



Financial sentiment analysis with Deep Ensemble Models (DEMs) for stock market prediction

Zeynep Hilal Kilimci

Department of Computer Engineering, Dogus University, Istanbul, 34722, Turkey

Department of Information Systems Engineering, Kocaeli University, Kocaeli, 41001, Turkey

Highlights:

- High performance sentiment analysis
- Identification of deep ensemble models for stock market prediction
- Evaluation of word embedding based document representation techniques

Keywords:

- Stock market prediction
- Deep ensemble model
- Sentiment analysis

Article Info:

Research Article

Received: 24.12.2018

Accepted: 10.04.2019

DOI:

10.17341/gazimmfd.501551

Correspondence:

Author: Zeynep Hilal

Kilimci

e-mail:

hkilimci@dogus.edu.tr

zhilal.kilimci@kocaeli.edu.tr

phone: +90 216 444 79 97 / 1232

Graphical/Tabular Abstract

In this study, it is proposed to estimate the direction of Borsa Istanbul 100 Index by using financial sentiment analysis. However, the performance of classification in social networks, which face the problem of size limitation in expressing user opinions such as Twitter, is significantly affected. Therefore, it is also proposed to enrich the datasets with various techniques from a semantic perspective and improve the classification performance of system by blending ensemble learning approach with deep learning algorithms. To the best of our knowledge, this is the first study in the literature in terms of both using the Twitter environment as a news source and forecasting the stock market direction with deep ensemble models. As a result, the extensive experiment results demonstrate that the proposed model is significantly outperforms the-state-of-art studies.

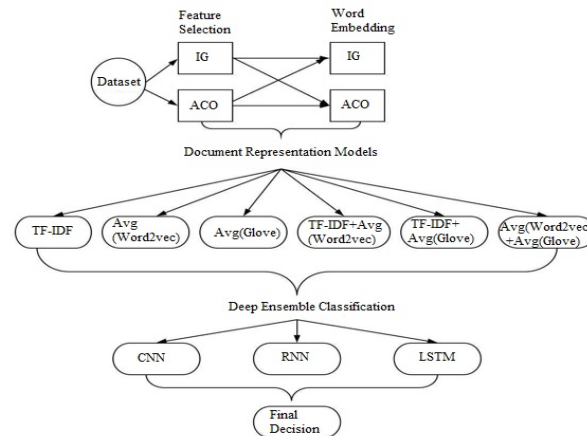


Figure A. Flowchart of the proposed system

Purpose: The aim of this study is to achieve a high classification performance for prediction of the stock market direction by utilizing financial sentiment analysis, deep ensemble models, and word embedding based document representation techniques on both English and Turkish datasets.

Theory and Methods:

It is aimed to get higher classification results by employing Avg(Word2vec), Avg(Glove), Avg(Word2vec)+Avg(Glove), TF-IDF+Avg(Word2vec), TF-IDF+Avg(Glove) models as document representation techniques, Word2vec and Glove as word embedding based models, and Convolutional Neural Networks, Recurrent Neural Networks, and Long Short Term Memory Networks as deep learning algorithms for ensemble approach which are unattempted before for prediction of the stock market direction.

Results:

Avg(Word2vec)+Avg(Glove) is the best document representation technique to get higher classification results for stock market estimation. Avg(Word2vec)+Avg(Glove) has also superior classification success for both individual deep learning algorithms and deep ensemble approach. Among the individual deep learners, the ideal classifier is LSTM algorithm with 78.42% accuracy value in Turkish and 79.46% accuracy result in English for stock market prediction. In deep ensemble models, stacking as a method of combining ensemble decisions outperforms both individual classifiers and majority voting.

Conclusion:

The combination of word embedding based document representation techniques and deep ensemble models exhibits a higher classification success compared to similar studies with 78.07% for Turkish dataset and 78.92% for English dataset. Instead of employing financial news, the usage of Twitter as a data source for BIST100 provides remarkable contribution to this study when high classification results are considered.



Borsa tahmini için Derin Topluluk Modelleri (DTM) ile finansal duygu analizi

Zeynep Hilal Kilimci*^{ID}

Doğuş Üniversitesi, Mühendislik Fakültesi, Bilgisayar Mühendisliği Bölümü, Kadıköy-İstanbul, 34722, Türkiye
Kocaeli Üniversitesi, Bilişim Sistemleri Mühendisliği Bölümü, Kocaeli, 41001, Türkiye

Ö N E Ç İ K A N L A R

- Yüksek performanslı duygu analizi
- Borsa yönü tahmini için derin topluluk modellerinin belirlenmesi
- Kelime gömme tabanlı belge gösterim tekniklerinin değerlendirilmesi

Makale Bilgileri

Araştırma Makalesi
Geliş: 24.12.2018
Kabul: 10.04.2019

DOI:

10.17341/gazimmfd.501551

Anahtar Kelimeler:

Borsa tahmini,
derin topluluk modeli,
duygu analizi

ÖZET

Borsa tahmini, analistler ve yatırımcılar için aktif bir araştırma konusu olmuştur. Bu çalışmada, finansal duygu analizi yapılarak Bist100 endeksinin yönünün tahminlenmesi amaçlanmıştır. Bildiğimiz kadarıyla bu çalışma, borsa yönü tahminlemede haber kaynağı olarak Twitter ortamını kullanması ve bunun derin topluluk modelleriyle yapılması açısından literatürdeki ilk çalışmadır. Bu çalışmanın literatüre katkısı dört aşamada özetlenebilir: Birincisi, Twitter ortamındaki boyut sınırlaması problemini ortadan kaldırmak amacıyla özellik kümesi anlamsal olarak zenginleştirilmiştir. İlk aşamada, veri kümesini ifade edebilecek anlamlı özellikler, bilgi kazanımı ve karınca kolonisi optimizasyonu yöntemleriyle seçilmiştir. Sonra, bu özelliklere veri kümesini anlam, bağlam, söz dizimi açısından ifade edebilecek Avg(Word2vec), Avg(Glove), Avg(Word2vec)+Avg(Glove), TF-IDF+Avg(Word2vec), TF-IDF+Avg(Glove) gibi farklı doküman gösterim teknikleri uygulanmıştır. İkincisi, sınıflandırmayı bir algoritmayla gerçekleştirmek üzere birden fazla öğrenme algoritmalarıyla yaparak sistem performansının iyileştirilmesi amaçlanmıştır. Burada, geleneksel sınıflandırma algoritmaları yerine Evrimsel Sinir Ağları, Tekrarlayan Sinir Ağları, Uzun Kısa Vadeli Hafıza Ağları gibi derin öğrenme mimarilerinin harmanlanmasıyla derin topluluk modeli (DTM) oluşturulmuştur. Üçüncüsü, derin topluluk modelinin nihai kararını elde etmek için çoğunluk oylaması ve yığıtma yöntemleri kullanılmıştır. Dördüncü olarak önerilen yaklaşımın sınıflandırma performansını iyileştirdiğini kanıtlamak amacıyla herkesin kullanımına açık Türkçe ve İngilizce Twitter veri kümeleri kullanılmıştır. Sonuç olarak, deney sonuçları önerilen modelin literatür çalışmalarlarıyla kıyaslandığında önceki çalışmalardan önemli ölçüde üstün olduğunu göstermektedir.

Financial sentiment analysis with Deep Ensemble Models (DEMs) for stock market prediction

H I G H L I G H T S

- Modeling high performance sentiment analysis
- Identification of deep ensemble models for the direction of stock market prediction
- Evaluation of word embedding based document representation techniques

Article Info

Research Article
Received: 24.12.2018
Accepted: 10.04.2019

DOI:

10.17341/gazimmfd.501551

Keywords:

Stock market prediction,
deep ensemble model,
sentiment analysis

ABSTRACT

The stock market forecasting is popular research topic for analysts. In this study, it is proposed to estimate direction of Bist100 index by financial sentiment analysis. To our knowledge, this is the first study in literature using Twitter for forecasting stock market direction and doing this with deep ensemble models. The contributions of study are fourfold: First, feature set is enriched semantically to eliminate size limitation problem in Twitter. In first stage, meaningful features that express dataset are selected by means of information gain and ant colony optimization. Next, features are enriched in meaning, context, syntax using document representation models such as Avg(Word2vec), Avg(Glove), Avg(Word2vec)+Avg(Glove), TF-IDF+Avg(Word2vec), TF-IDF+Avg(Glove). Secondly, it is proposed to improve system performance performing classification with multiple learning algorithms. Instead of traditional classification algorithms, a deep ensemble model (DTM) is constructed blending deep learning architectures such as Convolutional Neural Networks, Recurrent Neural Networks, Long Short-Term Memory Networks. Third, majority voting and stacking methods are used to obtain final decision of deep ensemble model. Fourthly, Turkish and English Twitter datasets are employed to demonstrate that proposed approach improves classification performance. Consequently, experimental results show that proposed model is significantly superior to previous studies when compared with literature studies.

*Sorumlu Yazar/Corresponding Author: hkilimci@dogus.edu.tr, zhilal.kilimci@kocaeli.edu.tr / Tel: +90 216 444 79 97/1232

1. GİRİŞ (INTRODUCTION)

Sosyal medya büyük miktarda veriyi analiz etmek ve internette yer alan birçok konuda fikirleri tespit etmek için kullanılan çok popüler bir kaynak haline gelmiştir. Bilinen sosyal medya platformlarından biri olan Twitter, 100 milyona kadar aktif kullanıcının fikirlerini ifade etmesi için tercih edilmektedir. Bu, Twitter ortamının pazar dinamikleri için etkili olabilecek değerli bilgiler içerdiği anlamına gelmektedir. Bu nedenle duygu analizi, kullanıcı taleplerini olumlu ve olumsuz yönler açısından anlamak için önemlidir. Kurumlara, çeşitli sosyal medya sitelerini gerçek zamanlı olarak gözlemlene ve ardından buna göre hareket etme yeteneği kazandırır.

Hisse senedi fiyat endeksinin tahmini, zaman serisi tahminlerinin en zorlu uygulamalarından biri olarak görülmektedir. Hisse senedi fiyat endeksinin tahmin etme konularını ele alan birçok çalışma olmasına rağmen çoğu, gelişmiş finansal piyasalarla ilişkilendirilmiştir. Özellikle, Türk borsaları gibi gelişmekte olan piyasalarda hisse senedi fiyat endeksinin yönünü tahmin etmek için literatürde az sayıda araştırma bulunmaktadır. Hisse senedi fiyat endeksleri ile yapılan tutarlı tahminler, etkin piyasa ticareti stratejilerinin geliştirilmesi bakımından önem arz etmektedir [1]. Böylelikle yatırımcılar, potansiyel piyasa risklerine ve spekülörlere karşı koruma sağlayabilirler. Bunun yanı sıra, hisse senedi endeksinde işlem yaparak kar elde etme fırsatı yakalayabilirler [2]. Borsa esas itibarıyla dinamik, doğrusal ve parametrik olmayan, karmaşık ve kaotik bir yapıya sahip olduğundan borsa tahmini, finansal zaman serileri tahmin sürecinin zorlu bir görevi olarak görülmektedir [3]. Buna ek olarak, borsa, politik olaylar, firma politikaları, genel ekonomik koşullar, yatırımcıların beklentileri, kurumsal yatırımcıların seçimleri, diğer borsaların hareketleri ve yatırımcıların psikolojisi gibi pek çok makro ekonomik faktörden etkilenmektedir [4]. Finansal metnin duygu analizi, hisse senedi fiyatlarının ya da yönlerinin tahmin edilmesinde analistler ve yatırımcılar için önemli ve aktif bir araştırma konusu olmuştur. Görüşlerin, piyasa dinamiklerini etkileyebileceği gözlemlenmiştir [5].

Bu açıdan bakıldığında duygu analizi, doğal dil işleme (NLP) ve metin sınıflandırma (TC) alanlarında hızla büyüyen bir araştırma alanıdır. Bu teknik, politika, iş dünyası, reklamcılık ve pazarlama gibi geniş bir uygulama yelpazesinin önemli bir parçası haline gelmiştir. Duygu analizi için kullanılan naif Bayes, destek vektör makineleri ve benzeri geleneksel makine öğrenme teknikleri bulunmaktadır. Ancak, son zamanlarda geleneksel makine öğrenme algoritmalarına kıyasla daha yüksek sınıflandırma performansı elde eden derin öğrenme modelleri tercih edilmektedir. Derin öğrenme modelleri, temel olarak karmaşık özelliklerin minimum dış destekle eğitilmesiyle otomatik özellik çıkarımı sağlamak, derin sinir ağları aracılığıyla verilerin anlamlı gösterimini elde etmek veya sınıflandırma yapmak amacıyla kullanılmaktadır. Evrimsel sinir ağları (CNN), tekrarlayan sinir ağları (RNN) ve derin

inanç ağları (DBN) gibi sinir ağlarının yanı sıra word2vec, Glove gibi çeşitli kelime gömülmesi yöntemleri de duygu sınıflandırmada yaygın olarak kullanılmaktadır. Bu teknikler, doğal dil işleme ve metin sınıflandırma alanlarının yanı sıra bilgisayar vizyonu, görüntü analizi, konuşma tanıma gibi farklı alanlarda da araştırmacılar tarafından uygulanmaktadır [55, 56]. Word2Vec ve GloVe, kelimeleri anlamlı vektörlere dönüştürebilen en doğru ve kullanışlı kelime gömülmesi yöntemleri arasındadır [6-9]. Bununla birlikte, bu yöntemler eğitim süreci için ve kelime vektörlerini üretmek amacıyla geniş bir metin kümesine ihtiyaç duyar. Araştırmacılar, sınırlı boyuttaki verilerle yaptıkları çalışmalarda kullanmak için genellikle Google haberler gibi büyük metin kümelerinde önceden eğitilmiş kelime gömülmelerinden faydalanırlar. Önceden eğitilmiş kelime gömülmeyle yapılan araştırmaların verimliliği, duygu analizi araştırmalarında büyük bir etkiye sahiptir.

