Report On

Stock Price Prediction

Submitted in partial fulfillment of the requirements of the Course project in Semester VII of Final Year Computer Engineering

by

Anish Patil (Roll No. 07)

Ajay Shitkar (Roll No.14)

Gaurav Singh (Roll No. 15)

Mentor
Dr. Megha Trivedi

University of Mumbai

Vidyavardhini's College of Engineering & Technology

Department of Computer Engineering

(2023-24)

Vidyavardhini's College of Engineering & Technology

Department of Computer Engineering

CERTIFICATE

This is to certify that the project entitled “Stock price Prediction” is a bonafide work of " Anish Patil(Roll No. 07), Ajay Shitkar (Roll No.14), Gaurav Singh(Roll No.15)" submitted to the University of Mumbai in partial fulfillment of the requirement for the Course project in semester VII of Final Year Computer Engineering.

Dr. Megha Trivedi
Mentor

Dr. Megha Trivedi
Head of Department

Dr. H.V. Vankudre
Principal

Vidyavardhini's College of Engineering & Technology

Department of Computer Engineering

Course Project Approval

This Course Project entitled **“Stock Price Prediction”** by **Anish Patil (Roll No. 07), Ajay Shitkar (Roll No. 14), Gaurav Singh (Roll No. 15)** is approved for the degree of **Bachelor of Engineering** in Semester VII of Final Year **Computer Engineering .**

Examiners

1………………………………………

(Internal Examiner Name & Sign)

2…………………………………………

(External Examiner Name & Sign)

Date:

Place:

Contents

Abstract

Stock price prediction is a pivotal field within financial markets, attracting extensive attention from investors, traders, and researchers alike. The ability to accurately forecast future stock prices is a complex and challenging task, given the multifaceted nature of financial markets, which are influenced by a multitude of factors, including economic indicators, market sentiment, and company-specific events. This abstract provides an overview of the key aspects of stock price prediction and the various methodologies employed in this domain.

The goal of stock price prediction is to leverage historical and real-time data to make informed decisions about buying or selling securities, thereby maximizing profitability or minimizing losses. Predictive models and techniques vary widely, encompassing both traditional and modern approaches.

Stock price prediction is a multifaceted field that combines data analysis, machine learning, and domain expertise to provide insights into future market movements. Accurate predictions can significantly impact investment decisions and financial success.

Acknowledgement

We would like to extend our heartfelt gratitude to several individuals who played pivotal roles in the successful completion of our project. First and foremost, we express our deepest appreciation to Dr. Megha Trivedi, who served as both our Mentor and Head of Department. Her guidance, expertise, and unwavering support were instrumental throughout our research journey. We also acknowledge the invaluable contributions of our esteemed Principal, Dr. H.V. Vankudre, for their encouragement and belief in our capabilities. Furthermore, we cannot overlook the significant influence of our friends and family, who provided both direct and indirect assistance, serving as pillars of strength during challenging times. Additionally, we are indebted to the numerous academicians who generously served as supervisors and assessors, offering their expertise and insights that greatly enriched our project. Their collective efforts have been indispensable, and we express our sincere gratitude to each of them for making this study a resounding success.

1. Introduction

1.1 Introduction

Stock price prediction has long been a subject of fascination and practical importance in the world of finance. The allure of accurately forecasting the future value of stocks has driven the development of numerous methodologies, ranging from traditional statistical techniques to cutting-edge artificial intelligence algorithms. This pursuit is fueled by the profound impact that precise predictions can have on investment decisions, risk management, and overall financial success. Financial markets are a complex ecosystem, influenced by an intricate web of factors, including economic indicators, geopolitical events, market sentiment, and company-specific developments. To navigate this complexity, investors and analysts turn to predictive models, leveraging a wealth of data sources and analytical tools. This multidisciplinary approach incorporates elements of data science, machine learning, natural language processing, and financial analysis.

1.2 Problem Statement

The problem of stock price prediction is marked by its inherent complexity and the ever-shifting landscape of financial markets. Investors and analysts grapple with the challenge of accurately foreseeing the future values of stocks, a task burdened by the dynamic interplay of numerous variables and uncertainties.

One key issue lies in the multifaceted nature of market dynamics. Predicting stock prices necessitates the consideration of a diverse array of data sources, including historical price and volume data, company financial reports, news sentiment, and social media chatter. The challenge is to extract valuable signals from this data amid the noise, and to model the intricate relationships between these factors and stock price movements.

