Stock Price Prediction using Machine Learning

Mr. Chiranjeevi M R[1], Mrs. Meghana C N[2], Ms. Hemalatha B M[3]
[1,2] *Assistant Professor CSE department Rit Hassan*
[3]*Assistant professor Department of Computer science and engineering Rajeev Institute of technology*

***Abstract*- In the current era stock price prediction plays a key role for prediction of future data with respect to training the past data by using machine learning or deep learning technologies. Building a model then passing the past data as input that is as training data to the model based on the results acquired needs to consider an algorithm which gives better accuracy and response time and segmentation. In this paper for estimating the stock values we are considering LSTM and regression models for machine learning. Factors considered are opening values of stock; closing values of stock; lower and higher values of stock and volume. Predicting stock market prices is a challenging task in the financial sector where the Efficient Market Hypothesis (EMH) posits the impossibility of accurate prediction due to the inherent uncertainity and complexity of stock price behaviour. This study employs advanced machine learning models that can predict stock price movements with the right level of accuracy if the correct parameter tuning and appropriate predictor models are developed.**

***Index terms* – Stock prediction, LSTM, EMH, predict stock price movements.**

I. INTRODUCTION

In recent times stock market predictions is gaining more attention, maybe due to the fact that if the trend of a the market is successfully predicted the investors may be better guided. The profits gained by investing and trading in the stock market greatly depends on the predictability.

If there is a system that can consistently predict the direction of the dynamic stock market will enable the users of the system to make informed decisions. More over which predicted trends of the market will help the regulators of the market in taking corrective measures.

Dataset Generation: The process begins with the creation of a data pre-processing technique. The pre-processing stage formally involves:

a. Data discretization: Reducing the data but with our importance and in accordance with the algorithms

b. Data transformation: Normalizing the data

c. Data Cleaning: Cleaning i.e. removing all the unnecessary elements from the dataset keeping the ones we actually needed.

d. Data Integration: Integration of data files
After the data-set is transformed into clean data-set, training and testing datasets are being taken from the dataset for further work and evaluation. Here, the training values are taken as the more recent values because we focus more on training the algorithms and model first. We have kept testing dataset as five to ten percent of the training dataset.

Model Training: Once the dataset is prepared, a machine learning model is trained to predict the stock market price value. For this purpose, two classifiers are employed:

- HOG Classifier: Utilizes handcrafted features to detect edges and gradients, which are then used for feature extraction. This method is computationally less intensive but might struggle with complex patterns.
- CNN Classifier: Leverages deep learning to automatically learn hierarchical features directly from raw image data making it robust to variations in pose, lighting, and expressions.

Stock price prediction: The stock market community has been attracted by the idea of a model that can forecast the future move of the market since the emergence of Data science and its main streaming in many industries. Stock market prediction is the process of examining numerous components of the stock market that can impact the price of a stock and then developing a model to predict the stocks price based on these potential factors.

A Machine Learning model estimates the value of an observation using numerous predictor inputs. The stock market operates similarly, in that the stock price swings based on a variety of variables. However,

unlike other problems that can be predicted, stock price projections are particularly tricky because they are frequently changing.

Anything, any incident in the outside world, can have an impact on the stock price. Political events, economic news, competitors or associated stock movements and other difficult to capture ideas such as rumor, worry and other psychological aspects are examples. Capturing all of these variables makes stock market prediction difficult, and Machine Learning is a possible solution. It can evaluate patterns and their relationship to a dependent variable using the vast number of inputs. Developing a decent stock price prediction model is especially difficult because it is non-linear. Machine Learning based algorithms can be exceeding as well as sophisticated, capturing the complex world of the stock market and how numerous factors influence stocks price.

Result Summary:
A stock market, equity market or share market is the aggregation of buyers and sellers (a loose network of economic transactions, not a physical facility or discrete entity) of stocks (also called shares), which represent ownership claims on businesses; these may include securities listed on a public stock exchange as well as those only traded privately. Examples of the latter include shares of private companies which are sold to investors through equity crowd funding platforms. Stock exchanges list shares of common equity as well as other security types, e.g. corporate bonds and convertible bonds.

Stock price prediction is one of the most widely studied problem, attracting researchers from many fields. The volatile nature of the stock market makes it really difficult to apply simple time-series or regression techniques.

This analysis highlights the critical need to balance accuracy and computational efficiency, offering valuable insights into selecting the most suitable model depending on specific application requirements, such as the need for real-time performance or high-accuracy outcomes.