Makine öğrenmesinde bir diğer popüler araştırma alanı ise topluluk öğrenmesidir. Topluluk öğrenmesi, literatürde öğrenciler komitesi, uzman karışimleri, sınıflandırıcı topluluklar, çoklu sınıflayıcı sistemler ve fikir birliği teorisi olarak da kullanılmaktadır [10, 58]. Temel fikir, tek bir sınıflandırıcı kullanmak yerine bir sınıflandırıcı grubu kullanarak sınıflandırma performansını arttırmaktır. Bir topluluk sisteminde, bir grup temel sınıflandırıcı kullanılır. Farklı sınıflandırıcılar temel öğrenci olarak kullanılıyorsa böyle bir sistem heterojen topluluk sistemi olarak adlandırılırken temel öğrenciler, tek tip sınıflandırıcıdan oluşuyorsa homojen topluluk sistemi olarak isimlendirilir. Bu çalışmada, heterojen topluluk sistemleri üzerine yoğunlaşmıştır.

Bir topluluk sistemi iki bölümden oluşur: Topluluk üretimi ve topluluk entegrasyonu [11-14]. Topluluk üretiminde, farklı temel sınıflandırıcılar kullanılarak farklı bir model kümesi oluşturulur. Topluluk entegrasyonunda ise sistemin nihai kararını elde etmek için temel sınıflandırıcıların kararlarını birleştirmede kullanılan birçok entegrasyon yöntemi bulunmaktadır [15-17, 57]. Topluluk üretimi için gerekli sınıflandırıcı seçimi ve topluluk entegrasyon yöntemlerinin yanı sıra, topluluk başarısını belirleyen iki temel faktör daha bulunmaktadır. Bunlar, topluluğun temel öğrencilerinin bireysel başarısı (düşük hata) ve temel öğrenci sonuçlarının birbirlerinden bağımsızlığıdır (yüksek çeşitlilik). Temel öğrencinin çeşitliliği arttıkça, sistemin sınıflandırma başarısı da artmaktadır. Çeşitlilik, homojen sistemler için torbalama, rastgele altuzaylar, rastgele ormanlar ve rotasyon ormanları gibi geleneksel topluluk algoritmaları ile sağlanır. Heterojen sistemler için çeşitlilik, farklı öğrenme algoritmalarıyla sağlanırken çoğunluk oylaması, yığılma, basamaklama gibi çeşitli karar verme teknikleriyle sonuçların harmanlanması sağlanır. Bu çalışmada, heterojen topluluk sisteminde çeşitliliği sağlamak amacıyla evrimsel sinir ağları (CNN), tekrarlayan sinir ağları (RNN), uzun kısa süreli hafıza ağları (LSTM) algoritmaları temel öğrenciler olarak çoğunluk oylaması

(MV) ve yığıtlama (STCK) ise topluluk kararlarını birleştirme yöntemleri olarak kullanılmıştır.

Bu çalışmada, literatürde sıkça kullanılan derin öğrenme yöntemleri, topluluk modeli olarak harmanlanmış veri kümesi, kelime gömülmesi yöntemleriyle anlam, bağlam ve söz dizimi açısından zenginleştirilmiştir. Böylece, Borsa İstanbul endeksi (BIST100) yönünün tahmin edilmesinde sınıflandırma performansının iyileştirilmesi hedeflenmiştir. Bu amaçla, veri kümesine bilgi kazanımı (IG) ve karınca kolonisi optimizasyonu (ACO) yöntemleri uygulanarak sadece bu yöntemlerden elde edilen özellikleri içeren yeni bir veri kümesi elde edilmiştir. Sonra, elde edilen bu özellikleri, anlamlı vektörlere dönüştürebilen kelime gömülmesi yöntemleri (Word2Vec, GloVe) kullanılarak orijinal veri kümesinden farklı; anlam, bağlam ve söz dizimi açısından zenginleştirilmiş veri kümeleri elde edilmiştir. Veri kümesinden öğrenilen kelimelerin vektör temsili sağlayan Word2vec ve GloVe gibi modellerin yanında TF-IDF ağırlıklı doküman-kelime matrisi, Word2vec ve GloVe kullanılarak elde edilen kelime gömülmesinin ortalamasıyla oluşturulan vektörler (Avg(Word2vec), GloVe (Avg(Glove)), Avg(Word2vec) ve Avg(Glove) vektörlerinin birleştirilmesinden elde edilen doküman matrisleri (Avg(Word2vec)+Avg(Glove)), Word2vec ve GloVe ile ayrı ayrı elde edilmiş kelime gömülmesinin eklenmesi ile geliştirilmiş TF-IDF ağırlıklı doküman matrisleri (TF-IDF+Avg(Word2vec), TF-IDF+ Avg(Glove)) de dâhil olmak üzere toplamda altı farklı doküman gösterim yöntemi kullanılmıştır. Sonra, derin topluluk modeli (DTM), üç farklı derin öğrenme algoritması (CNN, RNN ve LSTM) kullanılıp topluluk öğrenme metodolojisiyle harmanlanarak oluşturulmuştur. Önerilen yöntemin başarısını göstermek amacıyla deneyler, duygu analizinde sıkça kullanılan Twitter veri kümelerinde Türkçe ve İngilizce olmak üzere iki farklı dilde yürütülmüştür. Bildiğimiz kadarıyla bu çalışma, borsa yönü tahminlemesinde hem haber kaynağı olarak Twitter kullanması, hem bunun Türkçe ve İngilizce gibi yapısal olarak farklı iki dilde gerçekleştirilip anlam, bağlam ve söz dizimi açısından zenginleştirilmesi ile hem de derin öğrenme yöntemlerini topluluk modelleriyle harmanlayarak sınıflandırmada kullanması sebebiyle ilk çalışmadır. Diğer çalışmalarda, derin öğrenme ve topluluk metodolojileri özellik seçmede kullanılmış olup veri kümesi olarak çoğunlukla finansal haber kaynakları değerlendirilmiştir [18-20]. Sonuç olarak, farklı doküman gösterim yöntemleriyle de zenginleştirilen derin topluluk modelinin literatürdeki benzer çalışmalara kıyasla çok daha iyi sınıflandırma performansı sergilediği gözlenmiştir.

Makalenin geri kalan kısmı şu şekilde düzenlenmiştir: Bölüm 2, duygu analizinde topluluk sistemlerini, kelime gömülme yöntemlerini ve CNN, RNN, LSTM gibi derin öğrenme modellerini kullanan çalışmaların özetini sunmaktadır. Bölüm 3, önerilen sistem için kullanılan özellik seçme yöntemlerini, kelime gömülmesi metodlarını, doküman gösterim tekniklerini, heterojen topluluk modelinde kullanılan derin öğrenme algoritmalarını, topluluk kararları birleştirme yöntemlerini ve önerilen

modelin mimarisi sunmaktadır. Deney kurulumu ve sonuçları bölüm 4, 5 ve 6'da verilmektedir.

2. LİTERATÜR ÇALIŞMALARI (RELATED WORKS)

Bu bölüm, derin sinir ağları, topluluk sistemleri ve önceden eğitilmiş kelime gömülmesi ile ilgili duygu analizi ve metin sınıflandırma alanlarında yapılan çalışmaların özetini sunmaktadır.

Pek çok araştırmacı duygu analizinde daha doğru sınıflandırma modelleri sağlamak amacıyla derin öğrenme yaklaşımına odaklanmaktadır. Liao vd. [21], derin öğrenme modellerini kullanarak Twitter verileriyle duygu analizi yapmayı önermişlerdir. Basit bir evrişimsel sinir ağ modeli oluşturarak SVM ve naif Bayes sınıflandırıcılar gibi geleneksel öğrenme algoritmalarıyla bu ağı kıyaslamış ve daha iyi bir sınıflandırma performansı elde ettiklerini vurgulamışlardır. Kısa metinler üzerinde duygu analizi yapmak için karakterden cümle düzey bilgisine sahip yeni bir derin evrişimsel sinir ağı ise Santos ve Gatti tarafından önerilmiştir [22]. Yaklaşımlarının, literatürdeki sonuçları geride bıraktığını ve STS veri kümesi üzerinde %86,4 doğruluk sağladığını bildirmişlerdir. Başka bir çalışma [23], anahtar kelimelerin anlamlarını yorumlamanın önemine dikkat çekmiştir. Uzun kısa bellek ve kapılı tekrarlayan ünite ağları, IMDB ve SemEval-2016 veri kümeleri üzerinde anahtar kelime sözlüğü kullanılarak gerçekleştirilmiştir. Deney sonuçları, önerilen modelin verimliliğini %1-%2 doğruluk iyileşmesi ile kanıtlamıştır. Çince mikro bloglarının duygu sınıflandırması, [24]'de geliştirilmiş tekrarlayan sinirsel ağ modeli kullanılarak gerçekleştirilmiştir. Uzun süreli bağımlılığı çözmek için tekrarlayan sinir ağına gizli katmanını uzun-kısa süreli bellek yapısıyla değiştirerek alternatif bir çözüm sunmuşlardır. Sistemin sınıflandırma başarısının, geleneksel makine öğrenme algoritmalarından biri olan destek vektör makinesinin sınıflandırma başarısından çok daha üstün olduğunu vurgulamışlardır.

Heterojen mikroblog duygu sınıflandırması (MSC) olarak da adlandırılan duygu sınıflandırmasına ilişkin bir başka çalışma [25], kullanıcıların gönderilerinden ve bağlantılarından faydalanarak yeni bir tekrarlayan rastgele yürüyüş ağını önermişlerdir. Önerilen model, eğitim aşamasında geri yayılım yöntemini uygulayarak rastgele yürüme katmanına sahip derin sinir ağlarından oluşturulmuştur. Modelin başarısını göstermek amacıyla Twitter veri kümesi üzerinde deneyler yapılmıştır. Önerilen teknik, literatürdeki benzer çalışmalardan daha iyi bir sınıflandırma performansı sergilemiştir. Twitter veri kümesinde çok dilli duygu analizi yapmak için etkin bir derin sinir ağı mimarisi [26]'da kullanılmıştır. Önerilen model ile sırasıyla, uzun-kısa süreli bellek ağları ve evrişimsel ağlar kullanılarak kelime ve karakter düzeyinde gömülmeler elde edilmiştir. Karakter temelli mimari, uzun-kısa süreli bellek gömülmesi, evrişimsel gömme, evrişimsel gömme donma, evrişimsel karakter seviyesinde gömme ve geleneksel destek vektörü makinesi algoritmaları kullanılmış olup doğruluk ve F1-skoru değerlendirme metrikleri açısından sonuçlar

kiyaslanmıştır. Kapsamlı deney sonuçları ile önerilen tekniğin (evrimsel karakter temelli mimari), literatürde yer alan diğer derin sinir ağı modellerine kıyasla çok dilli duygu analizi için daha verimli olduğu gözlenmiştir. [27]'de Uysal ve Murphey, geleneksel özellik gösterim modellerinin karşılaştırılmasına ve derin öğrenme yaklaşımına yoğunlaşmışlardır. Bu karşılaştırmalı çalışmada, iki tip öznelik gösterim modeli kullanılmıştır. Birincisi, dokümandaki terimlerin sırasını dikkate almadan terim frekansı ile gerçekleştirilirken ikincisinde ise kelime gömülmesini kullanıp anlam açısından daha zengin bir vektör gösterimi amaçlanmıştır. Lineer çekirdekli SVM sınıflandırıcı, geleneksel tekniklerin sınıflandırma performansını göstermek için kullanılmıştır. Dahası, yazarlar genel olarak özellik seçiminde kullanılan derin öğrenme yaklaşımlarını bu çalışmada sınıflandırma yapmak amacıyla kullanmışlardır. Önerilen derin öğrenme temelli modeller ile geleneksel yöntemlere kıyasla daha iyi sınıflandırma sonuçları elde edildiğini bildirmişlerdir.

