

Mikro Bloglardaki Finans Toplulukları için Kullanıcı Ağırlıklandırılmış Duygu Analizi Yöntemi

Alpaslan Burak Eliaçık¹, Nadia Erdoğan¹
¹ Bilgisayar Mühendisliği Bölümü, İstanbul Teknik Üniversitesi,
İstanbul
{aeliacik, nerdogan}@itu.edu.tr

Özet. Duygu analizi bilgisayar bilimlerinde ele alınan popüler bir araştırma alanıdır. Duygu analizi, bireyler tarafından oluşturulan metinsel dokümanlardan yararlanarak bireylerin veya toplumların bir konu ile ilgili tutumlarını belirlemeyi amaçlamaktadır. Günümüzde sosyal mikro blog servislerinin yaygınlaşmasıyla, fikir beyan eden metin sayısında artış gözlenmektedir. Finans, ekonomi, politika gibi farklı sosyolojik alanlarda toplumun eğilimini analiz etmek için gerçekleştirilen duygu analizi çalışmalarında bu kamusal bilgiye sıklıkla başvurulmaktadır. Bu bağlamda bu çalışmada, mikro blog servisleri üzerinde analiz yapmak için yeni kullanıcı metriklerini kullanan özgün bir duygu analizi yöntemi önerilmektedir. Önerilen yöntem, mikro bloglardaki finans topluluklarının duygu polaritesi ölçümünde kullanılmaktadır. Ayrıca bu çalışmada hesaplanan finans topluluğuna ait duygu polarite değerleri ile Borsa İstanbul 100 endeks hareketleri arasında haftalık bazda korelasyon analizi yapılmaktadır. Aradaki bağdaşıklık oranı Pearson korelasyon katsayı yöntemi ile hesaplanmaktadır. Deneysel çalışmalar sonucunda önerilen duygu analizi yönteminin, borsa hareketleri ve mikro bloglardaki finans topluluklarına ait duygu polariteleri arasındaki bağdaşıklık oranını önceki duygu analizi yöntemlerine göre daha hassas hesapladığı görülmektedir.

Anahtar Kelimeler. Duygu Analizi, Sosyal Ağ Analizi, Mikro Blog, Borsa.

Abstract. Nowadays, sentiment analysis is a popular research area in computer science. which aims to determine a person's or a group's mood, behaviour and opinion about any topic with using textual documents. With the proliferation of social micro-blogging sites, digital opinion text data is increased. Thus, many sentiment analysis researches are performed on these public data in different sociological fields, such as finance, economy and politics. In this paper, a novel sentiment analysis method is proposed on micro-blogging sites which uses new user metrics. Proposed method is used to measure financial community's sentiment polarity on micro-blogging sites. In addition to that we analyze the correlation between the mood of financial community and the behavior of the Borsa İstanbul 100 index weekly by Pearson correlation coefficient method. Our test results show that this novel sentiment analysis method improves to the accuracy of linear relationship between the behavior of the stock market and the sentiment polarity of the financial community.

1 Giriş

Günümüzde, insanların iletişim ihtiyaçlarını pratik bir şekilde karşılayan mikro blog servisleri çok hızlı bir şekilde yaygınlaşmaktadır. Mikro blog servisleri sayesinde günlük olan olaylar ve haberler kamusal alanda hızlı bir şekilde yayılmaktadır. Bu imkândan yararlanmak isteyen bireyler ilgi duydukları alanlar ile alakalı bilgiye erişebilmek için birbirlerini mikro blog servisleri aracılığı ile takip etmektedir. Bu şekilde oluşan topluluklar sayesinde konu ile ilgili bilgi, topluluklar içerisinde hızlıca yayılmaktadır ve topluluğa üye olan kullanıcılara etkili bir şekilde tesir etmektedir.

Duygu analizi, bilişimsel dilbilimi ve veri madenciliği içerisinde yer alan bir alt disiplindir. Duygu analizinin ana amacı kişilerin ruh halini, davranışlarını ve fikirlerini metinsel dokümanlardan keşfetmektir. Mikro blog servislerinin artmasıyla, bu servislere ait kamusal veriler; politika, ekonomi ve finans gibi farklı sosyolojik alanları hedef alan duygu analizi çalışmalarında kullanılmaktadırlar.

Günümüzde, mikro blog servisleri üzerinden günlük çok sayıda finans haberi paylaşılmaktadır. Haberler mikro blog servisi üzerinden gönderildiği için piyasa hakkında önemli bilgiler içermektedir. Bu servisler sayesinde yatırımcılar finans ile alakalı önemli bilgilerden hızlıca haberdar olup bilgiler dâhilinde yatırım kararı almaktadır. Alınan yatırım kararları ile finansal piyasaların davranışı etkilenmektedir. Bu bağlamda bu iletiler üzerinde yapılan duygu analizi araştırmaları sayesinde finansal piyasaların davranışları analiz edilebilmektedir. Literatür incelendiğinde, araştırmacılar finansal piyasaların davranışını incelemek için birçok duygu analizi çalışması gerçekleştirmişlerdir [1], [3], [4], [11], [12], [13], [14].