II. AIMS AND OBJECTIVES

The aim of the project is to examine a number of different forecasting techniques to predict future stock returns based on past returns and numerical news indicators to construct a portfolio of multiple stocks in order to diversify the risk. We do this by applying supervised learning methods for stock price forecasting by interpreting the seemingly chaotic market data. Model Evaluation: The performance of the trained model is assessed using key metrics such as accuracy, precision, recall, and F1-score. These metrics provide a comprehensive evaluation of the model's effectiveness in recognizing faces and highlight areas that may require further refinement or optimization.

Stock price prediction is a classic and important problem. With a successful model for stock prediction, we can gain insight about market behavior over time, spotting trends that would otherwise not have been noticed. With the increasingly computational power of the computer, machine learning will be an efficient method to solve this problem.

Thus, our motivation is to design a public service incorporating historical data and users predictions to make a stronger model that will benefit everyone.
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.

III. METHODS

A. Linear Regression: Linear Regression can be used to forecast any continuous variable, including any stock price projections. The main advantage of this strategy is that it is highly interpretable since the user can determine which factors influence the stock price the most and how much. The downside is that it has very narrow scope. A large number of predictors cannot be used to solve the stock price prediction problem.

B. ARIMA: This time series forecasting method can be used to forecast a stock's price. Because it is a time series approach. It takes into account the time component and its impact on the stocks price. ARIMA is an acronym that stands for Auto-Regressive Integrated Moving Average and refers to a group of functions. If the hyper parameters p, d and q are considered it can be employed as a Machine Learning approach.

C. KNN: KNN appears to be the best candidate. It is a distance based strategy that looks for the most similar records for observation and then predicts the value based on the outcome of these records. In theory, for a market circumstance. KNN may hunt for the most similar previous event, determine how the market behaved, and forecast the stock price.

D. Python Technology:
Python is an interpreted, object- oriented programming language similar to PERL, that has gained popularity because of its clear syntax and readability. Python is said to be relatively easy to learn and portable, meaning its statements can be interpreted in a number of operating systems, including UNIX- based systems, Mac OS, MS- DOS, OS/2, and various versions of Microsoft Windows 98. Python was created by Guido van Rossum, a former resident of the Netherlands, whose favourite comedy group at the time was Monty Python's Flying Circus. The source code is freely available and open for modification and reuse. Python has a significant number of users.

E. Python platform:
Apart from Windows, Linux and MacOS, C Python implementation runs on 21 different platforms. Iron Python is a .NET framework based Python implementation and it is cap able of running in both Windows, Linux and in other environments where .NET framework is available.

F. Python Library:
Machine Learning, as the name suggests, is the science of programming a computer by which they are able to learn from different kinds of data. A more general definition given by Arthur Samuel is – "Machine Learning is the field of study that gives computers the ability to learn without being explicitly programmed." They are typically used to solve various types of life problems.

In the older days, people used to perform Machine Learning tasks by manually coding all the algorithms and mathematical and statistical formula. This made the process time consuming, tedious and inefficient. But in the modern days, it is become very much easy and efficient compared to the olden days by various python libraries, frameworks, and modules. Today, Python is one of the most popular programming languages for this task and it has replaced many languages in the industry, one of the reason is its vast collection of libraries. Python libraries that used in Machine Learning are:

- NumPy
- SciPy
- Scikit- learn
- Theano
- TensorFlow
- Keras
- PyTorch
- Pandas
- Matplotlib

NumPy:
NumPy is a very popular python library for large multi- dimensional array and matrix processing, with the help of a large collection of high- level mathematical functions. It is very useful for fundamental scientific computations in Machine Learning. It is particularly useful for linear algebra, Fourier transform, and random number capabilities. High- end libraries like TensorFlow uses NumPy internally for manipulation of Tensors.

SciPy:
SciPy is a very popular library among Machine Learning enthusiasts as it contains different modules for optimization, linear algebra, integration and statistics. There is a difference between the SciPy library and the SciPy stack. The SciPy is one of the core packages that make up the SciPy stack. SciPy is also very useful for image manipulation.

Skikit:
Skikit- learn is one of the most popular ML libraries for classical ML algorithms. It is built on top of two basic Python libraries, viz., NumPy and SciPy. Scikit- learn supports most of the supervised and unsupervised learning algorithms. Scikit- learn can also be used for data- mining and data- analysis, which makes it a great tool who is starting out with ML.

Theano:
Theano is a popular python library that is used to define, evaluate and optimize mathematical expressions involving multi- dimensional arrays in an efficient manner. It is achieved by optimizing the utilization of CPU and GPU. It is extensively used for unit- testing and self- verification to detect and diagnose different types of errors. Theano is a very powerful library that has been used in large- scale

computationally intensive scientific projects for a long time but is simple and approachable enough to be used by individuals for their own projects.

