



T.C.

GEBZE TEKNİK ÜNİVERSİTESİ

Bilgisayar Mühendisliği Bölümü

**GÜNLÜK HABER
VERİLERİNE GÖRE HİSSE
SENEDİ SEANS KAPANIŞI
FİYAT TAHMİNİ**

Adnan Uğur İNAÇ

Danışman

Yrd. Doç. Dr. Burcu YILMAZ

Aralık, 2017

Gebze, KOCAELİ

BIL495



T.C.

GEBZE TEKNİK ÜNİVERSİTESİ

Bilgisayar Mühendisliği Bölümü

**GÜNLÜK HABER
VERİLERİNE GÖRE HİSSE
SENEDİ SEANS KAPANIŞI
FİYAT TAHMİNİ**

Adnan Uğur İNAÇ

Danışman

Yrd. Doç. Dr. Burcu YILMAZ

Aralık, 2017

Gebze, KOCAELİ

BIL495

Bu çalışma/..... tarihinde aşağıdaki jüri tarafından Bilgisayar Mühendisliği Bölümünde Lisans Bitirme Projesi olarak kabul edilmiştir.

Bitirme Projesi Jürisi

Danışman Adı	Yrd. Doç. Dr. Burcu YILMAZ	
Üniversite	Gebze Teknik Üniversitesi	
Fakülte	Mühendislik Fakültesi	

Jüri Adı	Yrd. Doç. Dr. Burcu YILMAZ	
Üniversite	Gebze Teknik Üniversitesi	
Fakülte	Mühendislik Fakültesi	

Jüri Adı	Asst. Prof. Dr. Uraz Cengiz TÜRKER	
Üniversite	Gebze Teknik Üniversitesi	
Fakülte	Mühendislik Fakültesi	

ÖNSÖZ

Projenin tasarlanması, geliştirilmesi ve sonuç aşamalarında yardımlarından dolayı danışmanım Yrd. Doç. Dr. Burcu Yılmaz'a ve öğrenim hayatım boyunca beni destekleyip hiç yalnız bırakmayan aileme teşekkürlerimi sunuyorum.

Aralık/2017

Adnan Uğur İNAÇ

İÇİNDEKİLER

ÖNSÖZ.....	III
İÇİNDEKİLER	IV
ŞEKİLLER TABLOSU	VI
KISALTMA LİSTESİ	VII
ÖZET	VIII
SUMMARY	IX
1. GİRİŞ.....	1
1.1. PROJE TANIMI	1
1.2. PROJENİN NEDEN VE AMAÇLARI	1
2. KULLANILAN METODOLOJİ VE FORMÜLLER	2
2.1. ÇOK KATMANLI YAPAY SİNİR AĞLARI.....	2
2.2. UZUN KISA VADELİ HAFIZA	3
2.3. SİGMOİD AKTİVASYON FONKSİYONU	4
2.4. HİSSE SENEDİ NORMALİZASYONU.....	6
2.5. ARÇELİK A.Ş SEÇİLMESİNİN NEDENLERİ	6
3. PROJE ÖN YAPILANDIRMASI VE SİSTEM GEREKSİNİMLERİ	7
3.1. VERİ SETİNİN YAPILANDIRILMASI.....	7
3.2. SİSTEM GEREKSİNİMLERİ	8
4. SİSTEM MİMARİSİ	9
4.1. KULLANICI ARAYÜZÜ	9
4.2. ARAYÜZ KISIMLARI VE AMAÇLARI	10
4.2.1. Simülasyon Grafik Penceresi	10
4.2.2. Hisse ve Haber İçerik Penceresi	10

4.2.2.1. <i>Hisse Senedi Sembol Listesi</i>	10
4.2.2.2. <i>Haber İçerik Kutusu</i>	11
4.2.2.3. <i>Haber Tarih Etiketi Kutusu</i>	11
4.2.2.4. <i>Haber Kategorisi Tahmin Yazısı</i>	11
4.2.2.5. <i>Benzer Haber Tarih Yazısı</i>	12
4.2.3. <i>Tahmin Butonu</i>	12
5. YAZILIM SÜREÇLERİ VE GERÇEKLEŞTİRİLMESİ	12
5.1. HİSSE SENEDİ BİLGİLERİNİN VE HABERLERİN TOPLANMASI	14
5.2. HABER İÇERİĞİNİN VE TAHMİN ETKİSİNİN BULUNMASI	15
5.3. DERİN ÖĞRENME METODU KULLANILARAK SEANS KAPANIŞ FİYATININ TAHMİNİNİN YAPILMASI	19
5.4. KULLANICI ARAYÜZÜNÜN OLUŞTURULMASI	20
5.4.1. Çizgi Grafiklerinin Oluşturulması	20
5.4.2. Hisse ve Haber İçerik Penceresi ve Tahmin Butonun Kodlanması	21
6. BAŞARI KRİTERLERİNİN DEĞERLENDİRİLMESİ	22
7. ANALİZ	26
7.1. TREND TAHMİN SONUÇLARI	26
7.2. BAŞARI ORANIN HESAPLANMASI	27
8. TARTIŞMA VE SONUÇ	27
KAYNAKÇA	29
EKLER	30
A. GENEL USE-CASE DİAGRAMI GÖSTERİMİ	30
B. EN İYİ BAŞARININ YAKALANDIĞI AĞ YAPIPSİNİN VE TAHMİN ÇIKTILARININ SAYISAL GÖSTERİMİ	31
C. PROGRAM ARAYÜZÜNÜN BAŞLANGIÇ – ÇALIŞMA – BİTİŞ AŞAMALARI	32

ŞEKİLLER TABLOSU

Şekil 1:Yapay sinir ağlarının genel yapıları ve bağlantılarının gösterimi.	3
Şekil 2:Yapay sinir ağlarında kullanılan aktivasyon fonksiyonları ve grafikleri.	5
Şekil 3: Normalizasyon formülü.	6
Şekil 4:daset.csv dosyası örnek içeriği.	7
Şekil 5: Haber verilerinin etiket yapısı ve dizin yapısı gösterilmektedir.	8
Şekil 6:Kullanıcı arayüzünün bitiş görseli.	9
Şekil 7: Hisse senedi trend tahmini programı akış diyagramı.	13
Şekil 8: Gün Sonu Fiyat Hacim veri seti içeriği	14
Şekil 9: dataset.csv dosyasının içerik ve formatı	15
Şekil 10: Gelen haber ve kategori tahmini sonuç çıktısı.....	18
Şekil 11: Gelen haberin kategori içindeki en benzer haber ile eşleşmesi.	18
Şekil 12: Aynı parametrelere sahip 12 ve 5 katmanlı LSTM sonuçları.....	20
Şekil 13: Grafik özelliklerinin tanıtımı.	21
Şekil 14: Test için kullanılacak olan tarih aralıkları ve değerlerin gösterimi.	23
Şekil 15: 2011 yılı veri seti ile 5 katmanlı ağın tahmin ettiği trendlerin gösterimi.	24
Şekil 16: 01/2011'den 05/2012'ye kadar olan veri üzerinde 25/05/2012'ye ait haber grafiği.....	25
Şekil 17: Trend tahmini başarı oranı formülü gösterimi. [4].....	27
Şekil 18: Genel Use-Case Diyagramı.	30
Şekil 19: En yüksek başarıya sahip LSTM yapısı ve sayısal sonuçları.	31
Şekil 20: Arayüz Başlangıç Görseli.	32
Şekil 21: Arayüz Çalışma Görseli.....	32
Şekil 22: Arayüz Tahmin Bitiş Görseli.....	33

KISALTMA LİSTESİ

ARCLK	= Arçelik A.Ş hisse senedi sembolü
YSA	= Yapay Sinir Ağı
LSTM	= Long Short Term Memory (Uzun kısa vadeli hafıza)
BIST	= Borsa İstanbul
BIST30	= BIST 30 Endeksi, Borsa İstanbul pazarlarında işlem gören, piyasa değeri ve işlem hacmi en yüksek 30 payın bulunduğu bir performans sıralamasıdır.
KAP	= Kamu Aydınlatma Platformu
CSV	= Comma-separated values (Virgülle ayrılmış değerler)
TL	= Türk Lirası
FxPlus	= FxPlus finansal piyasalar için anlık veri ve haber izleme, alım satım emirlerinizi en hızlı biçimde iletmenize olanak tanıyan, teknik ve temel analiz özelliklerini barındıran geniş kapsamlı bir programdır.

ÖZET

Yapılan bu çalışmada şirketlerle ilgili olan haberlerin hisse senetlerinin kapanış fiyatlarına olan etkisi araştırılmıştır. Metin tanıma ve analiz etme yöntemleri kullanılmış olup, yapay sinir ağı kullanılarak da kapanış fiyatının değeri tahmin edilmeye çalışılmıştır. Yapılan araştırmada haberlerin yarattığı fiyat değişimleri geliştirilen bir algoritma ile kanıtlanmaya çalışılmıştır.

SUMMARY

In this study, the effect of the news about the companies on the closing prices of stocks was investigated. Text recognition and analysis methods have been used and artificial neural network has been used to estimate the value of the closing price. The price changes created by the news in the research done were tried to be proved with an improved algorithm.

1. GİRİŞ

Zaman serileri tahmini ve tahmini etkileyen etmenlerin tahmine olan etkileri literatürde uzun zamandan beri çalışılan bir konudur. Zaman serisi tahmininde tek aşamalı (single-step ahead) ve çok aşamalı (multi-step ahead) tahminleri gerçekleştirmek mümkündür. Tek aşamalı zaman serisi tahmininde, girdi değişkenleri, son gözlemden hemen sonraki gözlemi tahmin etmek için kullanılmaktadır. İki ya da daha fazla zaman sonraki gözlemlerin tahmin edilmesi ise çok aşamalı tahmin olarak adlandırılmaktadır. [1, p. 482]

Zaman serileri tahminin uygulama alanlarından biri de hisse senedi fiyatı, yükseliş ve düşüş trendinin bulunmasıdır. Fakat hisse senedi piyasalarının, politik ve sosyal kararlardan çok hızlı ve çok fazla etkilenmesinden dolayı sonuçları tahmin edebilmedeki başarıyı düşürmektedir. Bu çalışma hisse senetlerinin kendileri ile ilişkili olan haberlerin, seans kapanış değerlerini nasıl bir yönde etkilediği bulmak için ve zaman serisi tahminini güçlendirilmesi amaçlanmıştır.

1.1. PROJE TANIMI

Kamu Aydınlatma Platformu(KAP), şirketlerle ilgili haberlerin ve aracı kurum analizlerinin, derin öğrenme metotları kullanılarak hisse senetlerini seans bazlı ve periyotlar halinde nasıl etkilendiğinin bulunması ve değişimin hesaplanması.

1.2. PROJENİN NEDEN VE AMAÇLARI

Para kazanmak , değişen ve gelişen ekonomi içerisinde yerini almak ve daha büyük bir oluşumun parçası olmak için menkul kıymetlere yatırım yapılmaktadır. Bulunulan coğrafya, politik olaylar ve sosyal gelişmeler tercih edilen yatırımda değişimlere neden olabilmektedir. Bunun kestirilmesi şuanki piyasa süreçlerinde bir insanın aklının alabileceğinden çok daha üst seviyelerde değişkenlere sahiptir. Bu projenin amacında bilgili veya bilgisiz kişilere seçtikleri hisse senedinin günlük haberler karşısında ileriki süreçlerde nasıl bir dalgalanma izleyeceğinin grafik yoluyla gösterilmesidir.

Bu proje sayesinde:

- Haberin türü ve niteliği geçmiş haberler ile kıyaslanıp bulunup kullanıcı bilgilendirilecektir. Bu sayede kişi kendi başında bazı çıkarımlar yapabilmesi sağlanacaktır.
- Haberin çıktığı tarihten itibaren seans kapanış fiyatına olan yansımaları grafik ile görmesi tahmini değerleri görmesi sağlanacaktır.

