**Stock Market Prediction using Hybrid Machine Learning System**

**A PROJECT REPORT**

*Submitted by*

**SHRUSHTI GOVARDHAN [Reg No: RA1611008010099]**

**SHREYAS KULKARNI [Reg No: RA1611008010186]**

**PRATIK DHARAMDASANI [Reg No: RA1611008010298]**

**ASTHA MAHESH KUMAR [Reg No: RA1611008010752]**

*Under the Guidance of*

**Mr. S Deepan**

(Assistant Professor, Department of Information Technology)

*In partial fulfillment of the Requirements for the Degree*

*of*

**BACHELOR OF TECHNOLOGY**



**DEPARTMENT OF INFORMATION TECHNOLOGY**

**FACULTY OF ENGINEERING AND TECHNOLOGY**

**SRM INSTITUTE OF SCIENCE AND TECHNOLOGY**

**KATTANKULATHUR – 603203**

**MAY 2020**

**CHAPTER 1**

**INTRODUCTION**

**Abstract:**

Stock Markets are highly complex in nature and are dependent on so many factors that it is impossible for an individual to predict its course. It is curtail for investors to get an idea about the behavior of the stock price in order to generate notable profit. With the recent success of machine learning algorithms for solving a wide range of complex problems, a lot of attempts have been made for forecasting the returns of stock market. There currently exists no system that is accurate and reliable enough so as to be used by investors. This project aims at predicting the behavior of the stock prices at the National Stock Exchange of India. The proposed system makes use of two different machine learning approaches for determining the future stock prices. Important patters are extracted from the data using the Principal Component Analysis of the dataset. The extracted features are used as input for three non-linear machine learning models. The performance of these three models are then compared and analyzed for their accuracy in stock price prediction.

**Introduction:**

Stock market refers to a marketplace where investors can buy and sell shares of public companies. Companies who register in this market can sell a share of their businesses to investors and offer them corresponding parts of profits or losses in return. The market works by constant buying and selling of stocks by investors.

The problem arises when a person buy shares of a company and its value goes down. The investor suffers a loss due to the decline in the price of the share. This is such a huge problem that people end up losing all of their savings in the stock market. In our country, about half the people who invest in the stock market lose all their invested amount in about a year. The new generation is afraid to invest in the stock market due to its unpredictability and loss history.

Stock trading is one of the easiest ways in which companies can make use of public funds and expand without taking loans. For investors, it is an easy and fast way of earning good returns on their money. If there could be a system that could predict the price of the stocks, it would prove to be a very useful tool for the investors. It would enable everyone to invest in the stock market and earn good returns on their money while avoiding losses.

Stock market prediction is still an active area of research globally. No system has yet been made that is reliable enough to be used at commercially. By choosing this topic, we aim to make a contribution towards the study of stock market by using machine learning techniques. We hope that our effort would somehow prove to be a useful addition towards solving the problem of stock market prediction. We also hope to that our system would help investors in predicting the stock prices before investing and would stop losses at the share market.