Keras:
Keras is a very popular Machine Learning library for Python. It is a high- level neural networks API capable of running on top of TensorFlow, CNTK, or Theano. It can run seamlessly on both CPU and GPU. Keras makes it really for ML beginners to build and design a Neural Network. One of the best thing about Keras is that it allows for easy and fast prototyping.

PyTorch:
PyTorch is a popular open- source Machine Learning library for Python based on Torch, which is an open-source Machine Learning library which is implemented in C with a wrapper in Lua. It has an extensive choice of tools and libraries that supports on Computer Vision, Natural Language Processing(NLP) and many more ML programs. It allows developers to perform computations on Tensors with GPU acceleration and also helps in creating computational graphs.

Pandas:
Pandas is a popular Python library for data analysis. It is not directly related to Machine Learning. As we know that the dataset must be prepared before training. In this case, Pandas comes handy as it was developed specifically for data extraction and preparation. It provides high – level data structures and wide variety tools for data analysis. It provides many inbuilt methods for grouping, combining and filtering data.

Matplotlib:
Matpoltlib is a very popular Python library for data visualization. Like Pandas, it is not directly related to Machine Learning. It particularly comes in handy when a programmer wants to visualize the patterns in the data. It is a 2D plotting library used for creating 2D graphs and plots. A module named pyplot makes it easy for programmers for plotting as it provides features to control line styles, font properties, formatting axes, etc. It provides various kinds of graphs and plots for data visualization, viz., histogram, error charts, bar chats, etc.

TensorFlow:
TensorFlow is a very popular open- source library for high performance numerical computation developed by the Google Brain team in Google. As the name suggests, Tensorflow is a framework that involves defining and running computations involving tensors. It can train and run deep neural networks that can be used to develop several AI applications. TensorFlow is widely used in the field of deep learning research and application.

V. ARCHITECTURAL DESIGN & DEVELOPMENT

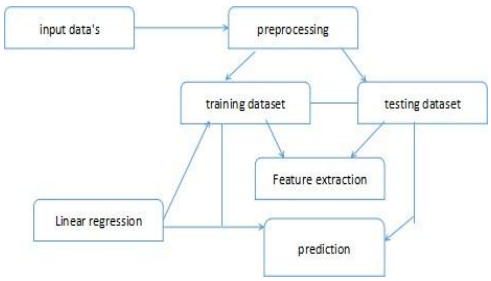
A. Data pre-processing:
The entries are present in the dataset. The null values are removed using df = df.dropna() where df is the data frame. The categorical attributes (Date, High, Low, Close, Adj value) are converted into numeric using Label Encoder. The date attribute is splitted into new attributes like total which can be used as feature for the model.

B. Feature selection:
Features selection is done which can be used to build the model. The attributes used for feature selection are Date, Price, Adj close, Forecast X coordinate , Y coordinate, Latitude , Longitude, Hour and month.

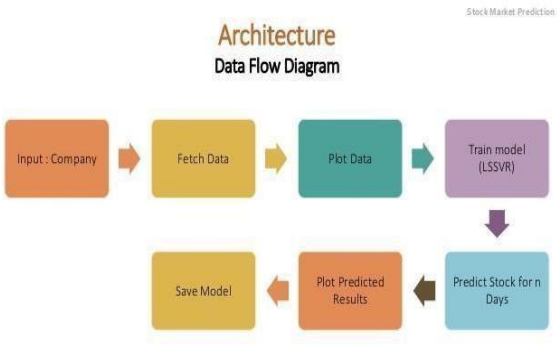
C. Building and training model:
After feature selection location and month attribute are used for training. The dataset is divided into pair of xtrain ,ytrain and xtest, y test. The algorithms model is imported form skleran. Building model is done using model. Fit (xtrain, ytrain). This phase would involve supervised classification methods like linear regression, Ensemble classifiers (like Adaboost, Random Forest Classifiers), etc.

VI. SYSTEM DESIGN

A. Architectural Design:
- The architectural design involves the following important three steps which is used the system design techniques.
- The system design architectural design steps are first step is the data pre-processing technique and second step is the feature selection and the next step is building and training a model.



