



Twitter Üzerindeki Finansal Tweetlerin LSTM Sinir Ağı Algoritması ile Duygu Analizi

Hakan YUSUFOĞLU¹, Hakan AYDIN¹, Ali ÇETİNKAYA^{2*}

¹*İstanbul Gelişim Üniversitesi, Mühendislik ve Mimarlık Fakültesi, Bilgisayar Mühendisliği Bölümü, İstanbul*

²*İstanbul Gelişim Üniversitesi, Teknoloji Transfer Ofisi Uygulama ve Araştırma Merkezi, İstanbul*

Özet

Bu çalışmanın amacı, duygu analizi kullanılarak LSTM modeli üzerinde EUR/USD değerlerini tahmin edebilen derin öğrenmeye dayalı bir model geliştirmek olarak belirlenmiştir. Bu amaç doğrultusunda çalışmada veri setlerinin belirlenmesi ve modelin eğitimi/testi olmak üzere iki farklı uygulama gerçekleştirilmiştir. Çalışmada gerçekleştirilen ilk uygulama veri setinin oluşturulmasını içermektedir. İlk defa bu çalışma kapsamında oluşturduğumuz ve kullandığımız söz konusu veri setinin oluşturulmasında ilk adım olarak 2005 yılından günümüze kadar Twitter üzerindeki “#eurusd” konu etiketli veya “eurusd” kelimesini barındıran filtrelenmiş İngilizce tweet mesajları içeriklerine göre duygu analizi yöntemiyle sınıflandırılmıştır. Bu maksatla her bir tweetin olumlu, olumsuz veya tarafsız (nötr) olup olmadığı belirlenmiş ve günlük olarak bu tweetlerin duygu oranlarının ortalamaları hesaplanmıştır. Veri setinin oluşturulmasındaki ikinci adım finance.yahoo.com sitesinden günlük olarak EUR/USD değerlerinin elde edilmesi aşamasıdır. Veri setinin oluşturulmasındaki son aşama ise elde edilen verilerin zaman serisine dayalı olarak “csv” formatında bir veri seti yapısında birleştirilmesidir. Böylelikle yüksek tahmin başarı oranına sahip LSTM modelinin eğitilmesi ve testi aşamalarında kullanılacak olan zaman serisi verilerine dayalı veri seti elde edilmiştir. Veri seti belirlendikten sonra LSTM tahmin modelinin eğitimi ve testi aşamasına geçilmiştir. Çalışmada yapılan ikinci uygulama LSTM modelinin eğitimi/testi aşamalarını içermektedir. Bu aşamada en yüksek başarı oranına sahip LSTM tahmin modelini belirlemek amacıyla eğitim adımlarının ve LSTM sinir ağı katman yapısının değiştirilmesi suretiyle farklı deneyler yapılmıştır. Yapılan bu deneyler sonucunda %94,48 başarı oranına sahip en iyi LSTM modeli belirlenmiştir. Çalışma neticesinde elde edilen sonuçlar geliştirilen LSTM tahmin modelinin para piyasalarında EUR/USD hareket yönünün yüksek doğrulukta tahmin edilmesinde yardımcı bir araç olarak kullanılabileceği yönündedir. Araştırmanın sonuçları yapay zeka algoritmaları ile duygu analizi açısından önemli bilgiler içermesinin yanı sıra, gelecekte farklı sektörlerde uygulanması açısından değer taşımaktadır.

Anahtar Kelimeler: Derin öğrenme, LSTM, Finansal Analiz, EUR/USD Tahmini, Duygu Analizi.

Makale Bilgisi

Başvuru:

16/09/2021

Kabul:

29/11/2021

* İletişim e-posta: alcetinkaya@gelisim.edu.tr

Sentiment Analysis of Financial Twitter Posts on Twitter with the LSTM Neural Network Algorithm

Abstract

The aim of this study is to develop a Deep Learning based model that can predict EUR/USD values based on sentiment analysis on LSTM model. For this purpose, two different applications were carried out in the study, namely the determination of the data sets and the training/testing of the model. The first application carried out in the study involves the creation of the dataset. As the first step in creating the above dataset, which we created and used for the first time in this study, filtered English tweet messages with the hashtag "#eurusd" or the word "eurusd" on Twitter from 2005 to the present were classified according to their content using the sentiment analysis method. For this purpose, it was determined for each tweet whether it was positive, negative or neutral (neutral) and the average of the emotion scores of these tweets was calculated on a daily basis. The second step in creating the dataset is to obtain daily EUR/USD values from finance.yahoo.com. The last step in creating the dataset is to combine the obtained data in a dataset structure in "csv" format based on the time series. In this way, a dataset based on time series data was obtained to be used in the training and testing phase of the LSTM model and has a high prediction success rate. Once the dataset was determined, the training and testing phase of the LSTM prediction model was started. The second application in the study involves the training/testing phases of the LSTM model. In this phase, various experiments were conducted by changing the training steps and the LSTM layer structure of the neural network to determine the LSTM prediction model with the highest success rate. As a result of these experiments, the best LSTM model was found to have a success rate of 94.48%. The results of the study show that the developed LSTM forecasting model can be used as a helpful tool to estimate the movement direction of EUR/USD in money markets with high accuracy. The results of the study not only contain important information about artificial intelligence algorithms and sentiment analysis, but also are valuable for future application in various sectors.

Keywords: Deep learning, LSTM, Financial Analysis, EUR/USD Forecast, Sentiment Analysis

1 Giriş

Günümüzde derin öğrenme algoritmalarının nesne tanıma, sınıflandırma, tespit, tanı gibi pek çok alanın yanında finans alanında da kullanıldığı görülmektedir. Finans piyasalarında başarılı olabilmek için aktif piyasa koşullarının iyi yorumlanması ve analiz edilmesi önemlidir. Bu maksatla yatırımcılar titiz bir araştırma ve inceleme süreci için derin öğrenme algoritmaları gibi bilgisayar alanındaki güncel teknoloji ve uygulamalardan yararlanarak geleceğe dönük piyasa analizleri yapmaktadırlar. Ancak finans piyasalarındaki belirsizlik ortamları, piyasanın sürekli hareketli olması gibi sebeplerle geleceğe dönük piyasa analizlerinin yapılması güçtür [1]. Serbest piyasa koşullarında, döviz kurları serbest piyasa koşullarına göre değerlendirildiğinden, gelecekteki döviz kurlarının belirsizliği döviz kurlarının tahmin edilmesini zorlaştırmaktadır [2]. Enflasyon, ekonomik büyüme, ithalat ve ihracat eğilimleri gibi döviz kurlarının zaman içinde artışı veya düşüşünü etkileyen birçok faktör vardır [3].

Ekonomik anlamda, piyasada yatırım yapmak veya alım satım işlemleri neticesinde elde edilecek karlar büyük oranda öngörebilme, yani tahmin işlemlerine bağlıdır. EUR/USD kur değerlerinin doğru tahmin edilmesi ihracat ve ithalat maliyet ve oranlarının hesaplanması, ödeme planlarının oluşturulması ve döviz gelirlerinin ulusal para birimine çevrilmesi ve böylelikle yatırım hamleleri için para piyasalarının iyi bir şekilde yorumlanması açısından yüksek faydalar sağlayabilecektir. Bu gibi finans tahmin işlemleri günümüzde derin öğrenme algoritmaları sayesinde gerçekleştirilebilmektedir. Günümüzde derin öğrenme algoritmalarından finans alanında da faydalandığı görülmektedir.

Derin öğrenme algoritmaları, geçmiş verilere bakarak gelecekteki değerlerini tahmin etme imkanını sağlamaktadır. Öğrenme verilerini ifade etmeyi sağlayan derin öğrenme, ileriye dönük tahmin işlemlerinde büyük oranda kullanılmaktadır. Bu yaklaşım, geçmiş verileri kullanarak oluşturulmak istenen profili karakterize eden bir model oluşturma yeteneği sunmaktadır.

Derin öğrenme algoritmaları kullanılmak suretiyle gelecekteki finans değerlerini tahmin edebilmek için geliştirilecek olan bir derin öğrenme modeli finans verileri kullanarak eğitilebilir ve böylelikle finans değerlerinin tahmini için çalışmalar yapılabilir. Uzun Kısa Süreli Bellek (LSTM), zaman serisine dayalı verileri öğrenebilen bir derin öğrenme algoritmasıdır. LSTM sinir ağını zaman serisine dayalı veriler kullanılarak ileriye dönük tahmin işlemlerinde kullanmak mümkündür.

Bu çalışmanın amacı, duygu analizi kullanılarak LSTM modeli üzerinde EUR/USD değerlerini tahmin edebilen derin öğrenmeye dayalı bir model geliştirmek olarak belirlenmiştir.

Çalışmanın başlıca katkıları şu şekilde ifade edilebilir:

- Zaman serisine dayalı EUR/USD verileri ile tweetlerin duyarlılık analizi verileri kullanarak geleceğe dönük olarak EUR/USD hareket yönünü %94.48 yüksek doğruluk oranı tahmin edebilen bir LSTM sinir ağı modeli geliştirilmiştir.
- Duygu analizi kullanılarak 2005 yılından günümüze kadar olan tweetlerin duygu analizi yapılmış, böylelikle tweetler olumlu, olumsuz veya tarafsız (nötr) olarak sınıflandırılmış, elde edilen bu veriler EUR/USD değerleri ile zaman serisine dayalı olarak "csv" formatında birleştirilerek ilk defa bu çalışma kapsamında kullanılan veri seti oluşturulmuştur.
- Geliştirilen LSTM modelinin doğruluk tahmin oranını belirlemek ve sonuçları analiz edebilmek amacıyla çalışmada değişik deneyler yapılmıştır. Her bir deneyin doğruluk değeri elde edilen başarı oranları dikkate alınarak değerlendirilmiş ve analiz edilmiştir.

Bu çalışma şu şekilde düzenlenmiştir: Bölüm 2'de konuyla ilgili olarak araştırılan çalışmalar hakkında bilgi verilmiştir. Bölüm 3'te çalışmada kullanılan materyal ve yöntemler anlatılmıştır. Bölüm 4'te çalışma kapsamında gerçekleştirilen uygulamalar anlatılmıştır. Bölüm 5'te deneysel çalışmalar ile bu çalışmalardan elde edilen bulgular değerlendirilmiştir. Bölüm 6'da ise sonuç bölümü yer almaktadır.

