

ONLINE WEB SCRAPPING TO ANTICIPATE THE STOCK MARKET PREDICTION

Prof. Jogi John, Saurabh Sonawne, Sagar Mankar, Dhanshri Wasu, Pranjali Rachchawar, Grishma Bhojar

Dept. of Computer Technology, Priyadarshini College Of Engineering, Nagpur, Maharashtra, India

ABSTRACT

In this project we attempt to implement machine learning approach to predict stock prices. Machine learning is successfully actualized in determining stock costs. The objective is to foresee the stock costs in advance to create more educated and precise speculation choices. We propose a stock cost expectation framework that coordinating numerical capacities, machine learning, and other outside components for the reason of accomplishing superior stock forecast precision and issuing productive exchanges. There are two types of stocks. You will know of intraday exchanging by the commonly utilized term "day exchanging." Intraday dealers hold securities positions from at slightest one day to the another and regularly for several days to weeks or months. LSTMs are exceptionally capable in arrangement forecast issues since they're able to store past data. This is often vital in our case since the past cost of a stock is pivotal in foreseeing its future cost. While predicting the actual price of a stock is an arduous task, ready to construct a demonstration that will foresee whether the cost will go up or down.

Keywords: Trade Open, Trade Close, Trade Low, Trade High, Python, Web Scrapping.

I. INTRODUCTION

The financial market may be a dynamic and composite system where people can purchase and sell currencies, stocks, equities and derivatives over virtual platforms supported by brokers. The stock exchange allows investors to have shares of public companies through trading either by exchange or over the counter markets. This market has given investors the prospect of gaining money and having a prosperous life through investing small initial amounts of cash, low risk compared to the danger of opening new business or the necessity of high salary career. Stock markets are suffering from many factors causing the uncertainty and high volatility within the market. Although humans can obey and submit them to the market, automated trading systems (ATS) that are operated by the implementation of computer programs can perform better and with higher momentum in submitting orders than any human. However, to gauge and control the performance of ATSs, the implementation of risk to strategies and safety measures applied. Many factors are incorporated and thought of when developing an ATS, as an example, trading strategy to be adopted, complex mathematical functions that reflect the state of a selected stock, machine learning algorithms that enable the prediction of the longer term stock value, and specific news associated with the stock being analysed.

Time-series prediction may be a common technique widely utilized in many real-world applications like meteorology and financial market prediction. It uses the continual data during a period of your time to predict the end in subsequent unit of time. Many time-series prediction algorithms have shown their effectiveness in practice. The foremost common algorithms now are supported Recurrent Neural Networks (RNN), also as its special type - Long-short Term Memory (LSTM) and Gated Recurrent Unit (GRU). Stock exchange may be a typical area that presents time-series data and lots of researchers study thereon and proposed various models.

during this project, LSTM model is employed to predict the stock price.

Research gaps: This project almost take 3 months to complete.

II. LITERATURE SURVEY

2.1 INTRODUCTION

"What other individuals think" has continuously been an critical piece of data for most of us amid the decision-making prepare. The Web and the Internet have presently (among other things) made it conceivable to discover out around the opinions and encounters of those within the endless pool of individuals that are neither our individual colleagues nor well-known proficient pundits — that's, individuals we have never listened of. And then again, increasingly individuals are making their suppositions accessible to outsiders through the Web. The intrigued that person clients appear in online suppositions approximately products and administrations, and the potential impact such suppositions use, is something that's driving drive for this zone of intrigued. And there are numerous challenges included in this prepare which must be strolled all over in advance to achieve legitimate results out of them. In this study we examined fundamental technique that more often than not happens in this prepare and measures that are to be taken to overcome the challenges being confronted.

2.2 EXISTING METHODS

2.2.1 Stock Market Prediction Using Machine Learning- The research work done by V Kranthi Sai Reddy Student, ECM, Sreenidhi Institute of Science and Technology, Hyderabad, India. within the finance world stock trading is

one among the foremost important activities. exchange prediction is an act of trying to work out the future value of a stock other financial instrument traded on a financial exchange. This paper explains the prediction of a stock using Machine Learning. The technical and fundamental or the statistic analysis is used by the foremost of the stockbrokers while making the stock predictions. the factitious language is used to predict the stock exchange using machine learning is Python. during this paper we propose a Machine Learning (ML) approach which is in a position to be trained from the available stocks data and gain intelligence then uses the acquired knowledge for an accurate prediction. during this context this study uses a machine learning technique called Support Vector Machine (SVM) to predict stock prices for the large and tiny capitalization.