BIST 100, Türkiye merkezli en iyi 100 kamuya açık şirketi içeren ve standart işlem seansında bu şirketlere ait hisse senetlerinin alınıp satıldığı bir borsa endeksidir. Bu çalışmada BIST 100 endeksinin hareketleri ile klasik duygu analizi yöntemi kullanılarak elde edilen Türk finans topluluğunun duygu polaritesi arasında doğrusal bir korelasyon olduğu görülmektedir. Daha sonra klasik duygu analizi yöntemine ek olarak önerilen özgün kullanıcı metrikleri ile birlikte topluluğa ait hesaplanan duygu polaritesinin, klasik yöntemlere nazaran BIST 100 endeksinin hareketleri ile daha doğrusal bir korelasyona sahip olduğu gösterilmektedir. Finans alanında gerçekleştirilen önceki duygu analizi yöntemlerinde topluluğun duygu polaritesi hesaplanırken, mikro blog kullanıcılarının ait olduğu topluluğa olan etkisi göz ardı edilerek sadece kısa ileti içeriklerine odaklanılmaktadır. Önceki çalışmalardan farklı olarak, bu çalışmada kullanıcıların topluluğa olan etkilerini hesaplamaya dâhil eden yeni bir duygu analizi yöntemi önerilmektedir. Her kullanıcı için bir "güven" değeri hesaplanmaktadır ve gönderilerine bu güven değeri ile orantılı bir katsayı değeri atanmaktadır. Kullanıcının, finans alanındaki uzmanlığı ve ait olduğu topluluktaki inandırıcılığı onun güven değerini oluşturmaktadır.

Bildirinin geri kalan bölümleri şöyle sıralanmaktadır: Bölüm 2'de finans alanında gerçekleştirilen duygu analizi yöntemleri ele alınmaktadır. Bölüm 3'te önerilen çalışma detaylı olarak açıklanmaktadır. Bölüm 4'te çalışma kapsamında toplanan veri kümesi ve yapılan deneysel çalışmaların sonuçları vurgulanmaktadır. Son bölümde ise çalışma, çıkan sonuçlar ışığında sonuca bağlanmaktadır.

2 Finans Alanındaki Mevcut Duygu Analizi Çalışmaları

Liu [7], insana ait görüşlerin hemen hemen tüm insani faaliyetlerin merkezinde yer aldığını belirtmektedir. Bu nedenle, birey bir konu hakkında karar alacağı zaman başkalarının o durum hakkındaki görüşlerini öğrenmek istemektedir. Bu bağlamda finansal piyasalarda, başkalarının düşünceleri çok önemlidir, çünkü finansal enstrümanların fiyatlarını finansal topluluktaki bireylerin görüşleri belirlemektedir.

Duygu analizinin temel motivasyonu diğer insanların düşüncelerini ortaya çıkarmaktır [8]. Literatürde, finansal piyasalar üzerinde mikro blog verilerini kullanarak gerçekleştirilen birçok duygu analizi araştırması gerçekleştirilmiştir. Bollen ve ark [4] Twitter servisi kullanıcılarına ait duygu polaritelerinin borsa davranışlarını tahmin etmekte kullanışlı olup olmadığını araştırmışlardır. Bir diğer çalışmada Rao ve Srivastava [11] çalışmalarında finansal market enstrümanları ile Twitter servisi kısa iletileri arasındaki ilişkiyi incelemişlerdir. Sprenger ve ark. [14] ise çalışmalarında finansal piyasalar ile birden fazla mikro blog servisi arasındaki ilişkiyi araştırmışlardır. Diğer bir önemli araştırmada Si ve ark. [12] Twitter servisi verilerini kullanarak borsa hareketlerini tahmin eden konu tabanlı bir duygu analizi yöntemi önermişlerdir. Smailović ve ark. [13] ise çalışmalarında hisse senedi piyasalarında kısa ileti tabanlı duygu analizi için akış tabanlı aktif öğrenme gerçekleştiren yeni bir yöntem önermişlerdir. Bir diğer duygu analizi araştırmasında Bing ve ark. [3] sosyal medyada duygu analizi kullanarak belirli şirketlerin borsada işlem gören hisse senedi fiyatlarını tahmin etmek istemişlerdir.

Finansal alanda yapılan duygu analizi çalışmaları artmasına rağmen, bu çalışmalar tanımlanmış bir sosyal ağın duygu polaritesini analiz etmek için sadece mikro blog servislerinden gönderilen kısa iletilere ait metinsel içerikleri göz önünde bulundurmaktadırlar. Böyle bir yaklaşımın tanımlanmış sosyal ağa ait duygu polaritesinin belirlenmesinde yetersiz olduğu varsayılmaktadır. Liu'dan esinlenerek bu çalışmada önerilen temel fikir, herhangi bir topluluğun gerçeğe yakın duygu polaritesini bulabilmek için kullanıcıların topluluk içerisindeki güven değerleri dikkate alınmalıdır. Bu nedenle bu çalışmada, mikro blog servisleri üzerinde analiz yapmak için yeni tanımlanan kullanıcı metrikleri kullanan özgün bir duygu analizi yöntemi önerilmektedir.

3 Yöntem

3.1 Veri Temizleme ve Dönüşümü

Finansal Sosyal Topluluk Paylaşımları. Toplanan Twitter servisi verisi duygu analizi için hazır değildir. Çalışmanın kapsamı doğrultusunda ilk önce Türkçe olmayan iletiler veri kümesinden çıkartılmaktadır. Bu adımdan sonra veriyi duygu analizi için daha anlamlı bir hale getirmek için bazı doğal dil işleme teknikleri uygulanmaktadır. Bu teknikler adım adım kısaca aşağıda ifade edilmektedir;

Cümleleri Ayırıştırma. Ana doğal dil işleme adıdır. Bu adımda girdi olarak alınan cümleler, kelimelere ve anlamlı ifadelere ayrıştırılmaktadır. Normal Türkçe metin ayrıştırma işlemlerine ek olarak mikro blog iletileri ayrıştırılırken bazı hususlar dikkate alınmalıdır. Bunlar; his simgeleri (":D", ":x", ":)"), Twitter işaret dili ("@", "#"), büyük harfle ve tekrarlı yazım ("HAYIR", "lüüütttfeenn").