**Literature Review**

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| S. No. | Paper Title | Year | Dataset | Preprocessing | Algorithm | Result and Inference | Metrics |
| 1 | Predicting the daily return direction of the stock market using hybrid machine learning algorithms | 2019 | SPDR S&P 500 ETF dataset | Handling the missing values, removing outliers and mismatched samples.  Transformation of data using PCA. | Deep Neural Network with different no of hidden layers (12, 16, 22, 28, 30 and 35). | Achieved highest accuracy of 59.9%  Accuracy stopped increasing as layers became larger than 30. | Training and Testing accuracy is used for comparison. |
| 2 | Stock Market Data Prediction Using Machine Learning Techniques | 2019 | Apple Inc. stock prices dataset. Created during the study. | Data was collected and following features were noted:  Date, Open, High, Low, Close and Volume. | Compared Random Forest and Multilayer Perceptron. | Random Forest Classifier outperformed Multilayered Perceptron in both correlation coefficient and error. | Comparison made using:  Correlation coefficient, Mean absolute error, Root mean squared error, Relative absolute error. |
| 3 | Stock Market Prediction Using Machine Learning | 2018 | BSE behavior dataset. Created during the study. | Feature Selection:  Commodity Prices (oil, gold, silver), Market History and Forex. | 4 Algorithms are compared:  SVM, Random Forest, Gradient Boosting and AdaBoost. | Test accuracy:  Random Forest – 68.4%,  SVM – 73%, Gradient Boosting – 73.2%, AdaBoost – 77% | Accuracy of the BSE value is used as comparison metric. |
| 4 | Stock Prediction Using Machine Learning Algorithms | 2018 | Yahoo Finance Dataset. | Feature Selection:  Open, Close, High, Low and Volume. | LSTM Neural Network and Regression Model are compared. | LSTM model achieved higher R-Square score than Regression Model. | R-Square confidence score was chosen as comparison metric. |
| 5 | Stock market prediction using machine learning techniques | 2016 | Data was collected during study. | Feature Selection:  Oil, Gold and Silver Prices, Forex, Interest Rates, News Feed.  News was labelled as positive or negative manually. | Single Layer Perceptron, Multilayer Perceptron, Radial Basis Function and Support Vector Machine algorithms are compared. | Test Accuracy:  SLP – 60%, MLP – 77%, RBF – 63%, SVM – 60%.  Multilayer Perceptron outperformed the other algorithms. | Market Behavior is predicted as positive or negative. Accuracy is used for comparison. |

**Problem Definition**

Every year investors lose millions in the stocks due to the unpredictability and complexity of the stock market. Sometimes the losses are so huge that people are forced declare bankruptcy. Due to these loss stories, new generation is afraid of putting money in the stock market which is affecting the growth of businesses.

The stock market works on such complex and interrelated factors that it is extremely difficult for a human being to make an accurate prediction about the trends in the market. There are a lot of factors that influence the stock market and due to the advancement of technology it is now possible to take in consideration a lot of data and process it for deriving insights.

There is no thing as a good stock or a bad stock, timing plays a major role while investing. If an investment is done at a bad time, then it is bound to bear losses to the investor. Hence it is very important to get the timing right while investing in the stock market. By knowing the future price of the stock, investors can make well timed investment decisions which will return good profits to them.

Since stock market is based on trends, we can extract the important features from historical stock market data and apply machine learning techniques to understand the underlying patterns. Using these patterns we can make predictions about how the stock price would be in the future.

When such a prediction would be made, it would enable investors to know in advance what the stock price would be in the future hence helping them towards making profitable investment decisions in the stock market.

**Requirement Analysis**

**Functional Requirements**

Functional requirements are the features that must be included in any system to satisfy the business needs and be useful to the user. The functional requirements of this system are:

1. The system must be able to generate approximate share price using the input data.
2. The input should be normalized before getting in the network.
3. The system should be capable of calculating the PCA of the data.

**Non Functional Requirements**

Non-functional requirements are description of a feature, characteristic and attribute of the system as well as any constraint that may define the boundaries of the proposed system. The non-functional requirements are essentially based on the performance, information, security, efficiency. The non-functional requirements of the proposed system are:

1. Accuracy of the price prediction system
2. The system should provide the results in short amount of time
3. High reliability
4. Low response time
5. Robustness
6. Maintainability

**Hardware Requirements**

These are the hardware requirements that the system should have in order to optimally run the proposed method. The proposed method makes use of non-linear machine learning algorithms and deep learning architecture. Therefore we need a system which has at least the following hardware specifications:

1. 8 Gigabytes of RAM
2. Intel i5 5th Generation Processor
3. 1 Gigabytes of Hard Disk space to store the dataset and trained models.