2.2.2 Forecasting the Stock Market Index Using Artificial Intelligence Techniques

The research work done by Lufuno Ronald Marwala A dissertation submitted to the school of Engineering and therefore the Built Environment, University of the Witwatersrand, Johannesburg, in fulfilment of the wants for the degree of Master of Science in Engineering. The weak sort of Efficient Market hypothesis (EMH) states that it's impossible to forecast the longer term price of an asset supported the knowledge contained within the historical prices of an asset. this suggests that the market behaves as a stochastic process and as a result makes forecasting impossible. Furthermore, financial forecasting may be a difficult task thanks to the intrinsic complexity of the economic system . the target of this work was to use AI (AI) techniques to model and predict the longer term price of a stock exchange index. Three AI techniques, namely, neural networks (NN), support vector machines and neuro-fuzzy systems are implemented in forecasting the longer term price of a stock exchange index supported its historical price information. AI techniques have the power to require into consideration economic system complexities and that they are used as financial statistic forecasting tools. Two techniques are wont to benchmark the AI techniques, namely, Autoregressive Moving Average (ARMA) which is linear modelling technique and stochastic process (RW) technique. The experimentation was performed on data obtained from the Johannesburg stock market . the info used was a series of past closing prices of the All Share Index. The results showed that the three techniques have the power to predict the longer term price of the Index with a suitable accuracy. All three AI techniques outperformed the linear model. However, the stochastic process method outperformed all the opposite techniques. These techniques show a capability to predict the longer term price however, due to the transaction costs of trading within the market, it's impossible to point out that the three techniques can disprove the weak sort of market efficiency. The results show that the ranking of performances support vector machines, neuro-fuzzy systems, multilayer perceptron neural networks depends on the accuracy measure used.

2.2.3 Indian stock market prediction using artificial neural networks on tick data

The research work done by Dharmaraja Selvamuthu, Vineet Kumar and Abhishek Mishra Department of Mathematics, Indian Institute of Technology Delhi, Hauz Khas, New Delhi 110016, India. A stock exchange may be a platform for trading of a company's stocks and derivatives at an agreed price. Supply and demand of shares drive the stock exchange . In any country stock exchange is one among the foremost emerging sectors. Nowadays, many of us are indirectly or directly associated with this sector. Therefore, it becomes essential to understand about market trends. Thus, with the event of the stock 5 market, people have an interest in forecasting stock price. But, thanks to dynamic nature and susceptible to quick changes available price, prediction of the stock price becomes a challenging task. Stock m Prior work has proposed effective methods to find out event representations which will capture syntactic and semantic information over text corpus, demonstrating their effectiveness for downstream tasks like script event prediction. On the opposite hand, events extracted from raw texts lacks of common-sense knowledge, like the intents and emotions of the event participants, which are useful for distinguishing event pairs when there are only subtle differences in their surface realizations. to deal with this issue, this paper proposes to leverage external common-sense knowledge about the intent and sentiment of the event. Experiments on three event-related tasks, i.e., event similarity, script event prediction and stock exchange prediction, show that our model obtains far better event embeddings for the tasks, achieving 78% improvements on hard similarity task, yielding more precise inferences on subsequent events under given contexts, and better accuracies in predicting the volatilities of the stock market¹. Markets are mostly a nonparametric, non-linear, noisy and deterministic chaotic system (Ahangar et al. 2010). because the technology is increasing, stock traders are moving towards to use Intelligent Trading Systems instead of fundamental analysis for predicting prices of stocks, which helps them to require immediate investment decisions. one among the most aims of a trader is to predict the stock price such he can sell it before its value decline, or buy the stock before the worth rises. The efficient market hypothesis states that it's impossible to predict stock prices which stock behaves within the stochastic process . It seems to be very difficult to exchange the professionalism of an experienced trader for predicting the stock price. But due to the supply of an interesting amount of knowledge and technological advancements we will now formulate an appropriate algorithm for prediction whose results can increase the profits for traders or investment firms. Thus, the accuracy of an algorithm is directly proportional to gains made by using the algorithm.

2.2.4 The Stock Market and Investment

The research work done by Manh Ha Duong Boriss Siliverstovs. Investigating the relation between equity prices and aggregate investment in major European countries including France, Germany, Italy, Netherlands and therefore the uk . Increasing integration of European financial markets is probably going to end in even stronger correlation between equity prices in several European countries. This process also

can cause convergence in economic development across European countries if developments available markets influence real economic components, like investment and consumption. Indeed, our vector autoregressive models suggest that the direct correlation between changes equity prices and investment is, generally, significant. Hence, 6 monetary authorities should monitor reactions of share prices to monetary policy and their effects on the trade cycle.