Gövdeleme. Doğal dil işlemede önemli bir adımdır. Gövdeleme işleminin ana görevi kelimenin çekim eklerini temizlemektir. Böylece duygu analizi yönteminin başarı oranı artmaktadır. Sadece çekim eklerini temizlediği için kelime anlam kaybına uğramamaktadır.

Finansal Sosyal Topluluk Ağ Bilgisi. Çalışmada kullanılan finansal sosyal topluluk, Türkiye'de yayın yapan bir ekonomi kanalının sunucuları ve bu sunucuların ortak arkadaşlarından oluşmaktadır. Bu çalışmanın amacı, finansal sosyal topluluğun Türkçe kısa iletileri üzerinde özgün bir duygu analizi yöntemi uygulamaktır. Bu yüzden elde edilen veri kümesi içerisindeki kişilerin paylaşımlarına bakılarak Türkçe kelime paylaşımı %50'den düşük olan kişiler bu topluluktan çıkartılmaktadır. Bir kelimenin Türkçe olup olmadığını anlamak için güncel Türk Dil Kurumu [15] sözlüğü kullanılmaktadır. Bu veri hazırlama adımı dışında daha güvenilir sonuçlar elde edebilmek için tanımlanan topluluk içerisinde 3 haftadan az kalan kişiler topluluktan çıkartılmaktadır. Topluluk verisinin üzerinde yapılan temizleme adımlarından sonra üyelerin topluluk içerisindeki kronolojik yaşam sürelerini içeren zaman çizelgeleri elde edilmektedir.

3.2 Özellik Çıkarma ve Seçme

Duygu analizi yaklaşımı temelde bir sınıflandırma problemidir. Bu nedenle, veri kümesi üzerinde duygu analizi gerçekleştirmek için öncelikle metin özellikleri çıkartılmalıdır. Daha sonra bu özellikler içerisinde belirlenen kriterler doğrultusunda uygun olan metin özellikleri seçilmelidir. Bu özellikler daha çok terimlerin varlığı veya sıklığı şeklinde durumlar hakkında görüş belirten kelimelerden oluşmaktadır. Özellik çıkarma işlemi dokümanlara bir kelime kümesi şeklinde davranarak n-grams (unigram, bigram, trigram) yaklaşımı uygulamaktadır. Uygun n-grams tekniği veri kümelerinin karmaşıklık derecesine göre belirlenmektedir.

Literatürde özellik seçimi için istatistiksel yöntemler kullanılmaktadırlar. Bu yöntemlerden bazıları Point-wise Mutual Information [20], Chi-square [2] ve Latent Semantic Index [5] yaklaşımlarıdır. Point-wise Mutual Information (PMI) yöntemi özellikler ve sınıflar arasındaki ilişkiyi bulan temel bir yaklaşım sunmaktadır. Sınıf c ve kelime a arasındaki birlikte gözlemlenebilirlik $P_c(w)$, sınıf c ve kelime a arasındaki bağımsızlık P_c ile tanımlanmaktadır (denklem 1).

$$V_c(w) = \log\left(\frac{P_c(w)}{P_c}\right) \quad (1)$$

Bu çalışmada, özellik çıkartmak için unigram ve bigram özellik temsilleri birleştirilmektedir. Özellik seçmek için ise PMI yöntemi kullanılmaktadır.

3.3 Duygu Sınıflandırması

Duygu sınıflandırması yöntemi iki ana dala ayrılmaktadır; makine öğrenmesi tabanlı yaklaşım ve sözlük tabanlı yaklaşım. Makine öğrenme tabanlı yaklaşım sınıflandırma sırasında makine öğrenme algoritmalarını ve dilbilimsel özellikleri kullanmaktadır. Sözlük tabanlı yaklaşım ise sınıflandırma sırasında önceden hazırlanmış duygu kavramlarından oluşan sözlüklerden yararlanmaktadır. Bazı çalışmalarda ise bu iki yönelimin melez bir yaklaşımı kullanılmaktadır.

Literatürde, duygu analizi için bazı makina öğrenmesi tabanlı sınıflandırma yöntemleri önerilmektedir. Ancak, duygu analizi için hangi yöntemin daha uygun olduğu tam olarak açık değildir [6]. Destek Vektör Makinası (DVM) yöntemi duygu analizi için kullanılan yöntemlerden biridir [9], [10]. Çalışmada DVM yöntemi duygu sınıflandırmasının karar mekanizmasında kullanılmaktadır. Çünkü DVM'in performansı diğer duygu analizi çalışmalarına benzer değerlere sahiptir ve ayrıca DVM metin madenciliği yaklaşımları için ideal bir çözüm sunmaktadır [9], [10].

3.4 Önerilen Duygu Analizi Yöntemi

Duygu analizinin amacı verilen metnin polaritesini sınıflandırmaktır. Mikro blog perspektifinde, duygu analizi sınıflandırması için kullanıcılarının kısa metin şeklindeki görüşleri kullanılmaktadır. Liu [7] çalışmasında görüşü beşli şekilde tanımlamaktadır (denklem 2). Bu denklemde e görüşü ifade edilen kavramı, a ise bu kavrama ait bir özelliği nitelemektedir. s hedef (e, a) için düşünülen duygu değerini, h görüşü ifade eden bireyi ve t ise ifadenin edildiği zamanı belirtmektedir.

$$Opinion(e, a, s, h, t) \quad (2)$$

Liu bir görüşün açık bir şekilde ifade edilebilmesi için bu beş parametrenin gerekli olduğunu belirtmektedir. Bu nedenle, başarılı bir duygu analizi gerçekleştirilebilmesi için bu beş kavramın görüş için kullanılması gerekmektedir.

