk, attığı tweetler ile kripto para piyasasında büyük dalgalanmalar yaratmıştır. Bir yatırım aracı olarak görülen bu piyasada etkili olmak isteyen yatırımcıların özellikle Twitter isimli sosyal medya mecrasından elde ettiği verileri bilgi haline getirmesi bir zorunluluk halini almıştır.

Sosyal medya kullanıcıların vakit geçirdiği sıradan alanlar olmaktan öteye geçerek, verilerin bir arada toplanmasını sağlayan önemli platformlar haline almıştır. Sosyal medya, kullanıcılar kendi ürettiği içeriği yayınladığı ve paylaştığı, makaleleri, haberleri, düşünceleri, günlük yaşanan olayları, fotoğrafları görebildiği, ayrıca kendi görüşlerini yansıtılabildiği etkileşimli ağlardır (Bozgül, 2017). Sosyal medyanın kullanımın yaygınlaşmasıyla birlikte kullanıcıların ürettiği veriyi, herhangi bir işletme ya da kişinin toplama imkânı oluşmuştur. Sosyal medyanın bu bilgiye ulaşma gücü ile insanlığın ürettiği veriyi çeşitli programlama dili ile birleştirmek ve anlamlı bilgiler çıkarmak oldukça kolay hale gelmiştir. Sosyal medyadan alınıp birleştirilen bu veriler insanların herhangi bir konu üzerinde hangi tercihlerinin olduğunu anlamak için kullanılabilir. Günümüzde çoğu şirket bu verileri bir araya getirip insanların tercihlerini öğrenmekte ve insanların tercihlerine göre hareket etmektedir. Şirketler bu verileri sosyal medyadan alırken verilerin çok olması sebebi ile işe yarayan verileri bulmakta zorlanmakta ve veri kirliliği yaşamaktadır. Bu sorunu büyük veri yığımına doğru soruları sorarak çözmek mümkündür. Bunun için veri madenciliği ve metin madenciliği yöntemi kullanılabilir. Özellikle kullanıcıların özgür fikirlerini sunabildikleri bir platform olarak kurulan Twitter isimli blog sitesi, gerek kullanıcılar gerekse şirketler açısından en popüler sosyal medya mecrası haline gelmiştir. Twitter'dan elde edilen veriler ile insanların tercihleri, düşünceleri ve önerileri öğrenilebilmektedir. Twitter'dan elde edilecek verilere duygu analizi ve içerik analizi yapılarak bir konu ya da ürün hakkında kullanıcıların görüşleri belirlenebilmektedir. Twitter'dan elde edilen büyük verinin analizi için açık kaynak kodlu R programlama dili ile analiz gerçekleştirilebilmektedir. Sosyal medyadan kullanıcıların verileri R programlama dili ile çekilerek veri kirliliğinden arındırılarak doğru veriye ulaşım kolaylaşmıştır. Sosyal medyada toplanan verilerin analiz edilmesinde, veri madenciliğinin en önemli çalışma alanları içerik analizi ve duygu analiz yöntemidir. İçerik analizi çekilen tweetlerin içerisindeki kelimelerin konu ile alakalı en çok tekrar eden kelimelerini bulmayı ve kelimeler arasındaki bağlantıyı tespit etmeyi bu sayede ilgili konuyla gerçekte bağlantılı olan olayları ve durumları tespit etmeyi amaçlar. Duygu analizi ise metinler içerisinde yer alan duygu ve düşünceleri çeşitli algoritmaları kullanarak kullanıcıların herhangi bir konu hakkındaki duygularının pozitif veya negatif olarak hangi kategoride olduğunu bulmayı hedefler. Kripto para birimlerinin günlük hayata girmesi ile sosyal medyada çok konuşulmaya

başlamış, özellikle Twitter’da neredeyse her gün hashtag olarak trendlere girmiştir. Bu çalışmanın amacı ise Twitter’deki kullanıcıların kripto para üzerindeki gerçek duygu ve görüşlerini öğrenmek, en çok konuşulan kripto para ile bağlantılı kelimeleri ortaya çıkarmak ve bu kelimeler arasındaki bağlantıyı yorumlamaktır. Bu çalışmada 24.12.2021 ile 14.01.2022 tarihleri arasında #kriptopara hashtagi ile 41949 tweet çekilmiştir. Tweetlerin çekildiği bu dönemde kripto para yükselişte olup Twitter Türkiye hashtaginde #kriptopara etiketi üst sıralarda yer almıştır. Bu sebeple bu tarihler arasında tweetler çekilip tepkiler ölçülmüştür.

2. LİTERATÜR TARAMASI

Çalışma kapsamında literatür incelenmiştir. Kripto para ve kripto para birimleri ile, metin madenciliği ve makine öğrenmesi algoritmaları ile Twitter üzerine yapılan çalışmalara yer verilmiştir.

Lamon vd. (2017), tarafından yapılan bu çalışmada haber ve sosyal medya verilerini kullanarak 3 kripto para birimi için fiyat dalgalanmaları analiz edilmiştir. Ele alınan kripto para birimleri: Bitcoin, Litecoin ve Ethereum’dur. Çalışma içerisinde birçok sınıflandırma algoritmaları kullanılmıştır. Bitcoin ve Litecoin için en iyi sonucu Logistic Regression, Ethereum fiyatları için en iyi sonucu Bernoulli Naive Bayes vermiştir.

Abraham vd. (2018) çalışmasında, Twitter verilerini ve Google Trends verilerini kullanarak Bitcoin ve Ethereum fiyatlarındaki değişiklikleri tahmin etmek için bir yöntem sunmuşlardır. Araştırmacılar, tweetleri analiz ederek, tweet duyarlılığından ziyade tweet hacminin fiyat yönünün bir tahmincisi olduğunu bulmuşlardır.