2.2.5 Automated Stock Price Prediction Using Machine Learning

The research work done by Mariam Moukalled Wassim El-Hajj Mohamad Jaber computing Department American University of Beirut. Traditionally and so as to predict market movement, investors want to analyse the stock prices and stock indicators additionally to the news associated with these stocks. Hence, the importance of stories on the stock price movement. Most of the previous add this industry focused on either classifying the released market news as (positive, negative, neutral) and demonstrating their effect on the stock price or focused on the historical price movement and predicted their future movement. during this work, we propose an automatic trading system that integrates mathematical functions, machine learning, and other external factors like news' sentiments for the aim of achieving better stock prediction accuracy and issuing profitable trades. Particularly, we aim to work out the worth or the trend of a particular stock for the approaching end-of-day considering the primary several trading hours of the day. to realize this goal, we trained traditional machine learning algorithms and created/trained multiple deep learning models taking into consideration the importance of the relevant news. Various experiments were conducted, the very best accuracy (82.91%) of which was achieved using SVM for Apple Inc. (AAPL) stock.

2.2.6 Stock Price Correlation Coefficient Prediction with ARIMALSTM Hybrid Model

The research work done by Hyeong Kyu Choi, B.A Student Dept. of Business Administration Korea University Seoul, Korea. Predicting the worth correlation of two assets for future time periods is vital in portfolio optimization. We apply LSTM recurrent neural networks (RNN) in predicting the stock price coefficient of correlation of two individual stocks. RNN's are competent in understanding temporal dependencies. the utilization of LSTM cells further enhances its long-term predictive properties. To encompass both linearity and nonlinearity within the model, we adopt the ARIMA model also. The ARIMA model filters linear tendencies within the data and passes on the residual value to the LSTM model. The ARIMA-LSTM hybrid model is tested against other traditional predictive financial models like the complete historical model, constant correlation model, single-index model and therefore the multi-group model. In our empirical study, the predictive ability of the ARIMA-LSTM model clothed superior to all or any other financial models by a big scale. Our work implies that it's worth considering the ARIMALSTM model to forecast coefficient of correlation for portfolio optimization.

2.2.8 Forecasting directional movements of stock prices for intraday trading using LSTM and random forests

The research work done by Pushpendu Ghosh, Ariel Neufeld, Jajati Keshari Sahoo Department of Computer Science & Information Systems, BITS Pilani K.K.Birla Goa campus, India bDivision of Mathematical Sciences, Nanyang Technological University, Singapore cDepartment of Mathematics, BITS Pilani K.K.Birla Goa campus, India. We employ both random forests and LSTM networks (more precisely CuDNNLSTM) as training methodologies to analyse their effectiveness in forecasting out-of-sample directional movements of constituent stocks of the S&P 500 from January 1993 till December 2018 for intraday trading. We introduce a multi-feature setting consisting not only of the returns with reference to the closing prices, but also with reference to the opening prices and intraday returns. As trading strategy, we use Krauss et al. (2017) and Fischer & Krauss (2018) as benchmark and, on each trading day, buy the 10 stocks with the highest probability and sell short the 10 stocks with the lowest probability to outperform the market in terms of intraday returns – all with equal monetary weight. Our empirical results show that the multi-feature setting provides a daily return, before transaction costs, of 0.64% using LSTM networks, and 0.54% using random forests. Hence, we outperform the single feature setting in Fischer & Krauss (2018) and Krauss et al. (2017) consisting only of the daily returns with reference to the closing prices, having corresponding daily returns of 0.41% and of 0.39% with reference to LSTM and random forests, respectively. 1
Keywords: Random forest, LSTM, Forecasting, Statistical Arbitrage, Machine learning, Intraday trading.