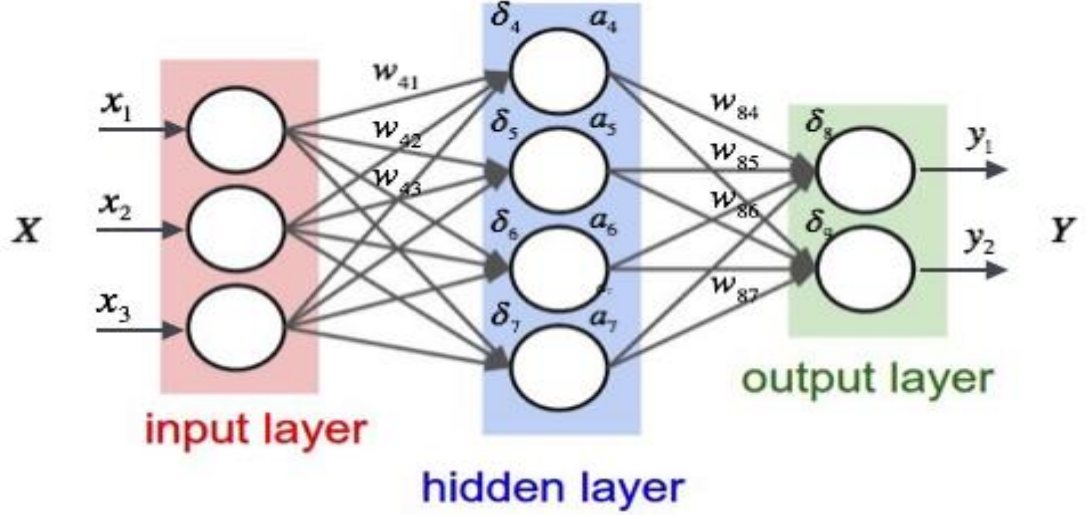
2. KULLANILAN METODOLOJİ VE FORMÜLLER

Bu bölümde çalışmada kullanılan temel analiz yöntemi olan çok katmanlı yapay sinir ağı(YSA), uzun kısa vadeli hafıza (Long short-term memory-LSTM) yapısını, hisse senedi normalizasyonu, LSTM içerisindeki aktivasyon fonksiyonu olan sigmoid fonksiyonun genel açıklaması ve test edilmek için neden Arçelik hissenin seçildiğinden bahsedilecektir.

2.1. ÇOK KATMANLI YAPAY SİNİR AĞLARI

İnsan beyninin bilgi işleme ve ilişkilendirme yönteminden ilham alınarak yapay sinir ağları geliştirilmiştir. Geçen süre zarfında öğrenme üzerine yapılan araştırmalar ile beyin ağ yapısı çözülmeye çalışılmış buda değişen ve kendini geliştirebilen yapay sinir ağlarının önlerini açmıştır.

Çok katmanlı yapay sinir ağları zaman serilerinin tahmini için ideal kabul edilmektedir. Bu ağ yapısı üç temel katmandan oluşur. Girdi katmanı, ara veya gizli katmanlar ve çıktı katmanı. Girdi ve çıktı katmanlarında kaç tane proses elemanının



Şekil 1:Yapay sinir ağlarının genel yapıları ve bağlantılarının gösterimi.

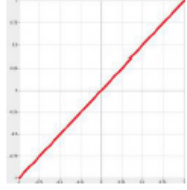
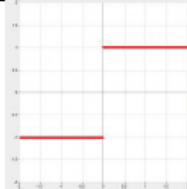
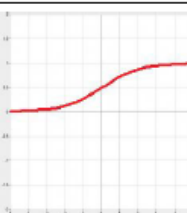
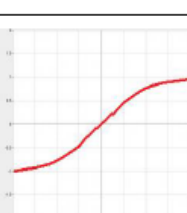
olması gerektiğine probleme bakılarak karar verilir. Ara katman sayısı ve her ara katmandaki proses elemanı sayısının kaç olması gerektiğini gösteren bir yöntem yoktur. Bu deneme yanılma yolu ile belirlenmektedir. Girdi katmanındaki proses elemanlarının her birisi ara katmandaki proses elemanlarının hepsine bağlıdır. Onlarda çıktı katmanındaki proses elemanlarının hepsine bağlıdır. Bilgi akışı girdi katmanından ara katmana oradan da çıktı katmanına ileri doğrudur. Şekil-1’de genel ağ yapısı gösterilmiştir. Bu çoklu katman yapısı ve içerisindeki ara katmanın artırılarak problemlere uygun hale getirilebilir olması çok katmanlı yapay sinir ağlarının gücünü ve başarısını arttırmaktadır.

2.2. UZUN KISA VADELİ HAFIZA

LSTM bir tekrarlayan sinir ağı çeşididir. Yapay sinir ağlarında uzun vadeli bağımlılıkları anlamaya, haliyle "bağlam" farkındalığına sahip sinir ağları elde etmeye yarar. Temelde düğümler arası daha karmaşık bir akış mekanizması kullanarak geri yayılım (*back propagation*) gibi modellerin yol açtığı "hataların katlanarak büyümesi"(*overfitting*) problemini elimine etmekte kullanılır.

2.3. SİGMOİD AKTİVASYON FONKSİYONU

Aktivasyon Fonksiyonu: Bu fonksiyon hücreye gelen net girdiyi işleyerek hücrenin bu girdiye karşılık üreteceği çıktıyı belirler. Aktivasyon fonksiyonu genellikle doğrusal olmayan bir fonksiyon seçilir. Yapay sinir ağlarının bir özelliği olan “doğrusal olmama” aktivasyon fonksiyonlarının doğrusal olmama özelliğinden gelmektedir. Aktivasyon fonksiyonu seçilirken dikkat edilmesi gereken bir diğer nokta ise fonksiyonun türevinin kolay hesaplanabilir olmasıdır. Geri beslemeli ağlarda aktivasyon fonksiyonunun türevi de kullanıldığı için hesaplamamanın yavaşlamaması için türevi kolay hesaplanır bir fonksiyon seçilir. Şekil-2 de yapay sinir ağlarında kullanılan aktivasyon fonksiyonları listelenmiştir. Günümüzde en yaygın olarak kullanılan “Çok katmanlı ağ” modelinde aktivasyon fonksiyonu olarak “Sigmoid fonksiyonu” kullanılır. [2]

Doğrusal (Lineer) Aktivasyon Fonksiyonu		$F(Net) = A * Net$ (A sabit bir sayı)	Doğrusal problemler çözmek amacıyla aktivasyon fonksiyonu doğrusal bir fonksiyon olarak seçilebilir. Toplama fonksiyonundan çıkan sonuç, belli bir katsayı ile çarpılarak hücrenin çıktısı olarak hesaplanır.
Adım (Step) Aktivasyon Fonksiyonu		$F(Net) = \begin{cases} 1 & \text{if } Net > \text{Eşik Değer} \\ 0 & \text{if } Net \leq \text{Eşik Değer} \end{cases}$	Gelen Net girdinin belirlenen bir eşik değerin altında veya üstünde olmasına göre hücrenin çıktısı 1 veya 0 değerini alır.
Sigmoid Aktivasyon Fonksiyonu		$F(Net) = \frac{1}{1 + e^{-Net}}$	Sigmoid aktivasyon fonksiyonu sürekli ve türevi alınabilir bir fonksiyondur. Doğrusal olmayışı dolayısıyla yapay sinir ağı uygulamalarında en sık kullanılan fonksiyondur. Bu fonksiyon girdi değerlerinin her biri için 0 ile 1 arasında bir değer üretir.
Tanjant Hiperbolik Aktivasyon Fonksiyonu		$F(Net) = \frac{e^{Net} + e^{-Net}}{e^{Net} - e^{-Net}}$	Tanjant hiperbolik fonksiyonu, sigmoid fonksiyonuna benzer bir fonksiyondur. Sigmoid fonksiyonunda çıkış değerleri 0 ile 1 arasında değişirken hiperbolik tanjant fonksiyonunun çıkış değerleri -1 ile 1 arasında değişmektedir.
Eşik Değer Fonksiyonu		$F(Net) = \begin{cases} 0 & \text{if } Net \leq 0 \\ Net & \text{if } 0 < Net < 1 \\ 1 & \text{if } Net \geq 1 \end{cases}$	Gelen bilgilerin 0 dan küçük-eşit olduğunda 0 çıktısı, 1 den büyük-eşit olduğunda 1 çıktısı, 0 ile 1 arasında olduğunda ise yine kendisini veren çıktılar üretilebilir.
Sinüs Aktivasyon Fonksiyonu		$F(Net) = \sin(Net)$	Öğrenilmesi düşünülen olayların sinüs fonksiyonuna uygun dağılım gösterdiği durumlarda kullanılır.

Şekil 2:Yapay sinir ağlarında kullanılan aktivasyon fonksiyonları ve grafikleri.

2.4. HİSSE SENEDİ NORMALİZASYONU

Normalizasyon, eşik değerler arasına sıkıştırma anlamı taşımaktadır. Yapılan işlem bir süre zarfındaki en küçük ve en büyük değere göre istenen değerin sıkıştırılmasından ibarettir.

Geniş veri aralığını göz önünde bulundurarak, tüm değişkenlerin değerinin sıfır ve bir arasında değişmek üzere ölçeklendiğinden emin olmak için verileri normalize ediyoruz. Normalleştirme şu şekilde gerçekleştirilir: burada R örnek bir veridir. RN, R'nin normalleştirilmiş değeri, Rmin, R'nin minimum değeri ve Rmax, R'nin maksimum değeri. [3]

$$RN = \frac{R - R_{min}}{R_{max} - R_{min}},$$

Şekil 3: Normalizasyon formülü.

2.5. ARÇELİK A.Ş SEÇİLMESİNİN NEDENLERİ

Arçelik A.Ş. halka arz bazında BIST içerisinde yüksek seviyelerde olduğundan hisse senedinin paralı kişilerce hareket ettirmesinin zor olmasından dolayı seçilmiştir. Aynı zamanda Türkiye kamuoyunda adından sıkça söz ettiren başarılı işler yapan uluslararası ticarete boy gösteren bir kuruluştur. Başarıları da sık sık haberlere konu olmuş insanların ilgisini çekmiştir. Gerçek kapasitesi olarak da ülke ekonomisine çok fazla katkı sağlamaktadır. Kurumsal bir şirket yönetimine sahip olması bakımından yatırımcılar tarafından güven duyulan bir şirket statüsündedir. İçinde bulunduğu beyaz eşya ve akıllı ev aletleri sektöründe yeniliklere çabuk adapte olabilmekte ve takip etmektedir. Bu yüzden yatırımcı kitlesi şirket ile ilgili güncel haberlere sık sık maruz kalmaktadır. Çalışma kapsamında daha iyi sonuçlar alabilmek için popüler ve gerçeği yansıtan Arçelik A.Ş şirketi seçilmiştir.

3. PROJE ÖN YAPILANDIRMASI VE SİSTEM GEREKSİNİMLERİ

Bu bölümde projede olması istenen gereksinimler ve daha fazla hisse için ileriki yıllarda da tahmin yapabilmesi için yapılması gereken yapılandırma süreçleri anlatılmıştır.

3.1. VERİ SETİNİN YAPILANDIRILMASI

1. Proje içerisinde Ocak 2011 ile Aralık 2014 arası BIST içerisinde bulunan bütün hisselerin seans bazlı olarak tarih ve isim sırasına göre verileri formatlı bir şekilde bulunmaktadır. Proje kapsamında Arçelik hissesi izlenmiş ve test edilmiş olduğu için Ocak 2012 ile Nisan 2012 arası yüz elli dört (154) haber tarih ve kategoriye göre etiketlenmiş bir şekilde bulunmaktadır.

Ayrıca haberlerin kategorize edilmesi ve Arçelik hissesinin seans kapanış fiyat tahmini yapabilmesi için eğitilmiş olan yapay sinir ağları zip formatında proje içerisinde bulunmaktadır.

2. Yeni bir analiz yapılmak istendiği takdirde:

1. Hisse senedi ile ilgili tahmin yapılacağı günden en az bir yıl geçmişteki “date, symbol, open, close, low, high, volume” değerlerinin doğru sırada virgül (,) ile ayrılmış olarak oluşturulması ve isminin “dataset.csv” olarak projenin “\dl4jMaster\dl4j-examples\dl4j-examples\src\main\resources” dizinine kaydedilmesi gerekiyor.