Erdoğan ve Canayaz (2018) tarafından çalışmada Twitter kullanıcılarının duyarlılık analizi yapılmıştır. Çalışmada tweetlerin polaritesi hesaplanmış, en çok olumlu tweet alan duygu Ethereum dijital para birimi, en çok olumsuz duygu alan Bitcoin Cash para birimi olduğu gözlemlenmiştir.

Kinderis vd. (2018) tarafından yapılan çalışmada başta Bitcoin olmak üzere kripto para biriminin fiyat hareketleri üzerindeki duyguları analiz edilmeye çalışılmıştır. Bu amaç için veri madenciliği teknikleri kullanılmıştır. Belirli tarihler arasında toplam 6000 tweet bir araya getirilmiş ancak makinenin RAM kapasitesi ile sınırlandırılmıştır. 2013’den itibaren 1 ay içerisinde 8620 makale ve 7.000.000’un üstünde tweet toplanmıştır. Doğal Dil İşleme (NLP) tekniğini kullanarak Bitcoin için fiyat yönünü tahmin edilmeye çalışılmıştır. Modelleme tahmini için K-Nearest Neighbour (k-En yakın komşu), Decision Trees (Karar Ağaçları), Support Vector Machine (Destek Vektör Makineleri), Gaussian Naive Bayes, Logistic Regression (Lojistik Regresyon) ve Linear Discriminant Analysis (Doğrusal Ayrıcılık Analizi) yöntemleri kullanılmıştır. Yapılan çalışmanın elde edilen sonucu Bitcoin fiyatlarındaki dalgalanmaların doğru ve gerçek zamanlı tahminler verdiğini ortaya koymuştur.

Karasu vd. (2018) tarafından yapılan çalışmada, 2012 ve 2018 yılları arasında Bitcoin’in günlük kapanış fiyatlarının serileri kullanılarak Bitcoin’in fiyat tahmini yapılmıştır. Bu tahmin için Doğrusal Regresyon ve Destek Vektör Makinesi yöntemleri kullanılmıştır. SVM’nin birbirinden farklı parametre kombinasyonları ile test edilerek çok az hata içeren tahmin modelini elde edilmiştir. Bitcoin veri seti için SVM modelinin fiyat tahmin performansının LR modeline göre daha yüksek olduğu gözlemlenmiştir.

Rahman vd. (2018) tarafından yapılan çalışmada makine öğrenimi algoritmalarını uygulayarak Bitcoin’in fiyat değişimi ile kullanıcın duyarlılığı arasındaki korelasyonu incelemiştir. İnsanların kripto para birimi alanındaki görüşlerinin önemini bulma umuduyla, Bitcoin fiyat değişimini tahmin etme iddialı sorununa duygu ile yaklaşmıştır.

Alghobiri (2019) tarafından yapılan çalışmada kullanıcıların attığı tweetler gerçek veri seti olarak kullanılmıştır. Bu amaçla, kullanıcıların tweetlerinin Bitcoin hakkındaki görüşlerinin duygu analizi yapılmıştır. Araştırmada farklı kullanıcıların 50.859 tweeti kullanılmıştır. Oluşturulan veri seti, daha iyi karşılaştırmalı analizler için harici linkler içeren tweetler, bahsetmeler, tweetler, retweetler ve Bitcoin para birimi ile ilgili tartışan kullanıcılar hakkında kapsamlı veri analizi de dahil

olmak üzere beş ayrı bölüme ayrılmıştır. Sınıflandırma algoritması için Naive Bayes kullanılmıştır. Çıkan sonuçlar değerlendirildiğinde veri setinin genel duyguları olumlu olduğu için ve kullanıcıların Bitcoin hakkında çok az olumsuz görüş olduğundan, duyarlılık sonuçları %87 ile %93 değerleri arasında kalmıştır.

Kraaijeveld ve De Smedt (2020), en büyük dokuz kripto para biriminin (Bitcoin, Ethereum, XRP, Bitcoin Cash, EOS, Litecoin, Cardano, Stellar ve TRON) fiyat getirilerini tahmin etmek için genel Twitter duyarlılığının ne ölçüde kullanılabileceğini incelemiştir. Kripto para birimine özgü bir sözlük tabanlı duygu analizi yaklaşımı, finansal veriler ve ikili Granger nedensellik testi kullanarak, Twitter duyarlılığının Bitcoin, Bitcoin Cash ve Litecoin getirileri için tahmin gücüne sahip olduğunu ortaya koymuştur. Ayrıca, elde edilen tweetlerin en az %1-14'ünün Twitter "bot" hesapları tarafından gönderildiğini keşfetmek için buluşsal bir yaklaşım geliştirmiştir. Bu araştırma, birden fazla kripto para birimi ortamında Twitter duyarlılığının tahmin gücünü ele alan ve kripto para birimiyle ilgili Twitter botlarının varlığını araştıran ilk çalışma olmasından dolayı önemlidir.

Erdinç ve Bursa (2021) tarafından yapılan çalışmada Twitter'da Ripple coin hakkında #xrp hashtagi ile atılan Türkçe tweetlerin üzerinde nedensellik ve duygu analizi incelemesi yapılmıştır. Çalışma içerisinde R programlama dili, Toda-Yamamoto yöntemleri ve Twitter materyali kullanılmıştır. Çalışma sonucunda tek yönlü nedensellik elde edilmiş, fakat fiyatların Türkçe olarak atılan tweetlerin etkilediğine dair yeterli kanıt bulunamamıştır.