Duygu sınıflandırma alanında topluluk stratejileri ve derin öğrenme yaklaşımlarının kombinasyonu üzerine sınırlı sayıda çalışma bulunmaktadır. Önerilen perseptron temelli ve çok katmanlı topluluk modeli, [19]'da, iyimser ya da kötümser olarak finansal metinlerin duygu puanlarını tahmin etmek için kullanılmıştır. Bu amaçla yazarlar, CNN, LSTM, vektör ortalaması ve özellik tabanlı dört model ile özellik vektörlerinin çeşitliliğini sağlamışlardır. Topluluk kararlarını birleştirme adımı uygulandıktan sonra, çok katmanlı perseptron ağı bir sınıflandırıcı olarak kullanılmıştır. Deney sonuçları, derin öğrenme algoritmalarıyla oluşturulan özellik tabanlı modellerin sınıflandırma performansını önemli ölçüde iyileştirdiğini göstermiştir. Bu çalışma, topluluk metodolojisini ve derin öğrenme tekniklerini yeni özellik vektörü oluşturmak için kullanırken bizim çalışmamız, hem topluluk hem de derin öğrenme modellerini sınıflandırmada kullanmaktadır. Nozza vd. [28] duygu sınıflandırması için derin öğrenme ve topluluk tekniklerini değerlendirerek alan adaptasyonu problemini ele almayı önermişlerdir. Naif Bayes, destek vektör makinesi, oylama perseptronu, karar ağacı, lojistik regresyon, k-en yakın komşu ve rastgele ormanlar temel öğreniciler olarak kabul edilmiştir. Torbalama, artırma, rastgele altuzay ve basit oylama topluluk metodları olarak kullanılmıştır. Derin öğrenme kısmı, bir yapay sinir ağı sınıfı olan otokodlayıcı ile gerçekleştirilmiştir. Yazarlar, önerilen yaklaşımın doğruluk sonuçlarını, literatürdeki çalışmalarla kıyasladıklarında daha yüksek sonuçlar elde edildiğini gözlemlemişlerdir. Topluluk teknikleri ve derin öğrenme ile duygu analizi üzerine yapılan bir başka çalışmada [29], derin öğrenme tekniklerinin başarısının artırılması önerilmiştir. Bu amaçla, bir kelime gömülmesi modeli ve topluluk sisteminin temel öğrenicisi olarak doğrusal bir makine öğrenme algoritmasını kullanmışlardır. Daha sonra, temel öğreniciyi ve diğer sınıflandırıcıları birleştirmek için topluluk stratejisi uygulamışlardır. Yazarlar, orijinal versiyonlara kıyasla önerilen yaklaşımın çok daha iyi bir sınıflandırma performansı sergilediğini vurgulamışlardır. Bu çalışma ise bildiğimiz kadarıyla, finansal duygu analizini

gerçekleştirmek amacıyla derin öğrenme modellerini topluluk yöntemleriyle harmanlayan literatürdeki ilk çalışmadır. Veri kümesi olarak Twitter kullanılmayan fakat finansal haberler kullanılarak Borsa İstanbul'un yönünü tahmin etmek üzere yapılan sınırlı sayıda çalışma bulunmaktadır. Bunlardan biri [30], finansal haber niteliği taşıyan metinlerin Borsa İstanbul endeksinin açılış fiyatlamasına etkisini ölçmeyi amaçlamıştır. Sınıflandırma yöntemi olarak naif Bayes, değerlendirme metrikleri olarak ise doğruluk, F-ölçümü, ortalama makro F-ölçümü kullanılmıştır. Özellik vektörlerini oluşturmak amacıyla Türkçe için metin madenciliği teknikleri veri kümelerine uygulanmış ve ortak bilgi (mutual information), ki-kare ve önerilen yeni dengeli ortak bilgi (balanced mutual information) metodları özellik seçme yöntemleri olarak kullanılmıştır. Önerilen yeni özellik seçme yöntemiyle birlikte farklı kaynaklardan toplanan metin içerikli veri kümesi kullanımının Borsa İstanbul endeksinin yönünü belirlemede önemli bir etken olabileceği kanıtlanmıştır. Bu çalışma, Borsa İstanbul endeksinin yönünü belirlemede farklı ve yeni özellik seçme metodlarını kullanarak tek bir sınıflayıcı kullanırken bizim çalışmamız, birden fazla derin sınıflandırıcıyı topluluk teknikleriyle harmanlamayı önermiş bunu yaparken özellik vektörünü özellik seçme yöntemleriyle elde etmek yerine son zamanlarda oldukça popüler olan Word2vec, Glove gibi kelime gömülmesi yöntemlerini kullanarak veri kümesini anlam, bağlam ve söz dizimi bakımından zenginleştirmiştir. Ayrıca bu çalışma, kelime oluşumlarını Türkçe metinler üzerinde sadece TF-IDF doküman gösterim tekniğinden faydalanarak incelerken bizim çalışmamızda, içinde TF-IDF yönteminin de olduğu altı farklı doküman gösterim tekniğinden faydalanılmış hem kelimeler arasındaki anlamsal ilişki kelime gömülmesi metodları kullanılarak göz ardı edilmemiş hem de veri kümesi olarak Türkçe ve İngilizce Twitter veri kümeleri kullanılmıştır. Bir diğer çalışma [18], Borsa İstanbul'da işlem gören banka hisselerinin yönlerini belirlemek için derin öğrenme yöntemlerinden biri olan uzun kısa dönem hafıza ağları (UKDH) yöntemini kullanmıştır. Kelime gömülmesi yöntemi olarak Fasttext'i kullanarak finansal duygu sözlüğünden faydalanılmıştır. Ayrıca, yazarlar rastgele ve naif olmak üzere iki farklı UKDH ağlarının performansını kıyaslamışlardır. Deney sonuçları, rastgele UKDH yönteminin naif yaklaşımından daha iyi bir sınıflandırma performansı sergilediğini göstermiştir.

3. ÖNERİLEN SİSTEM (PROPOSED SYSTEM)

Bu bölümde, çalışmada kullanılan özellik seçme yöntemleri, kelime gömülmesi teknikleri, doküman gösterim metodları, topluluk kararlarını birleştirme yöntemleri ve önerilen sistemin detaylarından bahsedilmektedir.

3.1. Özellik Seçme Yöntemleri (Feature Selection Methods)

Bilgi kazanımı, bir özelliğin gerçekleşmesini veya gerçekleşmemesini bilerek sınıf tahmini için elde edilen bilgi bitlerinin sayısını değerlendirir [31-33]. Diğer bir deyişle, bir özellik için verilen sınıflandırma sonuçlarının ne kadar değer

ile kazanılabileceğini gösterir. Aslında, genel özellik seçim süreci, belirli bir özellik seçim yöntemine göre her bir özelliği puanlamak ve en iyi k özelliği elde etmeye dayanır. Böylece, yüksek sınıflandırma başarısına sahip olabilecek en önemli özelliklerin kümesi elde edilir. (Eş. 1)

$$IG(t) = \sum_{i=1}^C P(C_i) \log P(C_i) + P(t) \quad (1)$$

$$\sum_{i=1}^C P(C_i | t) \log P(C_i | t) + P(t')$$

$$\sum_{i=1}^C P(C_i | t') \log P(C_i | t')$$

Burada, C sınıf sayısını, $P(C_i)$ C_i 'nin olasılığını, $P(t)$ ve $P(t')$ sırasıyla t teriminin olma ve olmama olasılığını gösterir.

Karınca kolonisi optimizasyonu, çeşitli alanlarda özellik seçimi için de kullanılabilen bir optimizasyon tekniğidir. Kokulu bir madde olan ve karıncalar tarafından salgılanan feromon yollar aracılığıyla yuvadan besin kaynağına en kısa yolları bulmaya dayanan bir yöntemdir. Belirli bir süre sonra feromon miktarının artması tercih edilen yolun en kısa yol olduğunu tespit etmede temel faktördür. Karıncalar, bir besin kaynağı keşfettiklerinde, yuvadan yemek kaynağına gidene kadar o yolu feromon ile işaretlerler. Böylece, yuvadan besin kaynağına gitmek isteyen ve daha önce bu yolu kullanmamış her bir karınca, feromonca zengin olan yolu takip ederek besin kaynağına ulaşır. Böylece, karıncalar tarafından bu kısa yol kullanıldıça yol üzerinde feromon miktarı artar ve bu feromon miktarı yeni karıncaların kısa yolu tercih etmelerinde önemli bir rol oynar. Karıncalar tarafından daha az tercih edilen yollarda ise feromon zamanla buharlaşır ve karınca geçişlerinin fazla olduğu yollar tercih edilir. Bu nedenle, geçilecek yolun seçilme olasılığını belirlemek için her karınca için bir geçiş olasılık kuralı vardır. Böylece, karınca kolonisi optimizasyonu tekniği, her seferinde optimum alt kümeyi aramayı yönlendirebildiğinden özellik seçim süreci için caziptir. Zaman t 'de yol (özellik) i 'den j 'ye gitmek isteyen bir karıncanın olasılık geçiş kuralı Eş. 2'deki gibidir.

$$p_{ij}^k(t) = \begin{cases} \frac{[\tau_{ij}(t)^\alpha][\eta_{ij}^\beta]}{\sum_{l \in J_i^k} [\tau_{il}(t)^\alpha][\eta_{il}^\beta]} & \text{eğer } j \in J_i^k, \\ 0 & \text{değilse} \end{cases} \quad (2)$$

Burada k karıncaların sayısını, η_{ij} özellik i 'deyken özellik j 'yi seçme isteğini, J_i^k , karınca k 'nin gezinmediği yolları (özellikleri), $\tau_{ij}(t)$ ise (i, j) yolu arasında salınan feromon miktarını simgeler. Feromon değerinin nispi önemi, global bilgi sağlayan α değişkeniyle gösterilirken β sezgisel ve yerel bilgi sağlar. Karınca kolonisi optimizasyonunda özellik seçim sürecinin ilk adımı, bir dizi k karınca üretmektir. Bu çalışmada karınca sayısı, veri kümesindeki özelliklerin sayısına eşit olarak belirlenmiştir. Böylece, her karınca rastgele bir özellik ile başlar ve durma şartı yerine getirilinceye kadar yolları gezinirler. Daha sonra, alt kümeler toplanır ve değerlendirilir. Algoritma belirli bir sayıyı gerçekleştirdikten veya optimal bir alt kümeye ulaştıktan

sonra genel özellik seçim süreci, en iyi özellik çıktısı elde edilerek sona erer. Her iki durumda da, feromonun yoğunluğunun güncellenmesi kaçınılmazdır. Sonrasında, yeni karıncalar üretilir ve özellik seçim süreci bir kez daha tekrar eder. Feromon güncellemesi, her yol için Eş 3'deki kural tarafından gerçekleştirilir:

$$\tau_{ij}(t+1) = (1-\rho) \tau_{ij}(t) + \rho \Delta \tau_{ij}(t) \quad (3)$$

Burada ρ feromon buharlaşma katsayısı, $\Delta \tau_{ij}(t)$ her bir karınca k tarafından biriktirilen feromon miktarını göstermektedir.

3.2. Kelime Gömülmesi Yöntemleri (Word Embedding Methods)

Kelime gömülmesi, bir dokümandaki kelimelerin daha iyi şekilde temsillerini bulmaya çalışan aktif bir araştırma alanı olmuştur. Tüm kelime gömülmesinin ardındaki düşünce, bir veri kümesindeki dokümanlardan olabildiğince bağlamsal, anlamsal ve sözdizimsel bilgi yakalamaktır. Kelime gömülmesi, her kelimenin, önceden tanımlanmış bir vektör uzayında gerçek değerli vektörler olarak temsil edildiği kelimelerin dağıtılmış bir gösterimidir. Dağıtılmış gösterim, benzer anlama sahip kelimelerin benzer bağlamlarda veya metinsel yakınlık meydana geldiği durumlarda dağılımsal hipotez kavramına dayanır. Dağıtılmış vektör temsiline nesne tanıma, anlam buğlaklığı, makine çevirisi ve ayrıştırma gibi birçok doğal dil işleme uygulamasında yararlı olduğu kanıtlanmıştır [34].