Additionally, the need for robust and adaptive predictive models is underscored by the rise of high-frequency trading, where decisions are made in milliseconds. Consequently, there is a pressing demand for models that can operate effectively in real-time, taking into account market microstructure and order book dynamics.

1.3 Objectives

The aims of this project are as to identify factors affecting share market, To generate the pattern from large set of data of stock market for prediction of NSE and to predict an approximate value of share price to provide analysis for users through web application The objective of the system is to give a approximate idea of where the stock market might be headed. It does not give a long term forecasting of a stock value. There are way too many reasons to acknowledge for the long term output of a current stock. Many things and parameters may affect it on the way due to which long term forecasting is just not feasible.

1.4 Scope

1. Time Horizons: Stock price prediction caters to a range of time horizons, from ultra-short-term high-frequency trading to long-term investment strategies. This scope accommodates the diverse needs of investors and traders.
2. Data Sources: The scope extends to a wide array of data sources, including historical price and volume data, economic indicators, company financial reports, news articles, social media sentiment, and order book data, enabling comprehensive analysis and modeling.
3. Techniques and Models: Stock price prediction incorporates a broad spectrum of techniques and models, such as time series analysis, regression models, machine learning algorithms, deep learning models, sentiment analysis, technical analysis, and fundamental analysis, allowing for flexibility in approach.
4. Market Segments: It applies to various financial instruments, including equities, commodities, cryptocurrencies, and forex markets. The scope extends to global markets, each with its unique dynamics and influencing factors.
5. Investor Profiles: Stock price prediction is relevant for a wide audience, ranging from individual retail investors to institutional investors, quantitative analysts, and algorithmic traders, making it applicable to various investment approaches and risk tolerances.
6. Risk Management: It encompasses risk assessment and management by providing tools for investors to understand and mitigate the risks associated with their investment decisions.

2 Literature Survey

Nowadays, as the connections between worldwide economies are tightened by globalization, external perturbations to the financial markets are no longer domestic. With evolving capital markets, more and more data is being created daily. The intrinsic value of a company's stock is the value determined by estimating the expected future cash flows of a stock and discounting them to the present, which is known as the book value. This is distinct from the market value of the stock, that is determined by the company's stock price. This market value of a stock can deviate from the intrinsic value due to reasons unrelated to the company's fundamental operations, such as market sentiment. The fluctuation of stock market is violent and there are many complicated financial indicators.

2.2 Limitation of Existing System or Research Gap

1. Lack of Real-Time Processing: Some existing systems lack real-time processing capabilities, which means that users might not receive immediate feedback on the classification of messages.
2. Resource Inefficiencies: Existing systems may suffer from resource inefficiencies due to the sheer volume of spam messages, leading to wasted server space, network bandwidth, and computational power.
3. Privacy and Security Concerns: Users might express concerns about privacy invasion and the presence of potentially harmful content in their inboxes, as existing systems struggle to protect them adequately.
4. Need for Continuous Model Improvement: Existing machine learning models used for stock improvement may not always adapt to emerging techniques. They require continuous updates and refinements.

2.3 Mini Project Contribution

The project can contribute to a more seamless and efficient user experience by automating the identification and management of stock prices. Users can engage with their time and effort with valuable content. Detecting and monitoring valuable stocks that aim to collect personal information can contribute to user privacy and data analysis. Users can trust that their sensitive information remains secure. The project can provide educational resources and tutorials to help users understand the system's features and best practices for spam management. This contributes to user awareness and digital literacy. The project can be designed to be scalable, capable of handling a variety of user volumes and message types. This scalability is an essential feature for addressing the dynamic nature of spam.

3 Proposed System

3.1 Introduction

In the field of finance, stock price prediction has long been a fascinating and useful topic. The need to predict stock values accurately has led to the creation of a multitude of approaches, from conventional statistical methods to state-of-the-art artificial intelligence algorithms. The significant influence that accurate forecasts can have on risk management, investment choices, and overall financial success is what motivates this endeavor. The financial markets are a multifaceted ecosystem that are impacted by a wide range of variables, such as company-specific advancements, market mood, geopolitical events, and economic indicators. Investors and analysts use prediction models, which make use of a multitude of data sources and analytical tools, to traverse this complexity. This multidisciplinary strategy combines aspects of finance, natural language processing, machine learning, and data science.