Koca (2021) tarafından yapılan çalışmada Twitter üzerinden elde edilen veriler ile Bitcoin üzerine duygu analizi yapılmıştır. Çalışma içerisinde Twitter Developer Hesabı kullanılarak #Bitcoin hashtagi ile 5000 tweet çekilmiştir. Tweetleri analiz etmek için Orange Programı kullanılmıştır. Çalışma sonucunda Bitcoin hashtagli tweetlerin %71'i sevinç, %15'inin şaşkınlık, %6'sının korku, %5'inin üzüntü, %2'sinin kızgınlık ve son olarak %0.0001'inin nefret içerdiği gözlemlenmiştir.

Köksal vd. (2021) tarafından yapılan çalışmada Twitter üzerinden elde edilen veriler Python programlama dili vasıtasıyla olumlu, nötr ve olumsuz tepkileri ayırarak etiket bulutu haline getirilmiştir. Lojistik Regresyon ve Naive Bayes algoritmaları kullanarak oluşturulan modellerin başarı oranları karşılaştırılmıştır. Naive Bayes yönteminin tweetlerdeki duyguları tahmin etmedeki başarı oranını %72,19, Lojistik Regresyon yönteminde ise bu oran %75,53 şekilde bulunmuştur. Bitcoin kelimesini içeren günlük pozitif tepkili tweet oranı ile Bitcoin günlük açılış değeri beraber kullanılarak Bitcoin'in kapanış değeri tahminlemesi yapılmıştır. Elde edilen sonuçlara göre Doğrusal Regresyon da r^2 değeri %88,97 çıkarken, Rastgele Orman Regresyonu yönteminde r^2 değeri %94,16 olmuştur.

Özer ve Torun (2022) tarafından yapılan çalışmada Twitter üzerinden #paribu etiketi ile 12627 tweet çekilerek duygu analizi yapılmıştır. Duygu analizinin bulgularında insanların #paribu etiketine tepkilerinin %75.72 oranında nötr, pozitif tepkiler oranının %15.25 ve negatif tepkilerin oranının %9.02 olduğu tespit edilmiştir. Elde edilen bulgularda nötr tepkilerin toplam tepkilerin %75.72'sini oluşturduğunu göz önünde bulundurularak yapılacak çıkarım, insanların doğrudan Paribu borsasına değil, Paribu borsası içerisinde yer alan kripto varlıklara -kripto paralara- yönelik olduğu düşünülmektedir sonucu gözlenmektedir.

Literatüre bakıldığında ethereum, xrp, riple coin, litecoin, cardano, tron, stellar ve yoğun bir şekilde de Bitcoin kripto para birimi gibi spesifik coin birimleri üzerine tweet çekilerek analiz yapıldığı görülmektedir. Bu çalışma Türkçe kriptopara etiketi ile çekilmiş tweetler ile yapılmıştır. Diğer çalışmalardan farklı olarak tüm kripto para birimlerini kapsayacak şekilde bağlantılı kelimelerin içerik analizi yapılmıştır. Kripto para birimlerini içeren kripto para piyasasına yönelik yapılan bu çalışmanın literatüre katkıda bulunacağı düşünülmektedir.

3. YÖNTEM

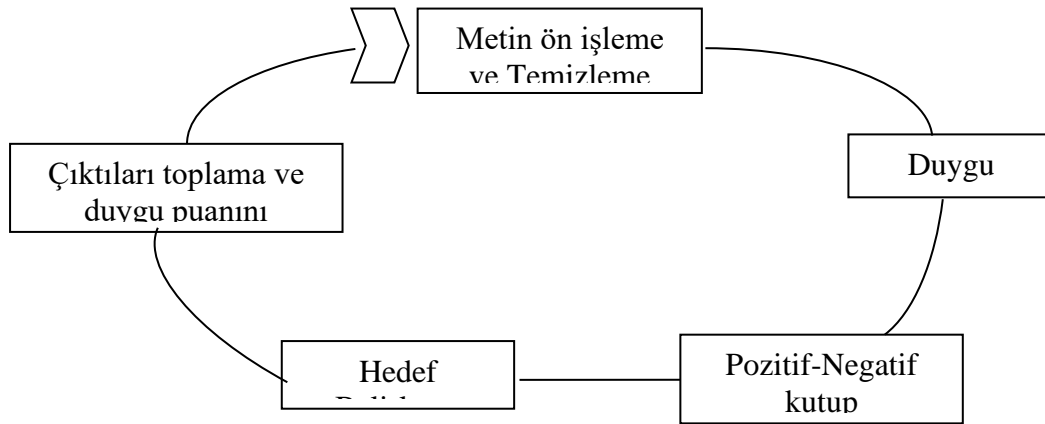
3.1 Metin Madenciliği

Metin madenciliği yapılandırılmamış olan verileri yapısal verilere dönüştürüp yararlı bilgiler elde etmeyi amaçlar (Yılmaz,2019). Metin madenciliğini, veri madenciliğinin diğer alt dallarından

ayırarak en önemli özelliği örüntülerin doğal dil metinlerinden çıkarılmasıdır. Metin madenciliği; metinlerin sınıflandırılması, metin özetleme, metinlerden konu çıkarılması ve metnin özü ile ilgili ilişkili modellemesi gibi çalışmaları hedefler. Metin madenciliği çalışmasının bu hedeflere ulaşması için; hece analizi, etiketleme, veri madenciliği, kelime frekans dağılımı ve görselleştirme gibi yöntemler kullanılmaktadır.