2 İlgili çalışmalar

Günümüzde derin öğrenme algoritmaları ile sosyal medya duygu analizi [4], fiyat tahmini [5, 6, 7, 8], hisse senedi tahmini [9, 10], deprem tahmini [11], talep tahmini [12] gibi pek çok alanda çalışmalar yapıldığı görülmektedir. Derin öğrenme çalışmaları

incelendiğinde matematiksel işlemlerin yoğun olarak kullanıldığı uygulamalarda tercih edildiği görülmektedir. Bu bağlamda matematiksel işlemlerin gerçekleştirilmesinde özellikle Python için bir derin öğrenme kütüphanesi olan "Keras" kütüphanesinin kullanıldığı görülmektedir. [13, 14, 15]. Geleceğe dönük tahmin işlemleri konusunda derin öğrenme algoritmalarını kullanan çalışmaların bir kısmında aktivasyon fonksiyonunu sağa veya sola ötelenmesini (shift) sağlayan "bias" değerinin tahmin performansını etkilediği belirtilirken, bir kısmında ise bu değer hisse senedi fiyatlarının tahmin edilmesi üzerine herhangi bir etkisinin olmadığı belirtilmektedir [16, 17]. Literatürde makine öğrenmesi ile bilgisayarlar üzerinde finans tahmin çalışmaları da yapıldığı görülmektedir [18]. Regresyon analizi kullanılarak gerçekleştirilen çalışmalar incelendiğinde başarılı sonuçların elde edildiği görülmektedir [19, 20, 21]. Literatür kapsamında araştırılan çalışmalarda, hisse senedi tahmin işlemleri için "Destek Vektör Makineleri (SVM)" algoritmalarının da kullanıldığı görülmektedir [22, 23, 24]. [25] tarafından yapılan çalışmada hisse senedi fiyatlarının tahmini konusunda bulanık mantık yönteminden yararlanıldığı görülmektedir. [26] tarafından yapılan çalışmada Yapay Sinir Ağları (YSA) kullanılarak Bitcoin tahmini yapıldığı görülmektedir. [27] tarafından çalışmada ise farklı regresyon modeli ve YSA kullanarak BIST 100 endeksinin tahmin uygulamasının yapıldığı görülmektedir. Emeklilik yatırım fonunun hisse senedi fiyatlarının üzerinde derin öğrenme algoritmaları kullanılarak tahmin işlemi yapılmıştır [28]. Literatürde zaman serisi analizi ile derin öğrenme algoritmalarının kullanımına yönelik çalışmalar incelenmiştir [29, 30]. Zaman serileri kullanılarak derin öğrenme algoritmaları ile geleceğe dönük tahmin işlemleri konusunda gerçekleştirilen çalışmalar incelendiğinde, yapılacak olan çalışmada en doğru algoritmayı seçmenin öngörülen doğru sonuca yaklaşmayı daha çok kolaylaştırdığı görülmektedir [31, 32]. Araştırmalarda, bulanık mantık ve regresyon yöntemlerinin derin öğrenme algoritmaları ile kullanıldığı, ancak derin öğrenme ve bulanık mantık yöntemleri ile yapılmış olan çalışmalarda regresyon yöntemine kıyasla daha yüksek tahmin sonuçlarına ulaşıldığı görülmektedir [33, 34]. İncelenen araştırmalarda, kira fiyat tahmini, borsa kapanış değeri tahmini, borsa hisse senedi tahmini, finans piyasası tahmini ve BİST30 İndeksinin tahmini gibi çalışmalarının LSTM derin öğrenme algoritması

kullanılarak gerçekleştirildiği görülmektedir [35, 36, 37, 38, 39].

[40] tarafından yapılan çalışmada hisse senedi fiyat hareketlerini tahmin etmeye yardımcı olmak için finansal haberlerden veya tweet'lerden elde edilen duygu tutumlarını (olumlu ve olumsuz) ve ayrıca duygu duygularını (sevinç, üzüntü vb.) kullanma potansiyeli araştırılmıştır. [41] tarafından yapılan çalışmada özellikle, yüksek frekanslı haber duyarlılığını 4 boyutlu bir zaman serisi olarak çıkarmak için FinBERT tabanlı bir model önerilmiştir. [42] tarafından yapılan çalışmada, borsada kullanılan tahmin tekniklerinin sınıflandırılması, karakterize edilmesi ve karşılaştırılması dahil olmak üzere güncellenmiş bir sistematik incelemesi yapılmıştır.

[43] tarafından yapılan çalışmada dikkate dayalı çift yönlü CNN-RNN Derin Modeli (ABCDM) önerilmiştir. Bu çalışmada iki bağımsız çift yönlü LSTM ve GRU katmanını kullanarak, her iki yöndeki zamansal bilgi akışını göz önünde bulundurarak hem geçmiş hem de gelecek bağlamları çıkartılmıştır. [44] tarafından yapılan çalışmada Arapça duygu analizi için kullanılan en etkili yaklaşımlar üzerine kapsamlı bir karşılaştırmalı çalışma yapılmıştır. Çalışmada ayrıca Arabic SA için mevcut yaklaşımların çoğu yeniden uygulanmış ve bunların etkinliği Arabic SA için en popüler üç karşılaştırma veri setinde test edilmiştir. [45] tarafından yapılan çalışmada araştırmacıların duygu analizi ve ilgili alanlar hakkında küresel bir anket sunmak için duygu analizi hakkında bir çalışma sunulmuştur.

Literatür taraması kapsamında incelenen derin öğrenme algoritmaları kullanılarak yapılmış bir kısım çalışmalara ilişkin başarı ve hata oranları Tablo 1'de sunulmuştur. Söz konusu veriler incelendiğinde derin öğrenme algoritmaları ile sosyal medya duygu analizi, finans tahmini, hisse senedi tahmini, deprem tahmini, hisse senedi tahmini, seçim sonuçlarının tahmini, kalitesiz videoların renk kalitesinin artırılması, kripto para değer tahmini gibi pek çok alanda yüksek başarı oranları elde edildiği görülmektedir.

Tablo 1. Literatür taraması bağlamında incelenen çalışmaların değerlendirilmesi

Referans Çalışma Numarası	Kullanılan Teknik	Alınan Başarı Sonucu	Alınan Hata Payı
Cakar (2017) [61]	Yapay Sinir Ağı	%99.5	%0.50
Yiğiter (2018) [31]	Makine Öğrenimi	%97	%3
Mazed (2019) [32]	Derin Öğrenme	%95	%5
Çelik (2019) [17]	Makine Öğrenimi	%81.15	%18.85
Sakarya (2019) [35]	Derin Öğrenme	%80	%20
Yasar (2020) [63]	LSTM	%79.01	%20.99
Hayrulloğlu (2017) [18]	Yapay Sinir Ağı	%79	%21
Çam (2018) [62]	Yapay Sinir Ağı ve Markov Zinciri	%70	%30
Mudinas (2019) [40]	Destek Vektör makinesi (SVM)	%70	%30
Farha (2021) [44]	Deep Learning	%84	%16
Birjali (2021) [45]	Machine Learning	%83,54	%16,46

3 Materyal ve yöntemler

Çalışmanın bu bölümünde çalışmada yararlanılan Yapay Sinir Ağları (YSA), derin öğrenme ve LSTM ağları ile ilgili araştırma sonuçları yer almaktadır.

3.1 Yapay sinir ağları (YSA)

YSA; girdi, çıktı ve ağırlık değerleri bir algoritmaya göre kendi kendine öğrenebilen, çıkarımlar yapabilen ve ayrıca kararlar alabilen doğrusal olmayan problemlerin de çözülebilmesine olanak sağlayan insan beynindeki biyolojik sinir yapısını modellemektedir [46, 47, 48]. Tek katmanlı algılayıcıların ilk sinir ağlarından biri olduğu söylenebilir. Bu yapı, birden fazla girdi alan ve bu girdilerden çıktı üreten bir sinir hücresine dayanmaktadır. Ve (and), Veya (or), Değil (not) durumları gibi doğrusal fonksiyonlarla ilgili problemlerde kullanılabilirler. Bu yapıda, giriş ağırlıklarının güncelleme sayısına Epoch denir. Çıkış değeri ile işlevden beklenen değer arasındaki farka hata denir.

Denklem 1'de gösterilen formül, eşik değerinin girdilerle toplandığı çıktıyı hesaplar. Denklem 1 üzerinde x_i girdi değerlerini, n toplam girdi sayısını, w_i ağırlık değerlerini, θ ise eşik değerini temsil etmektedir.

$$f = \sum_{i=1}^n (w_i x_i + \theta) \quad (1)$$

Tek katmanlı algılayıcı yapısında yalnızca doğrusal problemler çözülebildiğinden, Denklem 2'de gösterildiği gibi ağırlık çıktısı 1 veya -1'dir.

$$f(r) = \begin{cases} 1, & \zeta > 0 \\ -1, & \zeta \leq 0 \end{cases} \quad (2)$$

Çıktı değeri beklenen değeri vermiyorsa, ağırlıkların yeniden güncellenmesi gerekir. Denklem 3 üzerinde ağırlıkların güncellenmesi formülü gösterilmiştir. Denklem 3'deki formülde w_i girdiler sonucundaki ağırlık değerlerini, t zamanı, $\Delta\theta$ beklenen değer ile çıktının arasındaki değişimi temsil etmektedir.

$$w_i(t+1) = w_i(t) + \Delta\theta(t) \quad (3)$$

Ağırlıklar tekrar güncellendikten sonra Denklem 4'teki formül ile eşik değerinin güncellenmesi gerekir. Denklem 4'te θ eşik değerini, $\Delta\theta$ beklenen değer ile çıktının arasındaki değişimi temsil etmektedir.

$$\theta(t+1) = \theta(t) + \Delta\theta(t) \quad (4)$$

Denklem 5, aktivasyon fonksiyonu olan sigmoid fonksiyonunun matematiksel ifadesini içermektedir. Denklem 5'de Net ifadesi çıktıyı, $F(\text{Net})$ parametresi ise aktivasyon fonksiyonu sonucunda elde edilen değeri temsil etmektedir.

$$F(\text{Net}) = \frac{1}{1+e^{-\text{Net}}} \quad (5)$$

Denklem 5, aktivasyon fonksiyonu olan sigmoid fonksiyonunun matematiksel ifadesini içermektedir. Denklem 5'de Net ifadesi çıktıyı, $F(\text{Net})$ parametresi ise aktivasyon fonksiyonu sonucunda elde edilen değeri temsil etmektedir.

$$f(t) = \sigma(W_f[h_t - 1, x_t] + b_f) \quad (6)$$

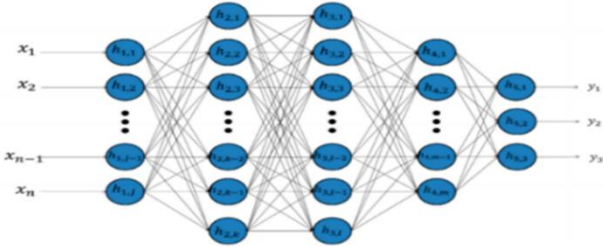
Ancak tek katmanlı algılayıcı yapısı, XOR problemi gibi doğrusal olarak sınıflandırılmayan problemlerin çözümünde başarısızdır. Bu problemin başarısızlığı nedeniyle Çok Katmanlı Algılayıcı (MLP) modeli geliştirilmiştir. Basit üç katmanlı bir ağ yapısı Şekil 1'de gösterilmektedir.