Word2vec Mikolov vd. tarafından önerilen en popüler kelime gömülmesi tekniğidir [7, 8]. Word2vec yöntemi, sınır ağlarını kullanarak bir eğitim kümesinden sözcüklerin vektör temsillerini öğrenir. Gömülü vektör uzayında birbirine yakın olacak vektörlere benzer anlam taşıyan kelimeleri eşler. Word2vec, kelimelerin dağıtılmış bir temsiline gerçekleştirmek için iki modelli mimari sunar: CBOW (sürekli torbalama) ve Skip-gram (sürekli atlama gramı) modeli. CBOW modeli belirli bir bağlamda bir sözcük tahmininde bulunurken, Skip-gram modeli belirli bir kelimenin bağlamını tahmin eder. Bir başka deyişle, CBOW modeli, kelime-anlam yaklaşımı gibi bağlam sırasını görmezden gelerek çevreleyen bağlam sözcükleri verilen bir kelimeyi tahmin etmektedir. Sürekli atlama-gram modeli ise verilen kelimeyi çevreleyen kelimeleri tahmin etmeyi amaçlamaktadır. Bu çalışmada, CBOW modeline kıyasla, kelime frekansı az olan kelimeler için kayda değer performansından dolayı sürekli atlama gramı modeline odaklanılmıştır. Word2vec, genellikle çok büyük veri kümelerinden sözcüklerin sürekli vektör temsillerini çıkarır. Belirli bir veri kümesinden kendi vektör temsili modelinizi oluşturmak mümkün olsa da, birçok çalışma, kelime gömülmesini elde etmek için gerekli yüksek hesaplama gücü ve eğitim süresi nedeniyle önceden eğitilmiş modelleri kullanmayı tercih etmektedir. Önceden eğitilmiş modeller, birçok doğal dil işleme uygulamasında faydalı bulunmuştur.

Glove, kelime yerleşimleri ya da gömülmesi olarak bilinen kelime anlam biliminin vektör uzay modellerini oluşturmak için kullanılan bir yöntemdir. Bu yöntem, küresel matris

faktörizasyonu (global matrix factorization) ve yerel bağlam penceresi (local context window) yaklaşımlarını kelime gömülmelemlerini elde etmek amacıyla birleştirir. Bu yöntemleri kullanan Glove, denetimsiz bir öğrenme algoritmasıdır. Eğitim, bir veri kümesinden toplanmış küresel kelime-kelime oluşum (word-word cooccurrence) istatistiği üzerinde gerçekleştirilir ve ortaya çıkan vektör temsili, kelime vektörü uzayının doğrusal altyapılarını gösterir. Stanford üniversitesinin doğal dil işleme ekibi tarafından açık kaynaklı bir proje olarak geliştirilmiştir [9].

3.3. Doküman Gösterim Yöntemleri (Document Representation Methods)

Dokümanların gösterim şekilleri, metin işleme uygulamalarının başarı performansında önemli rol oynamaktadır. Birçok metin sınıflandırma uygulamasında, veri kümesindeki her bir doküman reel sayıların bir vektörü olarak temsil edilir. Bir vektörün öğeleri, genellikle dokümanda görünen kelimelere karşılık gelir. Doküman kümelerinin vektör temsili ile doküman-kelime matrisi oluşturulur. Her kelimenin bir dokümandaki önemi, farklı kelime ağırlıklandırma yöntemleri kullanılarak hesaplanır. Geleneksel kelime ağırlıklandırma yöntemleri, boole (boolean), kelime frekansı, TF-IDF ağırlıklandırma şemalarını içerir. TF-IDF, metin işlemede kullanılan yaygın bir ağırlıklandırma yöntemidir. Bu gösterimde, her kelimenin frekansı ters doküman frekansı (IDF) ile çarpılmaktadır. Bu, toplamada ortak terimlerin önemini azaltmakta ve nispeten düşük frekanslara sahip nadir kelimelerin etkisini arttırmaktadır. Yukarıda açıklandığı gibi, kelime gömülmesi teknikleri de dokümanları temsil etmek için kullanılan diğer yöntemlerdendir.

Bu çalışmada aşağıdaki doküman gösterim yöntemleri kullanılmıştır:

- TF-IDF: Doküman vektörü, bir dokümandaki sözcüklerin TF-IDF doküman ağırlıklandırma yöntemiyle oluşturulmasına dayanır.
- Avg(Word2vec): Doküman vektörü, önceden eğitilmiş modellerle word2vec kullanılarak bir dokümanda görülen kelime gömülmelemlerinin tüm vektörlerinin ortalaması alınarak elde edilir.
- Avg(Glove): Doküman vektörü, önceden eğitilmiş modellerle Glove kullanılarak bir dokümanda görülen kelime gömülmelemlerinin tüm vektörlerinin ortalaması alınarak elde edilir.
- Avg(Word2vec) + Avg(Glove): Doküman vektörü, hem Avg(Word2vec) hem de Avg(Glove) vektörlerini içerir.
- TF-IDF + Avg(Word2vec): Doküman vektörü, hem TF-IDF hem de Avg(Word2vec) vektörlerini içerir.
- TF-IDF+ Avg(Glove): Doküman vektörü, hem TF-IDF hem de Avg(Glove) vektörlerini içerir.

Yukarıda bahsedilen kelime gömülmelemlerinin ortalamasını almak için kullanılan yöntem Eş. 4'deki gibidir.

$$e_d = \frac{1}{n} \sum_{i=1}^n e_{w_i} \quad (4)$$

Burada, kelimelerin 1'den n 'e kadar w ile temsil edildiği bir d dokümanı verildiğinde $d = w_1, w_2, \dots, w_n$ kelime gömülmelemlerini temsil eden vektörler $e_{w_1}, e_{w_2}, \dots, e_{w_n}$ ile gösterilmektedir.

3.4. Temel Sınıflandırıcılar (Base Classifiers)

Bu çalışmada, evrişimsel sinir ağları (CNNs), tekrarlayan sinir ağları (RNNs) ve uzun kısa vadeli hafıza ağları (LSTMs) temel sınıflandırıcılar olarak derin topluluk öğrenmesi yaklaşımında kullanılmıştır.

Evrişimsel sinir ağları (CNNs), görüntü tanıma ve sınıflandırmada dikkate değer sonuçlar elde etmiş bir derin öğrenme ağı sınıfıdır [35-37]. CNN modellerinin daha sonra doğal dil işleme alanında da etkili olduğu gösterilmiştir [6]. CNN, giriş-çıkış katmanına ve gizli katmanlara sahip ileriye dönük bir ağıdır. Gizli katmanlar, havuz katmanlarıyla birleştirilen kıvrımlı katmanlardan oluşur. CNN'nin en önemli bloğu, evrişim tabakasıdır. Evrişim katmanı, bir özellik haritası oluşturarak verileri ilgili verilerle birleştirmek ve veri girmek için bir Evrişim filtresi uygular. Evrişim katmanının son çıktısı haline gelen gelen özellik haritalarının bir yığını almak için giriş verilerine çoklu filtre uygulanır. Filtreler, eğitim sürecinde öğrenilir. Evrişim işlemi, orijinal veri bölgelerindeki yerel bağımlılıkları veya semantiği yakalar. Evrişim işleminden sonra, bir havuz katmanı her özellik haritasındaki örnek sayısını azaltır ve en önemli bilgileri saklar. Havuzlama katmanı, eğitim süresini kısaltır, verilerin boyutlarını azaltır ve ezberlemeyi ortadan kaldırır. Maksimum havuzlama (max-pooling), belirli bir komşu penceresinde en büyük değeri alan en yaygın havuz işlevi türüdür. CNN mimarileri, havuz katmanları ile serpiştirilmiş bir dizi evrişim katmanı ve bunu takiben bir dizi tam bağı tabakadan oluşur.

CNN orijinal olarak, giriş verilerinin iki boyutlu (2D) bir yapıda olduğu görüntü işleme ve tanıma alanında uygulanır. Görüntü işlemede CNN, 2D görüntülerin mekansal yerel korelasyonunu kullanır, 2D görüntülerin küçük bölgelerini öğrenir ve yanıt verir. Doğal dil işleme ve metin sınıflandırmasında, bir metin veya dokümanda bulunan kelimeler, tek boyutlu (1D) yapıya sahip sayıların vektörüne dönüştürülür. Metin işleme uygulamalarında CNN, bir belgenin tamamını temsil eden bir bağımsız kelime vektörleri havuzundan öğrenmenin yanı sıra, verilerin küçük metin bölgelerini öğrenerek 1D veri yapısını (kelime sırası) kullanabilir [38].

Genelde, bir sinir ağı modelinde tüm girdi ve çıktıların birbirinden bağımsız olduğu farz edilir. Ancak, metin işleme veya doğal dil işlemede girdilerin ve çıktıların birbirinden bağımsız düşünülmesi söz konusu değildir. Çünkü bir cümlede yer alan herhangi bir kelimedenden sonraki kelime tahmin edilmek isteniyorsa kelime bağımlılıkları önem arz eder. Bu sebeple, düğümler arasındaki bağlantılar sayesinde yönlendirilmiş bir döngü söz konusudur. RNN modelinin çalışma mantığında da buna benzer sıralı bilgi kullanma mantığı vardır. RNN modelinde, elde edilen çıktılar önceki

hesaplamalara bağlıdır ve girdilerin her biri için aynı işlevi yerine getirirler. Bu sebeple, tekrarlayan sinir ağları olarak adlandırılmışlardır. RNN'lerin konuşma tanıma [39], görüntü işleme [40] ve doğal dil işleme alanlarında [41] sistemin performansına katkı sağladığı kanıtlanmıştır.

Uzun kısa vadeli hafıza ağları (LSTMs), tekrarlayan sinir ağlarının özelleştirilmiş halidir. Tekrarlayan sinir ağlarında, bir önceki bilginin kullanımına dair bir yaklaşım söz konusudur. Metin işleme alanında RNN kullanılıyorsa kısa bir cümledeki bir kelimeyi tahmin etmek için önceki kelimenin de kullanımına ihtiyaç vardır. Fakat cümle uzadıkça ve kelimeler arasındaki bağlar arttıkça sadece önceki kelimenin kullanılması oldukça zordur. Bu amaçla, LSTM'lerde RNN'lerden farklı olarak hafıza bilgisi tutulmaktadır. Böylece, kelime tahmin edilirken hafızada tutulan diğer kelimelerden faydalanılarak tahmin edilecek kelimenin doğruluk oranı artırılmış olur. LSTM ağının çıktısı, kendini tekrarlayan bir şekilde girişlere bağlanır. LSTM, özellikle metin işleme ve doğal dil işleme konularında dikkate değer sonuçlar elde etmektedir [42-46].

3.5. Topluluk Kararlarını Birleştirme Yöntemleri (Ensemble Integration Methods)

Bu çalışmada kullanılan evrimsel sinir ağları (CNN), tekrarlayan sinir ağları (RNN), uzun-kısa vadeli hafıza ağları (LSTM) bireysel öğrencilerinin kararlarını birleştirmek için çoğunluk oylaması ve yığıtlama yöntemleri kullanılmıştır.

Çoğunluk oylaması yönteminde, etiketlenmemiş bir örnek, temel sınıflandırıcılardan toplanan ve en yüksek oyu alan sınıfa göre sınıflandırılır. İstifleme olarak da adlandırılan yığıtlama yönteminde, temel sınıflandırıcılarının kararını birleştirmek için bir meta-düzey sınıflandırıcı kullanılır [47]. İstifleme yöntemi iki adımdan oluşur. İlk adımda, bir dizi temel seviye sınıflandırıcı C_1, C_2, \dots, C_n bir dizi örnek kümesinden elde edilen $S_i = (x_i, y_i)$ özellik örneklerinden oluşur. Burada x_i özellik vektörü, y_i sınıf etiketidir. Temel düzey sınıflandırıcıların kararlarından bir meta veri kümesi oluşturulur. Meta veri kümesi, orijinal eğitim veri kümesindeki sınıflandırıcıların tahminleri için bir örnek içerir. Meta veri kümesi, $m_i = (d_i, y_i)$ biçimindedir; burada d_i , bireysel n temel sınıflandırıcılarının tahminidir. Meta veri kümesi, performansı geliştirmek için hem orijinal eğitim örneklerini hem de temel sınıflandırıcılarının $m_i = (x_i, d_i, y_i)$ biçimindeki kararlarını içerebilir. Meta-veri kümesinin oluşturulmasından sonra, bir meta-seviye sınıflandırıcı meta-veri kümesi ile eğitilir ve tahmin yapmak için kullanılır. Bu çalışmada, meta-veri kümesi hem orijinal eğitim örneklerini hem de temel seviye sınıflandırıcılarının kararlarını içermektedir.

3.6. Önerilen Modelin Mimarisi (Architecture of the Proposed Model)

Bu çalışmada, finansal metin kaynağı olarak Twitter kullanılarak borsa yönünün belirlenmesinde ne kadar etkili olabileceğinin ölçülmesi hedeflenmektedir. Sadece veri

kümesinin değil, aynı zamanda kullanılan özellik seçim yöntemlerinin farklı stratejilerle sunulması, birleştirilmesi ve sınıflandırmanın derin topluluk metodolojisiyle yapılarak önerilen yöntemin başarısının ölçülmesi de amaçlanmaktadır.