3.2 Duygu Analizi

Fikir madenciliği olarak da adlandırılan duygu analizi, bir yazarın metinde belirttiği duyguları anlamlandırmak için kullanılan doğal dil işleme yöntemidir. Bir metnin duygusu genellikle pozitif, nötr ve negatif olarak temsil edilir (Engüllü, 2018). Duygu analizi, işletmelerin sunmuş oldukları hizmet ve ürünlerin müşteri gözündeki yerini anlamak için kullanılabilir. İşletmeler duygu analizi yöntemini kullanarak marka görüşlerini takip edebilir. Buna yönelik pazarlama stratejileri geliştirebilir. Duygu analizi yapılırken herhangi bir metne tek başına bakarak değerlendirme yapmak yetersiz kalmaktadır. Çıkan sonuç doğruyu göstermeyebilir. Duygu analizinin homojen olması için sadece bir veri üzerinden gidilmesi yerine birçok veri kullanarak duygu analizi yapmak daha net sonuçlar verecektir. Duygu analizi çalışmaları doküman düzeyi duygu analizi, cümle düzeyi duygu analizi ve özellik tabanlı düzeydeki duygu analizi olmak üzere üç farklı düzeyde yapılmaktadır. Doküman düzeyinde duygu analizi dokümanın bütün olarak nötr/pozitif/negatif olarak sınıflandırılması çalışmasıdır. Bu şekilde nötr/pozitif/negatif şeklinde sınıflandırmaya duygu polarite tespiti denilmektedir (Kına,2022). Duygu analizi aşamaları Şekil 1'deki gibidir (Delen vd. , 2014; Kına, 2022).



Şekil 1: Duygu Akış Şeması.

3.3 Polarite Yöntemi

Polarite yöntemi bir dokümanın bütününü veya her bir kelimesinin pozitif, nötr veya negatif bir şekilde sınıflandırılmasını sağlayan bir duygu analizi çalışmasıdır (Özyurt,2021). -1 ile +1 arasında hesaplanmaktadır. 0 nötr değeri temsil etmektedir. +1 ne kadar olumlu olduğunu, -1 ise ne kadar olumsuz olduğunu göstermektedir. Bu çalışmada kelime bazlı polarite skoru hesaplanmış ve kelime bulutu oluşturulmuştur. Hesaplamanın matematiksel karşılığı formüldeki gibidir. Burada X tweetleri, Y ise polariteyi temsil etmektedir Eşitlik 1 (Erdoğan, 2019, 33). X değerini sınıflandırmak için kullanılan matematiksel formül Eşitlik 2'de verilmiştir.

$$X = \{x_0, x_1, x_2, x_3 \dots x_t \dots\} x, \in \mathbb{R}^n \quad (1)$$

$$Y = \{y_0, y_1, y_2, y_3 \dots y_t \dots\} y_i, \in \{-1, +1\}$$

$$W_{t+1} = W_T + \frac{\max(0, 1y_t(W^T \cdot X_t))}{\|x_t\|^2 + \frac{1}{2c}} \quad (2)$$

3.4 R Programlama Dili

R programlama dili veriyi işleme, grafiksel görüntüleme ve hesaplama yapmak için entegre edilmiş yazılım araçları paketi olarak ifade edilmektedir (Atalay, 2020). R programlama dili grafik oluşturma, istatistiksel hesaplama ve veri analizi için kullanılan açık kaynaklı bir yazılım dilidir. Bilimsel araştırmalar için oldukça önemli olan R programlama dili veriyi işlenebilir ve kullanışlı hale getirmede oldukça önemlidir. R programlama dili Twitter için semantik kümeleme, kullanıcı tecrübelerini gözetlemede ve veriyi görsel hale getirmede kullanılır.

3.5 Twitter Developer

Twitter'ın veri tabanındaki tweetler, beğeniler, retweetler gibi verileri kullanımımıza sunduğu bir uygulamadır. Bu uygulamanın veri tabanındaki bilgilere erişmek için uygulamaya ilk önce başvuru yapılması ve daha sonra başvurunun onaylanması gereklidir. Başvuru kabul olduktan sonra verilen dört anahtar ile veri çekilmeye başlanmıştır. Bu anahtarlar; consumer_key, consumer_secret, access_token ve access_secret'dir.

3.6 R Programlama Kütüphaneleri

R programlama dilinde Twitter'dan veri çekmek için hazır kodlar ile oluşturulan birçok kütüphane kullanılmıştır. Bu kütüphaneler Şekil 2'deki gibidir.

```
install.packages("twitter") #Twitterdan veri çekmek için kullanılır.
install.packages("ROAuth") #OAuth 1.0 spesifikasyonu için bir arayüz sağlar.
install.packages("tm") #Metin madenciliği yaparken kullanılır.
install.packages("RCurl") #Genel HTTP istekleri oluşturmaya izin veren işlevler sağlar.
install.packages("magrittr") #%%>% komutu ile zincirleme yapar.
install.packages("dplyr") #veri manipülasyon işlemlerini yapar.
install.packages("tidyverse") #veri bilimi için tasarlanmış paketleri içerir.
install.packages("ggplot2") #verileri görselleştirmeye yarar.
install.packages("furModeling") #tahmine dayalı modelleme yapar.
install.packages("lubridate") # zaman aralıklı işlevlerde çalışmak için kullanılır.
install.packages("stringr") #karakter yapılı veriler için kullanılır.
install.packages("tidytext") #Düzenli veri ilkelerini kullanmak birçok metin madenciliği görevini yapar.
install.packages("wordcloud") #kelime bulutu yapımı için kullanılır.
install.packages("RColorBrewer") #kelime bulutunun renklendirilmesi için kullanılır.
install.packages("tibble") #Modern bir veri çerçevesi oluşturma işleminde kullanılır.
install.packages("tidyr") #Verilerin analizine yönelik düzenlemeler için kullanılır.
install.packages("ggthemes") #Grafiklerin görünümünü kopyalayan 'ggplot2' temaları ve ölçekleri sağlar.
install.packages("readr") #csv, fwf gibi formattaki tablo yapısında veri içeren dosyaların okunmasını sağlar.
install.packages("readxl") #Excel dosyalarını içe aktarır.
install.packages("ggpubr") #Fonksiyonel programlama işlemlerini gerçekleştirir.
install.packages("formattable") #'Biçimlendirilebilir' Veri Yapıları oluşturur.
install.packages("ggstance") #Yatay 'ggplot2' bileşenlerini içerir.
install.packages("psych") #Psikolojik, Psikometrik ve Kişilik Araştırmaları için kullanılır.
install.packages("GGally") #ggplot2 uzantısıdır.
install.packages("rstatix") #İstatistiksel işlemler için kullanılır.
install.packages("sentimentr") #Metin polarite duyarlılığını hesaplar.
install.packages("webshot") #web sayfalarının ekran görüntülerini almak için kullanılır.
install.packages("htmlwidgets") #Çeşitli şekillerde işlenen HTML widget'ları oluşturmak için bir çerçevedir.
install.packages("syuzhet") #Metinden duygu ve duygudan türetilen plot yaylarını alır.
```