Çok katmanlı algılayıcı yapıda, ağırları uçtan uca eğitmek için geri yayılım kullanılmamıştır. Ancak önceki katmanlardan sonraki katmanlara en küçük kareler yöntemi kullanılmıştır [49]. Bu yapıda kullanılan denetimli öğrenme stratejisinde hem örnekler hem de örneklerden elde edilmesi gereken çıktılar verilerek sınıflandırma, tanıma ve genelleme gerektiren sorunlar çözülebilmektedir. Böylece verilen örneklerden genellemeler yapabilmek mümkündür. Böylelikle problem alanını temsil eden bir çözüm alanı oluşturulabilir. Bu ağ yapısında öğrenmeyi sağlamak için eğitim seti adı verilen ve örneklerden oluşan bir veri seti kullanılmaktadır. Tek bir nöron için birçok girdi yeterli olmayacağından, paralel işlem için birden fazla nörona ihtiyaç vardır. Veriler giriş katmanından alınır ve gizli katmana aktarılır. Ara katman, en az bir katman olarak farklılık gösterebilir. Her katmanın çıkışı, bir sonraki katmanın girişi olmaktadır. Her nöron, bir sonraki katmandaki nöronlara bağlanır. Çıktı katmanı, önceki katmandan alınan verileri işleyerek ağırlık çıktısını belirler. Çıktıların sayısı çıktı katmanındaki elemanların sayısına eşittir. Ağırlık öğrenmesi için örnek girdi ve çıktılardan oluşan bir eğitim seti gereklidir. Aktivasyon fonksiyonuna girecek girdi, ağırlıklar, transfer fonksiyonu ve net girdidir. Yani ağırlık giriş değerleri, ağırlıkları ve ağda uygulanan aktivasyon işlemi kullanılarak ağırlık çıktısı elde edilir. Bir çıktı katmanı, önceki katmandaki işlem öğelerine bağlı birden fazla işlem öğesi ve işlem öğesi içerir. Çok katmanlı ağlarda çeşitli aktivasyon fonksiyonları kullanılabilir. Sigmoid işlevi bunlardan birisidir. Etkinleştirme işlevi, girdi katmanları ve çıktı katmanları tarafından oluşturulan eğri eşitlemesini etkinleştirir. Ağırlık performansını etkilediği için uygun aktivasyon işlevinin seçilmesi önemlidir. Aktivasyon fonksiyonunu doğrusal olarak seçmek de mümkündür. Sigmoid aktivasyon fonksiyonu 0 ile 1 arasındaki değerleri hesapladığından, bir olayın meydana gelme olasılığını bulan modellerde kullanılır. Yapay sinir ağlarında, girdiyi işleyerek bu girdi sonucunda üreteceği çıktıyı belirleyen aktivasyon fonksiyonları kullanılmaktadır. Derin öğrenme modelinde, hata değerini minimize etmek için kullanılan geriye besleme işleminde türev hesabı kullanılmaktadır. Bu sebeple genellikle türevi rahatlıkla alınabilen aktivasyon fonksiyonları seçilmektedir. Bir aktivasyon fonksiyonu, yapay bir nörondaki, girdilere dayalı bir çıkış sağlayan fonksiyondur [50].

3.2 Derin öğrenme

Derin Öğrenme son yılların en önemli konuları arasındadır. Derin öğrenme, doğrusal olmayan işlem birimlerinin birçok katmanını kullandığı, birbirini izleyen her katmanın girdi olarak bir önceki katmanların çıktısını aldığı bir makine öğrenmesidir. Derin Öğrenme algoritmaları yapay sinir ağlarının (YSA) yapısal olarak daha karmaşık hali olarak düşünülebilir. Derin öğrenmenin tek katmanlı yapay sinir ağlarından farkı, lineer olmayan problemlerin çözülmesine olanak tanımasıdır [51]. Derin öğrenme yaklaşımında gizli katmanlar bulunmaktadır.

Derin Öğrenmenin kullanım alanları her geçen gün artmaktadır. Günümüzde GPU hızlarının artmasıyla birlikte, derin ağlar çok daha hızlı ve etkin olarak eğitilebilmekte ve bu sayede artan başarı oranları ile finans tahmin işlemlerinde dahil olmak üzere hemen hemen her alanda yaygınlaşarak kullanılmaktadır. Derin öğrenme modeli, makine öğreniminin birçok uygulamasını sağlamış ve yapay zeka alanındaki başarıyı artırmıştır. Derin ağlar, gizli katmanların sayısında yatar. Derin sinir ağının hücre yapısı aşağıdaki Şekil 1'de gösterilmektedir [52].



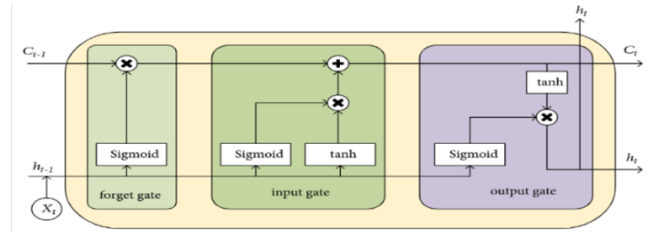
Şekil 1. Derin sinir ağı

Derin öğrenme mimarileri arasında Konvüsyonel Sinir Ağları (CNN), Tekrarlayan Sinir Ağları (RNN) ve Uzun Kısa Vadeli Hafıza Ağları (LSTM), Sınırlı Boltzmann makineleri (Restricted Boltzmann Machines-RBM), Derin inanç ağları (DBN), Diabolo ağı olarak adlandırılan Oto-kodlayıcılar (Autoencoders-AE) gibi mimariler bulunmaktadır [53]. CNN, bir tür çok katmanlı algılayıcıdır. Aynı zamanda ileri beslemeli bir sinir ağı algoritması olan bir derin öğrenme mimarisidir. RNN ise derin öğrenme uygulamalarında sıklıkla yer bulmuş olan bir diğer derin öğrenme mimarisidir. İleri beslemeli sinir ağlarının aksine, RNN'ler sıralı bilgileri kullanarak gelişigüzel girdi dizilerini işlemek için girdi belleğini kullanmaktadırlar.

3.3 LSTM Algoritması

Bu bölümde çalışmada kullanılan LSTM modeline ilişkin bilgiler yer almaktadır. LSTM, uzun vadeli bağımlılıkları öğrenebilen bir derin öğrenme algoritmasıdır. LSTM algoritması, siber güvenlik alanındaki sorunların çözümünün yanı sıra metin tanıma, konuşma tanıma, makine çevirisi gibi alanlarda da kullanılmaktadır. LSTM yapısı, özyinelemeli ağ yapısında olmayan girdi, çıktı ve unutmaya katmanlarına sahiptir. LSTM yapısındaki ilk adım, ağ üzerindeki verilerin unutulup unutulmayacağına karar vermektir. Örneğin, unutmaya katmanı bir ortamdan diğerine aktarıldığında, modelin seyri için eski ortam gerekli değilse, o ortam modelden çıkarılır. Derin öğrenme modelinde zaman serilerinin kullanılabilmesi için denetimli öğrenme yöntemine dönüştürülmesi gerekmektedir. Sürgülü pencere yöntemi, zaman serilerini denetimli bir öğrenme yaklaşımına dönüştürmek için kullanılabilir. Bu yöntem, sonraki zaman adımındaki değeri tahmin etmek için önceki zaman adımlarını kullanmaktan oluşur. Sürgülü pencere yöntemini kullanmanın temel amacı, zaman serilerine göre çıktının kesin olduğu durumlarda derin öğrenme algoritmalarını kullanmaktır. Bu yaklaşımın temel mantığı, mevcut zamandaki (t) değeri incelemek ve bir sonraki t anında (t + 1) değerini tahmin etmektir.

LSTM algoritmasının temel yapısı Şekil 2'de yer almaktadır.



Şekil 2. LSTM algoritmasının temel yapısı

RNN yapısında, büyük zaman serilerinin kullanımı sonucunda aktivasyon fonksiyonunun türevinin azalması veya çok büyük değerlere ulaşması problemi ortaya çıkmaktadır [54, 55]. Bu dezavantaja çözüm olarak 1997 yılında bilgisayar bilimcisi Hochreiter ve Schmidhuber LSTM yapısını ortaya koymuştur. LSTM yapısında girdi, çıktı katmanı ve özyinelemeli ağ yapısında olmayan unutmaya katmanları bulunmaktadır [56, 57]. LSTM yapısının ilk adımı ağıdaki verinin unutulup unutulmayacağına karar verilmesidir [56]. Örneğin, unutmaya katmanı bir ortamdan başka bir ortama geçildiğinde, eğer eski ortam modelin

gidişatı için gerekli değilse o ortam modelden kaldırılmaktadır [58, 59]. Denklem 6'da unutmama mekanizmasının matematiksel ifadesi gösterilmektedir. Denklem 6'da $f(t)$ unutmama mekanizmasının aktivasyon vektörünü, w ağırlık değerini, h_t mevcut durumu, x_t t anındaki girdiyi, b_f unutmama kapısının bias değerini temsil etmektedir.

$$f(t) = \sigma(W_f[h_t - 1, x_t] + b_f) \quad (6)$$

LSTM yapısının özinelemeli sinir ağlara göre daha geniş çalışma spektrumu bulunmaktadır. Örneğin, büyük zaman serilerindeki tahmin işlemlerinde, müzik bestesi yapabilme yeteneğinde, özinelemeli olarak gerçekleştirilen dil işleme gibi birçok alanda LSTM yapısı kullanılmaktadır [60].