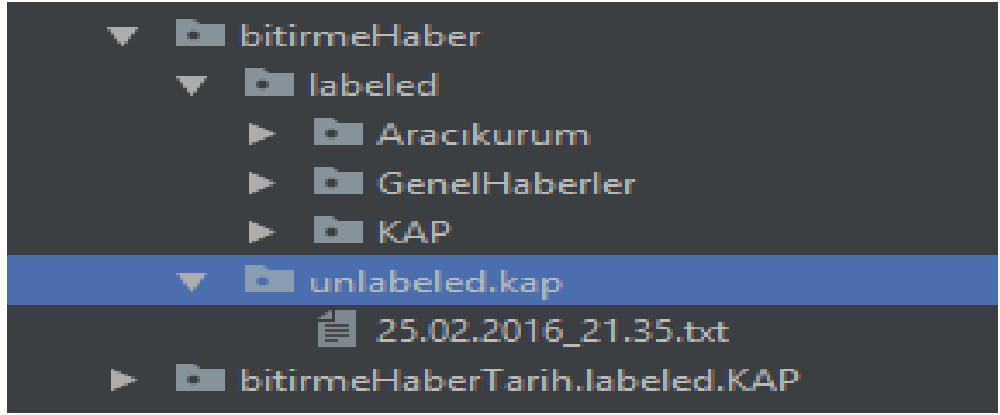
```
date,symbol,open,close,low,high,volume
3.01.2011,ACIBD,12.65,12.55,12.55,12.65,80080.45
3.01.2011,ACIBD,12.55,12.7,12.55,12.75,251900.45
3.01.2011,ADANA,5.6,5.62,5.58,5.64,835436.18
```

Şekil 4:dataset.csv dosyası örnek içeriği.

2. Haber veri setini değiştirmek veya yeni bir hisse ile ilgili haberler eklenmek istendiğinde, haberler “AracıKurum , GenelHaberler , KAP” olmak üzere üç gruba göre ayrılmalıdır. Etiketlenmiş bu haberler “hisse_senedinin_kısaltması_labeled” şeklinde bir klasör içerisine eklenmeli ve klasör proje içerisindeki

“\dl4jmaster\dl4j-examples\dl4j-examples\src\main\resources\bitirmehaber” kısmı içerisine kaydedilmelidir.

3. Haberlerin benzerlerinin bulunabilmesi içinde önceden etiketlenmiş kageti bazlı haberlerin “gün.ay.yıl” bazında klasörlere ayrılarak “\dl4jMaster\dl4j-examples\dl4j-examples\src\main\resources\bitirmeHaberTarih” klasörü içerisinde “hisse_senedinin_kısaltması_labeled” şeklinde eklenmesi gerekmektedir.



Şekil 5: Haber verilerinin etiket yapısı ve dizin yapısı gösterilmektedir.

Şekil-5’i inceleye bilirsiniz.

4. Program içerisindeki bir yıllık veriler ışığında minimum 11 günlük (22 seans) tahmin üretebiliyor. Tahmin süresi göz önüne alındığında programa verilecek olan ilişkili haber son 11 günlük aralıkta olursa haber tarihinden sonra yapılan tahminlerdeki değişim ve gerçek fiyat arası fark doğru şekilde bulunabilir. Kullanıcı son olarak arayüzünden istenilen hissenin kısaltması(sembolünü) seçerek programı çalıştırabilir.

3.2. SİSTEM GEREKSİNİMLERİ

Bu gereksinimlerin sağlanabilmesi için gerekli ihtiyaçlar şunlardır:

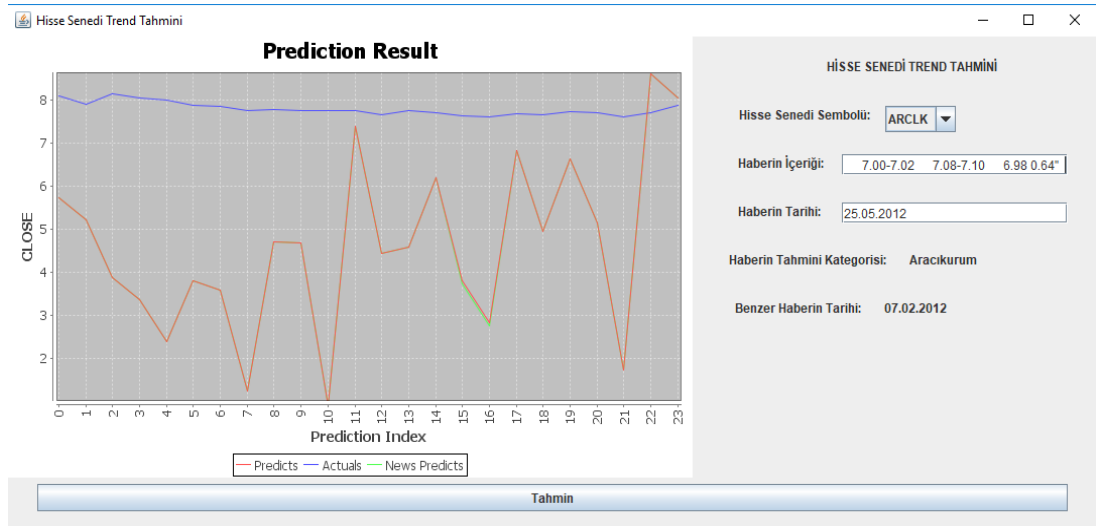
1. Windows 10 veya Linux işletim sistemine sahip bilgisayar.
2. Java Geliştirme Aracı (JDK).
3. IntelliJ IDEA tümleşik geliştirme ortamı (IDE).

4. SİSTEM MİMARİSİ

Bu bölümde projenin genel kullanıcı arayüzü ve arayüz parçalarının tanıtımı yapılmaktadır.

4.1. KULLANICI ARAYÜZÜ

Projenin kullanıcı arayüzü bitiş görünümü şu şekildedir:



Şekil 6:Kullanıcı arayüzünün bitiş görseli.

Şekil-6’da programın ilk çalıştırıldığında kullanıcıyı karşılayan arayüzü geliyor. Kullanıcının verileri girmesi gereken kısımlar arayüzün sağ tarafında yukarıdan aşağı dizilmiş durumdadır. Arayüz üzerinde toplamda 7 adet etkileşimde bulunulabilen yapı bulunmaktadır. Arayüz bilgiye kolay ulaşılması ve kullanım kolaylığı için üç parçadan oluşmaktadır. Parçalar şunlardır:

- “Simülasyon Grafik” penceresi
- “Hisse ve Haber İçerik” penceresi
- “Tahmin” Butonu

4.2. ARAYÜZ KISIMLARI VE AMAÇLARI

4.2.1. Simülasyon Grafik Penceresi

Simülasyon grafik penceresi, yatay ve dikey eksene sahip tahmin sonuçlarının çizgi grafik halinde gösterildiği kısımdır. Grafiğin yatay eksenini arka arkaya tahmin yapılan seans adetini göstermektedir. Dikey eksen ise hisse senedinin seans kapanış fiyatını gösterir. Proje kapsamında yükseliş ve düşüş trend tahminini gözlemlemek ve analiz edebilmek için grafik art arda yapılan tahminleri çizgi grafik halinde göstermektedir. Buda bize iki tahmin arasındaki negatif veya pozitif eğim ile trendin belirlenmesinde yardımcı olur. Grafik penceresinde iki adet sonuç üst üste gösterilmiştir. Programın doğruluğunu test etmek için yapılan bu işlemde kırmızı ile gösterilen çizgi Predicts(tahmin), mavi ile gösterilen Actuals(gerçek), yeşil ile gösterilen NewsPredicts(Haberlerin katkısı) gösterilmiştir. Grafikteki her dikey çizgilerle denk gelen kısımlar üzerine gelindiğinde tahmin ve gerçek sonuca ait değerler görünür olmaktadır. Şekil-13'ü inceleyebilirsiniz.

4.2.2. Hisse ve Haber İçerik Penceresi

Hisse ve Haber İçerik penceresi, tahmini yapılacak olan hissenin seçildiği, tahmine olan etkisinin gözlemleneceği haberin ve tarihinin girildiği ve haber tahmin sonuçlarının gözlemlendiği içinde aktif ve pasif olmak üzere toplam 5 component bulunduran kısımdır.

- Hisse Senedi Sembol listesi
- Haber İçerik kutusu
- Haber Tarih Etiket kutusu
- Haber Kategorisi Tahmin yazısı
- Benzer Haber Tarih yazısı

4.2.2.1. Hisse Senedi Sembol Listesi

Tahmin yapılması istenen hisse senedinin kısaltmasının veya sembolünün seçilmesi gereken yerdir. Bu bölüm Swing kütüphanesi kullanılarak JComboBox olarak tasarlanıp içine 15 seçilebilir öğe konulmuştur.

4.2.2.2.Haber İçerik Kutusu

Tahminlerin etkisini arttırmak ve gerçekçilik katmak için projenin ana konusu olan haberlerin değerlendirilmesi kısmı buradan başlamaktadır. Tahminin yapılacağı aralıkta gözlemlemek istenen haberin içeriğinin girilmesi gereken yerdir. Buradan alınan haber metni projenin kaynakları içinde bir dosya olarak tarihi ile birlikte etiketlenip işlem yapılacaktır. JTextField olarak tasarlanmış olan buradan bilgiler getText() metodu ile programın içine çekilmektedir. Program tahminleri yapmaya başladığında haberi okur ve hangi kategoriye ait olduğu çıkarımını yapar.

4.2.2.3.Haber Tarih Etiketi Kutusu

İşlenecek olan haberin kaçınıcı tahmine denk geleceğini bulmak kullanıcıdan alınan gerçek haber tarihi verisinin girildiği kısımdır. Çıkan grafikte gösterim yapabilmek için haberin gerçek tarihi önemlidir. Yapılacak tahmin ancak gerçek bir veri ile karşılaştırılırsa inandırıcı olacağı için bu kısımda kullanıcı tarafından doğru bir zaman girilmelidir. “gün.ay.yıl” formatında girdi alan kısımdır.

4.2.2.4.Haber Kategorisi Tahmin Yazısı

İçeriği girilmiş haberin hisselere olan etkisinin ölçülmesi için yapılan kategori bazındaki tahminin çıktısının gösterildiği yerdir. Burası pasif olduğu için *JLabel* yapısı kullanılmıştır. İşlem tamamlandığında kategori etiketi ekranda çıkmaktadır.

4.2.2.5. Benzer Haber Tarih Yazısı

Kategorisi bulunmuş olan haberin, geçmişteki en benzer içerikli haberin o gün nasıl bir ivmelendirme yaşattığını bulmak için yapılan kategori içi tarih araması sonucu bulunan en yakın haberin etiketidir. Bu kısımda pasif bir yapıya sahiptir.