Şekil 2: R Dilinde Kullanılan Kütüphaneler

3.7 R Programlamada Kullanılan Kodlar ve Analiz Aşamaları

R programlama dilinde aktif hale getirilen kütüphanelerden sonra R programlama dili ile Twitter Developer hesabını birbirlerine bağlanıp Twitter'dan veri çekmek için uygun ortam Şekil 3'teki gibi hazırlanmıştır.

```
api_key = [REDACTED]
api_secret = [REDACTED]
access_token = [REDACTED]
access_token_secret = [REDACTED]
setup_twitter_oauth(api_key,api_secret,access_token,access_token_secret)
```

Şekil 3: Twitter Developer Hesabı ile R Dilini Entegre Etme

Hangi hasthtagden veri çekileceği, nereden, hangi dille ve ne kadar veri çekileceği Şekil 4'teki gibi belirtilmiştir. Çalışma 21 gün süresince çekilen tweetlerden oluşmaktadır. Twitter geliştirici hesabı ile maksimum 7 günlük tweet çekilebilmektedir. O yüzden 21 gün boyunca aynı tweetlerin tekrar etmemesine dikkat edilerek veri seti oluşturulmuştur. Buradaki n sayısı talep edilen tweet miktarını göstermekte ve şekilde n= 10000 görülmektedir. Bu talep edilen tweet miktarını göstermektedir. Fakat bazı günler daha az bazı günler daha çok tweet gelmiş ve n sayısı ona göre düzenlenmiştir.

```
tweets <- searchTwitter("#kriptopara", n=10000, locale = "tr_TR", lang="tr")
tweets.df <- twListTODF(tweets)
tweet_clean <- tweets.df

tweet_clean$text <- stri_enc_toutf8(tweet_clean$text)
```

Şekil 4: Hashtag, Tweet Sayısı, Dil ve Lokasyon Bilgileri Girme

Çekilen tweetler analize uygun olması amacı ile temizleme işlemi yapılmıştır. Bunun için sırasıyla; retweetler ve hashtag sembolleri kaldırılmıştır, noktalama işaretleri temizlenmiştir. Tüm harfler küçük harfe dönüştürülmüştür. ASCII (bilgi değişimi için Amerikan standart kodlama sistemi) formatına uymayan karakterler temizlenmiş ve alfabetik olmayan karakterler çıkartılmıştır.

```
#####rtlerin kaldırılması
tweet_clean$text <- ifelse(str_sub(tweet_clean$text,1,2) == "RT",
                          substring(tweet_clean$text,3),
                          tweet_clean$text)

view(tweet_clean$text)
#URL linklerinin temizlenmesi
tweet_clean$text <- str_replace_all(tweet_clean$text, "http[^[:space:]]*", "")

#Hashtag "#" ve "@" işaretlerinin kaldırılması
tweet_clean$text <- str_replace_all(tweet_clean$text, "#\\s+", "")
tweet_clean$text <- str_replace_all(tweet_clean$text, "@\\s+", "")

#Noktalama işaretlerinin temizlenmesi
tweet_clean$text <- str_replace_all(tweet_clean$text, "[[:punct:][:blank:]]+", " ")

#Tüm harflerin küçük harfe dönüştürülmesi
tweet_clean$text <- str_to_lower(tweet_clean$text, "tr")

#Rakamların temizlenmesi
tweet_clean$text <- removenumbers(tweet_clean$text)
#stopwordsler
liste=c(stopwords("en"))

#Gereksiz tekrarlar ve bağlaçların temizlenmesi
tweet_clean$text = removewords(tweet_clean$text,liste)

#ASCII formatına uymayan karakterlerin temizlenmesi
tweet_clean$text <- str_replace_all(tweet_clean$text, "[<].* [>]", "")
tweet_clean$text <- gsub("\uFFFF", "", tweet_clean$text, fixed = TRUE)
tweet_clean$text <- gsub("\n", "", tweet_clean$text, fixed = TRUE)
view(tweet_clean)

#Alfabetik olmayan karakterlerin temizlenmesi
tweet_clean$text <- str_replace_all(tweet_clean$text, "[^[:alnum:]]", " ")
```

Şekil 5: Veri Setini Temizleme İşlemi

Temizlenmiş tweetlerden, stopwords denilen Türkçe de etkisiz kelimeler olarak kullanılan kelimeler çıkartılmıştır. Çekilen tweetler de çalışmanın bütünlüğünü bozan ben, sen, o, -de, -da, -ki gibi kelimeler, çalışmadan bağımsız olan reklamlar, popüler olan konu başlıkları gibi kelimeler Şekil 6'daki gibi ortadan kaldırılmıştır.

```
#stopwordsler
liste=c(stopwords("tr"),"value","", "tag", "it", "rt", "hediye", "retweet", "ben", "read", "will", "referral", "community", "project",
"prize", "map", "entry", "new", "hard", "don", "herkes", "global", "erbakanyeniden", "sen", "o", "takipçimize", "gereken", "olması",
"var", "daha", "ilgili", "yorum", "beğeni", "etkinlik", "hükümet", "vergi", "üye", "parti", "ki", "ile", "veya",
"ve", "meclis", "ile", "gönderi", "flood", "takipçi", "ilgi", "yıl", "siz", "için", "birisi", "alıyorum", "ayrıca",
"ediyorum", "gün", "elde", "üzere", "kupa", "sma", "takipleşiyoruz", "tamam", "de", "fakat",
"yönetim", "giriş", "gece", "grup", "dünya", "temel", "anket", "soru", "bence", "yorumlara", "teşekkür", "kişiye",
"şey", "dün", "gibi", "sıkı", "bugün", "kadar", "bu", "dahi", "ne", "like", "çekiliş", "ancak", "emekli", "da")

#Gereksiz tekrarlar ve bağlaçların temizlenmesi
tweet_clean$text = removewords(tweet_clean$text, liste)
```