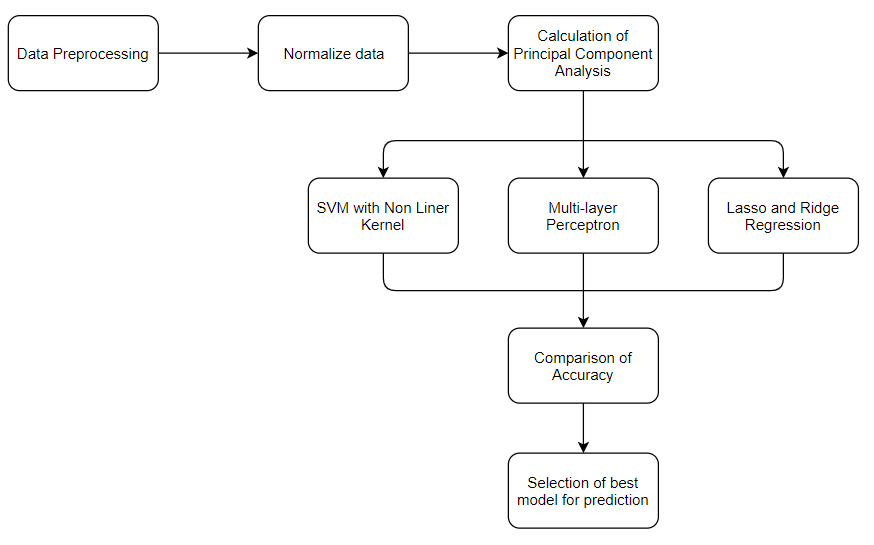
**Proposed Methodology**

**System Architecture**

We propose the a method for stock market prediction which makes use of non-linear machine learning algorithms and a deep neural network architecture achieving accurate results. The results obtained would be compared on the terms of accuracy of price prediction and the best model would be chosen.

The method makes use of preprocessing concepts for handling the missing values and organizing the data. The data is normalized and PCA is calculated for it. It is done for extracting the essential information that is responsible for defining the data pattern. Three non-linear models are trained and tested on different parts of dataset for obtaining reliable results. The accuracy is compared and best model is reported and selected.

The diagram below represents the flow of the proposed system.



**Description of System Architecture**

The proposed system makes use of machine learning and deep learning concepts for predicting the price of a stock. The system consists of the following steps:

1. Data Preprocessing: The data is collected from National Stock Exchange of India (NSE) from Kaggle. The missing values from the data are filled by taking the mean value of the column. The correlation of columns is computed and highly correlated columns are removed for decreasing the dimensionality.
2. Normalizing Data: Normalizing refers to the transformation of data in a specific range. All the values in the dataset is normalized in the range of 0 to 1. This step helps in increasing computation speed as all the data is converted to same scale. Additionally, normalization is necessary before calculation of Principal Component Analysis, which is done in the next step.
3. Calculation of PCA: Principal Component Analysis is a method of calculating Eigen vectors and Eigen values from the data. This calculation helps in understanding the fundamental properties that the data follows. From the existing research it has been observed that algorithms give significantly higher accuracy on PCA transformed dataset as compared to the non-transformed data. It also helps in dimensionality reduction.
4. Model Testing: Stock market data is highly non-linear in nature hence fitting linear models on this data would prove to be of less use. Therefore three non-linear models, namely, SVM (with non-linear kernel), Multilayer Perceptron and Lasso and Ridge Regression are chosen for this task. These models would be trained on the same transformed dataset. The dataset is split for training and testing in order to obtain correct accuracy scores.
5. Comparison of accuracy: The performance of all the models would be compared with accuracy taken as the metric. The accuracy obtained on the test dataset would be used for comparison. The models would be trained on a part of the dataset and would be tested on another part of the dataset. The aim would be to make the model predict the stock price for the next time interval. The error in prediction would be recorded and made to reduce for achieving high accuracy.
6. Selection of the best model: The model which outperforms the other models in terms of accuracy would be selected and reported as the most accurate model.

**Impact of solution**

Till date no reliable system exists that can used by economists to aid them for understanding the behavior of a stock price. The proposed solution might come up as an accurate and reliable system that could be used by general public and companies for making better investing decisions.

The proposed solution would make it easy for every person to invest in stock market without the fear of losing their money. The solution would have a great impact on the society as every year a lot of investors lose all of their money in the stock market in hopes of greater returns. This fear of loss has made the new generation afraid of the stock market so much that some even call it a gamble.