Literatürde, finansal piyasalar ve sosyal ağların duygu polaritesi arasındaki korelasyonu bulmayı hedefleyen birçok çalışma gerçekleştirilmiştir. Bu çalışmaların birçoğu [1], [3], [4], [11], [12], [13], [14] sadece sosyal medya aracılığı ile paylaşılan zaman damgasına sahip metinsel içeriği kullanarak sosyal ağın duygu polaritesine karar vermektedir. Metinsel içerik duygu polaritesini hesaplamak için, zaman değeri ise finansal enstrümanlar arasındaki korelasyonu bulmak için kullanılmaktadır. Araştırmacılar çalışmalarında sadece duygu (s), onun hedef aldığı kavramı (e, a) ve zaman değerini (t) göz önünde bulundurmaktadırlar. Görüş sahibi birey (h) duygu polaritesinin hesaplanmasında sağlıklı bir şekilde ele alınmamaktadır.

Bu bağlamda bu çalışma, bir konu üzerinde belirtilen görüşün sahip olduğu polarite değerinin, görüş sahibinin ait olduğu topluluğun ona duyduğu güven doğrultusunda değişeceğini iddia etmektedir. Bu iddia doğrultusunda toplulukların duygu polaritesini ölçümleyebilmek için görüş sahiplerini hesaplamaya dâhil eden özgün bir duygu analiz yaklaşımı önerilmektedir. Bu amaçla, mikro blog kullanıcıları için iki tane metrik tanımlanmaktadır. Bu metriklerin tanımlanma amacı kullanıcıların sosyal ağlardaki önemini ölçmek içindir. Bu metrikler; *Üyelik Derecesi (ü)* ve *İlgi Derecesi*

(*id*) olarak adlandırılmaktadır. *Üyelik Derecesi* kullanıcının grup içerisindeki inandırıcılığını ölçmektedir. *İlgi Derecesi* ise kullanıcının bir konu ile alakalı ilgi derecesini ölçmektedir.

Üyelik Derecesi. Bu çalışmada konu tabanlı mikro blog kullanıcılarının oluşturduğu sosyal ağ bir yönsüz çizelge olarak ifade edilmektedir (U, L). Bu ifadeye mikro blog kullanıcıları kümesi U ve bu kullanıcılar arasındaki arkadaşlık ilişkisi kümesi L olarak gösterilmektedir. Arkadaşlık ilişkisi l birbirini takip eden iki mikro blog kullanıcısı arasındaki ilişkiyi belirtmektedir. Bunlara ek olarak, H çalışma için belirlenen ekonomi kanalı sunucularının oluşturduğu kümeyi ve F ise çalışmada önerilen finansal ilgiye sahip kullanıcıların oluşturduğu sosyal ağı nitelemektedir. Bir mikro blog kullanıcısı olan x , F kümesine katılabilmesi için H kümesine ait iki kullanıcı h_1 ve h_2 ile arasında l_1 ve l_2 arkadaşlık ilişkisi olması gerekmektedir. Önerilen finansal sosyal ağa katılım tanımı denklem 3 ve 4'te ifade edilmektedir.

$$\forall x \in U, h_1, h_2 \in H, l_1, l_2 \in L; \quad (3)$$

$$\forall x \{l_1 = \{(x, h_1) = (h_1, x)\} \wedge l_2 = \{(x, h_2) = (h_2, x)\} \rightarrow x \in F\} \quad (4)$$

Bir kullanıcının gruba katılma zamanı t_{bm} ile gruptan ayrılma zamanı t_{lm} arasında geçen üyelik süresini t_m olarak tanımlanmaktadır. Ayrıca bütün sosyal ağa ait üyelerin ortalama üyelik süresini t_{m_ort} ile ifade edilmektedir. Üyelik Derecesi (*ü**d*), kullanıcının üyelik süresinin t_m , ait olduğu sosyal ağın üyelik süresi ortalamasına t_{m_ort} oranıdır. Bu tanımlar denklem 5, 6 ve 7'de gösterilmektedir. Çalışmada kullanıcının üye kalma süresi ile kullanıcının inandırıcılığı doğru orantılı düşünülmektedir. Bu nedenle üyelik zamanı uzun olan kullanıcıların görüşleri daha güvenilir bir kaynak olarak ele alınmaktadır.

$$t_m = t_{bm} - t_{lm} \quad (5)$$

$$t_{m_ort} = \frac{\sum_{i=1}^{|F|} t_{m_i}}{|F|} \quad (6)$$

$$\ddot{u}d = \frac{t_m}{t_{m_ort}} \quad (7)$$

İlgi Derecesi. Finans literatüründe, finans ile ilgili isim tamlamaları, kelimeler ve kısaltmalardan oluşan finans sözlükleri araştırmalarda kullanılmaktadır. Bu çalışmada, akademik çalışmalarda kullanılmış Türkçe finans sözlüklerinin birleştirilmesinden oluşan FW olarak ifade edilen bir sözlük kullanılmaktadır. Bu sözlük 1953 tane finans ile alakalı terim içermektedir, $FW = \{w_1, w_2, w_3, w_4, \dots\}$. Birleştirilen finans sözlüğü mikro blog kullanıcılarının ne kadar finans ile ilgili paylaşım yaptığını belirlemek için kullanılmaktadır.