3.2 Architecture/ Framework/Block diagram

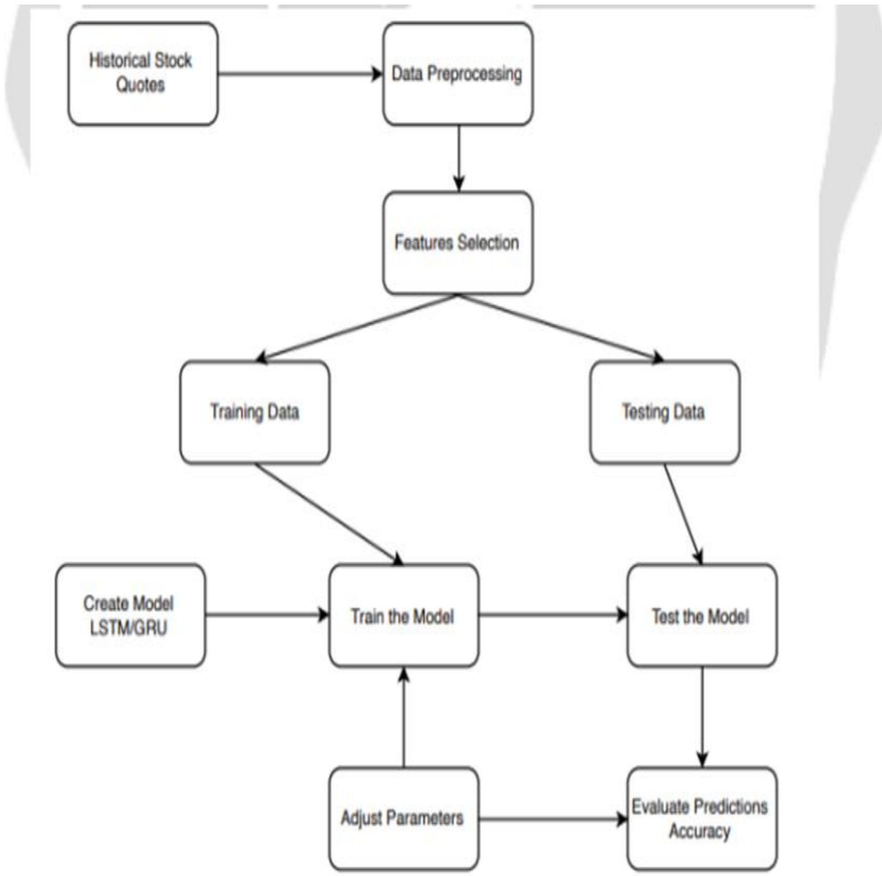


Fig -3: Proposed System Architecture

3.3 Algorithm and Process Design

The algorithm and process design for stock price prediction involves a combination of data preprocessing, feature engineering, model selection, training, and evaluation. Below is a simplified overview of the algorithm and process design:

1.Data Collection:

Gather historical price and volume data for the target stock or asset.

Collect additional relevant data sources, such as economic indicators, news sentiment, and social media sentiment.

2.Data Preprocessing:

Handle missing data and outliers through techniques like imputation or removal.

Normalize or scale the data to ensure that different features are on a similar scale.

Create a time series dataset by selecting a suitable time interval (e.g., daily, hourly, or minute-level data).

3.Feature Engineering:

Create informative features that capture relevant information, such as moving averages, technical indicators (e.g., RSI, MACD), and fundamental data (e.g., earnings per share).

Apply sentiment analysis to extract sentiment scores from news articles or social media posts.

Construct lag features to account for historical price and volume trends.

4.Model Selection:

Choose appropriate machine learning models or algorithms based on the problem requirements and data characteristics. Common choices include:

5.Training:

Split the data into training, validation, and test sets for model development and evaluation.Implement techniques such as cross-validation to prevent overfitting and assess model performance.

6.Model Evaluation:

Evaluate the model's performance using appropriate metrics such as Mean Absolute Error (MAE), Mean Squared Error (MSE), Root Mean Squared Error (RMSE), or more specific financial metrics like Sharpe ratio.

Ensure ethical and responsible use of predictive models in financial markets, adhering to relevant regulations and guidelines.