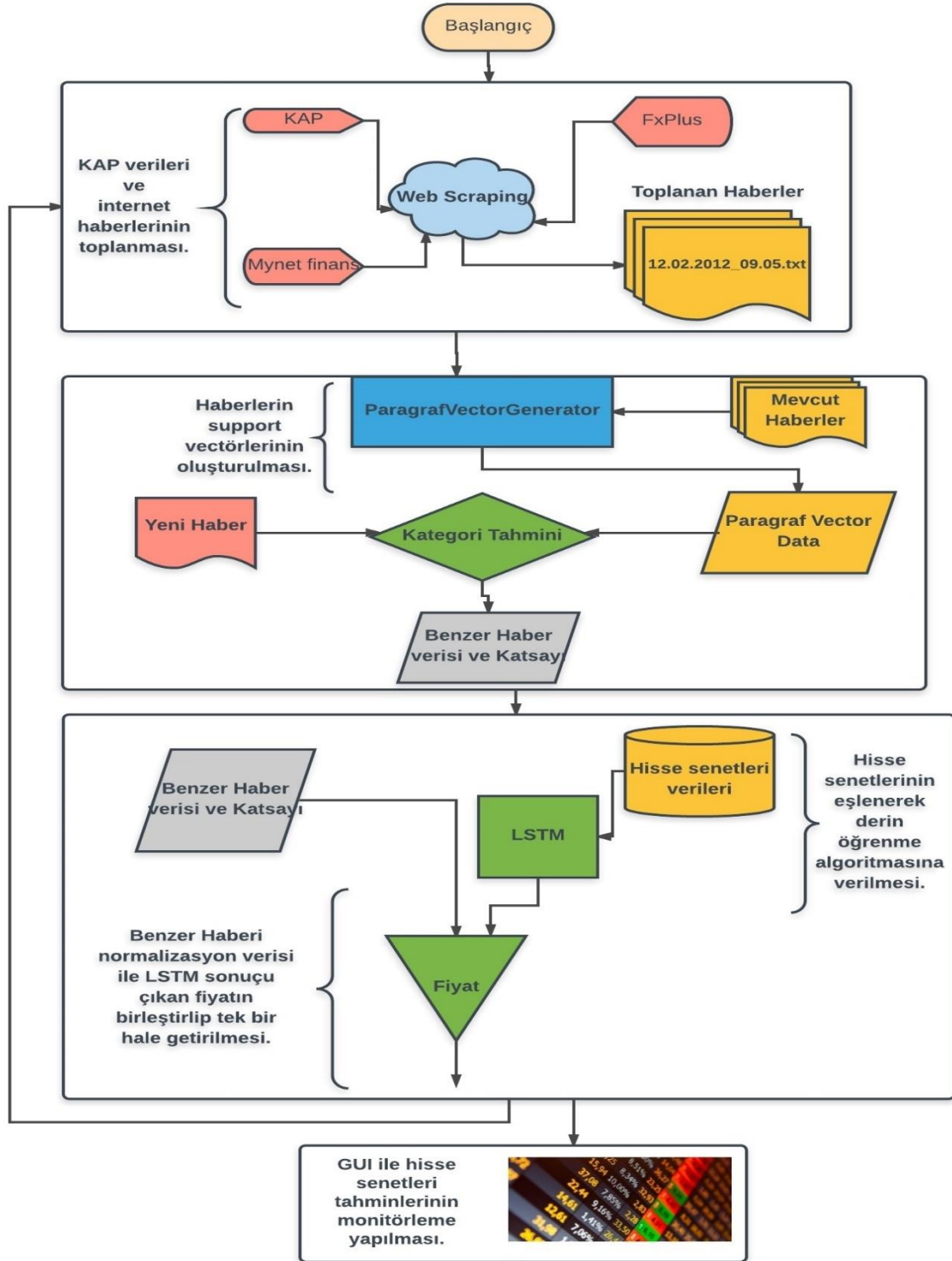
Şekil 23: Hisse ve Haber İçerik Penceresi

4.2.3. Tahmin Butonu

Programın verilen data sete ve haber içeriğine uygun olarak tahmin sürecini başlatan tetikleyici butondur.

5. YAZILIM SÜREÇLERİ VE GERÇEKLEŞTİRİLMESİ

Projenin tamamı art arda süreçlerin sıralı bir şekilde çalışarak en sonunda kullanıcı arayüzünde gösterilmesi ile tamamlanmaktadır. Bu yüzden programın tamamlanma süreçleri hisse senedi bilgileri, ilişki haberlerin toplanması ve ön işlemden geçirilmesi, yeni haberin içeriğinin belirlenmesi ve fiyata olan etkisinin bulunması, hisse senedinin derin öğrenme kullanılarak seans kapanış fiyatının tahmininin yapılması ve sonuçların arayüzde gösterimi olmak üzere dört etaba ayrılmıştır. Şekil-7’de programın akış diyagramı gösterilmiştir.



Şekil 7: Hisse senedi trend tahmini programı akış diyagramı.

5.1. HİSSE SENEDİ BİLGİLERİNİN VE HABERLERİN TOPLANMASI

Hisse senetleri üzerinde normalizasyon yaparak yapay sinir ağımızın katmanlarını besleyeceğimiz için gün içindeki bazı kritik verilere ihtiyacımız var. Program içerisinde gün sonu kapanış fiyatı tespiti yapıldığı için “tarih, sembol, seans açılış fiyatı, seans kapanış fiyatı, seans içi en yüksek fiyat, seans içi en düşük fiyat ve seans içi işlem miktarı” ihtiyaç vardı. Gerçekçi ve kaliteli tahmin yapabilmek adına bu veriler BIST DataStore üzerinden “Gün Sonu Fiyat Hacim” bilgilerinden oluşturulmuştur. Proje kapsamında Ocak 2011 ile Aralık 2014 yılları arası veriler kullanılmıştır. Şekil-8’da görüldüğü gibi indirilen veriler istenen formata uygun değildir ve çok fazla gereksiz bilgi içermektedir.

#	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T
1	TARİH;SEANS NO;PAZAR;PAY KODU;SERİ KODU;MIN FİYAT;MAX FİYAT;ONCEKI KAPANIS FİYAT;KAPANIS FİYAT;AO;İSLEM HACMI;İSLEM MİKTARI;SOZLESME SAYISI;SEANS SIRA NO;BAZ FİYAT;ALT LIMIT;UST LIMIT;N																			
2	DATE;SESSION NO;MARKET;STOCK SYMBOL;STOCK SERIES;MIN PRICE;MAX PRICE;PREVIOUS CLOSE PRICE;CLOSE PRICE;WAP;TRADED VALUE;TRADED QUANTITY;CONTRACT NUMBER;SESSION SEQUENCE NO;BASE F																			
3	02.11.2015;2;L;ACSEL;E;54.50;55.15;55.15;54.50;54.82;88704.60;1618;103;24573292;55.40;49.85;60.95;500;0;H;55.15;103;4;55.21;8;0;M;320;16;199;10;54.5;54.50;55																			
4	02.11.2015;1;L;ACSEL;E;54.80;58;54.40;55.15;55.38;210388.05;3799;193;24573291;54.25;48.80;59.70;500;0;H;58;11;2;55.38;8;0;M;0;0;0;0;54.80;55.15																			
5	02.11.2015;1;N;ADANA;E;6.69;6.80;6.57;6.76;6.74;846648.96;125527;161;24573291;6.56;5.90;7.22;5000;0;H;6.70;10023;12;6.74;A;0;M;0;0;0;0;6.75;6.76																			
6	02.11.2015;2;N;ADANA;E;6.75;6.77;6.76;6.76;144410.28;21350;91;24573292;6.74;6.06;7.42;5000;0;H;6.77;1000;6;6.75;A;0;M;301;3;1751;4;6.77;6.77;6.78																			
7	02.11.2015;1;N;ADBGR;E;4.50;4.58;4.45;4.57;4.55;61001.13;13413;76;24573291;4.42;3.97;4.87;500;0;H;4.58;100;2;4.55;A;0;M;0;0;0;0;4.57;4.58																			
8	02.11.2015;2;N;ADBGR;E;4.56;4.59;4.57;4.57;4.57;68911.19;15073;65;24573292;4.55;4.09;5.01;500;0;H;0;0;0;0;4.56;A;0;M;2289;10;26;1;4.57;4.57;4.59																			
9	02.11.2015;1;N;ADEL;E;66.35;68.50;64.85;67.20;67.29;3412495.65;50716;795;24573291;64.25;57.80;70.70;1000;0;E;68;2065;39;67.29;A;0;M;0;0;0;0;67.20;67.25																			
10	02.11.2015;2;N;ADEL;E;66.60;67.25;67.20;66.75;66.80;2076918.60;31091;599;24573292;67.30;60.55;74.05;1000;0;E;67.20;381;5;67.1;A;0;M;719;10;131;3;66.75;66.75;66.85																			
11	02.11.2015;1;L;ADESE;E;3.65;3.76;3.43;3.76;3.72;2695836.12;725246;1191;24573291;3.41;3.06;3.76;50000;0;H;3.66;21490;88;3.72;A;0;M;0;0;0;0;3.76;0																			
12	02.11.2015;2;L;ADESE;E;3.76;3.87;3.76;3.79;3.81;3171799.63;833420;1361;24573292;3.72;3.34;4.10;50000;0;H;3.80;11718;27;3.76;A;0;M;4706;12;1378;3;3.79;3.79;3.80																			
13	02.11.2015;1;N;ADNAC;E;0.68;0.70;0.67;0.69;0.69;2254886.59;3266674;446;24573291;0.67;0.60;0.74;50000;0;H;0.70;62400;13;0.69;A;0;M;0;0;0;0;0.69;0.70																			
14	02.11.2015;2;N;ADNAC;E;0.69;0.70;0.69;0.69;0.70;140920.04;202372;110;24573292;0.69;0.62;0.76;50000;0;H;0.69;2325;8;0.69;A;0;M;1097;10;2;1;0.69;0.69;0.70																			
15	02.11.2015;1;N;AEFES;E;23.80;23.90;23.90;23.45;23.64;2906419.15;122957;439;24573292;24.10;21.65;26.55;10000;0;E;23.85;526;5;23.81;A;0;M;28096;36;5000;4;23.45;23.45;23.55																			
16	02.11.2015;1;N;AEFES;E;23.80;23.90;23.90;24.08;1944837.20;80753;351;24573291;23.20;20.75;25.30;10000;0;E;24.50;13002;33;24.08;A;0;M;0;0;0;0;23.85;23.95																			
17	02.11.2015;2;N;AFYON;E;6.65;7.08;6.64;7.06;6.85;92014210.28;13441133;12143;24573292;6.58;5.92;7.24;10000;0;E;6.70;92402;89;6.76;A;0;M;168018;160;55889;64;7.06;7.05;7.06																			
18	02.11.2015;1;N;AFYON;E;6.64;6.64;6.13;6.64;6.58;42703562.38;6487344;7239;24573291;6.03;5.42;6.64;10000;0;E;6.64;227963;533;6.58;A;0;M;0;0;0;0;6.64;0																			
19	02.11.2015;2;K;AGYO;E;1.38;1.39;1.39;1.39;1.39;108827.77;78538;78;24573292;1.39;1.25;1.53;5000;0;H;1.38;5;1;1.39;8;0;M;5390;4;7593;10;1.39;1.39;1.40																			
20	02.11.2015;1;K;AGYO;E;1.37;1.40;1.36;1.39;1.39;198958.75;143574;146;24573291;1.36;1.22;1.50;5000;0;H;1.40;2800;6;1.39;8;0;M;0;0;0;0;1.38;1.39																			

Şekil 8: Gün Sonu Fiyat Hacim veri seti içeriği

Dosya csv formatına çevrilerek veri ayraçları değiştirilip gereksiz sütunlar silinmiştir. Eklenen her dosya için python dilinde yazılan bir kod ile yapılmıştır. Dosyalar istenen formata getirildikten sonra tarih sırasına göre satırlar alt alta eklenerek tek bir “dataset.csv” adında dosyada birleştirilmiştir.

1	date,symbol,open,close,low,high,volume
2	3.01.2011,ACIBD,12.65,12.55,12.55,12.65,80080.45
3	3.01.2011,ACIBD,12.55,12.7,12.55,12.75,251900.45
4	3.01.2011,ADANA,5.6,5.62,5.58,5.64,835436.18
5	3.01.2011,ADANA,5.62,5.6,5.54,5.76,1254009.2
6	3.01.2011,ADBGR,0,3.44,3.43,3.45,125178.53
7	3.01.2011,ADBGR,0,3.43,3.38,3.46,188882.98
8	3.01.2011,ADEL,20.2,20.5,20.2,20.65,553354.9
9	3.01.2011,ADEL,20.5,20.55,20.4,20.65,501372.15
10	3.01.2011,ADNAC,0.84,0.85,0.84,0.88,7930702.06
11	3.01.2011,ADNAC,0.85,0.86,0.85,0.86,2442525.13
12	3.01.2011,AEFES,0,23.7,23.4,23.8,654825.3

Şekil 9: dataset.csv dosyasının içerik ve formatı

Bu proje kapsamında yukarıda bahsedilen nedenlerden dolayı Arçelik hisse üzerine yoğunlaşarak onunla ilgili Ocak 2012 ile Nisan 2012 arası farklı konu ve içeriklere sahip yüz elli dört haber FxPlus uygulaması üzerinden elle kopyalanmıştır. Toplanan haberler içerik bakımından üç kategoriye ayrılmıştır. Bunlar kategoriler:

- Aracı kurumların analistlere hazırlattığı raporlar.-AracıKurum
- Medyada yayınlanan veya şirketin kendi yayınladığı basın bültenleri.-GenelHaberler
- Şirketlerin, devlet ve kendi mükellefleri ile paylaşmak zorunda olduğu bilgiler.-KAP

Bu kategoriler hisse senedinin fiyatına etki etmeleri bakımından farklı ağırlıklara sahip oldukları için ayrılmışlardır. Haberler kategorilere göre etiketlendikten sonra algoritmanın en benzer haberin olduğu günü seçebilmesi içinde kategoriler içerisinde gün bazlı klasörlere ayırma yapılmıştır. Bu işlem sayesinde geçmiş haber örnekleri içerisinde en benzer haberin çıktığı gün bulunarak o günkü değişiminde tahmin sonucuna yansıtılması hedeflenmiştir.