Şekil 6: Etkisiz Kelimelerin Çıkartılması

Günden güne çekip temizlenen tweetler Şekil 7’deki gibi kaydedilmiştir. Bu işlem her gün yapılmış 21 günün sonunda hepsi bir araya getirilmiştir. Ancak tekrar eden tweetler alınmamıştır.

```
write.csv(tweet_clean$text, file="kripto9.csv")
```

Şekil 7: Yazdırma İşlemi

CSV formatında bir araya getirilen veri seti, analiz yapmak için Şekil 8’deki gibi R programlama dilinde çalıştırılmıştır.

```
cryptopara = read.table(file.choose(), header=T, sep=";", encoding = "UTF-8")
str(cryptopara)
summary(cryptopara)
```

Şekil 8: Büyük Veriyi Analiz İçin Hazır Hale Getirme

Hazırlanan veriyi anlamlı bir hale getirmek ve en çok tekrar eden kelimeleri bulmak için Şekil 9’ daki kelime bulutunu oluşturmak için komut yazılmıştır.

```
#kelime bulutu
wordcloud(cryptopara$Tweets, min.freq=4, scale=c(4, 1, 8), max.words=100)
```

Şekil 9: Kelime Bulutu Oluşturma

Çekilen tweetlerin polarite skoru için Şekil 10’deki gibi çalıştırılmıştır.

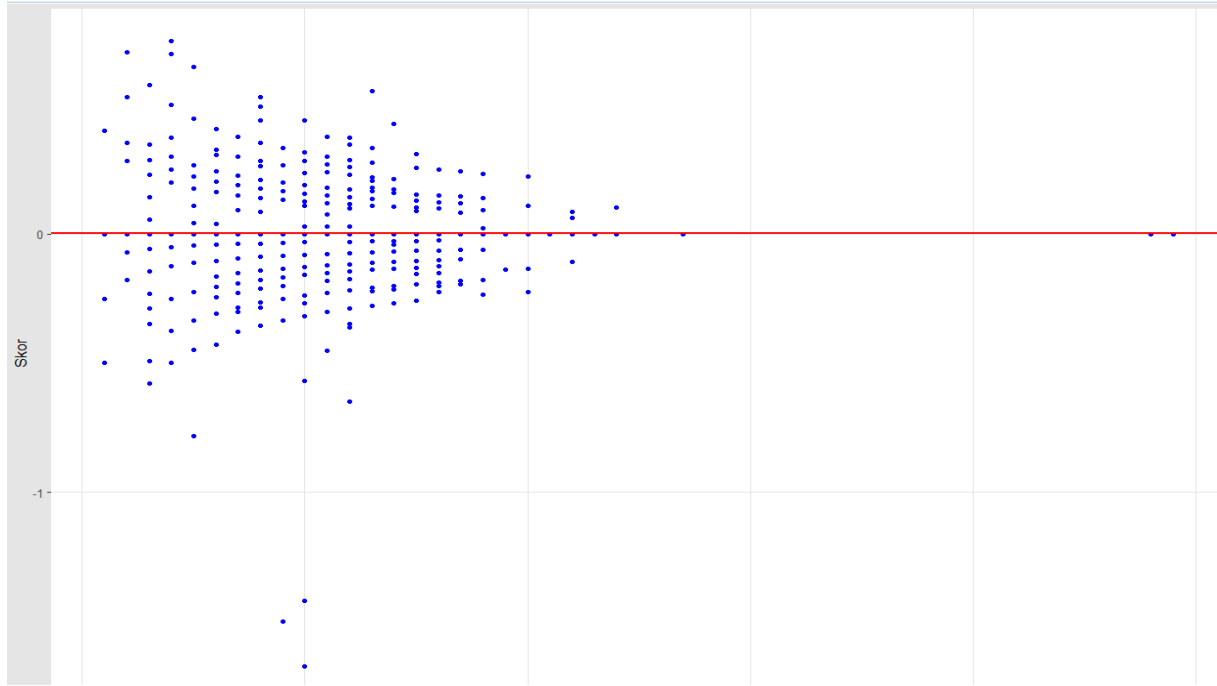
```
#polarite
polarite<-sentiment(cryptopara$Tweets$word)

tablo<-cbind(cryptopara$Tweets$word, polarite[,c(3,4)])

ggplot(tablo, aes(word_count, sentiment))+
  geom_point(color="blue")+
  geom_hline(yintercept = mean(tablo$sentiment), color="red", size=1)+
  labs(y = "Skor", x = "Kelimelerin Frekansı") +
  theme_igray()+
  labs() +
  theme(plot.caption = element_text(hjust = 0, face = "italic"))

stat.desc(polarite$sentiment, basic=T) %>% pander()
```

Şekil 10: Polarite Analizi



Şekil 12: Duygu Analiz Sonucu

Yapılan duygu analizinin matematiksel karşılığı Şekil 13’te verilmiştir. Bu tablodaki ‘mean’ kısmı duygu analizimizin matematiksel karşılığına denk gelmektedir. Mean pozitif ve negatif durumların ortalamasını alıp gerçek sonucu veren sütundur.