Giriş katmanı olarak nitelendirilen sigmoid katmanı hangi değerlerin güncellenip güncellenmeyeceğini belirtmektedir. Sigmoid katmanının yanındaki tanh katmanı ise yeni değerler vektörünü oluşturmaktadır. Denklem 7 ve 8'de, yeni bir EUR/USD paritesinin kapanış değerinin değer vektörü olarak seçilmesi işlemi matematiksel ifadeler ile gösterilmiştir [56]. Denklem 7'de i_t giriş ve güncelleme kapısının aktivasyon vektörünü, b_i giriş değerinin bias değerini temsil etmektedir. Denklem 8'de ise C_t durum vektörünü, b_c durum vektörünün bias değerini temsil etmektedir.

$$i_t = \sigma(W_i.[h_t - 1, x_1] + b_i) \quad (7)$$

$$C_t = \tan h(W_c.[h_t - 1, x_1] + b_c) \quad (8)$$

Denklem 9'da LSTM yapısında gerçekleştirilen işlemlerin matematiksel ifadesi verilmiştir [56]. Çıkış katmanında iki adet fonksiyon bulunmaktadır. Bu fonksiyonlar sigmoid ve tanjant fonksiyonudur.

$$C_t = f_t \cdot C_{t-1} + i_t * C_t \quad (9)$$

Sigmoid fonksiyonu hücre durumundaki görevi ile aynıdır. Tanjant fonksiyonu, tanjant katmanı olarak temsil edilmektedir. Bu fonksiyon -1 ile 1 aralığında bir değer almaktadır. Bu tanjant yapısının görevi bilginin ne kadarının kullanılacağını hesaplamaktır. Hesaplamanın formüle edilmiş hali Denklem 10 ve Denklem 11'de gösterilmektedir [56]. Denklem 10'da σ_t çıkış kapısının aktivasyon vektörünü temsil etmektedir.

$$\sigma_t = \sigma(W_i.[h_t - 1, x_1] + b_\sigma) \quad (10)$$

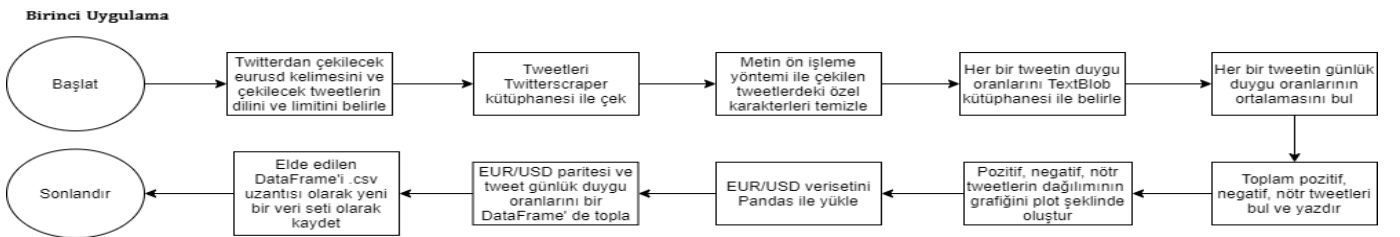
$$h_t = \sigma_t \cdot \tan h(C_t) \quad (11)$$

4 Uygulamalar

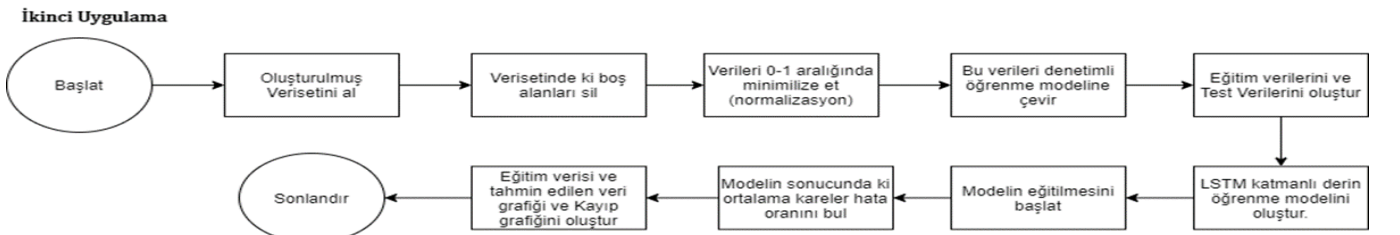
Çalışma kapsamında veri setlerinin oluşturulması ile modelin eğitilmesi/test edilmesi aşamalarını içeren iki ayrı uygulama geliştirilmiştir. Birinci uygulamada veri setinin oluşturulması gerçekleştirilmiştir. İkinci uygulamada ise geliştirilen LSTM tahmin modelinin eğitilmesi ve test edilmesine ilişkin işlemler gerçekleştirilmiştir.

4.1 Akış diyagramı

Çalışma genel itibarıyla veri setlerinin oluşturulması ile modelin eğitilmesi ve test edilmesi aşamalarını içeren akış diyagramı Şekil 3, Şekil 4'te sunulmuştur.



Şekil 3. Veri seti oluşturma akış diyagramı



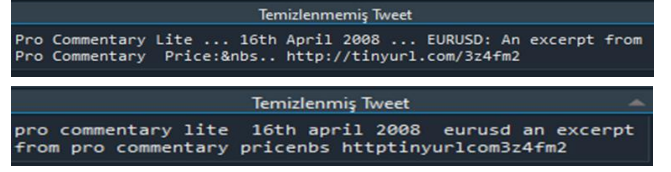
Şekil 4. Model eğitimi ve test edilmesi akış diyagramı

4.2 Birinci uygulama: Çalışmada kullanılan veri setinin oluşturulması

Bu aşama çalışmanın birinci uygulaması ile gerçekleştirilmiştir. Çalışmanın birinci uygulamasının amacı çalışmada kullanılan veri setinin oluşturulmasıdır. Çalışmada tweetlerin duyarlılık analizi ve EUR/USD oranlarını zaman serisine dayalı olarak ve birleştirilmiş bir yapıda içeren bir veri seti oluşturulmuş ve kullanılmıştır. Çalışma kapsamında öncelikle Twitter üzerinde tweet mesajlarının içeriklerine göre duyu sınıflandırması ile analiz edilerek olumlu veya olumsuz olarak sınıflandırılmıştır. 2005 yılından günümüze kadar Twitter üzerinden paylaşılan “#eurusd” konu etiketli veya “eurusd” kelimesini barındıran, İngilizce tweetler filtrelenmiştir. Bu filtreleme işleminin amacı, sadece EUR/USD ile ilgili olan tweetlerin kullanılması ve bu parite tweetlerinin olumlu, olumsuz ve tarafsız (nötr) olarak ayrıştırılabilmesidir. Bu amaç doğrultusunda, “eurusd” konu etiketli ya da bu ifade ile ilişkili paylaşım içeren tweetler, Twitter üzerinden belirli etiketlerin çekilmesini sağlayan TwitterScraper adlı bir kütüphane kullanılarak çekilmiştir [65]. Bu bağlamda TwitterScraper kütüphanesi kurulmuş ve Twitter sosyal medya platformundan İngilizce tweetler bu kütüphane yardımıyla elde edilmiştir. Bu tweetlerin günlük duyu oranı ortalaması için TextBlob kütüphanesinden faydalanılmıştır. Bu kütüphane, hedeflenen tweetlerin hangi tarih aralığından çekileceğini, maksimum çekilen veri miktarının ne olacağı ve hangi dildeki tweetlerin çekileceği gibi filtrelerin kullanılmasını sağlamaktadır. Çekilen tweetlerin içeriğinde, gerçekleştirilecek duyu analizini etkileyecek semboller, özel karakterler vs. bulunabilmektedir. Uygulamada, duyu analizini etkileyecek tweetlerin içeriğini temizlemek için metin ön işleme yapılmaktadır. Çalışmada, metin ön işleme için, tweetlerin hepsi, programlama dilinde bulunan “string” yapısına çevrilmiştir. Ardından, bu tweetlerin küçük harfe çevrilmesi ve özel karakterlerin silinmesi işlemleri gerçekleştirilmiştir.

Şekil 5’te, temizlenmiş ve temizlenmemiş tweetlere bir örnek gösterilmektedir. Temizlenmemiş tweet, platformdan direkt olarak çekilen tweettir. Temizlenmiş tweet ise, noktalama ve özel karakterler ayrıştırıldıktan sonra ortaya çıkan tweettir. Yukarıdaki şekilde, bir tane tweet için gerçekleştirilen metin ön işleme neticesinde nokta

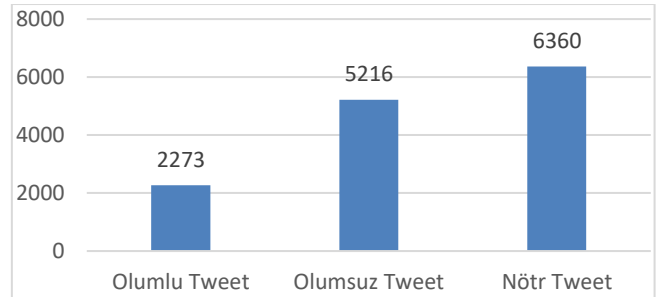
gibi özel karakterlerin nasıl temizlendiği görülmektedir.



Şekil 5. Metin ön işleme sonucu

Metin ön işleme veya metin içeriğinin temizlenmesi sürecinin önemi, geliştirilen uygulamanın performansını doğrudan etkileyebilmektedir [64]. Çünkü direkt olarak alınmış verilerle işlem yapılması uygulamanın yanlış sonuçlar çıkarma ihtimalini arttırmaktadır. Bu nedenle, metin üzerinde işlem yapılacağı zaman, öncelikle metin ön işleme gerçekleştirilmektedir. Çalışmada, metin ön işleme sonucunda veri setindeki tweetlerden özel karakterler temizlenmiş, ardından duyu analizi sürecine geçilmiştir.

Twitter üzerindeki tweetlerin duyu analizinden geçirilmesinin nedeni her bir tweetin olumlu olup olmadığını belirleyebilmek ve her gün sonunda bu tweetlerin duyu oranlarının ortalamasının alınmasıdır. Günlük duyu oranı ortalamasının alınmasındaki amaç, belirlenen gün zaman diliminde içerisinde atılmış olan olumlu ve olumsuz tweetlerin durumunu tespit etmektir. Uygulamada, bir tweetin olumlu ya da olumsuz olup olmadığını belirleyebilmek amacıyla TextBlob kütüphanesi kullanılmıştır [66]. Bu kütüphane yardımıyla, bir tweetin duyu oranı 0’dan büyük ise tweet olumlu, küçükse olumsuz ve 0’a eşit ise tweet nötr olarak kabul edilmiştir. TextBlob, metin verilerini işlemek için bir Python (2 ve 3) kitaplığıdır. Konuşma bölümü etiketleme, isim öbeği çıkarma, duyu analizi ve daha fazlası gibi yaygın doğal dil işleme (NLP) görevlerine dalmak için tutarlı bir API sağlar [66].



Şekil 6. Tweetlerin olumlu, olumsuz, nötr sayısı gösterimi

Şekil 6’da veri seti içerisindeki duygu oranlarının belirlenmesinin ardından 2273 adet olumlu, 5216 adet olumsuz ve 6360 nötr tweet verisi olduğu gösterilmiştir.

Veri setinin sayısal olarak ifade edilmesi amacıyla olumlu olarak değerlendirilen tweetler “1” ile, olumsuz olarak değerlendirilenler “-1” ile ve nötr

olarak değerlendirilenler ise “0” rakamı ile sayısal olarak temsil edilmektedir. Uygulamanın son adımında, gerçek EUR/USD paritesinin günlük kapanış değeri ile günlük duygu oranları ortalaması tek bir DataFrame’de satır ve sütunlardan oluşan bir yapıda ve “csv” formatında bir veri seti yapısında oluşturulmuştur. Söz konusu veri seti Tablo 2’de sunulmuştur.