5.2. HABER İÇERİĞİNİN VE TAHMİNE ETKİSİNİN BULUNMASI

Finans piyasalarının ve hisse senetlerinin fiyatlarının değişiminde yüzlerce farklı etmen vardır. Projemizin asıl amacı olan gerçekçi bir geçmiş analizinden yine gerçekçi bir gelecek tahmini yapmak olduğu için insan davranışlarının da buna uygun olması beklenmektedir. Yatırımcılar için hisselerin fiyatının belirlenmesinde

sosyal ve politik olayların yanında daha elle tutulur ve analiz edilebilir olan gerçek haberlerin niteliği ön plana çıkmaktadır. Çünkü haberlerin nitelikleri tartışılmaz bir gerçeği belirte bilir ve buda yatırımcıların gelecek ön görülerinde hissenin fiyatlanmasında önemli bir rol oynar. Finans sektöründe “beklenti” diye adlandırılan şey aslında para kazanmak için alınacak riskin miktarını göstermektedir. Araştırmada yatırımcıların realisttik olarak düşündüğü ve buna cevap verdiği düşünülmektedir. Ayrılan kategorilerde burada devreye girerek önem sıralarına göre katsayılar belirlenmiştir. “AracıKurum” kategorisi olarak bahsedilen haber verilerinde Türkiye ve Dünya üzerindeki Arçelik şirketi ile ilgili yapılan analizler, tahminler ve kredi notları haberleri yer almaktadır. Bu haber çeşidi yatırımcıları fazlası ile etkilemediği, çok sık yayınlandığı ve geniş bir ağı yayılamadığı için çok düşük bir dalgalanma yaratır. Bu yüzden “AracıKurum” kategorisinin katsayısı 0,05 olarak puanlanmıştır.

“GenelHaberler” kategorisi olarak bahsedilen haber verilerinde Arçelik şirketinin basın bültenleri, ulusal ve uluslararası basında yer alan haberleri, bağlı olduğu holding haberleri, şirketin boy gösterdiği sektör ile ilgili haberler bulunmaktadır. Bu kategori içerik bakımından çok daha anlaşılabilir ve son kullanıcı tarafından yorumlanabilir olduğu için yatırımcılar tarafından daha çabuk fiyatlanabilir. “GenelHaberler” kategorisinin katsayısı 0,15 olarak puanlanmıştır.

“KAP” kategorisi olarak belirtilen haberler verileri Arçelik şirketinin halka arzı sebebiyle tüm kamuoyu ile paylaşmak zorunda olduğu ve bir fiil yatırımcıları ilgilendiren konuların bahsedildiği haberlerdir. Kategorinin önem kazanmasının asıl nedeni bu haberlerin şirketin işleyişi, karlılık ve yapılan yatırımları doğrudan barındırıyor olmasından kaynaklanmaktadır. Yatırımcılar için fazlası ile önemli olan bu haber türü bütün insanların ulaşabilmesi içinde Kamu Aydınlatma Platformu üzerinden anlık olarak kayıta geçip yayınlanmaktadır. Bu yüzden “KAP” kategorisinin katsayısı 0,3 olarak puanlanmıştır.

Kategorilerin toplam katsayı puanı 0,5’tir. Şirketlerin hisseleri haberlerden etkilendiği gibi kişilerin parasal güçlerinden de etkilenebilir bu yüzden oranlama böyle yapılmıştır.

Çalışma kapsamında haber metinlerinden anlam analizi yapabilmek için *DeepLearning4J*'nin *ParagraphClassifier* ve *ParagraphVectors* özelliklerinden faydalandım. Paragraf vektörlerini ve bir model oluşturmak için *ParagraphClassifier* sınıfından bir obje oluşturuldu. Kendisine ait bir model oluşturmak için öğrenme işleminin gerçekleştiği *makeParagraphVectors(string labeledDocFileName)* metodu kullanıldı. Bu metod kendi içerisinde elle kategorilere ayrılmış olan haberleri kendi içlerinde parçalara ayırarak kelimelerin diziliş, tekrar etme ve anlamsal yapılarına göre haberleri öğrenmeye başlar ve kendine ait bir model oluşturur. Bu model ve haberlerin kelimeleri yeni gelecek olan haber ile ileride ilişki kurulması için saklanır. *ParagraphVectors* sınıfının modeli oluşturulurken şu parametreler dikkate alınmıştır:

learningRate = 0,025 , *minLearningRate* = 0,001 , *batchSize* = 1000 , *epochs* = 20

Oluşturulan model .zip uzantısı ile kaydedilip saklanmıştır. Model oluşturulduktan ve eğitim tamamlandıktan sonra kullanıcının girdiği kategorisi bilinmeyen haberin analizine geçilir. *CheckUnlabeledData(string unlabeledFile)* metodu ile haber içeriği ilk önce okunur. Daha sonra *MeansBuilder* ve *LabelSeeker* sınıfları önceden oluşturulan ve sınıflandırılan paragraf bilgileri ile ayağa kalkarlar. Bir döngü yardımı ile kategorilerin anlam vektörleri gerilerek yeni haberlerin anlam vektörü ile karşılaştırılır. En yüksek sonucun çıktığı kategori ismi ile haber etiketlenir ve katsayısı kaydedilir.

Dört adet test etmek amaçlı haber kullanılmıştır. Haberler daha rahat ayırt edilmesi adına okunarak “TestVerisi_kategori” şeklinde etiketlenmiştir. Şekil-10’de görüldüğü gibi sonuçlar üretmektedir.

```
Document '[TestVerisi_aracikurum]' ařağıdaki kategorilere girer:  
Aracikurum: 0.10445071756839752  
GenelHaberler: -0.17203302681446075  
KAP: -0.3377822935581207  
Tahmin, Aracikurum: 0.10445071756839752  
Document '[TestVerisi_genelhaberler]' ařağıdaki kategorilere girer:  
Aracikurum: -0.12481030076742172  
GenelHaberler: 0.5182886719703674  
KAP: -0.17525260150432587  
Tahmin, GenelHaberler: 0.5182886719703674  
Document '[TestVerisi_genelhaberler2]' ařağıdaki kategorilere girer:  
Aracikurum: -0.3447302579879761  
GenelHaberler: 0.5519798994064331  
KAP: 0.15251272916793823  
Tahmin, GenelHaberler: 0.5519798994064331  
Document '[TestVerisi_kap]' ařağıdaki kategorilere girer:  
Aracikurum: -0.2745513916015625  
GenelHaberler: 0.21073710918426514  
KAP: 0.656518280506134  
Tahmin, KAP: 0.656518280506134
```

řekil 10: Gelen haber ve kategori tahmini sonu ıktısı.

Daha sonra kategorisi belirlenmiř olan haber gemiřteki en yakın anlam ieriğıne sahip haberi bulabilmek adına bu iřlemleri kategorisi ierisinde “gün.ay.yıl” řeklinde etiketlenmiř veriler iinde tekrardan yapmaktadır.

```
Document '[TestVerisi_kap]' ařağıdaki kategorilere girer:  
Tahmin, 12.04.2012: 0.5584744811058044
```

řekil 11: Gelen haberin kategori iindeki en benzer haber ile eřleřmesi.

Test iin seilen haberin ieriğı “ARELİK A.ř. (Genel Kurul Toplantısı ağırsı)” iken, buna karřılık olarak algoritmanın benzerliğı en yksek olarak setiğı KAP kategorisi haberinin ieriğı de “Ynetim kurulu toplantısı” olmuřtur.

İkinci yapılan derinlemesine arama iřlemi ile de gemiřte benzeřtiğı haberin o gnk normalizasyon deėerini bulabiliyor. Elde edilen bu iki farklı katsayı, LSTM sonucuna ekleniyor.

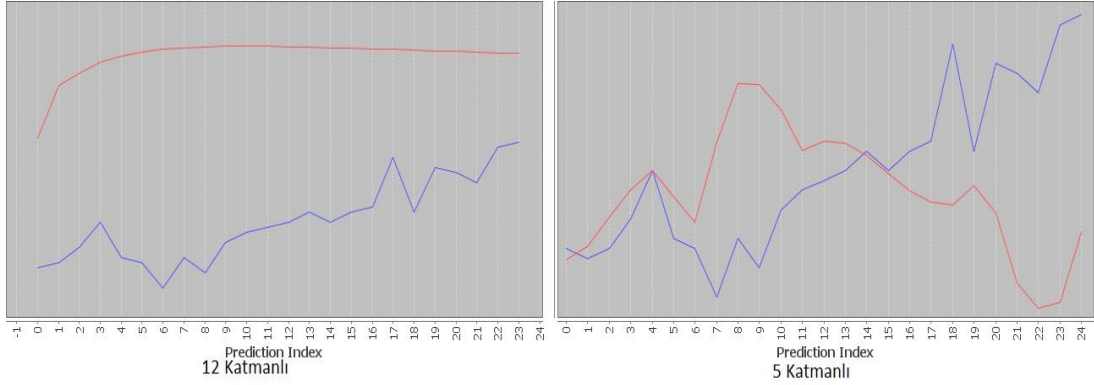
5.3. DERİN ÖĞRENME METODU KULLANILARAK SEANS KAPANIŞ FİYATININ TAHMİNİNİN YAPILMASI

Çalışma kapsamında derin öğrenme yöntemlerinden olan LSTM kullanılmıştır. Yapay sinir ağlarının eğitimi sırasında hiçbir şekilde parametrelerin bulunma formülü olmadığı için parametrelere verilen değerler deneme yanılma yolu ile veri setine en uygun hale getirilmeye çalışılarak bulunmuştur. Çalışmada kullanılan yapay sinir ağlarına ilişkin bazı parametreler şu şekildedir: Giriş, çıkış ve gizli katmanlar olmak üzere toplam 6 katmana sahiptir. Katmanlar sırayla 256,256,256,64,32 adet düğümden oluşmaktadır. Dört ve beşinci katmanda RELU fonksiyonu kullanılmıştır. LSTM ağının backpropagation(geri yayılım) yapması sağlanmıştır. Aktivasyon fonksiyonları olarak Sigmoid ve Tanjant hiperbolik kullanılmıştır. Epochs sayısı 80, pencere genişliği 26 olarak belirlenmiş ve veri setinin %90'ı öğrenme için %10'u test için ayrılmıştır.

Program ilk kez bir model oluşturulması istenildiği zaman “dataset.csv” dosyası içerisinde istenen hisse senedinin sembolüne ait tüm satırlar bulunur ve vektörlere sıra ile yerleştirilir. Normalizasyon yapabilmek için data set içerisindeki en yüksek ve en düşük fiyat bulunur ve kaydedilir. Vektörlere doldurulan satırların her biri normalizasyondan geçirilerek LSTM ağını beslemek için etiketlenir. Oluşturulan ağ Rolling window yöntemini kullanarak kendini geçmişteki veriler ışığında eğitir.

Eğitim tamamlanıp model oluştuktan sonra tekrar kullanabilmek için .zip formatında kaydedilir. Test aşamasına geçildiğinde %10'luk kısım için Rolling window yöntemi tekrar kullanılarak, $(\text{test verisi adeti} - (\text{pencere uzunluğu} + 1))$ formülüyle tahmin adeti belirler. Daha sonra pencere uzunluğu kadar öğrenme yapıp bir sonraki seansı tahmin ederek işlemi sürdürür. Tahminler grafik oluşturabilmek için bir vektörde toplanır.

Şekil-12'te farklı yapay sinir ağı yapısına sahip ağların tahmin sonuçları gösterilmektedir.