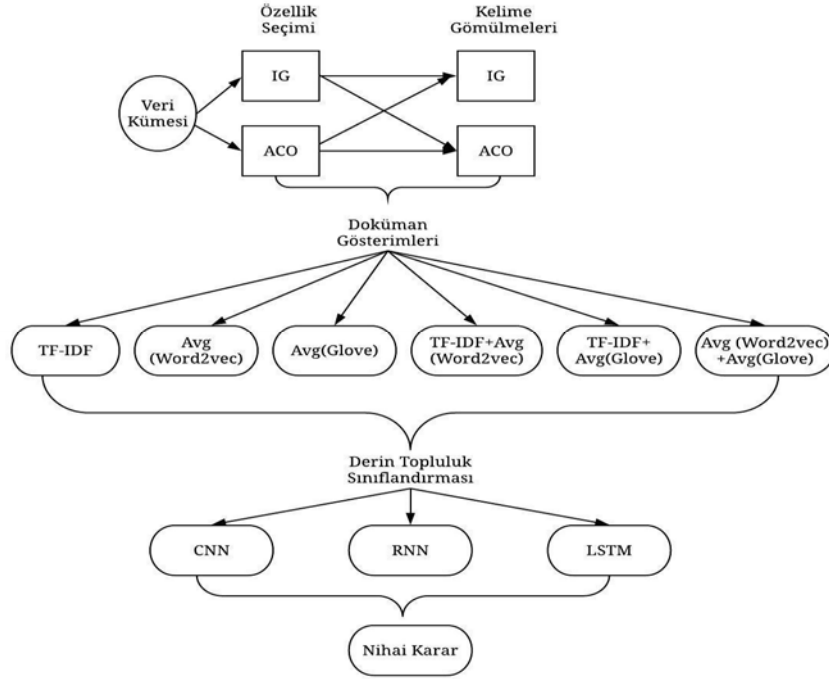
Bu amaçla, ilk aşamada veri kümesindeki orijinal özelliklerin kullanılmasındansa bilgi kazanımı (IG) ve karınca kolonisi optimizasyonu (ACO) yöntemleriyle veri kümesini en iyi ifade eden özelliklerin seçilmesi sağlanmıştır. Sonra, IG ve ACO ile elde edilen özelliklerin kelime gömümleri word2vec(IG) , glove(IG) , word2vec(ACO) , glove(ACO) olmak üzere ayrı ayrı elde edilmiştir. Bu aşamada veri kümesinin özelliklerinin yarısı IG ile seçilen özelliklerin kelime gömümlerinden kalan yarısı ise ACO ile seçilen özelliklerin kelime gömümlerinden oluşur. Böylece, özellik seçim aşamasından sonra veri kümesinin $\text{word2vec(IG)+word2vec(ACO)}$, $\text{word2vec(IG)+glove(ACO)}$, $\text{glove(IG)+word2vec(ACO)}$, $\text{glove(IG)+glove(ACO)}$ olmak üzere dört farklı versiyonu bulunmaktadır.

Derin topluluk modelinin gerçekleştirebilmesi için dokümanların vektör temsillerinin elde edilmesi gerekmektedir. Bildiğimiz kadarıyla bu çalışma, borsa yönünün belirlenmesi çalışmalarında kullanılan TF-IDF doküman gösterim yöntemi dışında, Avg(Word2vec) , Avg(Glove) , $\text{Avg(Word2vec)+Avg(GloVe)}$, $\text{TF-IDF+Avg(Word2vec)}$, TF-IDF+Avg(GloVe) doküman gösterim yöntemlerini öneren ilk çalışmadır.

Doküman gösterimleri elde edilen veri kümeleriyle CNN, RNN ve LSTM derin algoritmaları kullanılarak topluluk sınıflandırması adımı gerçekleştirilir. Her bir derin temel öğrenciden elde edilen sonuçlar, sistemin nihai kararını elde etmesi amacıyla çoğunluk oylaması ve yığıtlama topluluk stratejileri yöntemleriyle birleştirilir. Bu çalışma, derin öğrenme algoritmalarını topluluk öğrenmesi yöntemiyle harmanlaması, önerdiği doküman gösterim teknikleriyle sınıflandırmaya sağladığı katkısı ve veri kaynağı olarak sosyal medyayı kullanmasıyla borsa yönü tahminlemesi alanında yapılan diğer çalışmalardan farklılık göstermektedir. Önerilen sistemin mimarisi Şekil 1'de gösterilmiştir.

4. DENEY KURULUMU (EXPERIMENT SETUP)

Deneylerde, Türkçe ve İngilizce olmak üzere iki farklı dilde veri kümesi kullanılmıştır. Bu veri kümeleri, Twitter ortamından kullanıcıların iki farklı dilde BIST100 ve XU100 etiketleriyle yaptıkları yorumların toplanmasıyla elde edilmiştir. Veri kümesi, Şubat 2018 ve Ağustos 2018 (185 gün) tarihleri arasındaki bireysel kullanıcıların ve organizasyonların yorumlarını kapsamakta olup her bir yorum pozitif, negatif veya nötr etikete sahiptir. Pozitif ve negatif yorumlar, bireysel kullanıcıların yorumlarından oluşurken sosyal medya kullanıcılarını yönlendirmemek adına sadece bilgilendirme amaçlı yorumlarda bulunan



Şekil 1. Önerilen sistemin akış diyagramı (Flowchart of the proposed system)

organizasyonların sınıf etiketlerinin çoğunun nötr olduğu gözlenmiştir. Organizasyon yorumlarının pozitif veya negatif etiketine sahip yorumlarının ise gün sonu, saatlik, dört saatlik BIST100 grafiklerinin gerçekleşen değerler üzerinden “yükseldi” veya “düştü” gibi yorumlandığı dolayısıyla herhangi bir tahmin yorumu içermediği saptanmıştır. Bu sebeple, organizasyon yorumlarından pozitif ve negatif etiketine sahip olanlar veri kümesine dahil edilmeyip sadece nötr olanlar dikkate alınmıştır. Sonuç olarak, Türkçe veri kümesi için 20,172 negatif, 7,844 pozitif ve 2130 nötr olmak üzere toplamda 30,146 adet kullanıcı yorumu elde edilirken İngilizce veri kümesi için 18,357 negatif, 4,203 pozitif ve 1346 nötr olmak üzere toplamda 23,906 adet kullanıcı yorumu toplanmıştır. Sistemin performansını etkilememek için veri kümesi üzerinde sık kullanılan bağlaçları yok etme (stop-word elimination), kelime kökü elde etme (stemming) gibi herhangi bir ön işleme adımı uygulanmamıştır.

Deneylerde, tekrarlanan tutma (repeated holdout) yöntemini kullanılmıştır. Veriler ise eğitim için %80, test için %20 olan iki veri kümesine rastgele olarak ayrılmıştır. Güvenilir bir tahmin elde etmek için, tutma işlemi 10 kez tekrarlanmış ve her bir yinelemenin ortalamaları alınarak genel bir doğruluk hesaplanmıştır. Temel öğrencilerin sayısı, daha önceki çalışmalar [48,49] dikkate alınıp 100 olarak ayarlanmıştır. Özellik seçme metodlarından faydalanılarak seçilecek özelliklerin sayısı, her bir özellik seçme metodu için özellik sayısının yarısı kadardır. Yani, toplam d sayıda özelliğe sahip bir veri kümesi için IG yöntemi kullanılarak elde edilecek özellik sayısı $d/2$ ACO kullanılarak elde edilecek özellik sayısı yine $d/2$ 'dir. Ayrıca, ACO yöntemi ile özellik seçmek için parametrelerin belirtilmesi gereklidir. İlk olarak,

karınca sayısı, her veri kümesi için özelliklerin sayısına eşittir. Bu nedenle karınca sayısı veri kümesine göre değişir. Dahası, ACO algoritması belirli bir sayıda gerçekleştirilmelidir. Bu sayı, çalışmada kullanılan sistemde temel öğrencilerin sayısına karşılık gelmektedir. Temel öğrencilerin sayısı 100 olduğuna göre algoritma 100 kez uygulandıktan sonra feromon yoğunluğu güncellenir. Sonrasında, yeni bir karınca kümesi oluşturulup işlem bir kez daha yinelenir. Her bir özelliğin ilk feromon yoğunluğu 1 olarak ayarlanmıştır. Karıncaların geçişi hakkında yerel ve küresel iki önemli bilgi α ve β parametreleriyle belirlenir. Bu parametrelerin seçimi deneysel olarak sırasıyla 1 ve 0,1 olarak ayarlanmıştır. Feromon izi buharlaşma katsayısı ($\rho = 0.2$) feromon yollarını güncelleyen ve 0 ile 1 arasında bulunan bir parametredir.

Bu çalışmada, özellik seçme süreci dâhil önerilen sistemin tüm süreçleri için Python programlama dili kullanılmıştır. Ayrıca, Python 3 versiyonunu Pycharm ortamı ile Gensim tema modelinde kullanarak sürekli atlama-gramı (continuous skip-gram) modeli üzerine çalışılmıştır. Bu model, kelimeleri göstermek için 200 boyutlu bir vektör uzayını kullanmakta olup eğitim penceresi 5 olarak ayarlanmıştır. Ayrıca, Google, Word2Vec Skip-gram algoritmasıyla önceden eğitilmiş vektörler elde etmek için yaklaşık 100 milyar kelime içeren Google Haberler veri kümesinden faydalanılmıştır [50-52]. Ön eğitilmiş model, yaklaşık 3 milyon kelime ve kelime öbeği için kelime vektörlerini içermektedir. Türkçe yazılmış bütün Wikipedia makaleleri ile de önceden eğitilmiş Türkçe Word2vec modeli oluşturulmuştur [53]. Önceden eğitilmiş bu modeller, İngilizce ve Türkçe olarak dokümanları temsil etmek amacıyla kullanılmaktadır.

5. DENEY SONUÇLARI (EXPERIMENT RESULTS)

Bu bölümde, Türkçe ve İngilizce veri kümelerinde gerçekleştirilen deneylerin, sınıflandırma doğrulukları doküman gösterim tekniklerine göre Tablo 1, Tablo 2, Tablo 3 ve Tablo 4'te gösterilmiştir. Tablo 1 ve Tablo 2'de bireysel derin öğrencilerin sınıflandırma doğrulukları, Tablo 3 ve Tablo 4'te ise derin topluluk öğrencilerin sınıflandırma doğrulukları sunulmuştur. Tablolar içerisindeki kalın yazı tipi ile belirtilen değerler, en iyi sonuçları göstermektedir. Tablolarda kullanılan bazı kısaltmaların açıklaması şu şekildedir: CNN: Evrimsel sinir ağı, RNN: Tekrarlayan sinir ağı, LSTM: Uzun kısa vadeli hafıza ağı, MV: Çoğunluk oylaması, STCK: İstifleme, TF-IDF: TF-IDF yöntemiyle elde edilen doküman vektörü, Avg(Word2vec): Word2vec ile elde edilen kelime gömülmesinin ortalamalarının alınmasıyla oluşturulan doküman vektörü, Avg(Glove): Glove ile elde edilen kelime gömülmesinin ortalamalarının alınmasıyla oluşturulan doküman vektörü, TF-IDF+Avg(Word2vec): TF-IDF ve Avg(Word2vec) doküman vektörlerinin birleşiminden oluşan doküman vektörü, TF-IDF+Avg(Glove): TF-IDF ve Avg(Glove) doküman vektörlerinin birleşiminden oluşan doküman vektörü, Avg(Word2vec)+Avg(Glove): Avg(Word2vec) ve Avg(Glove) doküman vektörlerinin birleşiminden oluşan doküman vektörüdür.

Tablo 1. Türkçe veri kümesi için bireysel derin öğrencilerin doküman gösterim yöntemlerine göre sınıflandırma doğrulukları (Classification accuracies of deep base learners for Turkish dataset in terms of document representation methods)

Doküman Gösterim Yöntemi	Bireysel Sınıflandırıcılar		
	CNN	RNN	LSTM
TF-IDF	71,15	69,44	73,84
Avg(Word2vec)	74,87	71,36	76,02
Avg(Glove)	73,24	70,85	75,37
TF-IDF+Avg(Word2vec)	76,41	74,13	77,56
TF-IDF+Avg(Glove)	75,38	73,01	77,03
Avg(Word2vec)+Avg(Glove)	77,13	75,40	78,42
Ortalama	74,70	72,37	76,37

Türkçe veri kümesinde bireysel derin öğrencilerin sınıflandırma başarısı, her bir doküman gösterim tekniğine göre Tablo 1'de sunulmuştur. Bu tabloya göre ortalama doğruluk değerleri dikkate alındığında en iyi sınıflandırma performansının %76,73 ile LSTM sınıflandırıcıya ait olduğunu gözlenmiştir. LSTM'yi sırasıyla CNN ve RNN temel öğrencileri takip etmektedir. Bunun yanı sıra, doküman gösterim tekniklerinin sınıflandırıcının performansını önemli ölçüde etkilediği gözlenmektedir. TF-IDF yöntemine dayanan bir gösterim ile LSTM en kötü sınıflandırma performansına sahipken Avg(Word2vec)+Avg(Glove) gösterim tekniği ile LSTM sınıflandırıcı harmanlandığında sınıflandırma başarısının yaklaşık %5 oranında arttığı görülmektedir. Dahası, doküman gösterim tekniklerinin TF-IDF+Avg(Word2vec),

TF-IDF+Avg(Glove), Avg(Word2vec)+Avg(Glove) şeklinde birleştirilmesi tek başına kullanılan doküman gösterim tekniklerine göre daha iyi bir sınıflandırma başarısı elde edilmesini sağlamaktadır. Böylece, bireysel sınıflandırıcıları en iyi performansa sahip doküman gösterim tekniğiyle harmanlamak, sistemin sınıflandırma performansını arttırmada kayda değer bir öneme sahiptir. Doküman gösterim yöntemlerinin Türkçe veri kümesi için genel sınıflandırma başarısı şu şekildedir: Avg(Word2vec)+Avg(Glove)>TF-IDF+Avg(Word2vec)>TF-IDF+Avg(Glove)>Avg(Word2vec)>Avg(Glove)>TF-IDF.