Tablo 2. Csv formatlı veri seti

Tarih	Açılış	Yüksek	Düşük	Kapanış	Duygu Oranı Ortalaması	Pozitif Sayısı (Tweet)	Negatif Sayısı (Tweet)	Nötr Sayısı (Tweet)
2008-08-27	1.4647	1.4771	1.4641	1.4725	-5	199	874	4
2008-08-29	1.4711	1.4763	1.4640	1.4670	14.375	380	779	3
2008-09-02	1.4577	1.4596	1.4474	1.4515	2.0370	120	98	1
2008-09-03	1.4518	1.4531	1.4388	1.4486	0	191	316	1
2008-09-04	1.4486	1.4540	1.4227	1.4264	0	131	447	1

Bu aşamada internet üzerinde herkese açık olarak sunulan EUR/USD değerleri kullanılarak oluşturulmuştur. Bu veri seti, “finance.yahoo.com” web sitesinde herkese açık olarak Yahoo! şirketinin sağlamış olduğu EUR/USD paritesine ait değerler kullanılarak oluşturulmuştur. Söz konusu veri seti 2005 yılından günümüze kadar olan EUR/USD sayısal değerlerini içermektedir. Veri setinin oluşturulmasında son olarak ise anlatılan adımlarda elde edilen zaman serisine dayalı veriler “csv” formatında bir veri seti yapısında birleştirilmiştir. Böylelikle yüksek tahmin başarı oranına sahip LSTM modelinin eğitilmesi ve testi aşamalarında kullanılacak olan zaman serisi verilerine dayalı veri seti elde edilmiştir.

4.3 İkinci uygulama: modelin eğitilmesi ve test edilmesi

Çalışmanın ikinci uygulamasının amacı veri setleri üzerinde yapılan değişik deneyler ile yüksek başarı oranına sahip LSTM modeli belirlenmesidir. Veri seti belirlendikten sonra LSTM tahmin modelinin eğitimi ve testi uygulama aşamasına geçilmiştir. Bu uygulama bir önceki bölümde oluşturulan veri seti kullanılarak modelin eğitilmesi ve test edilmesi ile ilgili konuları içermektedir. Bu aşamada en yüksek başarı oranına sahip LSTM tahmin modelini belirlemek maksadıyla epoch sayılarının ve LSTM sinir ağı katman yapısının değiştirilmesi suretiyle

farklı deneyler yapılmıştır. Bu deneyler ile en yüksek başarı oranına sahip LSTM modelini belirlemek hedeflenmiştir. Her bir deneyin doğruluk değeri elde edilen başarı oranları dikkate alınarak değerlendirilmiş ve analiz edilmiştir. LSTM modelinin hazırlanması için kayan pencere metoduyla kaydırma işlemlerinin gerçekleştirilmesi gerekmektedir. Burada amaç, denetimli öğrenme algoritmasını LSTM modeline çevirmektir. Bu işlemlerin ardından LSTM modelinin eğitilebilmesi için denetimli öğrenme modelinden çıkan $var(t)$, $var(t-1)$ gibi değerlerin LSTM modelinde girdi, test girdisi ve çıktı verisi olarak ayarlanması gerekmektedir. Verilerin denetimli öğrenme modelinden LSTM modeline çevrilmesi işleminin gerçekleştirilmesinin ardından LSTM modelinin oluşturulması gerekmektedir. Bu çalışmada, model geliştirme için, yaygın olarak tercih edilen Keras kütüphanesi kullanılmıştır. Eğitim verisi ve test verilerinin miktarı uygulama boyunca sürekli değiştirilebilmektedir.

Çalışmanın veri seti, 110’u eğitim verisi ve 83’ü test verisi olmak üzere 193 veriden oluşmaktadır. Modelin ilk katmanında 10 girişli bir LSTM modeli oluşturulmuştur ve bir sonraki katmanın sadece çıkış katmanı olması hedeflenmiştir. Modelin ezberlemesini engellemek için 0.2 değerinde dropout kullanılmıştır. Kayıp fonksiyonu olarak, bir eğimin gerçek değere ne kadar yakın olduğunu gösteren Kök Ortalama Kare Hata fonksiyonu seçilmiştir. Optimizasyon yöntemi olarak ise

“adam” seçilmiştir. Epoch döngüsü başlangıçta 20, sonrasında ise 500 ve 5000 epoch olarak belirlenmiştir. Epoch değerlerinin değiştirilmesi sonrasında modelin eğitim kalitesi ile ilgili farklar ortaya çıkmıştır.

5 Bulgular ve değerlendirme

Çalışmada oluşturulan veri seti üzerinde yapılan değişik deneyler ile yüksek başarı oranına LSTM modelinin belirlenmesi hedeflenmiştir. LSTM modelinin hazırlanması için kayan pencere metoduyla kaydırma işlemlerinin gerçekleştirilmesi gerekmektedir. Burada amaç, denetimli öğrenme algoritmasını LSTM modeline çevirmektir. LSTM algoritmasını kullanabilmek için, girdi ve çıktıları makine öğreniminin bir algoritması olan denetimli öğrenme algoritması kullanılmıştır. Bu işlem sonunda eğitim verisi ve test verisi belirlendikten sonra, LSTM katmanlı bir derin öğrenme modeli geliştirilmiştir. Bu işlemlerin ardından LSTM modelinin eğitilebilmesi için denetimli öğrenme modelinden çıkan $var(t)$, $var(t-1)$ gibi değerlerin LSTM modelinde girdi, test girdisi ve çıktı verisi olarak ayarlanması gerekmektedir. Verilerin denetimli öğrenme modelinden LSTM modeline çevrilmesi işleminin gerçekleştirilmesinin ardından LSTM modelinin oluşturulması gerekmektedir.

Veri setleri üzerinde işlem yapmak için Pandas kütüphanesi ve söz konusu modelin eğitimi için de Keras kütüphanesi kullanılmıştır. Uygulamada Pandas kütüphanesinde bulunan Shift() fonksiyonu da kullanılmıştır. Bu fonksiyon kullanılarak zaman serisini ileriye veya geriye hareket ettirilebilmek mümkündür.

Çalışma içerisinde, girdi ve çıktılar için geriye hareket yöntemi kullanılmıştır. Bu yöntem, genel olarak çok değişkenli girdilerin kullanılmasında tercih edilmektedir. Eğer n zaman aralığındaki veriler tahmin edilmekteyse bu verileri n kez geriye doğru kaydırmak gerekmektedir.

Tablo 3 üzerinde EUR/USD kapanış değerinin girdi gözlemleri sırasıyla $var1(t-1)$ ve $var1(t-2)$ olarak ve duyu oranları girdileri ise sırasıyla $var2(t-1)$ ve $var2(t-2)$ gösterilmiştir. Çıktı zaman adımı, yani tahmin edilecek değer olan EUR/USD kapanış değeri ise $var1(t)$ olarak ifade edilmiştir. Çalışmada geliştirilen LSTM uygulaması, eğitim ve test işlemleri ile 110 adet eğitim verisi ve 83 adet test verisi olmak üzere toplamda 193 veri üzerinde yazılmıştır. İlk önce model içinde aynı anda gösterilen eğitim sayısı 20, 500 ve 5000 epoch olacak şekilde değiştirilmiştir. Bu değiştirilen değerlerin sonuçlarını karşılaştırabilmek adına her bir eğitim adımı değeri sırasıyla 2, 4 ve 6 katmanlı modeller üzerinde tekrar eğitilmiştir.

Tablo 3. Girdi ve çıktıların denetimli modele dönüştürülmesi

Var1(t-3)	Var2(t-3)	Var1(t-2)	Var2(t-2)	Var1(t-1)	Var2(t-1)	Var1(t)
0.5937	0.8041	0.6312	0.6733	0.6381	0.7555	0.6511
0.6312	0.6733	0.6381	0.7555	0.6511	0.6666	0.6462
0.6381	0.7555	0.6511	0.6666	0.6462	0.7654	0.6541
0.6511	0.6666	0.6462	0.7654	0.6541	0.5694	0.6884
0.6462	0.7654	0.6541	0.5994	0.6884	1	0.7070
0.6541	0.5694	0.6884	1	0.7070	0.6666	0.7347
0.6884	1	0.7070	0.6666	0.7374	0.7059	0.6970
0.7070	0.6666	0.7347	0.7059	0.6970	1	0.6923

Bu modellerin eğitim verileri ile tahmin edilen veriler arasındaki kayıp şekil olarak gösterilmektedir. Gerçekleştirilen deneylerde, modelin katman sayısı ve epoch değerlerinin değiştirilmesi sonucunda alınan başarı oranlarında belirgin farklar gözlemlenmiştir. Her bir modelin katman sayısı ve epoch değerlerinin değiştirilmesi sonucunda farklı başarı sonuçları elde edilmiştir

Tablo 4 üzerinde detayları verilen deneyler sonucunda en doğru tahmin işleminin iki numaralı deneyde gerçekleştiği görülmüştür. 2 katmanlı 500 epoch ile gerçekleştirilen bu deney EUR/USD paritesinin kapanış değerini %94.48 gibi bir oranla başarılı şekilde tahmin etmiştir.

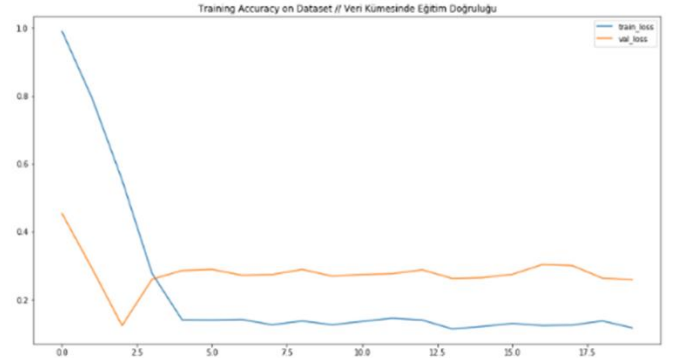
Şekil 7’den Şekil 14’te kadar verilen grafiklerde, bu deneylerden elde edilen sonuçlar gösterilmektedir.