III. PROPOSED METHODOLOGY

I. WEB SCRAPPING -

Here Scraping refers to the gathering of knowledge ethically without violating the overall norms set by the web site from which we are collecting data. We are collecting data from the online platforms available to all or any to be used hence it's referred to as Web Scapping. Now we will see the overall samples of web Scapping in day to day life like copying any important educational data from educational website for study and research, copying the lyrics of any song you wish from website and lots of more. But still as discussed all of those data should be legal to use. Similar concept is carried once we mention Machine Learning. on predict the longer term stock prices, we'd like to think about the past data of the corporate. wish to train our machine supported the past data set, a machine learning library called Pandas_Datareader must be imported in header of python code. We are Scapping data from the web site called Yahoo Finance(<http://in.finance.yahoo.com>)

II.TOOLS/PACKAGES AND PLATFORM USED -

To begin with the project, we have implementation language as **Python 3** and some of dependencies as,

```
In [ ]: import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import pandas_datareader as data
import streamlit as st
from tensorflow.keras.models import load_model
```

NUMPY:

NumPy is a Python library used for working with arrays. It also has functions for working in domain of algebra , fourier transform, and matrices. NumPy was created in 2005 by Travis Oliphant. it's an open source project and you'll use it freely.

PANDAS:

Pandas is an open source Python package that's most generally used for data science/data analysis and machine learning tasks. it's built on top of another package named Numpy, which provides support for multi-dimensional arrays.

MATPLOTLIB:

Matplotlib is a cross-platform, data visualization and graphical plotting library for Python and its numerical extension NumPy. As such, it offers a viable open source alternative to MATLAB. Developers can also use matplotlib's APIs (Application Programming Interfaces) to embed plots in GUI applications.

PANDAS_DATAREADER:

Pandas Datareader may be a Python package that permits us to make a pandas DataFrame object by using various data sources from the web . it's popularly used for working with realtime stock price datasets.

STREAMLIT:

Streamlit is an open-source python library for creating and sharing web apps for data science and machine learning projects. The library can assist you create and deploy your data science solution during a jiffy with a couple of lines of code.

TENSORFLOW:

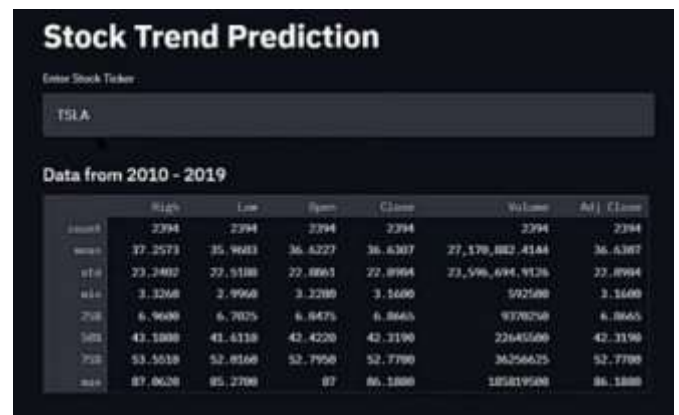
Tensorflow is a Python library for fast numerical computing created and released by Google. it's a foundation library which will be wont to create Deep Learning models directly or by using wrapper libraries that simplify the method built on top of TensorFlow.

KERAS:

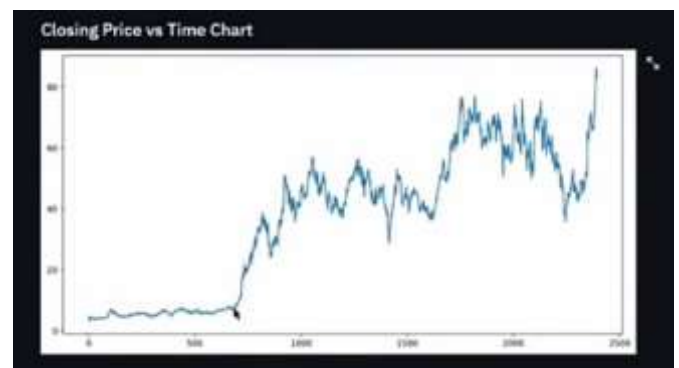
Keras may be a high-level, deep learning API developed by Google for implementing neural networks. it's written in Python and is employed to form the implementation of neural networks easy. It also supports multiple backend neural network computation.