Tablo 2. İngilizce veri kümesi için bireysel derin öğrencilerin doküman gösterim yöntemlerine göre sınıflandırma doğrulukları

(Classification accuracies of deep base learners for English dataset in terms of document representation methods)

Doküman Gösterim Yöntemi	Bireysel Sınıflandırıcılar		
	CNN	RNN	LSTM
TF-IDF	72,54	68,97	74,09
Avg(Word2vec)	75,60	71,24	76,70
Avg(Glove)	74,84	70,96	76,10
TF-IDF+Avg(Word2vec)	76,91	73,27	78,07
TF-IDF+Avg(Glove)	76,44	72,80	77,62
Avg(Word2vec)+Avg(Glove)	78,12	75,00	79,66
Ortalama	75,74	72,04	77,04

Tablo 2'de, İngilizce veri kümesi üzerinde doküman gösterim tekniklerinin ve bireysel sınıflayıcıların performansı sunulmuştur. Avg(Word2vec)+Avg(Glove) yöntemi ile LSTM sınıflandırıcı kullanıldığında %79,66 doğruluk değeri, CNN sınıflandırıcıda %78,12, RNN sınıflandırıcıda ise %75,00 doğruluk değeri elde edilmiştir. Herhangi bir doküman gösterim tekniği seçildikten sonra bireysel derin öğrencilerin sınıflandırma performansı gözlemlendiğinde LSTM > CNN > RNN sıralaması elde edilmiştir. Bu durumda, en yüksek sınıflandırma performansına sahip bireysel derin öğrenci (LSTM) ile en düşük sınıflandırma başarısına sahip sınıflandırıcı (RNN) arasındaki fark, %4 ila %6 arasında değişiklik göstermektedir. Ortalama doğruluk değerlerine bakıldığında, %77,04 ile en yüksek sınıflandırma başarısına LSTM sınıflandırıcının sahip olduğu gözlenmektedir. Doküman gösterim yöntemlerinin sınıflandırma başarısı Türkçe veri kümesi üzerinde gözlenen sınıflandırma performansı ile aynıdır. Bireysel derin sınıflandırıcıların İngilizce veri kümesi üzerinde yapılan deney sonuçları dikkate alındığında Türkçe veri kümesi ile gerçekleştirilen deneylere nazaran yaklaşık %1 daha iyi bir sınıflandırma performansı sergilediği görülmüştür.

Tablo 3 ve Tablo 4'te derin topluluk sınıflandırıcıların sırasıyla Türkçe ve İngilizce dillerinde doküman gösterim tekniklerine göre sınıflandırma performansları verilmiştir. Türkçe veri kümesi için Tablo 3 incelendiğinde, derin topluluk sisteminin bireysel sınıflandırıcılara oranla çok

Tablo 3. Türkçe veri kümesi için derin topluluk öğrencilerin doküman gösterim yöntemlerine göre sınıflandırma doğrulukları (Classification accuracies of deep base learners for Turkish dataset in terms of document representation methods)

Doküman Gösterim Yöntemi	Bireysel Sınıflandırıcılar			Topluluk Sınıflandırıcılar	
	CNN	RNN	LSTM	MV	STCK
TF-IDF	71,15	69,44	73,84	74,88	75,06
Avg(Word2vec)	74,87	71,36	76,02	77,43	78,31
Avg(Glove)	73,24	70,85	75,37	76,36	77,44
TF-IDF+Avg(Word2vec)	76,41	74,13	77,56	78,91	79,17
TF-IDF+Avg(Glove)	75,38	73,01	77,03	77,50	78,65
Avg(Word2vec)+Avg(Glove)	77,13	75,40	78,42	78,44	79,82
Ortalama	74,70	72,37	76,37	77,25	78,07

Tablo 4. İngilizce veri kümesi için derin topluluk öğrencilerin doküman gösterim yöntemlerine göre sınıflandırma doğrulukları

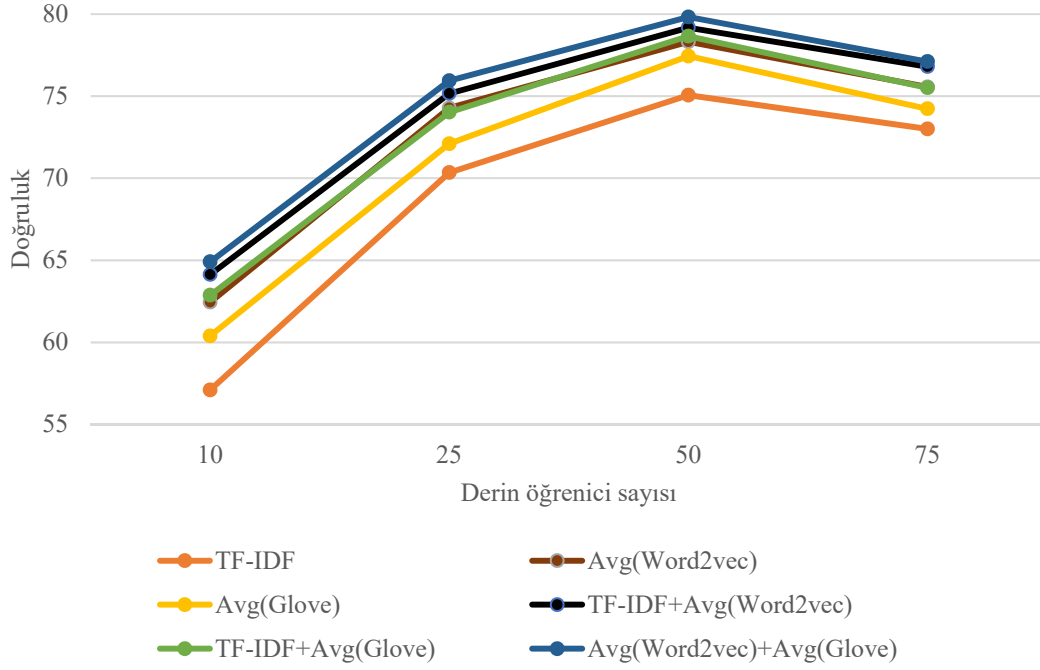
(Classification accuracies of deep base learners for English dataset in terms of document representation methods)

Doküman Gösterim Yöntemi	Bireysel Sınıflandırıcılar			Topluluk Sınıflandırıcılar	
	CNN	RNN	LSTM	MV	STCK
TF-IDF	72,54	68,97	74,09	75,13	76,81
Avg(Word2vec)	75,60	71,24	76,70	77,70	78,37
Avg(Glove)	74,84	70,96	76,10	76,94	77,25
TF-IDF+Avg(Word2vec)	76,91	73,27	78,07	79,42	79,96
TF-IDF+Avg(Glove)	76,44	72,80	77,62	78,12	78,85
Avg(Word2vec)+Avg(Glove)	78,12	75,00	79,66	80,24	82,27
Ortalama	75,74	72,04	77,04	77,93	78,92

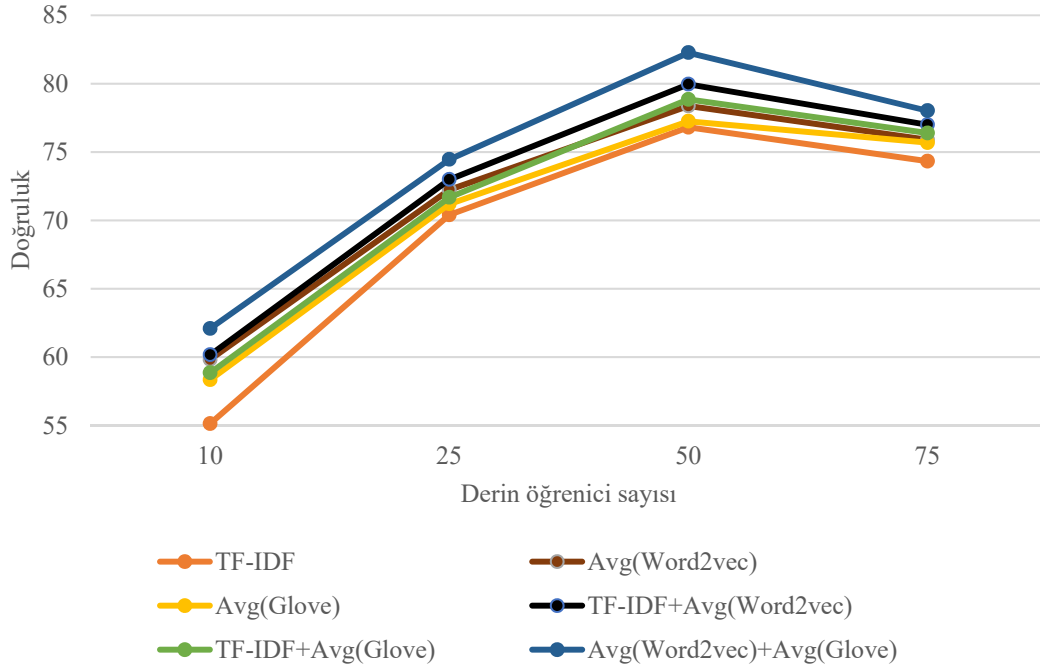
daha iyi bir sınıflandırma performansı sergilediği gözlenmektedir. Topluluk sınıflandırmasını elde etmede kullanılan yöntemlerden biri olan MV ortalama değerlere bakıldığında CNN ve RNN'ye göre daha iyi bir performans sunarken LSTM ile çok yakın bir sınıflandırma başarısına sahip olduğu gözlenmektedir. Özellikle, Avg(Word2vec)+Avg(Glove) tekniği seçildiğinde MV ile LSTM arasındaki sınıflandırma başarısı neredeyse aynıdır. STCK yöntemi ile heterojen derin sınıflandırıcıların kararlarını harmanlamak, sistemin performansının ciddi anlamda arttırmaktadır. STCK hem ortalama doğruluk değerleri hem de doküman gösterim yöntemlerine göre sınıflandırma başarısı dikkate alındığında en yüksek sınıflandırma başarısına sahip olan yöntemdir. MV'ye göre %1 daha iyi bir performans sergilerken bireysel sınıflandırıcılar düşünüldüğünde %2 - %4 arasında sistemin performansını iyileştirdiği görülmektedir. Derin topluluk modeli sınıflandırıcılarının kararlarını birleştirmek için kullanılan yöntemlerin performansları incelendiğinde çoğunlukla Avg(Word2vec)+Avg(Glove) doküman gösterim tekniği, bireysel derin öğrencilerde olduğu gibi burada da en başarılı yöntem olarak gözlenmektedir. Genelden farklı olarak TF-IDF+Avg(Word2vec) doküman gösterim tekniği MV için %78,91 ile en yüksek sınıflandırma başarısına sahip olup onu, Avg(Word2vec)+Avg(Glove) %78,44 ile takip etmektedir. Tablo 4, İngilizce veri kümesinde derin topluluk sınıflandırıcılarının doküman gösterim tekniklerine göre sınıflandırma performanslarını

göstermektedir. Ortalama sınıflandırma başarıları dikkate alındığında STCK %78,92 doğruluk değeri ile ilk sıradadır. STCK, yaklaşık olarak MV'ye göre %1, LSTM'ye göre %1, CNN'e göre %3, RNN'e göre %5 daha iyi bir sınıflandırma başarısına sahiptir. Avg(Word2vec)+Avg(Glove) yönteminin, diğer doküman gösterim teknikleri ile kıyaslandığında önemli ölçüde üstün olduğu gözlenmektedir. Topluluk sınıflandırıcılar dikkate alındığında sistemin sınıflandırma performansı şu şekilde sıralanabilir: STCK > MV > LSTM > CNN > RNN. Doküman gösterim tekniklerinin sınıflandırma başarısı İngilizce veri kümesi için Avg(Word2vec)+Avg(Glove) > TF-IDF+Avg(Word2vec) > TF-IDF+Avg(Glove) > Avg(Word2vec) > Avg(Glove) > TF-IDF şeklinde olup Türkçe veri kümesi için bu sıralama sadece MV yönteminde değişiklik göstermektedir.