İlgi Derecesi (*id*) bir kullanıcının ne kadar finans ile alakalı paylaşım yaptığını göstererek aslında o kullanıcının konu hakkındaki uzmanlık derecesi hakkında ipucu vermektedir. İlgi derecesini hesaplarken ilk önce kullanıcının mikro blog iletileri kelime kümesine dönüştürülmektedir; $UTW = \{w_1, w_2, w_3, w_4, \dots\}$. Daha sonra hem

FW kümesinin hem de UTW kümesinin üyesi olan elemanlar için UFW kümesi oluşturulmaktadır. Kesişim kümesi, kullanıcıların kullandığı finans ile alakalı kelimeleri içermektedir, $UFW = \{w_1, w_2, w_3, w_4, \dots\}$. Kümeler arasındaki ilişki denklem 8’de gösterilmektedir.

$$\forall x \{x \in FW, x \in UTW; \rightarrow x \in UFW\}, FW \cap UTW = UFW \quad (8)$$

Bir kullanıcının tanımlanan finansal kelime kullanma yüzdesi ilgi değeri iv olarak tanımlanmaktadır. Ayrıca bütün F sosyal ağına ait kullanıcıların ortalama ilgi değeri iv_{ort} olarak ifade edilmektedir. İlgi Derecesi (id), kullanıcının ilgi değerinin iv ait olduğu sosyal ağın ortalama ilgi değerine iv_{ort} oranıdır. Bu tanımlar denklem 9, 10 ve 11’de gösterilmektedir

$$iv = \frac{|UFW|}{|UTW|} \quad (9)$$

$$iv_{ort} = \frac{\sum_{k=1}^{|F|} iv_k}{|F|} \quad (10)$$

$$id = \frac{iv}{iv_{ort}} \quad (11)$$

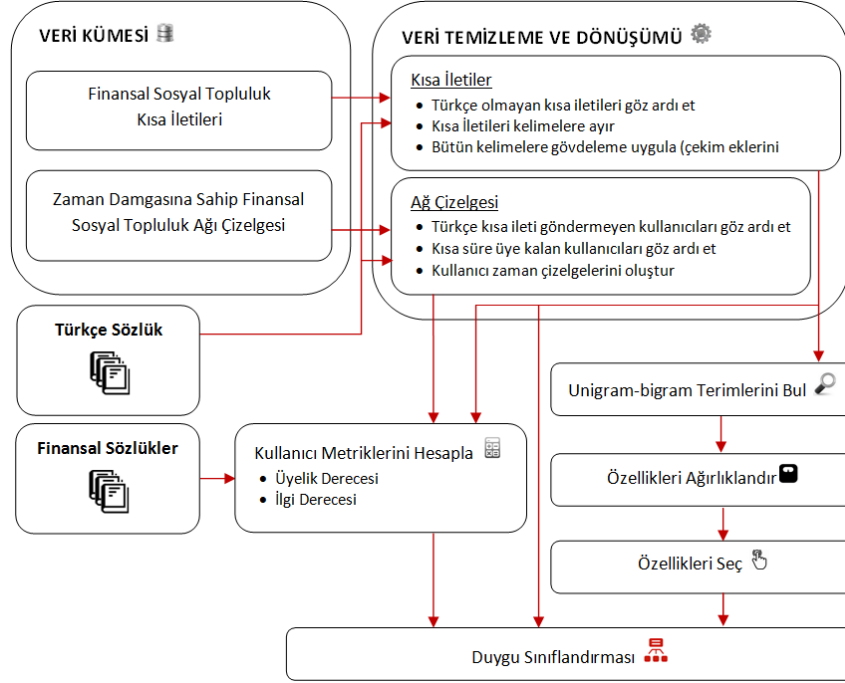
Önerilen özgün duygu analizi yönteminde, her kullanıcı için finansal alan ile ilgili duygu polaritesi ölçümünün başarımını artırmak için $üid$ ve id olarak adlandırılan iki metriğin belirlenmesi gerekmektedir.

Önceki finans alanı üzerinde yapılan duygu analizi çalışmalarında, topluluğun duygu polaritesi hesaplanırken kullanıcıların o konuya olan ilgi seviyesi ve grup içindeki inandırıcılık seviyesi göz önünde bulundurulmamaktadır. Araştırmacılar sadece mikro blog içeriğini hesaplamaya dâhil etmektedirler. Bir başka ifadeyle aslında araştırmacılar gerçekleştirilen çalışmalarında kullanıcı kavramının yani kısa ileti sahibinin katsayı değerini 1 almaktadırlar. Önceki çalışmalarda toplulukların duygu polaritesini (tdp) hesaplamak için kullanılan yöntem denklem 12’de gösterilmektedir. SP kümesi her kısa ileti için duygu polaritesini içeren küme olarak ifade edilmektedir, $SP = \{sp_1, sp_2, sp_3, sp_4, \dots\}$.

$$tdp = \frac{\sum_{i=1}^{|SP|} sp_i}{|SP|} = \frac{\sum_{i=1}^{|SP|} sp_i * 1}{|SP|} \quad (12)$$

Bu çalışmada, bir kullanıcının bir konu üzerindeki alaka düzeyinin ve dâhil olduğu gruptaki inandırıcılık düzeyinin duygu analizindeki değerlendirme sürecine dâhil edilmesi gerektiği iddia edilmektedir. Yukarıda tanımlanan metrikler sayesinde, topluluğun duygu polaritesi hesaplamasında her bir kullanıcı için $üid$ ve id değerleri eklenerek klasik duygu polaritesi hesaplama yaklaşımları geliştirilmektedir. Denklem 12’de ifade edilen hesaplama yöntemi her kısa ileti için kısa ileti sahibinin $üid_{ks}$ ve id_{ks} değerleri eklenerek denklem 13’deki ifadeye dönüştürülmektedir. Önerilen yaklaşımdaki adımlar blok diyagram olarak Şekil 1’de tasvir edilmektedir.