mean	SE. mean	CI. mean. 0.95	var	std. dev	coef. var
0.003639	0.0003172	0.0006217	0.00422	0.06496	17.85

Şekil 13: Duygu Analizi Matematiksel Sonucu

5. SONUÇ VE TARTIŞMA

İnsanlar tarafından kripto paraya olan ilgi ve alaka gün geçtikçe artmıştır. Kripto parayı bir yatırım aracı olarak gören yatırımcılar, daha çok gelir elde etmek ve riskleri minimize hale getirmek için sosyal medya araçlarını takip etmeye başlamıştır. Bu sosyal medya araçları içerisinde yer alan veriyi anlamlı bilgiler haline getirerek yatırımlarına yön vermeye çalışmaktadırlar. Twitter, kripto para ile özdeşleşen bir platform haline gelmiştir. Twitter’da kripto para ile ilgili bir çok etiket açılmakta ve açılan etiketler aracılığıyla fikir ve görüşler bildirilmektedir. Bu amaçla araştırmada kullanıcıların attığı tweetler analiz edilerek kullanıcıların kripto paraya karşı duygu ve düşünceleri incelenmiştir.

Çalışmada birçok kullanıcının attığı tweetlerden yola çıkılarak en çok ön plana çıkan kelimenin ‘Bitcoin’ olduğu görülmüştür. Bitcoin’in kelimelerde en üst sırada çıkmasının nedeni diğer altcoinlerin durumunun Bitcoin’in artması veya azalması durumuna göre belirlenmektedir. Bu sayede insanlar Bitcoin’in durumuna göre altcoinleri tercih etmekte ve insanlara yol gösterici nitelikte olmaktadır. Bitcoin’in ardı sıra “direnc” kelimesi karşımıza gelmektedir. Direnc genel piyasada kripto paranın beklediği artışın durmasını veya azalışın durması beklenen seviyeye denir. Atılan tweetlerin çekildiği zaman da Bitcoin’in düşüşünün beklediği direnc seviyesinde olduğu

gözlemlenmektedir. Diğer kelime ise “altcoin” olarak çıkmaktadır. Altcoin Bitcoin’in durumuna göre artış veya azalış yaşayan diğer dijital para birimleridir. Kelime bulutunda “dot”, “dogecoin”, “axs”, “ath”, “btp” “pitbull” gibi coinler karşımıza çıkmaktadır. Bu altcoinler çekilen tarihler arasında büyük beklenti ve yükseliş beklentisi yaşamışlardır. Kelime bulutunda çıkan “mana” kelimesi ise şu zamanlarda en çok revaçta olan altcoindir. Bir diğer kelime ise “Metaverse” olarak karşımıza çıkmıştır. Metaverse kelimesi sanal evrendir. Bu sanal olan evrende para birimi olarak kripto para kullanılır. Mana coin’in kendi sanal dünyası olduğundan revaçta olduğu gözlemlenmektedir.

Yapılan duygu analizine göre kullanıcıların kripto paraya karşı duygu yönü nötr olarak çıkmıştır. Polarite skoru 0.003639 çıkmıştır. Pozitif yönlü duygusal bir gelişme başladığı söylenebilir. Dünyaca ünlü markaların kripto paraya olan yatırımı insanların kripto paraya olan güvenini arttırmıştır. Kullanıcıların çoğunluğu bu dijital para birimlerinden kar elde etmiştir. Web 3.0 kavramının yaygınlaşması ile birlikte insanlar dijital dünyayı daha iyi anlamış ve çoğu kullanıcı Metaverse dünyasına kendisini hazırlamaya başlamıştır. Gelecekte dünyanın bu evrenlerde geçecek olması ve para kullanımının dijital para olacak olması ile birlikte insanların gün geçtikçe daha çok dikkatini çekmektedir. Bu açıdan şirketler için tüketiciler ve de işletme arasında etkileşim yaratabilmek adına bir fırsat doğmaktadır. Farkındalık yaratmak isteyen ve sosyal medyada konuşularak etkisini arttırmak isteyen şirketler için kripto para önemli bir araç olabilecektir. Ayrıca kripto paraya yönelik duygunun pozitive doğru dönüşmeye başlaması, marka bağlılığı yaratmak isteyen şirketler için de bir avantaj oluşturabilecektir. Kripto para piyasasında şirketlerin doğru bilgi kaynağı oluşturması ve yatırımcılarına kazandırması ile güven unsuru oluşturulabilir. Diğer yandan R programlama ile çekilen verilerin temizlenmesi aşamasında, hashtag içerisinde yer alan; ancak kripto para ile ilgisi bulunmayan tweetler de görülmüştür. Bu durum bir çok kullanıcının dikkatini çeken kripto para ile ilgili etiketin bir tanıtım ve reklam aracı olarak kullanıldığını göstermektedir. Bu yönüyle kripto para bir dikkat çekme aracı olarak kullanılabilir. Ayrıca tweetler içerisinde aynı amaç için yazılmış bir birini takip eden farklı kullanıcılar tarafından atılan tweetlere de rastlanılmıştır. Bu durum bot hesaplar ile açıklanabilmektedir. Yatırımcıları bekleyen önemli bir tehdit bu bot hesaplarının bilgiyi kirlenmesi ve manipüle etmesidir. Bu açıdan yatırımcıların elde ettikleri verileri bot ve troll diye tabir edilen hesaplardan arındırmaları önemli bir adımdır.