The algorithm and process design for stock price prediction is an iterative and evolving process, and it requires a deep understanding of both financial markets and data science techniques. The specific implementation may vary based on the chosen model, data sources, and the time frame for predictions.

3.4 Details of Hardware & Software

Hardware Specifications:

- System type: x64-based processor, 64-bit operating system.
- Memory (RAM) installed: 8.00 GB (7.34 GB Usable)
- Total size of Hard disk: 1 TB

Software Specifications:

- Operating system: Microsoft Windows 10
- Integrated Development Environment: PyCharm
- Streamlit, Scikit-Learn, NLTK(Natural Language Toolkit)
- Programming language: Python

3.5 Experiment and Results for Validation and Verification:

Exploratory Data Analysis :

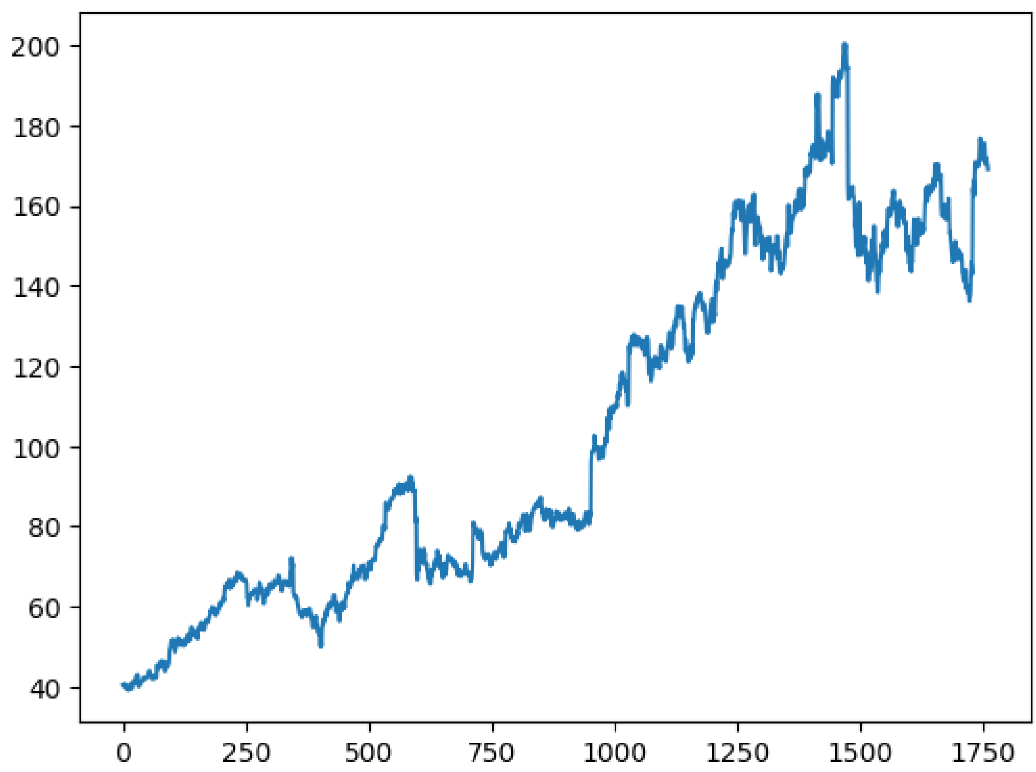
Code:

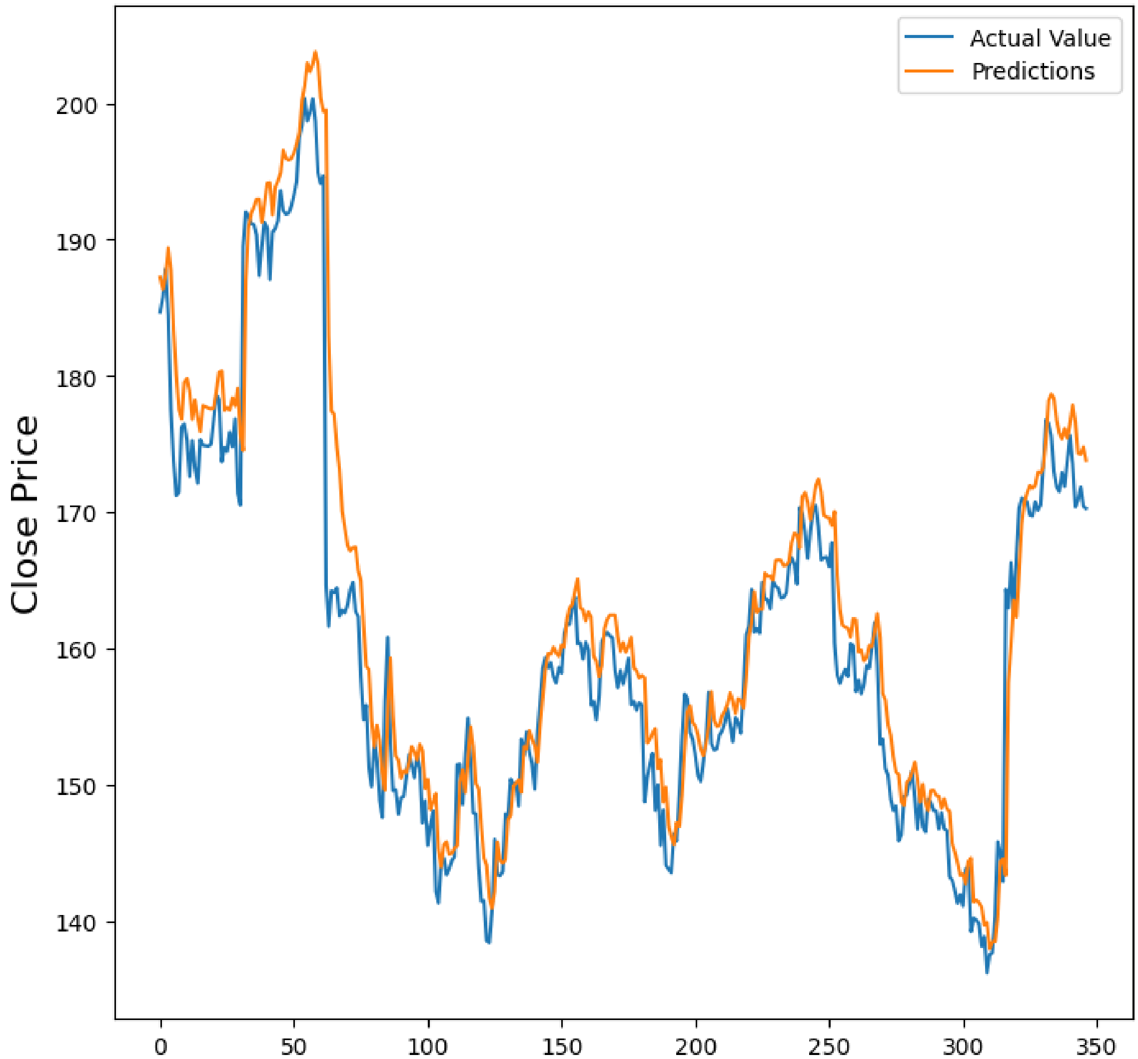
```
#Importing all the required libraries


import numpy as np

import pandas as pd

import matplotlib.pyplot as plt

import seaborn as sns

# Importing dataset

data = pd.read_csv("prices-split-adjusted.csv")

df = pd.DataFrame(data)

df.head()

df.describe()

# showing column wise %ge of NaN values they contains

for i in df.columns:

  print(i,"\t-\t", df[i].isna().mean()*100)

df = df[df['symbol']=='AAP']   # Choosin stock values for any company

cormap = df.corr()

fig, ax = plt.subplots(figsize=(5,5))

sns.heatmap(cormap, annot = True)

def get_corelated_col(cor_dat, threshold):

  # Cor_data to be column along which corelation to be measured

  #Threshold be the value above which of corelation to considered

  feature=[]

  value=[]

  for i ,index in enumerate(cor_dat.index):

    if abs(cor_dat[index]) > threshold:
```

```
      feature.append(index)

      value.append(cor_dat[index])

  df = pd.DataFrame(data = value, index = feature, columns=['corr value'])

  return df

top_corelated_values = get_corelated_col(cormap['close'], 0.60)

top_corelated_values

df = df[top_corelated_values.index]

df.head()

df.shape

sns.pairplot(df)

plt.tight_layout()

X = df.drop(['close'], axis=1)

y = df['close']

from sklearn.preprocessing import MinMaxScaler

scaler = MinMaxScaler()

X = pd.DataFrame(scaler.fit_transform(X), columns=X.columns)

X.head()

close = df.reset_index()['close']

close.head()

plt.plot(close)

plt.show()

time_step = 30

X, y = [], []

for i in range(len(close)-time_step-1):

    X.append(close[i:(i+time_step)])

    y.append(close[(i+time_step)])
```

```
X = np.array(X)

y = np.array(y)

X[:5]

y[:5]

#now lets split data in test train pairs

from sklearn.model_selection import train_test_split

X_train, X_test, y_train, y_test = train_test_split(X, y, test_size = 0.2, shuffle=False)

Acc = []

X_train_ = X_train.reshape(X_train.shape[0],X_train.shape[1],1)

X_test_ = X_test.reshape(X_test.shape[0],X_test.shape[1],1)

from tensorflow.keras.layers import LSTM

def Reg():

    model = Sequential()

    model.add(LSTM(70, return_sequences=True, input_shape=(30,1)))

    model.add(LSTM(70, return_sequences=True))

    model.add(LSTM(70))

    model.add(Dense(1))

    model.compile(loss='mean_squared_error', optimizer='adam')

    return model

# Model Training

model_1 = reg()

model_1.fit(X_train_, y_train, epochs=100, validation_split=0.2)

# Prediction

y_pred_1 = model_1.predict(X_test_)

# Measure the Accuracy Score

from sklearn.metrics import r2_score
```

```
print("Accuracy score of the predictions: {0}".format(r2_score(y_test, y_pred_1)))

Acc.append(r2_score(y_test, y_pred_1))

plt.figure(figsize=(8,8))

plt.ylabel('Close Price', fontsize=16)

plt.plot(pred_df)

plt.legend(['Actual Value', 'Predictions'])

plt.show()
```