$$tdp = \frac{\sum_{i=1}^{|SP|} sp_i * üid_{ks} * id_{ks}}{|SP|} \quad (13)$$



Şekil 1. Önerilen Duygu Analizi Yöntemi

4 DeneySEL ÇALIŞMA

Önerilen yöntemi değerlendirmek için, yukarıda tanımlanan kullanıcı metriklerini kullanan özgün bir duygu analizi uygulaması geliştirilmiştir. Uygulamada kullanılacak modeli oluşturabilmek için Twitter4j kütüphanesi [17] yardımıyla Türkçe kısa iletiler Twitter servisinden elde edilmiştir. Elde edilen kısa iletiler, finans konusunda uzman üç kişi tarafından etiketlenmiştir. Etiketlenen kısa iletilerin çalışmaya dâhil edilebilmesi için üç kişinin onayı alınmıştır. Eğer bu üç kişi aynı kısa ileti için finans ile alakalı olduğu ve aynı polariteye sahip olduğu ortak kararını vermiş ise o kısa ileti çalışmada kullanılmıştır. Bu etiketleme yöntemi ile 1501 negatif kısa ileti ile 907 pozitif kısa ileti içeren bir Türkçe veri kümesi oluşturulmuştur.

Özellikleri çıkartmak için unigram ve bigram özellik temsilleri birleştirilmiştir. Özellik seçmek için ise PMI yöntemi kullanılmıştır. Özellik çıkarma ve seçme işlemlerinden sonra 35,030 özellikten oluşan özellik kümesi elde edilmiştir. Bu çalışmada, Destek Vektör Makinası tabanlı bir duygu sınıflandırma yöntemi seçilmiştir. Etiketlenmiş veri kümesinin sınıflandırma performansı artırmak için 10-katlamalı çapraz doğrulama yöntemi kullanılmıştır. Yöntem sonucunda elde edilen duygu sınıflandırma modelinin başarı oranı %73,63'dir.

Çalışmada kullanılan veri Twitter servisinden elde edilmiştir. Twitter sosyal ağı içerisindeki finans ile alakalı bireyleri keşfedilmek için bir yaklaşım belirlenmiştir. İlk olarak, Türkiye'de yayın yapan bir ekonomi kanalının spikerleri belirlenmiştir. Sonra

bu belirlenen spikerler ile çift yönlü izleme ilişkisine sahip bireyler belirlenmiştir. Bu belirlenen grup içerisinde eğer iki spiker aynı kişi ile çift yönlü izleme ilişkisine sahipse, bu bireyin de finans ile ilgili olduğu varsayılmıştır. Bu öngörülü yaklaşım sayesinde finans hakkında konuşan bir sosyal ağ topluluğu oluşturulmuştur. Algoritma 1 finans topluluğu oluşturmak için kullanılan yöntemin çalışma adımlarını göstermektedir. FINANSALSOSYALTOPLULUKHAZIRLA prosedürünün zaman karmaşıklığı $O(k*m) = O(n^2)$; m bu prosedürde sunucuların ortak arkadaş sayısını, k ise sunucuların ortak arkadaşlarının arkadaşlarını tanımlamaktadır.

Algoritma 1: Finansal Sosyal Topluluğu Hazırlama

```

1: procedure FINANSALSOSYALTOPLULUKHAZIRLA(sunucular)
2:   finansalSosyalToplulukÜyeleriİlişkileri ← []
3:   arkadaşlar ← []
4:   for all sunucu ∈ sunucular do
5:     arkadaşlar.Ekle(ÇİFTYÖNLÜİLİŞKILERESAHİPKULLANICILARIBUL(sunucu))
6:   end for
7:   ortakArkadaşlar ← ORTAKARKADAŞBUL(arkadaşlar,2) // İki'den fazla arkadaşlar'a eklenen kişiler
8:   for all ortakArkadaş ∈ ortakArkadaşlar do
9:     finansalSosyalToplulukÜyeleriİlişkileri.Ekle(ÇİFTYÖNLÜİLİŞKILERIBUL(ortakArkadaş))
10:  end for
11:  return finansalSosyalToplulukÜyeleriİlişkileri
12: end procedure

```

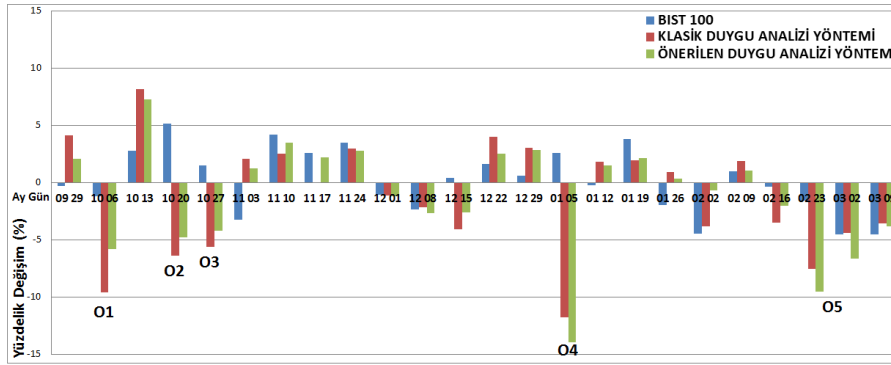
Çalışmada kullanılan verileri Twitter servisinden elde etmek için Twitter4j kütüphanesi [17] kullanan bir araç geliştirilmiştir. Bu araç finansal sosyal ağın o anki görüntüsünü alabilmek için her Pazartesi çalıştırılmıştır. Test verisi bu çalışma için 22 Eylül 2014 ile 14 Mart 2015 tarihleri arasında haftalık toplanmıştır. Veri kümesi, 475 kullanıcının topluluk içerisindeki zaman çizelgesini yani hangi haftalarda topluluğun üyesi kalmış bilgisi ve bu kullanıcılara ait 619091 tane kısa iletiyi içermektedir.