This solution would make it possible for everyone to make an investment on the correct time so that nobody would lose money in the stock market. This would also help economists for predicting market crisis before it happens and would enable them for taking decisions that would help stabilizing the market. This solution, if developed further to perfection, could be useful by countries for making better financial decisions and improving the GDP of the nation by making smart financial decisions.

**Ethical and Social Issues**

When it comes to Stock Market, there are many ethical issues that come into play. Many professionals in the past have been found guilty of using unethical practices for increasing the estimated worth of their company and fooling people into investing in their businesses. Such people have been accused of selling shares to people just to fool people into giving them more money than the company’s actual worth.

Over the years, the stock exchange market in our country (NSE and BSE) have set up rules and policies for all the transactions that take place in order to stop fraudulent exchanges. These stock exchange markets are governed by Securities and Exchange Board of India (SEBI), which is always governing the transactions and punishing the companies which get involved in unethical ways for increasing the worth of their stock.

The stock market is now a safe place for transactions however, if the use of the proposed method is not governed, it might lead to unethical use by some professionals.

The social issues around Stock Market are well known by everybody. People who are new to stock market think of it as a quick way of getting rich and invest heavily in the high risk businesses without balancing the portfolio. This leads to big losses for these people and they blame the stock market for its unpredictability. In reality however, the people who take calculated decisions on the stock market and balance the risk according to their financial situation are the ones who are able to obtain profits.

The people should be aware that the proposed system would not be a tool for getting rich quickly on the stock market. The investors would still be required to take calculated risks and balance their portfolio in order to see profit returns on the stock market. The proposed method would lessen the fear of stock market from the new investors. The removal of fear would lead to more people studying about stock market and making calculated investments. This would increase the earnings of general public and would increase the living standards and would reduce poverty from the country.

**Work Distribution**

|  |  |
| --- | --- |
| SHRUSHTI GOVARDHAN  [Reg No: RA1611008010099] |  |
| SHREYAS KULKARNI  [Reg No: RA1611008010186] |  |
| PRATIK DHARAMDASANI  [Reg No: RA1611008010298] |  |
| ASTHA MAHESH KUMAR  [Reg No: RA1611008010752] |  |

**Work list:**

Introduction, Abstract, Literature Review (can be done by two or three people), Defining problem statement, Analyzing requirements (can be done by two people), Proposing the solution (can be done by two or three people), Analyzing the impact of the solution, Analyzing ethical and social issues.

**References**

[1] A. Nayak, M. Pai, R. Pai, Prediction models for Indian stock market. Proced. Comput. Sci. (2016)

[2] Z. Hu, J. Zhu, K. Tse, Stocks market prediction using support vector machine, in *2013 6th International Conference on InformationManagement, InnovationManagement and Industrial Engineering* (2013)

[3] WeiHuang,YoshiteruNakamori, Shou-YangWang, Forecasting stock market movement direction with support vector machine. Comput. Oper. Res. (2005)

[4] C. Cortes, V. Vapnik, Support vector networks. Mach. Learn. (1995)

[5] J. Ali, R. Khan, N. Ahmad, I. Maqsood, Random forests and decision trees. Int. J. Comput. Sci. Issues (2012)

[6] S. Yutong, H. Zhao, Stock selection model based on advanced AdaBoost algorithm, in *2015 7th International Conference on Modelling, Identification and Control (ICMIC)* (2015)

[7] P. Wu, H. Zhao, Some analysis and research of the AdaBoost algorithm, in *Communications in Computer and Information Science* (2011)

[8] L. Zhao, L. Wang, Price trend prediction of stock market using outlier data mining algorithm, in *2015 IEEE Fifth International Conference on Big Data and Cloud Computing* (2015)

[9] M. Usmani, S. Hasan Adil, K. Raza, S. Ali, Stock market prediction using machine learning techniques, *ICCOINS* (2016)

[10] J. Ali, R. Khan, N. Ahmad, I.Maqsood, Random forest and decision trees. Int. J. Comput. Sci. (2012)