3.6 Conclusion and Future work.

Stock price prediction represents a vital and dynamic field within the financial industry, offering a broad spectrum of techniques, methodologies, and tools for investors, traders, and financial professionals. It is driven by the relentless pursuit of accurate forecasts and the ever-changing nature of financial markets. This abstract has shed light on the multifaceted aspects of stock price prediction, including data sources, predictive models, sentiment analysis, and risk management, underscoring its importance in investment decision-making.

As the financial world continues to evolve, the scope of stock price prediction widens, adapting to new data sources, market dynamics, and technological advancements. Yet, it remains a challenge, as the unpredictability of events and the interplay of numerous variables persist. Future developments in this field will likely focus on improving model accuracy, accommodating high-frequency trading, incorporating increasingly diverse data sources, and enhancing sentiment analysis techniques.

Future Work:

1 .Enhanced Data Integration: Future work will involve integrating more diverse and unstructured data sources, such as satellite imagery, web scraping, and real-time IoT data, to provide a comprehensive view of market dynamics.

2.Advanced Machine Learning Models: The development of more sophisticated machine learning models, particularly deep learning architectures, will be a focal point to capture complex patterns and relationships in financial data.

3.Explainability and Interpretability: Efforts to make predictive models more interpretable and explainable will continue, especially in areas where regulatory compliance is crucial.

4.Real-time Analysis: Developing models that can operate in real-time and adapt rapidly to market changes will remain a priority, aligning with the needs of high-frequency traders.

5.Market Sentiment Analysis: The advancement of natural language processing and sentiment analysis will aid in understanding market sentiment and its influence on stock prices more accurately.

6.Behavioral Finance Integration: Future research may integrate insights from behavioral finance to better understand the psychological factors driving market behavior.

In sum, the future of stock price prediction holds the promise of more accurate, adaptable, and ethical models that empower market participants to navigate the ever-changing landscape of financial markets with greater confidence and success.

3.7 References

[1] X. Li, Y. Li, X.-Y. Liu, D. Wang, “Risk management via anomaly circumvent: mnemonic deep learning for midterm stock prediction.” in Proceedings of 2nd KDD Workshop on Anomaly Detection in Finance (Anchorage ’19), 2019.

[2] P. Chang, C. Fan, and C. Liu, “Integrating a piece-wise linear representation method

and a neural network model for stock trading points prediction." IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews) 39, 1 (2009), 80–92.

[3] Akita, Ryo, et al. "Deep learning for stock prediction using numerical and textual information." IEEE/ACIS 15th International Conference on Computer and Information Science (ICIS). IEEE, 2016.