Tablo 4. Çalışma sürecinde yapılan deneyler

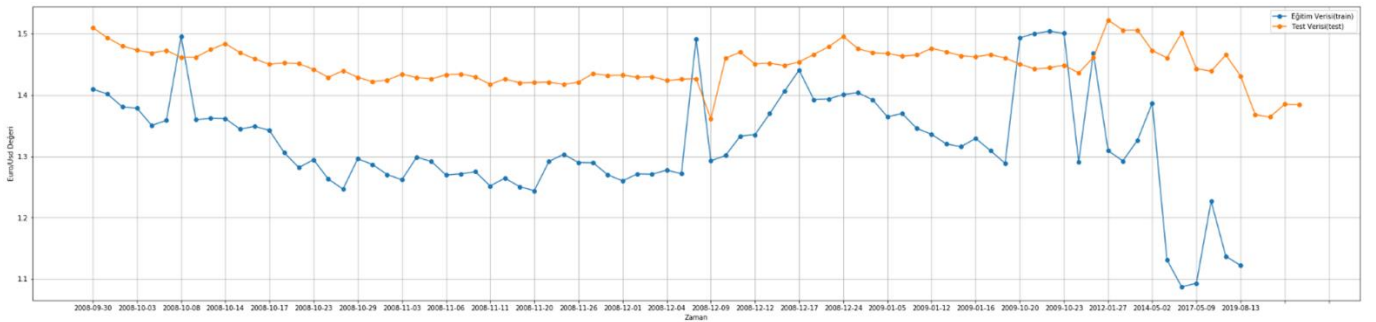
Deney Numarası	Modelin Katman Sayısı	Epoch Değeri	Modelin Eğitim Verisi	Modelin Test Verisi	RMSE Hata Oranı	Alınan Başarı Oranı	Alınan Hata Oranı
Deney 1	2	20	110	83	0.144	%89.87	%10.13
Deney 2	2	500	110	83	0.087	%94.48	%5.52
Deney 3	2	5000	110	83	0.137	%91.06	%8.94
Deney 4	4	20	110	83	0.149	%89.68	%10.32
Deney 5	4	500	110	83	0.089	%94.40	%5.60
Deney 6	4	5000	110	83	0.177	%88.12	%11.88
Deney 7	6	20	110	83	0.156	%89.19	%10.81
Deney 8	6	500	110	83	0.117	%92.06	%7.94
Deney 9	6	5000	110	83	0.140	%90.45	%9.55

Şekil 7’de, 2 katmanlı, 20 epoch değerine sahip bir modelin eğitim verileri ile tahmin edilen veriler arasındaki kayıp gösterilmektedir.

Şekil 8’de ise 2 katmanlı modelin 20 epoch sonrasındaki verileri gösterilmektedir. Şekil 8’deki grafik içerisinde mavi çizgi eğitim değerlerini, turuncu çizgiler ise tahmin edilen verilere işaret etmektedir. Şekilde görüldüğü gibi modelin ilk eğitimi sırasında tahmin edilen veri ile gerçek veri arasında büyük bir fark meydana gelmiştir. Grafik, modelin orta kısımlarında ise aynı oranda artış veya azalış olduğunu göstermektedir. Tahmin işlemi son ana geldiğinde tahmin edilen değerle, gerçek değer arasında bir zıtlık meydana gelmiştir. 20 epoch ile eğitilen bu modelin kök ortalama kare hataları toplamı 0.144 olarak tespit edilmiştir.



Şekil 7. 20 epoch ve 2 katmanlı modelin veri kümesi eğitim doğruluğu



Şekil 8. 2 katmanlı 20 epoch ile eğitilen model

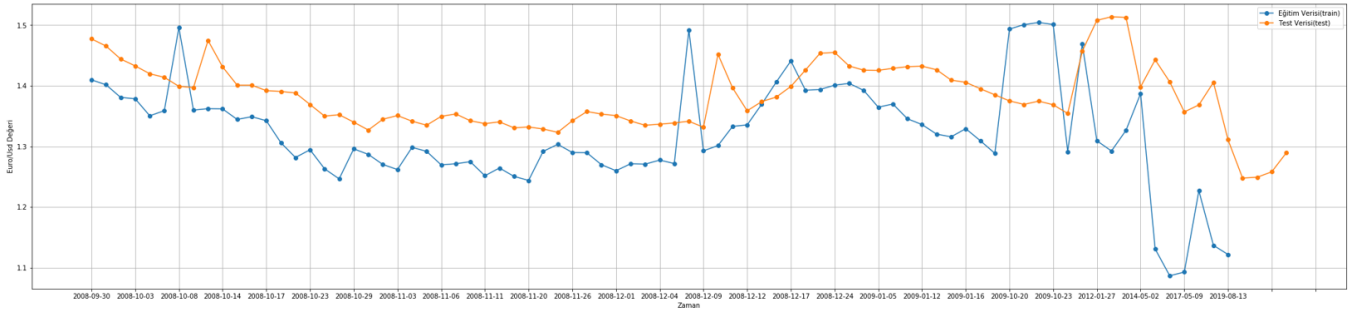
Şekil 9’da 2 katmanlı, 500 epoch değerine sahip modelin kaybı grafiksel olarak gösterilmektedir. Bu modelin kök ortalama kare hataları toplamı 0.087 olarak tespit edilmiştir. Şekil 12, geliştirilen modelin 500 defa eğitildiğini ifade etmektedir. Geliştirilen bu model, Şekil 8, Şekil 10’da olduğu gibi bir izlenim vermektedir. Ancak modelin ortalarına

ve sonlarına gelindiğinde gerçek veriler ile tahmin edilen veriler arasında çok yakın bir ilişki olduğu gözlenmektedir. 20 epoch ile 500 epoch arasındaki fark karşılaştırıldığında modelin 500 epoch ile daha başarılı sonuç verdiği gözlenmiştir.

Epoch sayısının artmasıyla modelin başarı oranının da arttığı görülmektedir. 2 katmanlı 500 epoch değerine sahip modelin 2 katmanlı 20 epoch değerine sahip modele göre daha başarılı olduğu tespit edilmiştir (kök ortalama kare hataları toplamı 0.057).



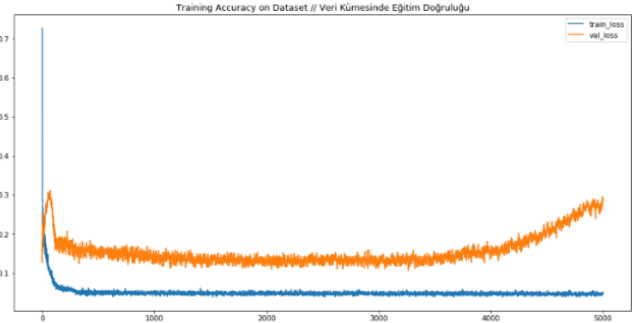
Şekil 9. 2 katmanlı ve 500 epoch modelin veri kümesi eğitim doğruluğu



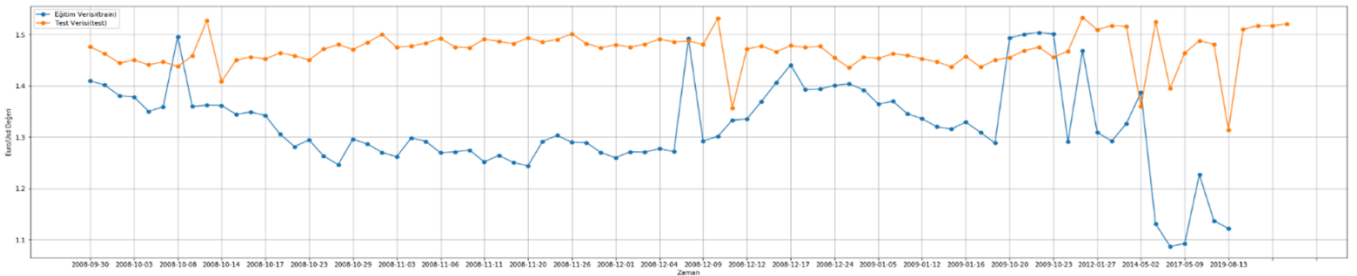
Şekil 10. 2 katmanlı 500 epoch ile eğitilen model

Şekil 11'de 4 katmanlı, 5000 epoch değerine sahip bir model üzerindeki eğitim verileri ile tahmin edilen veriler arasındaki kayıp oranı gösterilmektedir. Bu modelin kök ortalama kare hataları toplamı ise 0.177 olarak ölçülmüştür. Şekil 12, deney 6'da 4 katmanlı 5000 epoch ile eğitilen modeli göstermektedir. Bu modelin analizi sonucunda, 2 katmanlı 5000 epoch ile eğitilen modelle aynı yönde eğilim gösterdiği anlaşılmıştır.

4 katmanlı farklı epoch değerine sahip modellerin karşılaştırılması sonucunda, en küçük kök ortalama kare hataları toplamı 0.089 olarak belirlenmiş ve bu modelin 4 katmanlı 500 epoch değerine sahip model olduğu anlaşılmıştır.



Şekil 11. 4 katmanlı ve 5000 epoch modelin veri kümesi eğitim doğruluğu



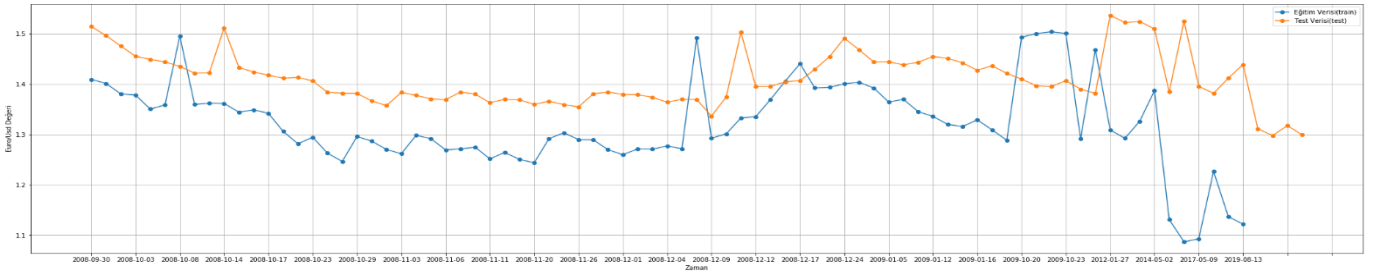
Şekil 12. 4 katmanlı 5000 epoch ile eğitilen model

Şekil 13'te 6 katmanlı, 500 epoch değerine sahip bir modelin kayıp oranı gösterilmektedir. Şekil 14, deney 8'de ise 6 katmanlı 500 epoch ile eğitilen modelin grafiği gösterilmektedir. Bu grafikten, söz konusu modelin 500 Epoch ile eğitilen diğer modeller ile aynı eğilimde olduğu anlaşılmaktadır.

Deney 8 üzerinden denemesi yapılan modelin kök ortalama kare hataları toplamı 0.117 olarak tespit edilmiştir. Epoch sayısının artmasıyla modelin başarı oranının da arttığı görülmektedir. 6 katmanlı 500 epoch değerine sahip bir modelin 6 katmanlı 20 epoch değerine sahip modelin kök ortalama kare hataları toplamı 0.039 olduğundan diğerlerine göre daha başarılı olduğu belirlenmiştir.