Her kullanıcı için haftalık pozitif ve negatif kısa ileti sayısını Weka kütüphanesi [18] kullanılarak geliştirilen bir araç yardımıyla elde edilmiştir. Daha sonra, bir kullanıcının bir konu hakkındaki ilgisinin ve topluluk içerisindeki inandırıcılığının duygu polaritesine nasıl etki ettiği gözlemlenmek için kullanıcının *ü*d ve *id* değerleri hesaplanıp, normal duygu analizi sonucunda çıkan polarite sayıları ile çarpılmıştır. Böylece o kullanıcıya ait ağırlıklandırılmış polarite değeri bulunmuştur.

Haftalık karşılaştırma sırasında yüzdesel değişimi izleyebilmek için, BIST 100 kapanış ve açılış değerleri Türkiye Merkez Bankası internet sitesinden [16] elde edilmiştir. Diğer taraftan BIST 100 değerleri ile karşılaştırma yapabilmek için, normal (klasik hesaplama yöntemi) ve ağırlıklandırılmış (önerilen hesaplama yöntemi) duygu polaritesinin haftalık değişimi (*dphd*) hesaplanmıştır. Duygu polaritesinin haftalık yüzdesel değişimi, önceki hafta duygu polarite değeri ve mevcut haftadaki duygu polarite değeri arasındaki fark alınıp önceki hafta duygu polarite değerine bölerek hesaplanmıştır. Böylelikle haftalık duygu polariteleri yüzdesel değişimi elde edilmiştir. Bu hesaplama denklem 14'te gösterilmektedir.

$$dphd = \frac{\text{haftalık duygu değeri}_{i+1} - \text{haftalık duygu değeri}_i}{\text{haftalık duygu değeri}_i} \quad (14)$$

Şekil 2’de 29 Eylül 2014 ile 14 Mart 2015 tarihleri arasında, klasik yöntem ile hesaplanan duygu polarite değeri, önerilen yöntem ile hesaplanan duygu polarite değeri ve BIST 100 haftalık değişim değeri görülmektedir. Şekil 2 incelendiği zaman, çoğu hafta topluluğun duygu polaritesi ile BIST 100 arasında doğrusal bir ilişki olduğu açıkça fark edilmektedir. Ancak bu değerler incelendiği zaman bazı haftalarda bu ilişkinin bozulduğu dikkat çekmektedir. Bu haftalar araştırıldığında sosyal topluluk üzerinde büyük etki yaratan olağandışı olayların bu haftalar içerisinde gerçekleştiği görülmektedir. Bu olağandışı olayları listesi Tablo 1’de verilmektedir.



Şekil 2. 29 Eylül 2014 ile 14 Mart 2015 Tarihleri Arası Değişim Yüzdeleri

Tablo 1. Olağandışı Olaylar

Gün (Y A G)	Olağandışı Olaylar
2014 10 06	Kobani Protestoları (O1)
2014 10 20	Hakkâri’de Silahlı Çatışma (O2)
2014 10 27	Ermenek Maden Kazası (O3)
2015 01 05	Petrol Fiyatlarının \$50’ın Altına Düşmesi (O4)
2015 02 23- 2015 03 02	Merkez Bankası Faiz Kararı Tartışmaları (O5)

Şekil 2’de açıkça görüldüğü gibi önerilen genişletilmiş duygu analizi yönteminin yüzdesel değişim oranları, klasik duygu analizi yönteminin yüzdesel değişim oranlarına göre BIST 100’un haftalık yüzdesel değişim oranlarına daha yakın seyretmektedir. Buna ek olarak, ayrıca eğer olağandışı olay ekonomi ile alakalı ise önerilen yöntemle ait sonuçların klasik yöntemle ait sonuçlara göre daha çok etkilendiği görülmektedir. Önerilerimizi daha netleştirmek ve haklı çıkarmak için, iki değer serisi arasındaki doğrusal ilişki ölçümünde sıklıkla kullanılan Pearson korelasyon katsayı [19] yöntemi kullanılmıştır (denklem 15). Pearson korelasyon katsayı "r" değeri -1 ile 1 arasında değişmektedir. Yöntemi kullanmadan önce verinin normal dağılımı "Shapiro-Wilk" testi ile sınanmıştır. Test sonucunda üç grubun da önem olasılık değerleri (sig.) 0.05' den büyük olduğundan tüm gruplar için %95 güvenle veriler normal dağılımlı olduğu gözlemlenmiştir ve Pearson korelasyon katsayısı yöntemine uygun olduğu görülmüştür.

$$PCC(X,Y) = r = \frac{\Sigma XY - (\Sigma X)(\Sigma Y)/n}{\sqrt{(\Sigma X^2 - (\Sigma X)^2/n)(\Sigma Y^2 - (\Sigma Y)^2/n)}} \quad (15)$$

BIST 100 ile topluluğun duygu polaritesi arasındaki güçlü doğrusal ilişkiyi bulmak ve önerilen duygu analizi yönteminin başarımını ölçmek için Pearson korelasyon katsayı yöntemi ile dört farklı analiz yapılmıştır.