Çalışmanın kapsamına bakıldığında sınırlılıkları ve geliştirilmesi gereken yönleri olduğu görülmektedir. Yapılan bu araştırma 21 gün içerisinde atılan tweetler üzerinden yapılmıştır. Araştırmanın daha da fazla geliştirilmesi için 21 günden daha uzun zaman diliminde tweet çekilebilir. Araştırmada 21 gün içerisinde çekilen 41949 adet tweet kullanılmıştır; ancak ileride yapılacak olan çalışmada bu sayı daha da artırılabilir. Ayrıca araştırma sadece Türkiye’den, Türkçe olarak atılan tweetler kullanılmıştır. Farklı bölgelerden ve/veya farklı dillerden atılan tweetler kullanılarak araştırmanın kapsamı daha da geliştirilebilir. Sonuç olarak bu çalışma, kripto para etiketi ile atılan tweetlerin analiz edilerek kullanıcıların duygu ve düşüncelerini analiz etmesi ve elde edilen verilerde diğer bütün kripto para birimlerini içerisinde barındırmasından dolayı önemli ve literatüre katkı sağlayacak bir çalışmadır.

KAYNAKÇA

- Abraham, J., Higdon, D., Nelson, J., Ibarra, J. (2018). “Cryptocurrency Price Prediction Using Tweet Volumes and Sentiment Analysis”. *SMU Data Science Review*, 1 (3), 1-21.
- Alghobiri M. (2018). “Using Data Mining Algorithm for Sentiment Analysis of Users Opinions About Bitcoin Crypto Currency”. *Journal of Theoretical and Applied Information Technology*, 97(8), 2195-2205.
- Atalay İ. E. (2020). R programlama Dili ile Türkiye Finansal Risk Verilerinin Animasyonları. Yüksek Lisans Tezi. Eskişehir Osmangazi Üniversitesi Fen Bilimler Enstitüsü. İstanbul.
- Bozgül F. (2017). “Sosyal Medya Nedir?”, [https://www.pazarlamasyon.com/sosyal-medya-nedir,\(08.02.2022\).](https://www.pazarlamasyon.com/sosyal-medya-nedir,(08.02.2022).)

- Delen D., Sharda R., Turban E., (2014). “Business Intelligence and Analytics : System for Decision Support”, Pearson Education Limited (10), 322-326.
- Duygu Analizi. Doktora Tezi. Van Yüzüncü Yıl Üniversitesi Fen Bilimler Enstitüsü İstatistik Anabilim Dalı. Van.
- Engüllü B. (2018). Twitter Sentiment Analysis. Yüksek Lisans Tezi. Bahçeşehir Üniversitesi Fen Bilimleri Enstitüsü ve Bilgisayar Mühendisliği. İstanbul.
- Erdoğan U., Bursa N. (2021). “Covid-19 Pandemi Sürecinde Twitter Yorumları ile Altcoin
- Erdoğan M. C., Canayaz M. (2018). "Crypto-Currency Sentiment Analyse on Social Media," 2018 International Conference on Artificial Intelligence and Data Processing (IDAP), 2018, 1-5.
- Erdoğan M.C. (2019). Büyük Veri Araçları Kullanarak Sosyal Medyada His Analizi Yapma. Yüksek Lisans Tezi. Van Yüzüncü Yıl Üniversitesi Fen Bilimleri Enstitüsü ve Elektrik-Elektronik Mühendisliği. Van.
- Karasu S., Altan A., Saraç Z., Hacıoğlu R. (2018). “Prediction of Bitcoin Prices with Machine Learning Methods Using Time Series Data”.26th Signal Processing and Communications Applications Conference (SIU), 1-4.
- Kına E. (2022). Makine Öğrenmesi Algoritmaları Kullanılarak Twitter Mobil Oyun Verilerinde
- Kinderis M., Bezbradica M., Crane M. (2018). “Bitcoin Currency Fluctuation”. Future Information System and Risk, 31-41.
- Koca, G., “Bitcoin Üzerine Twitter Verileri ile Duygu Analizi”. Anadolu Üniversitesi İktisadi ve İdari Bilimler Fakültesi,22(4),19-30.
- Köksal, B., Erdem, G., Türkeli, C., & Öztürk, Z. K. “Twitter'da Duygu Analizi Yöntemi Kullanılarak Bitcoin Değer Tahminlemesi”. Düzce Üniversitesi Bilim ve Teknoloji Dergisi, 9(3), 280-297.
- Kraaijeveld, O. ve De Smedt (2020). “The Predictive Power of Public Twitter Sentiment for Forecasting Cryptocurrency Prices”. Journal of International Financial Markets, Institutions and Money, 65.
- Kripto”. Yönelim ve Ekonomi Araştırmaları Dergisi (19),362-381.
- Lamon, C., Nielsen, E., Redondo, E. (2017). “Cryptocurrency Price Prediction Using News and Social Media Sentiment”. SMU Data Sci. Rev (1), 1-27.
- Özer E. ve Torun N.K. (2022). “R İle İçerik Analizi Ve Duygu Analizi Paribu Borsası”. Uluslararası Bilişim Kongresi (IIC 2022),175-180.
- Özyurt B. (2021). Heterojen Veri Kaynaklarında Fikir Madenciliği, Otomatik Özellik Çıkarımı ve Duygu Analizi. Doktora Tezi Gazi Üniversitesi Bilgisayar Bilimleri Ana Bilim Dalı. Doktora Tezi. Ankara.
- Rahman S., Hemel, J. N., Anta, S. J. A., Muhee A., H., Uddin J. (2018). “Sentiment Analysis Using R: An Approach to Correlate Cryptocurrency Price Fluctuations with Change in User Sentiment Using Machine Learning”. 2nd International Conference on Imaging, Vision & Pattern Recognition (icIVPR), 492-497.
- Yılmaz C.M. (2019). Spam Detection By Using Network and Text Embedding Approaches. Yüksek Lisans Tezi. Ortadoğu Teknik Üniversitesi Fen Bilimleri Enstitüsü. Ankara.