Şekil 13. 6 katmanlı ve 500 epoch modelin veri kümesi eğitim doğruluğu



Şekil 14. 6 katmanlı 500 epoch ile eğitilen model

Tablo 5. Çalışma sonucunda yapılan karşılaştırma.

Tarih	Kapanış Değeri	Tahmin Edilen Kapanış Değeri
2007-09-12	1.3908	1,3140
2007-10-03	1.41	1,3321
2008-01-04	1.4754	1,3939
2008-01-11	1.4790	1,3973
2009-01-02	1.3922	1,3153
2009-01-05	1.3645	1,2892
2010-07-23	1.2913	2,5113
2011-04-27	1.4688	2,8565
2015-11-05	1.0868	1,0268
2018-11-09	1.13684	2,2054
2019-08-13	1.12183	1,0599

Tablo 5 üzerinde çalışma sonucunda elde edilen başarı yüzdesi ile tarihler arasında kapanış değerlerinin karşılaştırılması yapılmıştır. Bu karşılaştırmada çalışma sonucunda tahmin edilen veri ile gerçek veri karşılaştırılarak başarı ölçümü yapılmıştır.

6 Sonuç

Bu çalışmada duygu analizi kullanılarak EUR/USD'nin yön hareketini tahmin etmek için bir derin öğrenme algoritmalarından LSTM ile tahmin modeli geliştirilmiştir. Bu kapsamda çalışmada EUR/USD değerlerini yüksek doğrulukta tahmin edebilen bir LSTM tahmin modeli yapılan değişik deneyler ile geliştirilmiş ve modelin doğruluk oranı analiz edilmiştir. Çalışmada veri setlerinin belirlenmesi ile modelin eğitimi/testi olmak üzere iki uygulama yapılmıştır. Çalışmada yapılan ilk uygulama kullanılacak veri setinin oluşturulması aşamalarını içermektedir. Söz konusu veri seti ilk defa bu çalışma kapsamında oluşturulmuştur. Bu veri seti duygu analizi ile belirlenen duygu ortalamaları verilerinin ve EUR/USD değerlerinin zaman serisine dayalı olarak birleşiminden oluşturulmuştur. Veri setinin oluşturulmasında ilk olarak 2005 yılından günümüze kadar Twitter üzerindeki "#eurusd" konu etiketli veya "eurusd" kelimesini barındıran filtrelenmiş İngilizce tweet mesajları içeriklerine göre duygu analizi yöntemiyle sınıflandırılmıştır. Bu maksatla her bir tweetin olumlu, olumsuz veya tarafsız (nötr) olup olmadığı belirlenmiş ve günlük olarak bu tweetlerin duygu oranlarının ortalamaları hesaplanmıştır. Her bir tweetin olumlu, olumsuz veya tarafsız (nötr) olup

olmadığı belirlenmiş ve günlük olarak bu tweetlerin duygu oranlarının ortalamaları hesaplanmıştır. Böylelikle 2273 adet olumlu, 5216 adet olumsuz ve 6360 nötr tweet verisi elde edilmiştir. Veri setinin oluşturulmasında ikinci olarak finance.yahoo.com sitesinden EUR/USD değerleri günlük bazda elde edilmiştir. Veri setinin oluşturulmasında son olarak ise birinci ve ikinci adımlarda elde edilen zaman serisine dayalı veriler "csv" formatında bir veri seti yapısında birleştirilmiştir. Böylelikle yüksek tahmin başarı oranına sahip LSTM modelinin eğitilmesi ve testi aşamalarında kullanılacak olan zaman serisi verilerine dayalı veri seti elde edilmiştir.

Veri seti belirlendikten sonra LSTM tahmin modelinin eğitimi ve testi uygulama aşamasına geçilmiştir. Bu aşamada en yüksek başarı oranına sahip LSTM tahmin modelini belirlemek amacıyla epoch sayılarının ve LSTM sinir ağı katman yapısının değiştirilmesi suretiyle farklı deneyler yapılmıştır. Bu deneyler ile en yüksek başarı oranına sahip LSTM modelini belirlemek hedeflenmiştir. Her bir deneyin doğruluk değeri elde edilen başarı oranları dikkate alınarak değerlendirilmiş ve analiz edilmiştir. Yapılan bu deneyler sonucunda LSTM modeli üzerinde %94.48 başarı oranına sonuç alınmıştır. Ayrıca sistem üzerinden ölçümlenen hata sonucu %5.52 olarak tespit edilmiştir.

Çalışmamızda ilk defa oluşturulan veri seti ile 2005 yılından günümüze kadar olan Twitter üzerindeki "#eurusd" konu etiketli veya "eurusd" kelimesini barındıran filtrelenmiş İngilizce tweet mesajları içeriklerine göre duygu analizi yöntemiyle sınıflandırılmıştır. Çalışmamızda elde edilen sonuçlar, geliştirilen LSTM tahmin modelinin finans açısından EUR/USD hareket yönünün yüksek doğrulukla tahmin edilmesinde güvenilir sonuçlar ortaya koyduğunu göstermektedir. Geliştirdiğimiz uygulama ile Twitter gibi sosyal medya platformlarındaki yorumlar üzerinde duygu analizi yapmak suretiyle EUR/USD paritesinin kapanış değeri tahminini yüksek doğruluk oranı ile yapılabilir.

Gelecekteki çalışmalarda bu makalede geliştirilen LSTM modelinin finans alanında hisse senedi ile tahvil ve bono tahminini içeren farklı veri seti üzerinde kullanılmasına yönelik çalışmalar planlanmaktadır.

Kaynaklar

- [1] Altumur N., Karaca S., "Türkiye için finansal belirsizlik endeksi önerisi". Uluslararası Yönetim Eğitim ve Ekonomik Perspektifler Dergisi. 9(1), 64-77, 2021.
- [2] Şak N., "Finansal piyasalar ve risk algısı: Türkiye üzerine bir inceleme". Finansal Araştırmalar ve Çalışmalar Dergisi. 13(25), 717-731, 2021.
- [3] Doru Ö., Düşünceli F., "Türkiye'de ticari dışa açıklık ve enflasyon ilişkisi: ardl sınır testi ve nedensellik analizi". Finansal Araştırmalar ve Çalışmalar Dergisi Kafkas Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi. 12(23), 37-54, 2021.
- [4] Çoban Ö., Özyer GT., "Twitter duygu analizinde terim ağırlıklandırma yönteminin etkisi". Pamukkale Üniversitesi Mühendislik Bilimleri Dergisi. 24(2), 283-91, 2018.
- [5] Namli E., Ünlü R., Gül E., "Fiyat tahminlemede makine öğrenmesi teknikleri ve doğrusal regresyon yöntemlerinin kıyaslanması; Türkiye'de satılan ikinci el araç fiyatlarının tahminlenmesine yönelik bir vaka çalışması". Konya Mühendislik Bilimleri Dergisi, 7(4), 806-821, 2019
- [6] Aydın C., Darıcı B., Çelik U., "Tekstil fiyatları tahminlemede YSA ve KNEA model karşılaştırması". International Conference on Empirical Economics and Social Sciences (ICEESS'18), Bandırma, Türkiye, 27-28 Temmuz 2018.
- [7] Çelik U., Başarı Ç., "ARPS düşüş eğrisi modelinden genişletilen çok değişkenli doğrusal olmayan model ile altın fiyatlarının analiz edilmesi". International Conference on Empirical Economics and Social Sciences (ICEESS'18), Bandırma, Türkiye, 27-28 Temmuz 2018.
- [8] Taş T., Selim S., "Türkiye'de vadeli işlem ve opsiyon piyasasının etkinliği ve sözleşmelerin karşılaştırmalı fiyat öngörüsü". Ege Akademik Bakış Dergisi, 19(4), 469-485, 2019.
- [9] Özçalıcı M., "Aşırı öğrenme makineleri ile hisse senedi fiyat tahmini". Hacettepe Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi, 35(1), 67-88, 2017.
- [10] Özçalıcı M., "Yapay sinir ağları ile çok aşamalı fiyat tahmini: BIST30 senetleri üzerine bir araştırma". Dokuz Eylül Üniversitesi İktisadi İdari Bilimler Fakültesi Dergisi, 31(2), 209-227, 2016.
- [11] Kayaalp K., Süzen A. A., "Derin Öğrenme ve Türkiye'deki Uygulamaları". Türkiye, IKSAD International Publishing House, 2018.
- [12] Ervural B. Ç., Sarı İ. U., Koçyiğit B., "Kural tabanlı bulanık yaklaşımla talep tahmini ve hızlı tüketim sektöründe bir uygulama". Pamukkale Üniversitesi Mühendislik Bilimleri Dergisi. 24(1), 83-93, 2018.
- [13] İnik Ö., Ülker E., "Derin öğrenmede kullanılan veri setleri ve yazılım kütüphaneleri". 1st International Symposium on Multidisciplinary Studies and