1. Bütün veri üzerinde klasik duygu analizi yöntemi (BV-KDAY).
2. Bütün veri üzerinde önerilen duygu analizi yöntemi (BV-ÖDAY).
3. Olağandışı olayların gerçekleştiği haftalar çıkarılmış veri üzerinde klasik duygu analizi yöntemi (OO-KDAY).
4. Olağandışı olayların gerçekleştiği haftalar çıkarılmış veri üzerinde önerilen duygu analizi yöntemi (OO-ÖDAY)

Analizler sonucunda elde edilen Pearson korelasyon katsayı değerlerinin gösterdiği üzere önerilen duygu analizi yöntemiyle, haftalık borsa fiyat değişimleri ile haftalık finansal sosyal topluluk duygu polaritesi değişimleri arasında önceki yaklaşımlara göre daha yakın bir doğrusal ilişki elde edilmiştir. Analiz sonuçları Tablo 2'de gösterilmektedir.

Tablo 2. Pearson Korelasyon Katsayı Sonuçları

Analiz Tipi	Pearson Korelasyon Katsayı
BV-KDAY	0,1986
BV-ÖDAY	0,2283
OO-KDAY	0,5701
OO-ÖDAY	0,6662

5 Sonuç

Literatürde gerçekleştirilen araştırmalarda finans ve duygu analizi ilişkisi çokça ele alınmasına rağmen sosyal ağ kullanıcıları yeterince göz önünde bulundurulmamaktadır. Bu çalışmada, finans ile alakalı kullanıcı verileri göz önünde bulundurularak o kullanıcının finans ile alakalı ilgi düzeyini ve finansal topluluk içerisindeki inandırıcılığını kullanan, yeni bir topluluk düzeyi duygu analizi yöntemi önerilmektedir. Yapılan analizler sonucu önerilen yöntemin önceki yöntemlere göre finansal sosyal toplulukların duygu polaritesi ile borsa fiyatları arasındaki bağıdaklık oranını daha hassas hesapladığı görülmektedir. Bu nedenle bir topluluğun duygu polaritesi hesaplamasında önerilen yöntemin önceki yaklaşımlardan daha etkin olduğu düşünülmektedir. Son olarak, gelecekte yapılacak çalışmalarda farklı sınıflandırma yöntemleri ve farklı veri kümeleri kullanılarak önerilen yaklaşımın daha detaylı ele alınması planlanmaktadır.

Kaynaklar

1. Agarwal, A., Xie, B., Vovsha, I., Rambow, O., Passonneau, R.: Sentiment analysis of twitter data. In: Proceedings of the Workshop on Languages in Social Media. Association for Computational Linguistics, pp. 30-38 (2011)
2. Aggarwal, C. C., Zhai, C.: Mining text data. Springer Science & Business Media (2012)
3. Bing, L., Chan, K.C.C., Ou, C.: Public Sentiment Analysis in Twitter Data for Prediction of a Company's Stock Price Movements. In: e-Business Engineering (ICEBE), 2014 IEEE 11th International Conference, pp.232-239 (2014)
4. Bollen, J., Mao, H., Zeng, X.: Twitter mood predicts the stock market. Journal of Computational Science, vol. 2, no. 1, pp. 1-8 (2011)
5. Deerwester, S. C., Dumais, S. T., Landauer, T. K., Furnas, G. W., Harshman, R. A.: Indexing by latent semantic analysis. JASIS, vol. 41, no. 6, pp. 391-407 (1990)
6. Go, A., Bhayani, R., Huang, L.: Twitter sentiment classification using distant supervision, CS224N Project Report, Stanford, pp. 1-12 (2009)
7. Liu, B.: Sentiment Analysis and Opinion Mining, Morgan & Claypool Publishers, (2012)
8. Pang, B., Lee, L.: Opinion mining and sentiment analysis. Foundations and Trends in Information Retrieval, vol. 2, no. 1-2, pp. 1-135 (2008)
9. Pang, B., Lee, L., Vaithyanathan, S.: Thumbs up?: sentiment classification using machine learning techniques. In: Proceedings of the ACL-02 Conference on Empirical Methods in Natural Language Processing, pp. 79-86 (2002)
10. Phientrakul, T., Kijirikul, B., Takamura, H., Okumura, M.: Sentiment classification with support vector machines and multiple kernel functions'. In: Proceedings of the 2009 International Conference on Neural Information Processing, pp. 583-592 (2009)
11. Rao, T., Srivastava S.: Analyzing stock market movements using twitter sentiment analysis. In: Proceedings of the 2012 International Conference on Advances in Social Networks Analysis and Mining, pp. 119-123, Washington, DC, USA (2012)
12. Si, J., Mukherjee, A., Liu, B., Li, Q., Li, H., Deng, X.: Exploiting Topic based Twitter Sentiment for Stock Prediction. In: Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics, Association for Computational Linguistics, pp. 24-29 (2013)
13. Smailovic, J., Grcar, M., Lavrac, N., Znidarsic, M.: Stream-based active learning for sentiment analysis in the financial domain. Information Sciences, vol. 285, pp. 181-203 (2014)
14. Sprenger, T. O., Tumasjan, A., Sandner, P. G., Welp, I. M.: Tweets and trades: The information content of stock microblogs. European Financial Management, vol. 20, no. 5, pp. 926-957 (2014)
15. Türk Dil Kurumu, <http://www.tdk.gov.tr>
16. Türkiye Merkez Bankası, <http://evds.tcmb.gov.tr>
17. Twitter4j, <http://twitter4j.org/en/index.html>
18. Waikato Üniversitesi Makine Öğrenme Grubu, <http://www.cs.waikato.ac.nz/ml/weka/>
19. Wright Gillham, N. : A Life of Sir Francis Galton: From African Exploration to the Birth of Eugenics, Oxford University Press (2001)
20. Yu, L. C., Wu, J. L., Chang, P. C., Chu, H. S.: Using a contextual entropy model to expand emotion words and their intensity for the sentiment classification of stock market news. Knowledge-Based Systems, vol. 41, pp. 89-97 (2013)