IV. RESULT

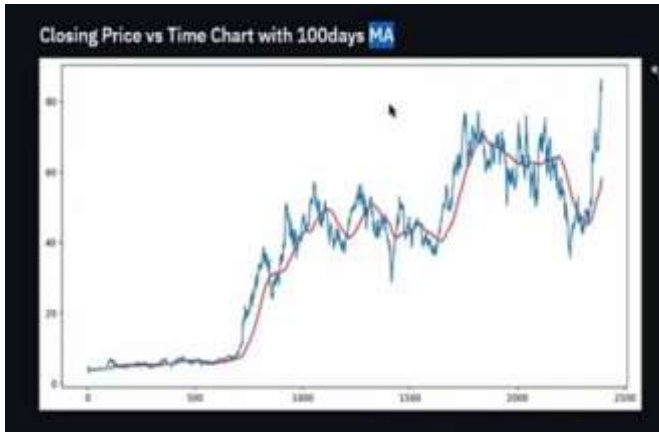
The first step on this prediction is first we web scrap the data of any company .As you can see in the below snippet we have consider the TSLA company. After web scrapping we have some stock data such as highs,lows,opening and closing price of stock of TSLA.In this data we take stock data from 2010-2019,but we can update the date as per our requirement.



Here we consider the closing price of stock,and after that we code to draw Closing price VS Time graph/chart,we can also called this graph as a original graph as shown in below snippet.

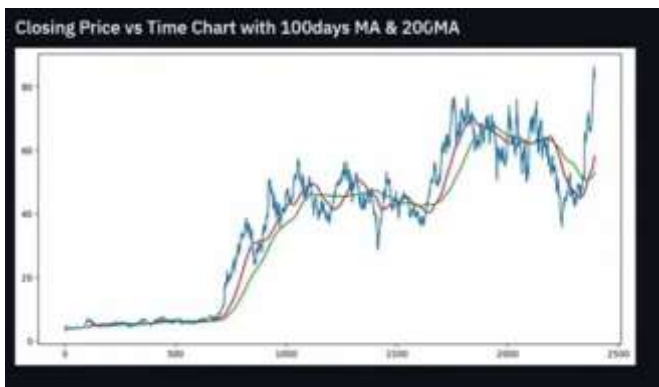


After that we take similar data of stock and we have done the coding for the similar Closing price VS Time chart but with the 100 days moving average. That means we have consider the data of previous 100 days from the date we want analyze the stock.



Similarly we take data of 200 days moving average that means data of previous 200 days and done coding to draw the graph/chart of Closing price VS Time with 200 days moving Average.

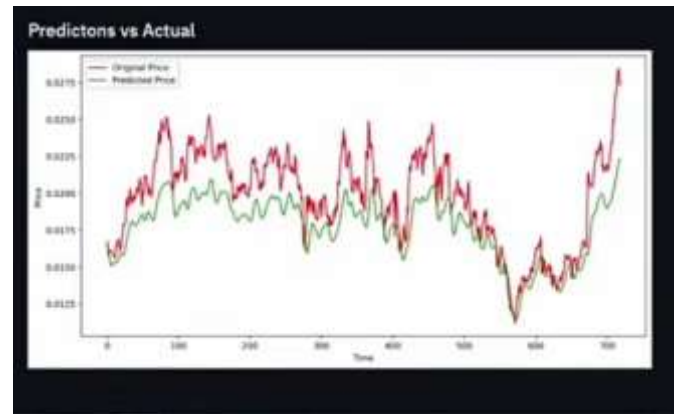
And then we merged both the graph with 100days moving average and 200 days moving average.



The red line showing the 100 days moving average and green line showing the 200 days moving average stocks.

Here if the 100 days MA line (red line) is above the 200 days MA (green line), then it indicates that the stock price is increasing. This technique is manually done by the company to calculate stock, but we implement it practically using graphs.

After this we merged the predicted stock price graph and original stock price graph, and we got our final result.



V. CONCLUSION

In this project, we are predicting closing stock price of any given organization, we developed a web application for predicting close stock price using PYTHON, Web scrapping, some python libraries. for prediction. We can also predict stock of other industries such as Google, Nifty50, TCS, Infosys and Reliance Stocks and can achieved above 80% accuracy for these datasets.

Future work

- We want to extend this application for predicting cryptocurrency trading.
- We want to add sentiment analysis for better analysis.

REFERENCES

Datasets: Nifty 50 data, tw_spydata_raw(trading data)

- [1] Stock Price Prediction Using LSTM on Indian Share Market by Achyut Ghosh, Soumik Bose, Giridhar Maji, Narayan C. Debnath, Soumya Sen
- [2] S. Selvin, R. Vinayakumar, E. A. Gopalkrishnan, V. K. Menon and K. P. Soman, "Stock price prediction using LSTM, RNN and CNN-sliding window model," in International Conference on Advances in Computing, Communications and Informatics, 2017.
- [3] Murtaza Roondiwala, Harshal Patel, Shraddha Varma, "Predicting Stock Prices Using LSTM" in Undergraduate Engineering Students, Department of Information Technology, Mumbai University, 2015.
- [3] Murtaza Roondiwala, Harshal Patel, Shraddha Varma, "Predicting Stock Prices Using LSTM" in Undergraduate Engineering Students, Department of Information Technology, Mumbai University, 2015.
- [5] Ishita Parmar, Navanshu Agarwal, Sheirsh Saxena, Rida Arora, Shikhin Gupta, Himanshu Dhiman, Lokesh Chouhan Department of Computer Science and