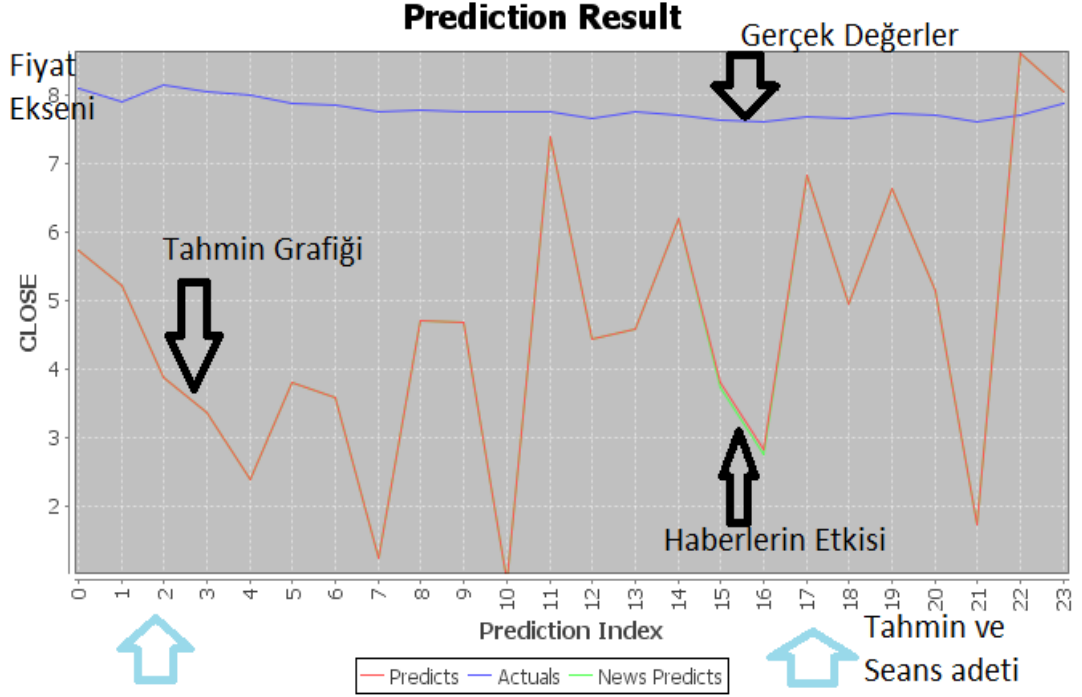
Şekil 12: Aynı parametrelere sahip 12 ve 5 katmanlı LSTM sonuçları.

5.4. KULLANICI ARAYÜZÜNÜN OLUŞTURULMASI

Arayüz Java 8 Swing kütüphanesi kullanılarak kodlanmıştır. Ana ekran *MainFrame* isimli *JFrame* sınıfında türemiş penceredir. Burası doldurulacaktır.

5.4.1. Çizgi Grafiklerinin Oluşturulması

Grafik penceresi oluşturulmadan önce içinde gösterimini yapacağımız tahmin ve test sonuçlarını *XYSeriesCollection* sınıfından oluşturduğumuz dataset objesinin içine ekliyoruz. İçinde bulunan veri sıralı ve tahmin veya test olarak ayrılmış şekilde bulunuyor. Grafik gösterimini yapabilmek için Java Swing kütüphanesindeki *JFreeChart* sınıfından bir obje türetiyoruz. *JFreeChart* sınıfından obje oluştururken hazırladığımız dataset'i, pencerenin dikey ve yatay gösterimdeki isimlerini objenin *ChartFactory.createXYLineChart()* metoduna giriyoruz. Oluşturulan objenin *getXYPlot()* metodu ile x ve y eksenin orantısı , içlerindeki bölmelerin boyutu gibi ayarlamaları yapılabilir. Bu işlemlerden sonra hazırlanan grafik kullanıcı arayüzüne *JFrame* objesine *add(grafik)* metodu ile ekleniyor. Çizelge üzerinde üç adet çizgi grafik oluşuyor. Kırmızı çizgi, sadece LSTM sonucu çıkan verileri simgelerken, mavi çizgi gerçek verileri ve yeşil çizgi ise haber katsayılarının LSTM sonucu ile birleştirilmiş olan değerleri göstermektedir. Yapılan grafik dikeyde Türk Lirası cinsinden olurken, yatayda tahmin edilen seans sıraları bulunmaktadır.



Şekil 13: Grafik özelliklerinin tanıtımı.

5.4.2. Hisse ve Haber İçerik Penceresi ve Tahmin Butonun Kodlanması

Tahmin butonu *JButton* yapısı ile oluşturulmuştur. Program içerisinde *addActionListener()* metodu ile tuşa basılıp basılmadığını anlayamaya çalışmaktadır. Butona basıldıktan sonra bir dizi olay gerçekleşir. İlk önce kullanıcının girmiş olduğu bilgiler program içerisine çekilerek kontrol edilir. Daha sonra gerekli yerlere kaydedilerek programın çalıştığını simgeleyen bir *gif* ekrana gelerek program işlemleri yapısına kadar kalmaya devam eder. Sistem içerisinde bu sırada senkronize çalışma yapılabilmesi için bir *thread* açılarak ana yapılar harekete geçirilir.

Aynı anda hem arayüze hem de sisteme erişmek için *thread* içerisinde *SwingUtilities.invokeLater()* kullanılarak ekranda *gif* yapısının görünür hale getirilmesi sağlanır. Thread içindeki ana yapıyı çalıştırıp bittikten sonra *thread* beraber *gif*'te *setVisible(false)* ve *dispose()* yapılarak ekrandan kaybolması ve kapanması sağlanır. Arayüz tekrardan *invalidate()* , *validate()* ve *repaint()* edilerek yenilenmesi sağlanır. Ve ekran kapatılana kadar görünmeye devam eder.

6. BAŞARI KRİTERLERİNİN DEĞERLENDİRİLMESİ

Projenin sonunda yapılması hedeflenen başarılar ve sonuçları yer almaktadır.

BAŞARI KRİTERİ 1

Elde edilen haberlerin kategorilerinin belirlenmesi ve benzer haberler ile eşleştirilebilmesi. (En az 20 farklı haberde %100 başarı)

Sonuç

Tasarlanan yapı sayesinde veri setinin içeriğinden 4 yıl sonraki 30 adet haber kategori bazında %100 başarı sağlanmıştır. Haberler anlam benzerliği konusunda bir gözlemciye ihtiyaç duyduğu için yakaladığı başarı iyi olarak nitelendirilebilir.

BAŞARI KRİTERİ 2

Farklı LSTM ağları kullanılarak trend tahminindeki başarıyı %55'den yukarı çıkarabilmek ve ağ çeşitlerinin başarıyı nasıl etkilediğini gözlemlemek.

Sonuç

Programda 4, 5 ve 12 katmanlı LSTM ağları oluşturulup farklı aktivasyon fonksiyonları ile 1 yıllık veri için tahmin sonuçları gözlenmiştir. En iyi sonuç 4 ve 5 katmanlı ağlarda gözlenmiştir. 12 katmanlı ağda overfitting olmuştur. Katman yapısı 5 olan denemede %69,5 yükseliş ve düşüş trendini doğru olarak tahmin etmiştir. 23 Adet tahmin içerisinde 12 yükseliş 4 adet alçalışı toplamda 16 ivme yönünü doğru tahmin etmiştir.

BAŞARI KRİTERİ 3

Yapılan yükseliş ve alçalış trend tahminlerin belirli bir miktar para ile sağlayacağı karın tahminlerin son gününe kadar başlangıç miktarının kazandırdığı paradan daha fazla kar miktarının hesaplanması.

Sonuç

Arçelik hissesi için veri setinden alınan değerler şekil-14’de gösterilmiştir.

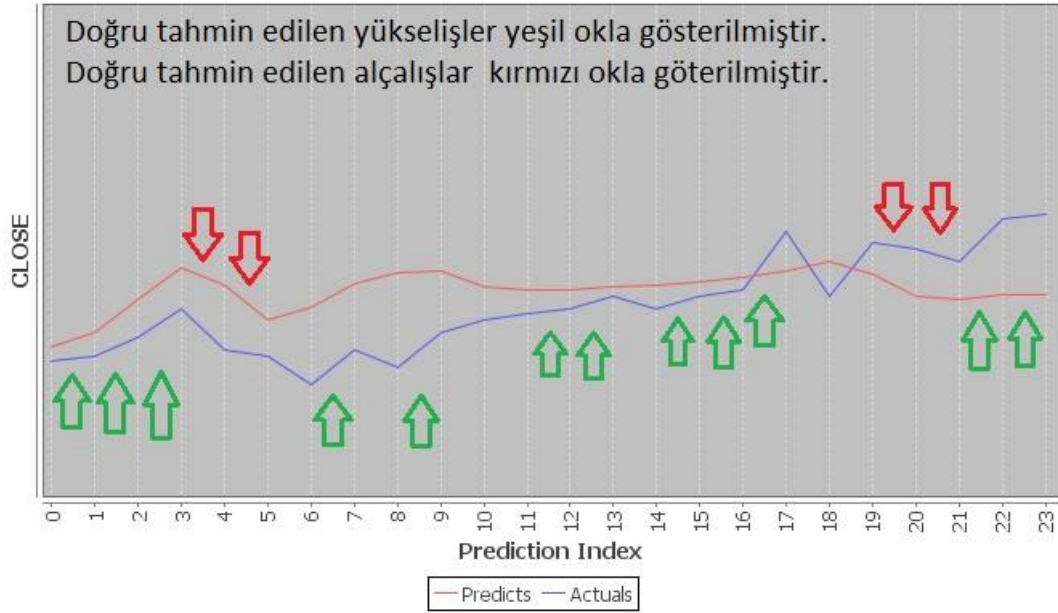
```
Test data date=>14.12.2011 Close value 5.780  
Test data date=>30.12.2011 Close value 6.280
```

Şekil 14: Test için kullanılacak olan tarih aralıkları ve değerlerin gösterimi.

Test doğru olarak tahmin edilmiş trendler üzerinden yapılacaktır ve yükseliş görüldüğünde hisse senedi elde tutulacak, alçalış görüldüğü zamansa satılacaktır. Eğer bir alçalış sonrası yükseliş görülmüşse eldeki para miktarı ile alına bilecek kadar hisse senedi alınacaktır. Hisse senedi fiyatında virgülden sonra iki hane göz önüne alınacaktır.

Testin 1000(bin) adet lot ile başlamasına karar verilmiştir, 14.12.2011 tarihi ile 1000 lotun toplam fiyatı 5780TL’dir. Toplamda geçen 24 seans sonucunda 1000 lotun gerçek değeri 6280TL olmuştur. Ana para üzerinden 500TL yani %8,6 kar elde edilmiştir.

Prediction Result



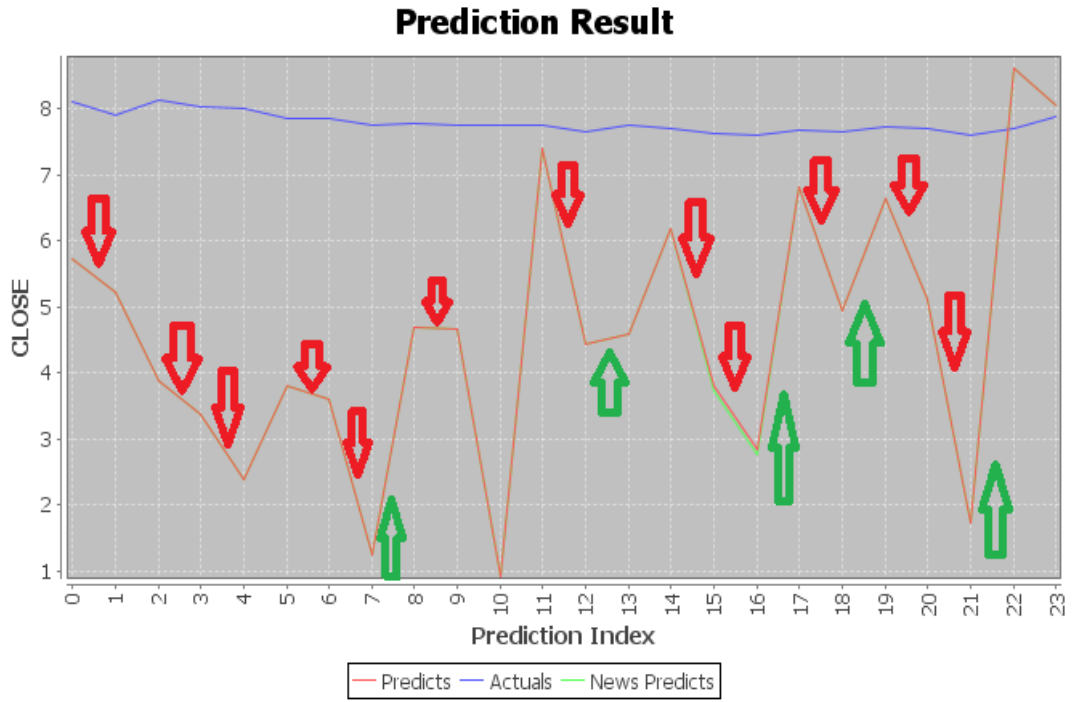
Şekil 15: 2011 yılı veri seti ile 5 katmanlı ağı tahmin ettiği trendlerin gösterimi.