Şekil 2 ve 3, doküman gösterim yöntemlerine göre derin öğrenci sayılarının değişmesi durumunda doğruluk sonuçlarını yansıtmaktadır. Şekil 2, Türkçe veri kümesi için Şekil 3 ise İngilizce veri kümesi için sınıflandırma performansını sunmaktadır. Şekil 2 ve 3 detaylı incelendiğinde derin öğrenci sayısının artması sınıflandırma başarısını olumlu yönde etkilemektedir. Derin öğrenci sayısının 50 olduğu durumda her bir doküman gösterim tekniğinde en iyi sınıflandırma performansı elde edilmektedir. Derin öğrenci sayısının 50'den sonra da artması durumunda ise doğruluk değerlerinin düştüğü



Şekil 2. Türkçe veri kümesi için derin öğrenici sayılarının doküman gösterim yöntemleri açısından doğruluk sonuçları (Accuracy results of the number of deep learners in terms of document representation methods for Turkish dataset)



Şekil 3. İngilizce veri kümesi için derin öğrenici sayılarının doküman gösterim yöntemleri açısından doğruluk sonuçları (Accuracy results of the number of deep learners in terms of document representation methods for English dataset)

gözlenmektedir. Bu sebeple, deneylerimizde derin öğrenici sayısı 50 olarak belirlenmiştir. İki şekilde de açıkça görüldüğü üzere derin öğrenici sayısının artması ve azalması durumunda Avg(Word2vec)+Avg(Glove) yöntemi diğer yöntemlere nazaran çok daha iyi bir sınıflandırma performansına sahiptir.

Onu sırasıyla TF-IDF + Avg(Word2vec), TF-IDF + Avg(Glove), Avg(Word2vec), Avg(Glove) ve TF-IDF yöntemleri izlemektedir. Hem İngilizce hem de Türkçe veri kümesinde TF-IDF+Avg(Glove) ve Avg(Word2vec) yöntemleri birbirine çok yakın sınıflandırma başarıları sergilemektedirler. Dolayısıyla, dilden bağımsız olarak TF-

IDF yöntemi Avg(Glove) ile zenginleştirildiğinde sadece Avg(Word2vec) kullanımına göre ekstra bir sınıflandırma performansı gösterememiştir. Özellikle derin öğrenici sayısı 25'in üzerindeyse bu fark, yok denecek kadar azdır. Bu sebeple, sistem performansı düşünüldüğünde Avg(Word2vec) yöntemini doküman gösterim tekniği olarak kullanmak daha avantajlı olacaktır.

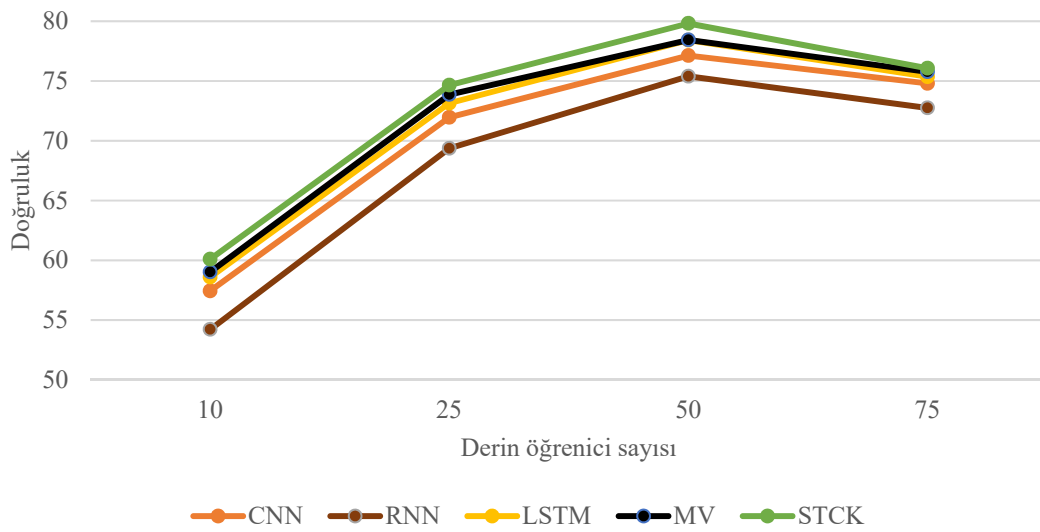
Bir diğer dikkat çeken sonuç ise Türkçe veri kümesinde derin öğrenici sayısı arttıkça doküman gösterim teknikleri arasındaki sınıflandırma performansı farkı artmasına rağmen İngilizce veri kümesinde bu durumun tersi gözlenmektedir. Şekil 3'te de gözlendiği gibi özellikle derin öğrenici sayısı 25'in üzerinde iken Avg(Word2vec)+Avg(Glove) yöntemi hariç doküman gösterim tekniklerinin başarısı birbirine yaklaşmaktadır. Sonuç olarak, derin öğrenici sayısı sınıflandırma performansını önemli ölçüde etkilemektedir. Öğrenici sayısının değişkenlik gösterdiği aralıkta en yüksek sınıflandırma performansına sahip doküman gösterim yöntemi Avg(Word2vec)+Avg(Glove) iken en düşük sınıflandırma başarısı TF-IDF yöntemiyle elde edilmektedir.

Şekil 4 ve 5, sırasıyla Türkçe ve İngilizce veri kümeleri için derin öğrenici sayılarının değişmesi durumunda her bir derin öğrenicinin sınıflandırma performansını değerlendirmektedir. Her iki veri kümesi için de en iyi sınıflandırma performansı STCK ile sağlanmaktadır. En düşük sınıflandırma başarısı ise RNN derin öğrenicisiyle elde edilmektedir. Bu sebeple, sistemi bireysel sınıflandırıcılarla ayrı ayrı eğitmektense topluluk sınıflandırıcılarla eğitmek, sınıflandırma performansı açısından avantaj sağlar. Dahası, Şekil 2 ve 3'te olduğu gibi burada da en iyi sınıflandırma performansı derin öğrenici sayısı 50 olarak ayarlandığında elde edilmektedir. Derin öğrenici sayısının değiştiği aralıkta genellikle bireysel derin öğrenicilerin sınıflandırma başarısı STCK > MV > LSTM > CNN > RNN şeklindedir.

Şekil 4'te derin öğrenici sayısı arttıkça özellikle LSTM ve MV'nin performanslarının birbirlerine çok yakın olduğu gözlenmektedir. Dolayısıyla, topluluk kararlarını birleştirmek için sadece MV yöntemi kullanılacaksa sistem performansı düşünüldüğünde sergilediği sınıflandırma başarısından ötürü LSTM'yi bireysel sınıflayıcı olarak seçmek uygun olacaktır. Böyle bir durumda, topluluk öğrenmesi yerine bireysel öğrenme, sistemin performansı için idealdir. Bunun dışında, Türkçe veri kümesinde derin öğrenicilerin sınıflandırma performansı, az sayıda öğreniciyle gerçekleştirildiğinde birbirine daha yakinken derin öğrenici sayısı belli bir yere kadar artırıldığında sınıflandırma performansları arasındaki fark genellikle artmaktadır.

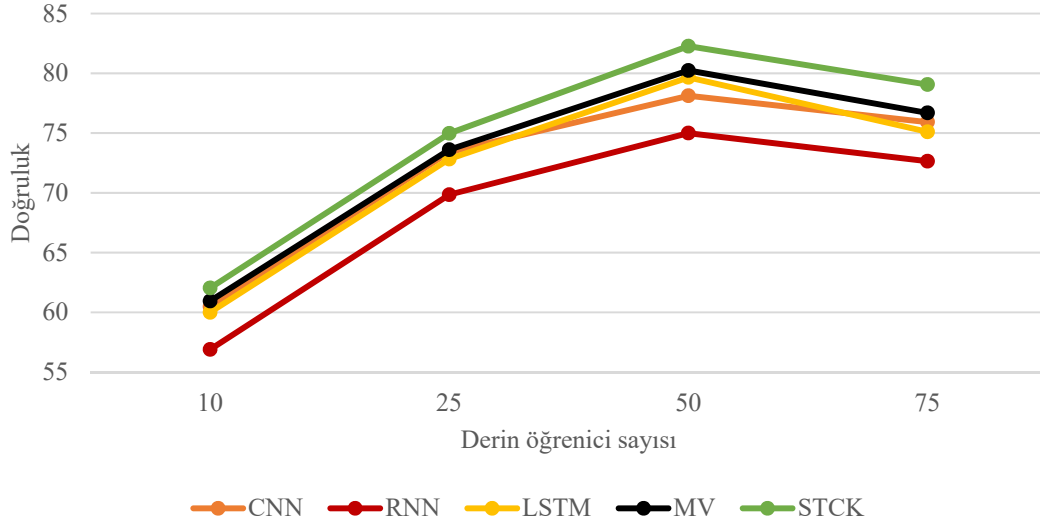
Şekil 5'te, Şekil 4'ten farklı olarak az sayıda derin sınıflandırıcı kullanıldığında öğrenici sayısı genellikle artsa da sınıflandırma performansları en yüksek ve en düşük olanlar hariç birbirine çok yakındır. Aralarındaki sınıflandırma başarısı farkının en belirgin olduğu durum, derin öğrenici sayısı 50 olarak ayarlandığında sağlanmaktadır. Sınıflandırma performansı açısından CNN ve LSTM arasındaki fark, özellikle derin öğrenici sayısı 50 haricinde bir değer olarak belirlendiğinde yok denecek kadar azdır. MV, STCK kadar belirgin bir sınıflandırma başarısına sahip olmamakla birlikte özellikle öğrenici sayısı 10 ve 50 aralığındayken CNN ve LSTM ile çok daha iyi bir sınıflandırma performansı sergilemektedir.

Borsa yönünü tahmin etmek için kullanılan farklı veri kümeleri, doküman temsil yöntemleri ve topluluk yaklaşımlarının farklılığından ya da olmamasından dolayı sonuçlarımızın performansını diğer çalışmalarla karşılaştırmak zordur. Borsa İstanbul'un yönünü belirlemek için finansal haberleri kullanan bir çalışmada [30], derin öğrenme ve topluluk algoritmaları kullanılmamasına rağmen



Şekil 4. Türkçe veri kümesi için derin öğrenici sayılarının bireysel derin öğreniciler ve topluluk derin öğrenicileri açısından doğruluk sonuçları

(Accuracy results of the number of deep learners in terms of individual deep learners and deep ensemble learners for Turkish dataset)



Şekil 5. İngilizce veri kümesi için derin öğrenici sayılarının bireysel derin öğreniciler ve topluluk derin öğrenicileri açısından doğruluk sonuçları

(Accuracy results of the number of deep learners in terms of individual deep learners and deep ensemble learners for English dataset)

çalışmamıza yakınlık göstermektedir. Bu çalışmada, doküman gösterim tekniği olarak sadece TF-IDF kullanılmış olup sınıflandırma işlemi yalnızca naif Bayes algoritması ile gerçekleştirilmiştir. Ayrıca, bu çalışmada herhangi bir topluluk stratejisi uygulanmamış olup veri kümesi Türkçe finansal haberlerden oluşmaktadır. Bu açıdan bakıldığında yazarların önerdiği yöntem, %75,00 doğruluk başarısı sağlarken bizim çalışmamızda önerilen yöntem ile Türkçe veri kümesinde %78,07, İngilizce veri kümesinde ise %78,92 doğruluk değerlerine ulaşılmaktadır. Diğer bir çalışma [18], borsa yönünü tahmin etmek için derin öğrenme algoritmalarından LSTM'nin performansını Türkçe finansal haberler ile ölçmeyi amaçlamıştır. Bu çalışmadan farklı olarak derin öğrenme algoritmalarına uygulanan topluluk stratejileri ve farklı doküman gösterim teknikleri ile çalışmamız, Türkçe veri kümesinde %77,81 F-ölçüm değerine ulaşırken bu çalışma [18] ile %53,4 F-ölçüm değeri elde edilmiştir.

Bir diğer çalışmada [54] ise metin içerikli veriler yerine sayısal veriler kullanılarak borsa yönünün tahmin edilmesi hedeflenmiştir. Bu amaçla, destek vektör makineleri ve yapay sinir ağları kullanarak önerilen yöntemle %75,74 sınıflandırma başarısı elde etmişlerdir. Bizim çalışmamız, Türkçe veri kümesi için %78,07, İngilizce veri kümesi için %78,92 doğruluk değerleriyle daha iyi bir sınıflandırma başarısı elde etmektedir. Derin öğrenicilerin topluluk sınıflandırmasında kullanılması, gerekli deneysel ayarlamalar yapıldığında hem bireysel derin öğrenicilerden hem de bilinen geleneksel topluluk sınıflandırma yöntemlerinden çok daha iyi sonuçlar sunmaktadır. Böylece, bu çalışmada kullanılan derin topluluk modeli, hem duygu analizi çalışmalarına hem de borsa tahmini alanında yapılan çalışmalara sınıflandırma performansını iyileştirmek amacıyla katkı sağlayacak ve araştırmacılara ilham kaynağı olacaktır.