- Innovative Technologies Proceedings Book, Tokat, Turkey, 2-4 Kasım 2017.
- [14] Özkişi H., Topaloğlu M., "Fotovoltaik hücrenin verimliliğinin yapay sinir ağı ile tahmini". *Bilişim Teknolojileri Dergisi*, 10(3), 247-253, 2017.
- [15] Ser G., Bati C. T., "Derin sinir ağları ile en iyi modelin belirlenmesi: mantar verileri üzerine Keras uygulaması". *Yüzüncü Yıl Üniversitesi Tarım Bilimleri Dergisi*, 29(3), 406-417, 2019.
- [16] Şenocak F., "Elektrik Piyasa Takas Fiyatı Ağırlıklı Ortalamasının Anfis ve Yapay Sinir Ağları ile Tahmini". Doktora Tezi, Karadeniz Teknik Üniversitesi, Trabzon, Türkiye 2018.
- [17] Özçalıcı M., "Eşik değerinin hisse senedi fiyat tahmin performansına etkisinin incelenmesi". *Hacettepe University Journal of Economics & Administrative Sciences/Hacettepe Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi*, 35(4), 97-114, 2017.
- [18] Yılmaz V., "Finansmanı öğrenen makineler". 4th International Symposium on Innovative Approaches in Social, Human and Administrative Sciences, Samsun, Türkiye, 22-24 Kasım 2019.
- [19] Afşar A., Yılmazel Ö., Yılmazel S., "konut fiyatlarını etkileyen faktörlerin Hedonik model ile belirlenmesi: Eskişehir örneği". *Selçuk Üniversitesi Sosyal Bilimler Enstitüsü Dergisi*, 37, 195-205, 2017.
- [20] Çelik Ö., Osmanoğlu U.Ö., "Prediction of the prices of second-hand cars". *Avrupa Bilim ve Teknoloji Dergisi*, 16, 77-83, 2019.
- [21] Hayrullahoğlu G., Aliefendioğlu Y., Tanrıvermiş H, Hayrullahoğlu A. C., "Konut piyasalarında Hedonik değerlendirme modeli tahmini: Ankara ili Çankaya ilçesi Çukurambar bölgesi örneği". *Proceedings of 2 nd International Conference on Scientific Cooperation for the Future in the Economics and Administrative Sciences*, Selanik, Yunanistan, 6-8 Eylül 2017.
- [22] Özçalıcı M., "Hisse senedi fiyat tahmininde genetik algoritma ile değişken seçimi". *Anadolu Üniversitesi Sosyal Bilimler Dergisi*, 17(3), 109-124, 2017.
- [23] Akşehir Z. D., Kılıç E., "Makine öğrenmesi teknikleri ile banka hisse senetlerinin fiyat tahmini". *Türkiye Bilişim Vakfı Bilgisayar Bilimleri ve Mühendisliği Dergisi*, 12(2), 30-39, 2019.
- [24] Pabuççu H., "Borsa endeksi hareketlerinin tahmini: trend belirleyici veri". *Selçuk Üniversitesi Sosyal Bilimler Meslek Yüksekokulu Dergisi*, 22(1), 246-256, 2019.
- [25] Kanat E., Dilek Ş., "Pay senedi fiyatlarının bulanık mantık yaklaşımı ile tahmin edilmesi". *Uluslararası Yönetim İktisat ve İşletme Dergisi*, 14(4), 977-1002, 2018.
- [26] Gencer A. H., Sakız B., "Yapay sinir ağları ile bitcoin fiyatını tahminleme". *International Conference On Eurasian Economies*, Taşkent, Özbekistan, 18-20 Haziran 2018.
- [27] Sarıkaya G., "Yapay sinir ağları ve regresyon modelleri ile Bist Ulusal-100 endeksinin tahmini". 21. Yüzyılda Eğitim ve Toplum Eğitim Bilimleri ve Sosyal Araştırmalar Dergisi, 8(23), 325-340, 2019.
- [28] Onocak D., Koç S., "Yapay sinir ağları ile emeklilik yatırım fonu hisse senedi fiyatlarının tahmini". *Finans Ekonomi ve Sosyal Araştırmalar Dergisi*, 3(3), 590-600, 2018.
- [29] Erdoğan N. K., "Finansal zaman serilerinin fraktal analizi". *Aksaray Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi*, 9(4), 49-54, 2017.
- [30] Büyükaşahin Ü. Ç., Ertekin Ş., "Tek değişkenli zaman serileri tahmini için öznitelik tabanlı hibrit ARIMA-YSA modeli". *Journal of the Faculty of Engineering & Architecture of Gazi University*, 35(1), 467-478, 2020.
- [31] Can Ş., Gerşil M., "Manisa pamuk fiyatlarının zaman serisi analizi ve yapay sinir ağı teknikleri ile tahminlenmesi ve tahmin performanslarının karşılaştırılması". *Yönetim ve Ekonomi: Celal Bayar Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi*, 25(3), 1017-1031, 2018.
- [32] Çam S., Ballı E., Sigeze Ç., "Petrol fiyatlarındaki oynaklığın Arch/Garch modelleri ve yapay sinir ağları algoritması ile tahmini". *Uluslararası Yönetim İktisat ve İşletme Dergisi*, 13(5), 588-597, 2017.
- [33] Yiğiter Ş. Y., Sarı S. S., Başakın E. E., "Hisse senedi kapanış fiyatlarının yapay sinir ağları ve bulanık mantık çıkarım sistemleri ile tahmin edilmesi". *Kahramanmaraş Sütçü İmam Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi*, 7(1), 1-22, 2017.
- [34] Yiğiter Ş. Y., Sarı S. S., Karabulut T., Başakın E. E., "Kira sertifikası fiyat değerlerinin makine öğrenmesi metodu ile tahmini". *Uluslararası İslam Ekonomisi ve Finansı Araştırmaları Dergisi*, 4(3), 74-82, 2018.
- [35] Mazed M., *Stock Price Prediction Using Time Series Data*. PhD Thesis, Dakka, Bangladeş, Brac University, 2019.
- [36] Kilimci Z. H., "Borsa tahmini için derin topluluk modelleri (DTM) ile finansal duygu analizi". *Gazi Üniversitesi Mühendislik Mimarlık Fakültesi Dergisi*, 35(2), 635-650, 2020.
- [37] Sezer Ö. B., Güdelek M. U., Özbayoğlu A. M., "Financial time series forecasting with deep learning: a systematic literature review: 2005-2019". *Applied Soft Computing*, 90(106181), 2020.
- [38] Sakarya Ş., Yılmaz Ü., "Derin öğrenme mimarisi kullanarak BİST30 indeksinin tahmini". *European Journal of Educational and Social Sciences*, 4(2), 106-121, 2019.
- [39] Çayır A., Yenidoğan İ., Dağ H., "Konutların günlük elektrik güç tüketimi tahmini için uygun model seçimi". *Fırat Üniversitesi Mühendislik Bilimleri Dergisi*, 30(3), 15-21, 2018.
- [40] Mudinas A., Zhang D., Levene M., (2019). Market trend prediction using sentiment analysis: lessons learned and paths forward. arXiv preprint arXiv:1903.05440.

- [41] Xing F., Hoang D. H., Vo D. V., "High-frequency news sentiment and its application to forex market prediction". In Proceedings of the 54th Hawaii International Conference on System Sciences (HICSS), 2020.
- [42] Bustos O., Pomares-Quimbaya A., "Stock market movement forecast: A Systematic review". Expert Systems with Applications, 156, 113464, 2020.
- [43] Basiri M. E., Nemati S., Abdar M., Cambria E., Acharya U. R., "ABCDM: An attention-based bidirectional CNN-RNN deep model for sentiment analysis". Future Generation Computer Systems, 115, 279-294, 2021.
- [44] Farha I. A., Magdy W., "A comparative study of effective approaches for arabic sentiment analysis". Information Processing & Management, 58(2), 102438, 2021.
- [45] Birjali M., Kasri M., Beni-Hssane A., "A comprehensive survey on sentiment analysis: Approaches, challenges and trends". Knowledge-Based Systems, 107134, 2021.
- [46] Sucu İ., "Yapay Zekânın toplum üzerindeki etkisi ve yapay zekâ (AI) filmi bağlamında yapay zekâyâ bakış". Uluslararası Ders Kitapları ve Eğitim Materyalleri Dergisi, 2(2), 203-215, 2019.
- [47] Ömrüuzun B., Saldanlı A., "Yapay Sinir Ağları ile Kripto Paraların Fiyat Modellemesi". MSc Thesis, Istanbul University, İstanbul, Türkiye, 2019.
- [48] Öztemel E., Yapay Sinir Ağları. Üçüncü baskı. İstanbul, Türkiye, Papatya Yayıncılık, 2012.
- [49] Ivakhnenko A., Grigor'evich O., "Cybernetic predicting devices." Purdue Univ Lafayette Ind School Of Electrical Engineering, 1966.
- [50] Şenkal S., Emeksiz C., "Gauss fonksiyonları ile pencerelenmiş polinom aktivasyon fonksiyonu kullanan dalgacık sinir ağı ile yapay sinir ağlarının, çok kısa vadeli rüzgâr hızı tahmininde, tahmin başarımlarının karşılaştırılması". 3rd International Symposium on Innovative Approaches in Scientific Studies, Ankara, Türkiye, 19-21 Nisan 2019.
- [51] Brownlee J., "Develop Sequence Prediction Models With Deep Learning". Machine Learning Mastery, 2017.
- [52] Berman D., "A survey of deep learning methods for cyber security". Information, 10(4),122, 2019.
- [53] Şeker A., Diri B., Balık H. H., "Derin öğrenme yöntemleri ve uygulamaları hakkında bir inceleme". Gazi Mühendislik Bilimleri Dergisi (GMBD), 3(3), 47-64, 2017.
- [54] Eşref Y., "Türkçe Dizi Etiketleme için Sinir Ağ Modelleri". Master's Thesis, Hacettepe Üniversitesi, Ankara, Türkiye, 2019.
- [55] Işık G., "Türkçe Ağızların Tanınmasında Derin Öğrenme Tekniğinin Kullanılması". Doktora Tezi, Hacettepe Üniversitesi, Ankara, Türkiye, 2019.
- [56] Güdelek M. U., "Zaman Serisi Analiz ve Tahmini: Derin Öğrenme Yaklaşımı". Yüksek Lisans Tezi. TOBB Ekonomi ve Teknoloji Üniversitesi, Ankara, Türkiye, 2019.
- [57] Kızrak M. A, Bolat B., "Derin öğrenme ile kalabalık analizi üzerine detaylı bir araştırma". Bilişim Teknolojileri Dergisi, 11(3), 263-286, 2018.
- [58] Latıfoğlu L., Nuraşam K. B., "Tekil spektrum analizi ve uzun-kısa süreli bellek ağları ile nehir akım tahmini". Avrupa Bilim ve Teknoloji Dergisi, 376-381, 2020
- [59] Liang X., Ge Z., Sun L., He M., Chen H., "LSTM with wavelet transform based data preprocessing for stock price prediction." Mathematical Problems in Engineering, 1-8, 2019.
- [60] Keçeli A.S., Kaya A., "Video görüntülerinde şiddet içeren aktivitelerin LSTM ağı ile tespiti". Dokuz Eylül Üniversitesi Mühendislik Fakültesi Fen ve Mühendislik Dergisi, 21(63), 933-939, 2019.
- [61] Cakar T., "Otomotiv endüstrisinde yapay sinir ağı kullanarak maliyet tahmin modeli geliştirme". Karaelmas Science and Engineering Journal, 7(1), 237-249, 2017.
- [62] Çam S., Kılıç S. B., "Altın fiyatı günlük getirilerinin yapay sinir ağları algoritması ve Markov Zincirleri modelleri ile tahmini". Uluslararası İktisadi ve İdari İncelemeler Dergisi, 681-694, 2018.
- [63] Yasar H., Kilimci Z. H., "US Dollar/Turkish lira exchange rate forecasting model based on deep learning Methodologies and Time Series Analysis". Symmetry. 12(9), 1-18, 2020.
- [64] Bilen B., Horasan F., "LSTM Network based Sentiment Analysis for Customer Reviews". Politeknik Dergisi. 1(1). 2021.
- [65] TwitterScraper. "Twitter Scraper". <https://github.com/taspinar/twitterscraper> (11.12.2020).
- [66] TextBlob. "TextBlob: Simplified Text Processing". <https://textblob.readthedocs.io/en/dev/> (07.07.2021).