LSTM sonuçlarına göre işlem yapıldığı zaman 3.index tahmininde ilk satış emri verilerek ana para 5960tl olmuştur. Altıncı index tahmininde 5,69tl'den bütün para ile tekrar alındığında 1047 hisseye yükselmiş ve 2,57tl para kasamızda bulunmaktadır. Tekrardan 19.indexte satış yaptığımızda kasamızda 6462,56tl bulunmaktadır. Tekrar 21.indexte 6,11tl'den alış gerçekleştirdiğimizde 1057 lot + 4,29tl kasamızda bulunmaktadır. Son kapanış gününde 6,28tl'den tüm hisseleri sattığımızda 6642,25tl kasamızda bulunmaktadır. Ana paradan toplam 24 seans sonra 862,25tl %15 kar sağlamış bulunmaktayız. Bu hesaplamalardan yola çıkarak 5 katmanlı bir LSTM ağının vermiş olduğu sonuçlar ile %6,3 daha fazla kar elde edebiliyoruz. Bu hesaplamalar şekil-15'e göre yapılmıştır.

BAŞARI KRİTERİ 4

Oluşturulan haber algoritması ile LSTM sonuçlarının desteklenerek elde edilen başarının iyileştirilmesi ve parasal karın artığının kanıtlanabilir olması.

Sonuç



Şekil 16: 01/2011'den 05/2012'ye kadar olan veri üzerinde 25/05/2012'ye ait haber grafiği

Test doğru olarak tahmin edilmiş trendler üzerinden yapılacaktır ve yükseliş görüldüğünde hisse senedi elde tutulacak, alçalış görüldüğü zamansa satılacaktır. Eğer bir alçalış sonrası yükseliş görülmüşse eldeki para miktarı ile alına bilecek kadar hisse senedi alınacaktır. Hisse senedi fiyatında virgülden sonra iki hane göz önüne alınacaktır. Arçelik hissesinin 15.05.2012 tarihi ile değeri 8.1 Türk lirasıdır. Bugün ile birlikte 24 seans sonrası 31.05.2012 tarihinde değeri 7.88 Türk lirasıdır.

Testin 1000(bin) adet lot ile başlamasına karar verilmiştir, 15.05.2012 tarihi ile 1000 lotun toplam fiyatı 8100TL'dir. Toplamda geçen 24 seans sonucunda 1000 lotun gerçek değeri 7880TL olmuştur. Ana para üzerinden 220TL azalma olmuş ve %2,7 zarar elde edilmiştir.

- 0.index tahmininde ilk satış emri verilerek ana para 8100tl elde tutulmuştur.
- 7.index tahmininde 7.73 liradan alış yapıldı ve 1047 lot + 6.69tl kaldı.
- 8.index tahmininde 7.78 liradan satış yapıldı ana para 8152.35tl oldu.
- 12.index tahmininde 7.65 liradan alış yapıldı ve 1065 lot + 7.33tl kaldı.
- 14.index tahmininde 7.69 liradan satış yapıldı ana para 8227.94tl oldu.
- 16.index tahmininde 7.59 liradan alış yapıldı ve 1084 lot + 0.38tl kaldı.
- 17.index tahmininde 7.67 liradan satış yapıldı ana para 8314.66tl oldu.
- 18.index tahmininde 7.65 liradan alış yapıldı ve 1086 lot + 6.76tl kaldı.
- 19.index tahmininde 7.71 liradan satış yapıldı ana para 8379.82tl oldu.
- 21.index tahmininde 7.59 liradan alış yapıldı ve 1104 lot + 0.46tl kaldı.
- 23.index 7.88 liradan son satış yapıldı ve ana para 8699.98tl oldu ve test sonlandı.

LSTM sonuçlarına göre işlem yapıldığı zaman 600tl, %7,4 kar elde edebiliyoruz. Bu hesaplamalar şekil-16'e göre yapılmıştır.

Şekil-16'ya göre 23 tahmin üzerinden 17 tane doğru bilinmiştir. Buda başarı %74 başarı anlamına gelmektedir.

7. ANALİZ

Bu bölümde trend tahmin sonuçlarının ve başarı oranının nasıl hesaplandığına yer verilecektir.

7.1. TREND TAHMİN SONUÇLARI

Şekil-15'daki tahmin ve gerçek fiyat verileri baz alınarak açıklanacaktır.

“Trend“olarak adlandırdığımız olgu aslında arka arkada gelen iki fiyat arasındaki farkın durumudur. Bu fark pozitif oluyorsa “Yukarı Yönlü Trend” yükseliş anlamında kullanırken, negatif olursa da “Aşağı Yönlü Trend” düşüş anlamında kullanılır. İki nokta arası farka bakıldığı için program toplamda 24 seanslık bir tahmin çıktısı üretirken buradaki başlangıç tahmini, trende bakılırken es geçilmiş olur ve 23 adet trend tahmini kalır.

(Trend Çizgileri miktarı = Toplam tahmin miktarı – 1) olarak formüle edilebilir.

Şekil-15’da yeşil ok ile gösterilen çizgiler pozitif yönlü farkı gösterirken kırmızı ok ile gösterilen çizgiler ise negatif yönlü farkın doğru tahmin edildiği aralıkları göstermektedir.

7.2. BAŞARI ORANIN HESAPLANMASI

Şekil-15’da oluşturulan trend tahminlerinin başarı oranı aşağıdaki formül yardımıyla hesaplanabilir.

$$DTO = \frac{p + n}{t}$$

Şekil 17: Trend tahmini başarı oranı formülü gösterimi. [4]

Formülde, bir sonraki gün fiyat artışının doğru bir şekilde tahmin edildiği seans sayısını, bir sonraki gün fiyat azalışının doğru bir şekilde tahmin edildiği seans sayısını ve ise test periyodunu temsil etmektedir. [4, pp. 209-227]

8. TARTIŞMA VE SONUÇ

Proje kapsamında planlanan haber metinlerini sınıflandırma olayı yüksek bir başarı sağlayarak gerçekleştirilmiştir. Kategoriler daha fazla çeşitlendirilip yeni bir katsayı dağıtımı yapılabildiği taktirde daha kesin ve şüpheyeye yer bırakmadan sınıflandırma yapılabilir. Aynı zamanda haber verileri daha kolay bir yöntem ile temin edilebilirse öğrenme yüzdesi arttırılarak kategorilere göre ayırma başarısı aynı şekilde artmış olacaktır.

Yapay sinir ağlarının oluşturulması sırasında sistemin üzerinde bulunduğu makinanın donanımsal yetersizliklerinden dolayı çok fazla ağ oluşturulamamış, farklı parametre ve veri setleri ile testlere tabi tutulamamıştır. Fakat bu durumda bile yaklaşık 7-8 dakika içerisinde 1 yıllık veri setinden %69,5 başarı oranı ile 23 adet trend tahmin üretip kullanıcıya sunmayı başarmıştır. Daha güçlendirilmiş cihazlar yardımı ile yapay sinir ağları oluşturulup öğrenme, tahmin yakınlığı gibi kısımlarda iyileşme olup olmadığı gözlemlenebilir. Yapılan çalışma zarfında literatür

araştırmasında, 10 katmanlı LSTM ağı ile bir sonraki gün tahmininde direk fiyat bazında %70 oranlarında başarı sağladığı ile ilgili makaleler bulunmuştur fakat kullanılan ağ yapısının birçok özelliği belirtilmediği için deneme yanılma yoluyla bu parametreler bulunamamıştır. Şekil-12’de görüldüğü üzere denemelerde giriş ve çıkış katmanı dahil olmak üzere 12 katmanlı ağda overfitting gözlemlendiği için denemelere son verilmiştir. Daha ufak ama başarısı yüksek olan 5 katmanlı bir yapı baz alınarak şu anki sonuçlar elde edilmiştir.

Yapılan çalışma yatırım tavsiyesi kapsamına girmemektedir. Amaç gözlemciye geçmişe dönük bazı hisse davranışlarının tekrar karşılaştığında aynı şekilde fiyatlandığını göstermek amaçlı yapılmıştır.

İleride geliştirilen birden çok ağ modeli ile BIST30 endeksindeki hisseler üzerinde test edilmesi hedeflenmektedir. Hisselere daha iyi ve tekil bir yaklaşım sergilemek için haberlerin olumlu veya olumsuz anlam taşıyıp taşımadığına bakılması planlanmaktadır. Bu sayede haberlere daha öznel katsayı etiketleri vermek amaçlanmaktadır.

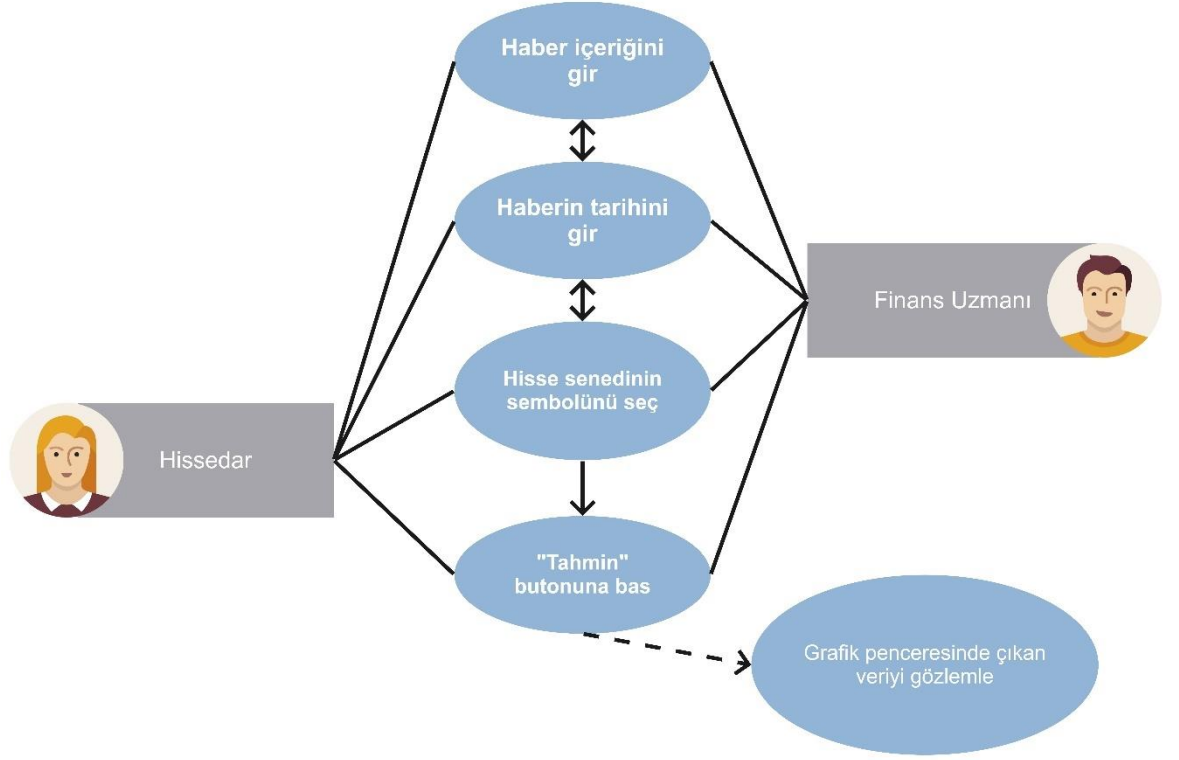
Proje Java’da tasarlandığı için işletim sisteminden bağımsız olarak çalışmakta fakat gelecekte web’e taşınıp kullanıcıların kendi bilgisayarlarına bir yükleme işlemi yapmadan online olarak analiz gerçekleştirmesi hedeflenmektedir.