- Engineering National Institute of Technology, Hamirpur – 177005, INDIA – Stock Market Prediction Using Machine Learning.
- [6] Pranav Bhat Electronics and Telecommunication Department, Maharashtra Institute of Technology, Pune. Savitribai Phule Pune University - A Machine Learning Model for Stock Market Prediction.
- [7] Anurag Sinha Department of computer science, Student, Amity University Jharkhand Ranchi, Jharkhand (India), 834001 - Stock Market Prediction Using Machine Learning.
- [8] V Kranthi Sai Reddy Student, ECM, Sreenidhi Institute of Science and Technology, Hyderabad, India - Stock Market Prediction Using Machine Learning.
- [9] Asset Durmagambetov currently works at the mathematics, CNTFI. Asset does research in Theory of Computation and Computing in Mathematics, Natural Science, Engineering and Medicine. Their current project is 'The Riemann Hypothesis-Millennium Prize Problems' Stock price prediction.
- [10] Mariam Moukalled Wassim El-Hajj Mohamad Jaber Computer Science Department American University of Beirut - Automated Stock Price Prediction Using Machine Learning.
- [11] Manh Ha Duong Boriss Siliverstovs June 2006 - The Stock Market and Investment.
- [12] Dharmaraja Selvamuthu, Vineet Kumar and Abhishek Mishra Department of Mathematics, Indian Institute of Technology Delhi, Hauz Khas, New Delhi 110016, India - Indian stock market prediction using artificial neural networks on tick data.
- [13] Lufuno Ronald Marwala A dissertation submitted to the Faculty of Engineering and the Built Environment, University of the Witwatersrand, Johannesburg, in fulfillment of the requirements for the degree of Master of Science in Engineering - Forecasting the Stock Market Index Using Artificial Intelligence Techniques.
- [14] Xiao-Yang Liu¹ Hongyang Yang, Qian Chen⁴, Runjia Zhang, Liuqing Yang Bowen Xiao Christina Dan Wang Electrical Engineering, ²Department of Statistics, ³Computer Science, Columbia University, ^{3A}AI4Finance LLC., USA, Ion Media Networks, USA, Department of Computing, Imperial College, ⁶New York University (Shanghai) - A Deep Reinforcement Learning Library for Automated Stock Trading in Quantitative Finance.
- [15] Pushpendu Ghosh, Ariel Neufeld, Jajati Keshari Sahoo Department of Computer Science & Information Systems, BITS Pilani K.K.Birla Goa campus, India ^bDivision of Mathematical Sciences, Nanyang Technological University, Singapore ^cDepartment of Mathematics, BITS Pilani K.K.Birla Goa campus, India - Forecasting directional movements of stock prices for intraday trading using LSTM and random forests.
- [16] Xiao Ding, Kuo Liao, Ting Liu, Zhongyang Li, Junwen Duan Research Center for Social Computing and Information Retrieval Harbin Institute of Technology, China - Event Representation Learning Enhanced with External Commonsense Knowledge.
- [17] Huicheng Liu Department of Electrical and Computer Engineering Queen's University, Canada - Leveraging Financial News for Stock Trend Prediction with Attention-Based Recurrent Neural Network.
- [18] Hyeong Kyu Choi, B.A Student Dept. of Business Administration Korea University Seoul, Korea = Stock Price Correlation Coefficient Prediction with ARIMA-LSTM Hybrid Model.
- [19] M. Nabipour Faculty of Mechanical Engineering, Tarbiat Modares University, 14115-143 Tehran, Iran; Mojtaba.nabipour@modares.ac.ir - Deep Learning for Stock Market prediction.
- [20] Lavanya Ra SRM Institute of Science and Technology | SRM · Department of Computer Science - Stock Market Prediction.
- [21] M. Mekayel Anik · M. Shamsul Arefin (B) Department of Computer Science and Engineering, Chittagong University of Engineering and Technology, Chittagong, Bangladesh - An Intelligent Technique for Stock Market Prediction.