6. SONUÇLAR (CONCLUSIONS)

Bu çalışma kapsamında borsa yönünün tahmin edilmesi amacıyla kullanılacak derin topluluk modelleri geliştirilmiştir. Geliştirilen sistemde metin içerikli veri kümeleri, farklı doküman gösterim teknikleri kullanılarak anlam, bağlam ve söz dizimi açısından zenginleştirilmiştir. Dahası, topluluk öğrenmesi yaklaşımı derin öğrenme teknikleriyle harmanlanarak sınıflandırma performansı açısından önemli bir avantaj elde edilmektedir. Bu çalışma, Borsa İstanbul endeksinin yönünü belirlemek için Twitter ortamını kullanması, farklı doküman gösterim tekniklerini önerilen metodolojilerle harmanlaması ve topluluk sınıflandırmasını derin öğrenme algoritmalarıyla değerlendirmesi açısından bildiğimiz kadarıyla literatürdeki ilk çalışma niteliğindedir. Bu çalışma, Türkçe veri kümesi için %78,07, İngilizce veri kümesi için %78,92 sınıflandırma sonuçları ile benzer çalışmalara kıyasla daha yüksek bir sınıflandırma başarısı sergilemektedir.

KAYNAKLAR (REFERENCES)

1. Leung M.T., Daouk H., Chen A.S., Forecasting stock indices: A comparison of classification and level estimation models, *International Journal of Forecasting*, 16, 173-190, 2000.
2. Manish K. ve Thenmozhi M., Forecasting stock index movement: A comparison of support vector machines and random forest, *Indian Institute of Capital Markets Conference*, Hindistan, 20-36, 2005.
3. Abu-Mostafa Y.S. ve Atiya A.F., Introduction to financial forecasting, *Applied Intelligence*, 16 (3), 205-213, 1996.
4. Tan T.Z., Quek C., See N.G., Biological brain-inspired genetic complementary learning for stock market and

- bank failure prediction, *Computational Intelligence*, 23 (2), 236-261, 2007.
5. Goonatilake R. ve Herath S., The volatility of the stock market and news, *International Research Journal of Finance and Economics*, 3 (11), 53-65, 2007.
 6. Young T., Hazarika D., Poria S., Cambria, E., Recent trends in deep learning based natural language processing, *IEEE Computational Intelligence Magazine*, 13 (3), 55-75, 2018.
 7. Mikolov T., Chen K., Corrado G., Dean J., Efficient estimation of word representations in vector space, *International Conference on Learning Representations*, Arizona, 1-12, 2013.
 8. Mikolov T., Sutskever I., Chen K., Corrado G, Dean J., Distributed representations of words and phrases and their compositionality, *Neural Information Processing Systems Conference*, Lake Tahoe, 3111-3119, 2013.
 9. Pennington J., Socher R., Manning C., GloVe: Global vectors for word representation, *Empirical Methods in Natural Language Processing Conference*, Katar, 1532-1543, 2014.
 10. Brown G., Wyatt J.L., Tino P., Managing diversity in regression ensembles, *Journal of Machine Learning Research*, 6, 1621-1650, 2005.
 11. Rokach L., Ensemble-based classifiers, *Artificial Intelligence Review*, 33, (1-2), 1-39, 2010.
 12. Polikar R., Ensemble based systems in decision making, *IEEE Circuits and Systems Magazine*, 6 (3), 21-45, 2006.
 13. Gopika D. ve Azhagusundari B., An analysis on ensemble methods in classification tasks, *International Journal of Advanced Research in Computer and Communication Engineering*, 3 (7), 7423-7427, 2014.
 14. Ren Y., Zhang L., Suganthan P. N., Ensemble classification and regression-recent developments, applications and future directions, *IEEE Computational Intelligence Magazine*, 11 (1), 41-53, 2016.
 15. Mangai U. G., Samanta S., Das S., Chowdhury P. R., A survey of decision fusion and feature fusion strategies for pattern classification, *IETE Technical Review*, 27 (4), 293-307, 2010.
 16. Woźniak M., Graña M., Corchado E., A survey of multiple classifier systems as hybrid systems, *Information Fusion*, 16, 3-17, 2014.
 17. Tsoumakas G., Angelis L., Vlahavas I., Selective fusion of heterogeneous classifiers, *Intelligent Data Analysis*, 9 (6), 511-525, 2005.
 18. Gündüz H., Yaslan Y., Çataltepe Z., Finansal haberler kullanılarak derin öğrenme ile borsa tahmini, *IEEE Sinyal İşleme Ve İletişim Uygulamaları Kurultayı*, İzmir, 1-4, 2018.
 19. Ghosal D., Bhatnagar S., Akhtar M.S., IITP at SemEval-2017 Task 5: An ensemble of deep learning and feature based models for financial sentiment analysis, *International Workshop on Semantic Evaluations*, Canada, 899-903, 2017.
 20. Warikoo N., Chang Y.C., Dai H.J., Hsu W.L., An ensemble neural network model for benefiting pregnancy health stats from mining social media, *Asia Information Retrieval Symposium*, Taiwan, 3-15, 2018.
 21. Liao S., Wang J., Yu R., Sato K., Cheng Z., CNN for situations understanding based on sentiment analysis of twitter data, *Procedia Computer Science*, 111, 376-381, 2017.
 22. Santos C. N., Gatti M., Deep convolutional neural networks for sentiment analysis of short texts, *International Conference on Computational Linguistics*, Ireland, 69-78, 2014.
 23. Hu F., Li L., Zhang Z., Wang J., Xu X., Emphasizing essential words for sentiment classification based on recurrent neural networks, *Journal of Computer Science and Technology*, 32 (4), 785-795, 2017.
 24. Chen Q., Guo Z., Sun C., Li W., Research on Chinese micro-blog sentiment classification based on recurrent neural network, *International Conference on Computer Science and Technology*, China, 859-867, 2017.
 25. Zhao Z., Lu H., Cai D., He X., Zhuang Y., Microblog sentiment classification via recurrent random walk network learning, *International Conference on Artificial Intelligence*, Australia, 3532-3538, 2017.
 26. Becker W., Wehrmann J., Cagnini H.E.L., Barros R.C., An efficient deep neural architecture for multilingual sentiment analysis in Twitter, *International Conference on Florida Artificial Intelligence Research Society*, Florida, 246-251, 2017.
 27. Uysal A.K., Murphey Y.L., Sentiment classification: Feature selection based approaches versus deep learning, *IEEE International Conference on Computer and Information Technology*, Finland, 23-30, 2017.
 28. Nozza D., Fersini E., Messina E., Deep learning and ensemble methods for domain adaptation, *International Conference on Tools with Artificial Intelligence*, USA, 184-189, 2011.
 29. Araque O., Corcuera-Platas I., Sánchez-Rada J.F., Iglesias C.A., Enhancing deep learning sentiment analysis with ensemble techniques in social applications, *Expert Systems and Applications*, 77, 236-246, 2017.
 30. Gündüz H., Çataltepe Z., Borsa İstanbul (BIST) daily prediction using financial news and balanced feature selection, *Expert Systems with Applications*, 42, 9001-9011, 2015.
 31. Chaurasia V., Pal S., Data mining techniques: To predict and resolve breast cancer survivability, *International Journal of Computer Science and Mobile Computing*, 3 (1), 10-22, 2014.
 32. Uysal A.K., Gunal S., The impact of preprocessing on text classification, *Information Processing and Management*, 50 (1), 104-112, 2014.
 33. Zheng Z., Wu X., Srihari R., Feature selection for text categorization on imbalanced data, *SIGKDD Explorations*, 6 (1), 80-89, 2004.
 34. Young T., Hazarika D., Poria S., Cambria E., Recent Trends in Deep Learning Based Natural Language Processing, *IEEE Computational Intelligence Magazine*, 13 (3), 55-75, 2018.

35. Lecun Y., Bottou L., Bengio Y., Haffner P., Gradient-based learning applied to document recognition, *Proceedings of the IEEE*, 86 (11), 2278-2324, 1998.
36. Schmidhuber J., Deep learning in neural networks: An overview, *Neural Networks*, 61, 85–117, 2015.
37. LeCun Y., Bengio Y., Hinton G., Deep learning, *Nature*, 521, 436–444, 2015.
38. Johnson R. ve Zhang T., Effective use of word order for text categorization with convolutional neural networks, *Annual Conference of the North American Chapter of the Association for Computational Linguistics, USA*, 20-30, 2015.
39. Graves A. ve Jaitly N., Towards end-to-end speech recognition with recurrent neural networks, *International Conference on Machine Learning, China*, 1764–1772, 2014.
40. Karpathy A. ve Fei-Fei L., Deep visualsemantic alignments for generating image descriptions, *IEEE Conference on Computer Vision and Pattern Recognition, USA*, 3128–3137, 2015.
41. Wang P., Xu B., Xu J., Tian G., Liu C.L., Hao H., Semantic expansion using word embedding clustering and convolutional neural network for improving short text classification, *Neurocomputing*, 174, 806-814, 2016.
42. Graves A. ve Schmidhuber J., Framewise phoneme classification with bidirectional LSTM and other neural network architectures, *Neural Networks*, 18 (5-6), 602–610, 2005.
43. Graves A., Mohamed A., Hinton G., Speech recognition with deep recurrent neural networks, *IEEE International Conference on Acoustics, Speech and Signal Processing, Canada*, 6645–6649, 2013.
44. Fernández S., Graves A., Schmidhuber J., An application of recurrent neural networks to discriminative keyword spotting, *International Conference on Artificial Neural Networks, Portugal*, 220–229, 2007.
45. Baccouche M., Mamalet F., Wolf C., Garcia C., Baskurt A., *Sequential deep learning for human action recognition*, Springer, Berlin, Heidelberg, 29–39, 2011.
46. Schmidhuber J., Gers F., Eck D., Learning nonregular languages: A comparison of simple recurrent networks and LSTM, *Neural Computation*, 14 (9), 2039–2041, 2002.
47. Džeroski S. ve Ženko B., Is combining classifiers with stacking better than selecting the best one?, *Machine Learning*, 54 (3), 255-273, 2004.
48. Adnan M.N., Islam M.Z., Comprehensive method for attribute space extension for random forest, *International Conference on Computer and Information Technology, Bangladesh*, 25–29, 2014.
49. Amasyalı M.F., Ersoy O.K., Classifier ensembles with the extended space forest, *IEEE Transactions on Knowledge and Data Engineering*, 26 (3), 549–562, 2014.
50. Kilimci Z.H., Akyokus S., Omurca S.İ., The evaluation of heterogeneous classifier ensembles for Turkish texts, *IEEE International Conference on INnovations in Intelligent Systems and Applications, Poland*, 307-311, 2017.
51. Kilimci Z.H., Akyokus S., Deep Learning-and Word Embedding-Based Heterogeneous Classifier Ensembles for Text Classification, *Complexity*, 2018, 1-10, 2018.
52. Kanakaraj M. ve Guddeti R.M.R., Performance analysis of Ensemble methods on Twitter sentiment analysis using NLP techniques, *IEEE International Conference on Semantic Computing, USA*, 169-170, 2015.
53. Turkish Pre-trained Word2vec Model, <https://github.com/akoksal/Turkish-Word2Vec>
54. Kara Y., Boyacıoğlu M.A., Baykan Ö.K., Predicting direction of stock price index movement using artificial neural networks and support vector machines: The sample of the Istanbul Stock Exchange, *Expert Systems with Applications*, 38 (5), 5311-5319, 2011.
55. Arı A. ve Hanbay D., Tumor detection in MR images of regional convolutional neural networks, *Journal of the Faculty of Engineering and Architecture of Gazi University*, 34 (3), 1395-1408, 2019.
56. Kaya A., Keçeli A.S., Can A.B., Examination of various classification strategies in classification of lung nodule characteristics, *Journal of the Faculty of Engineering and Architecture of Gazi University*, 34 (2), 709-725, 2019.
57. Bulut F., A new clinical decision support system with instance based ensemble classifiers, *Journal of the Faculty of Engineering and Architecture of Gazi University*, 32 (1), 65-76, 2017.
58. Bulut F. ve Amasyalı M.F., Katı kümeleme ve yeni bir geçiş fonksiyonuyla uzman karışımlarında sınıflandırma, *Journal of the Faculty of Engineering and Architecture of Gazi University*, 31 (4), 1017-1025, 2016.