KAYNAKÇA

- [1] Y. Bao , T. Xiong ve Z. Hu, «“Multi-step-ahead time series prediction using multiple-output support vector regression,» %1 içinde *Neurocomputing*, Amsterdam, Netherlands, Elsevier B.V, 2014, pp. 482-493.
- [2] N. şengöz, «yapay-sinir-aglari,» 4 Mart 2017. [Çevrimiçi]. Available: <http://www.derinogrenme.com/2017/03/04/yapay-sinir-aglari/>.
- [3] M. Qiu ve Y. Song, «Predicting the Direction of Stock Market Index Movement Using an Optimized Artificial Neural Network Model,» *MYQ YS.*, 2016.
- [4] M. ÖZÇALICI, «Yapay Sinir Ağları ile Çok Aşamalı Fiyat Tahmini: BIST30 Senetleri Üzerine Bir Araştırma,» *İktisadi ve İdari Bilimler Fakültesi Dergisi*, pp. 209-227 , 2016.

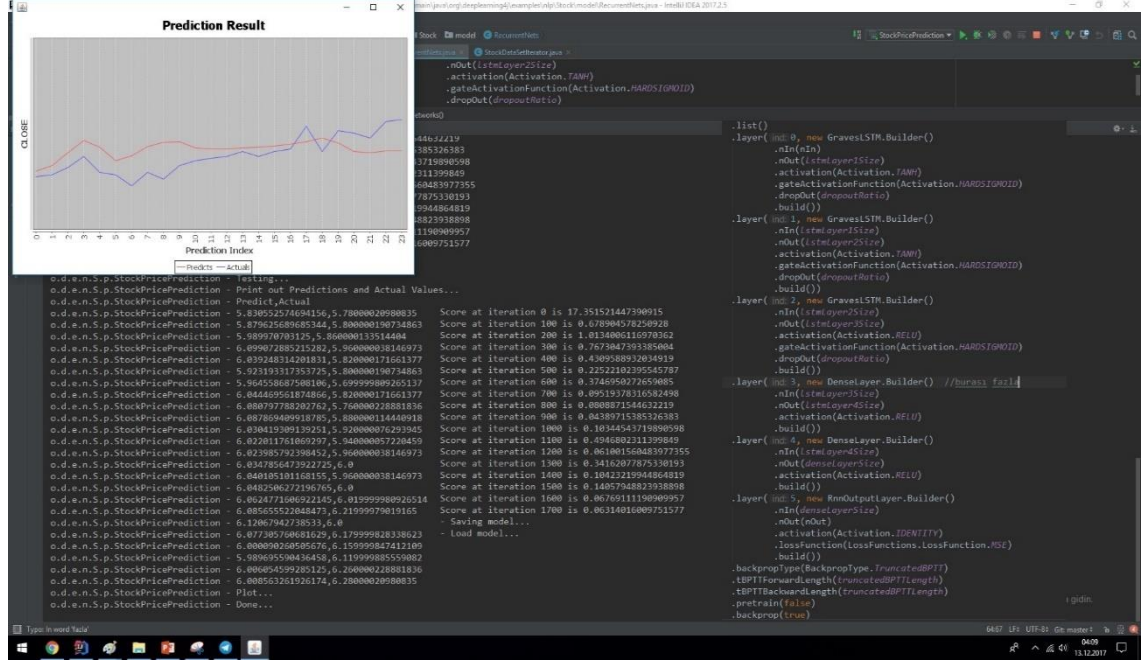
EKLER

A. Genel Use-Case Diagramı gösterimi



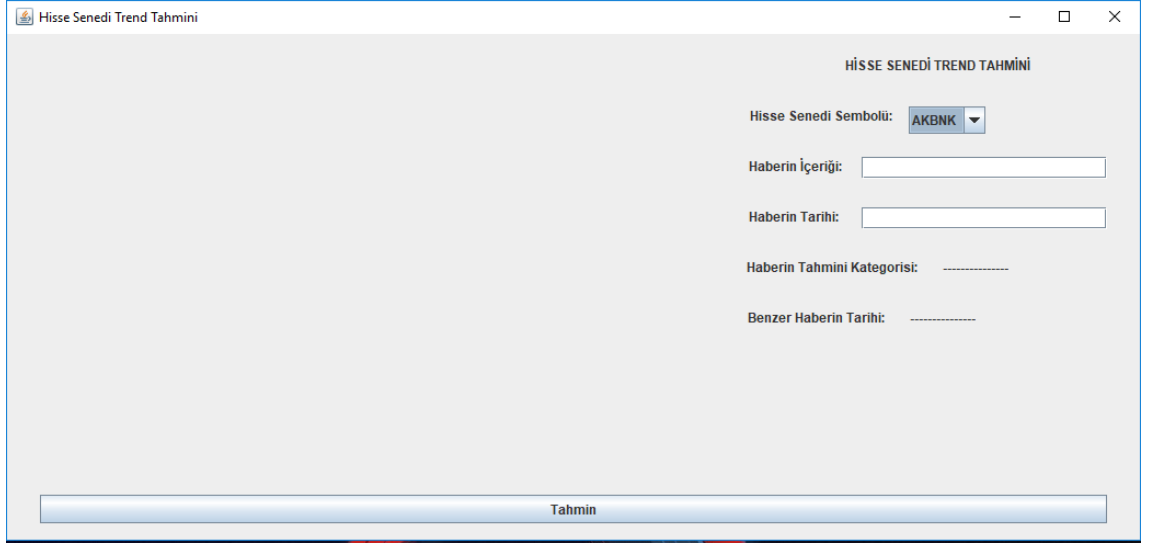
Şekil 18: Genel Use-Case Diyagramı.

B. En iyi başarımın yakalandığı ağ yapısının ve tahmin çıktıların sayısal gösterimi



Şekil 19: En yüksek başarıya sahip LSTM yapısı ve sayısal sonuçları.

C. Program Arayüzünün Başlangıç – Çalışma – Bitiş aşamaları

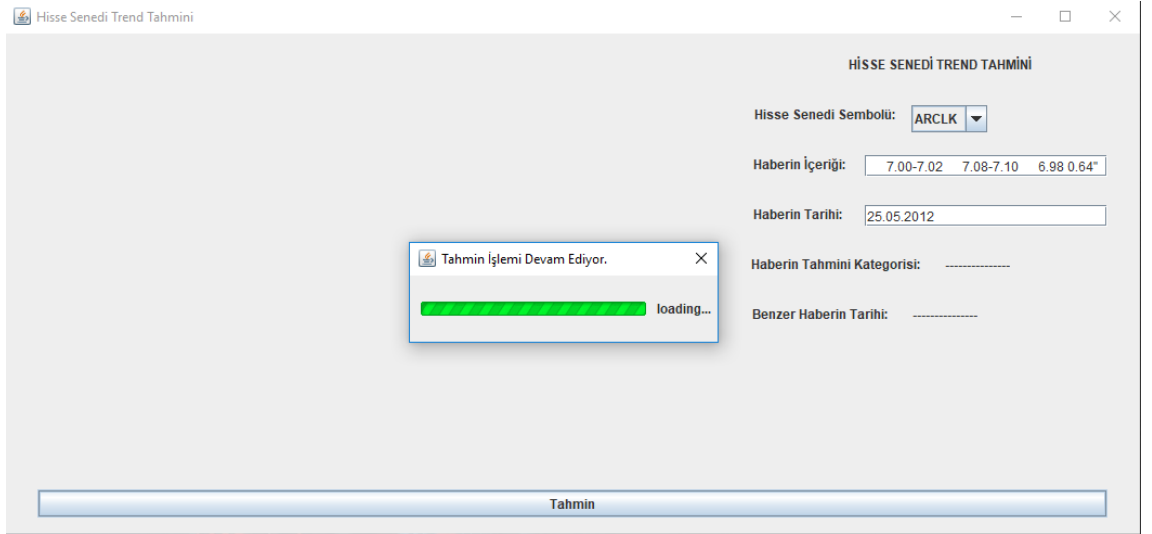


The screenshot shows the 'Hisse Senedi Trend Tahmini' application window. The title bar reads 'Hisse Senedi Trend Tahmini'. The main area is a light gray. On the right side, there is a form titled 'HİSSE SENEDİ TREND TAHMİNİ'. The form contains the following fields:

- Hisse Senedi Sembolü: A dropdown menu with 'AKBNK' selected.
- Haberin İçeriği: A text input field.
- Haberin Tarihi: A text input field.
- Haberin Tahmini Kategorisi: A text input field.
- Benzer Haberin Tarihi: A text input field.

At the bottom of the window, there is a large blue button labeled 'Tahmin'.

Şekil 20: Arayüz Başlangıç Görseli.



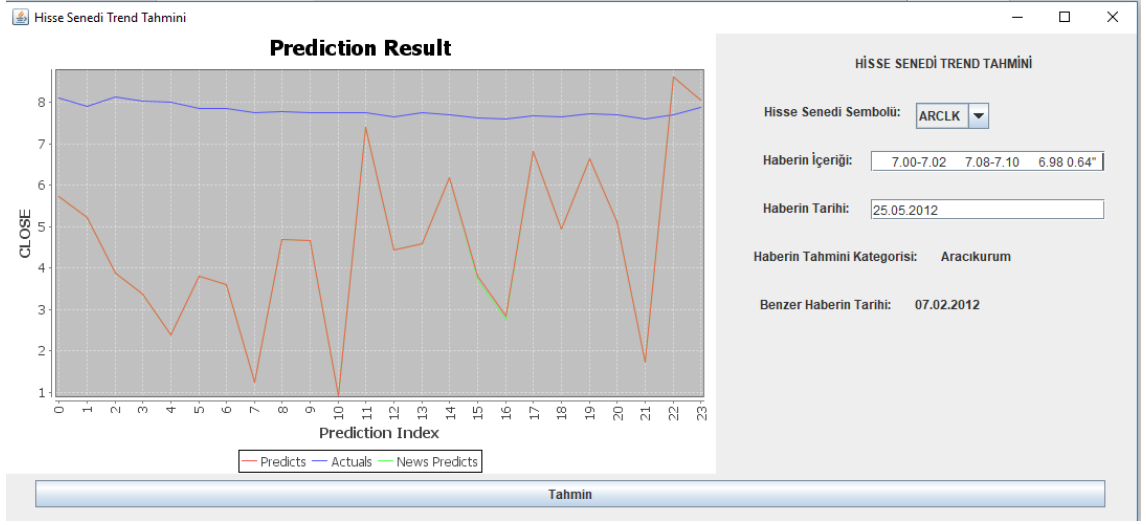
The screenshot shows the 'Hisse Senedi Trend Tahmini' application window during the calculation phase. The title bar reads 'Hisse Senedi Trend Tahmini'. The main area is a light gray. On the right side, there is a form titled 'HİSSE SENEDİ TREND TAHMİNİ'. The form contains the following fields:

- Hisse Senedi Sembolü: A dropdown menu with 'ARCLK' selected.
- Haberin İçeriği: A text input field containing '7.00-7.02 7.08-7.10 6.98 0.64*'. Below this field, there is a small text 'loading...'.
- Haberin Tarihi: A text input field containing '25.05.2012'.
- Haberin Tahmini Kategorisi: A text input field.
- Benzer Haberin Tarihi: A text input field.

In the center of the window, there is a small dialog box titled 'Tahmin İşlemi Devam Ediyor.' with a close button (X). The dialog box contains a green progress bar and the text 'loading...'.

At the bottom of the window, there is a large blue button labeled 'Tahmin'.

Şekil 21: Arayüz Çalışma Görseli.



Şekil 22: Arayüz Tahmin Bitiş Görşeli.