B. Data Flow Diagram (DFD):



VII. RESULT

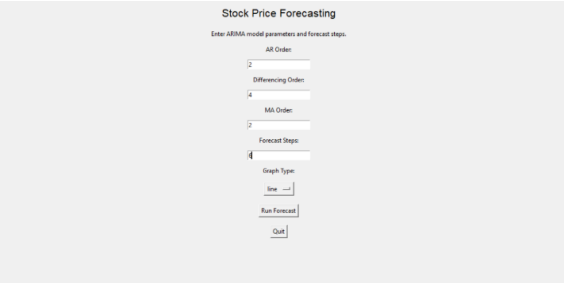


Fig 1: Output of stock price forecast

Fig 2: Result shown using line graph

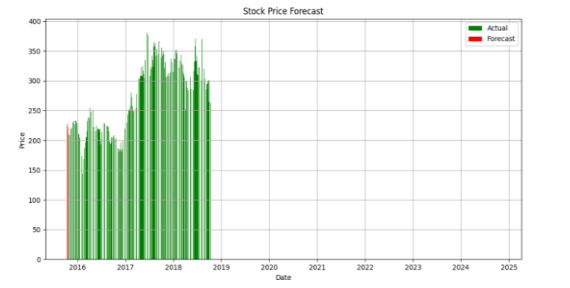


Fig 3: Result shown using bar graph

CONCLUSION

By measuring the accuracy of the Linear Regression algorithms, we found that the most suitable algorithm for predicting the market price of a stock based on various data points from the historical data. The algorithm will be a great asset for brokers and investors for investing money in the stock market since it is trained on a huge collection of historical data and has been chosen after being tested on a sample data. The project demonstrates the machine learning model to predict the stock value with more accuracy as compared to previously implemented machine learning models. Future scope of this project will involve adding more parameters and factors like the financial ratios, multiple instances, etc. The more the parameters are taken into account more will be the accuracy. The algorithms can also be applied for analyzing the contents of public comments and thus determine patterns/relationships between the customer and the corporate employees. Internet has a growing rate and the rate with which the data is being generating, it has become almost impossible for us to handle and take care of such data. Such an enormous amount of information is processing nowadays that it becomes difficult for us to study their behavior or to conclude anything from them, thus making it so hard to summarize it. Then comes Machine Learning algorithms that helps us in understanding such datasets. Technical and fundamental analysis have showed a little work in the experiments carried out. Machine learning algorithm was applied to various data sources of different companies. Report highlights that stock market is prone to differences. Report also concludes that predicting stock prices is extremely tough job. The main objective of this system is providing ways to heal the stock market. Our task is such that it can't be used for official model because of its limitedness. We have reached to a certain degree of accuracy by incorporating the limited number of parameters. Since stock market is highly fluctuating so to predict everything with great or large accuracy can't be taken into account. So our model that we have created has only depends on the selected number of parameters and their relationship with the share price.

REFERENCES

[1] Survey of stock market prediction using machine learning approach, Ashish Sharma ; Dinesh Bhuriya ; Upendra Singh 2017 International conference of Electronics, Communication and Aerospace Technology (ICECA).

[2] Short-term prediction for opening price of stock market based on self-adapting variant PSO-Elman neural network,Ze Zhang ; Yongjun Shen ; Guidong Zhang ;Yongqiang Song ; Yan Zhu, 2017 8th IEEE International Conference on Software Engineering and Service Science (ICSESS).

[3] Combining of random forest estimates using LSboost for stock market index prediction,Nonita Sharma ; Akanksha Juneja,2017 2nd International Conference for Convergence in Technology (I2CT).
[4] Using social media mining technology to assist in price prediction of stock market,Yaojun Wang ; Yaoqing Wang,2016 IEEE International Conference on Big Data Analysis (ICBDA).
[5] Stock market prediction using an improved training algorithm of neural network,Mustain Billah ; Sajjad Waheed ; Abu Hanifa,2016 2nd International Conference on Electrical, Computer & Telecommunication Engineering (ICECTE).
[6] S. D. Bekiros, "Sign Prediction and Volatility Dynamics With Hybrid Neurofuzzy Approaches," in IEEE Transactions on Neural Networks, vol. 22, no. 12, pp. 2353-2362, Dec. 2011.
[7] K. Raza, "Prediction of Stock Market performance by using machine learning techniques," 2017 International Conference on Innovations in Electrical Engineering and Computational Technologies (ICIEECT), Karachi, 2017.
[8] Z. Hu, J. Zhu and K. Tse, "Stocks market prediction using Support Vector Machine," 2013 6th International Conference on Information Management, Innovation Management and Industrial Engineering.
[9] Z. Wang, S. Ho and Z. Lin, "Stock Market Prediction Analysis by Incorporating Social and News Opinion and Sentiment," 2018 IEEE International Conference on Data Mining Workshops (ICDMW) Singapore 2018.
[10] S. Sarode, H. G. Tolani, P. Kak and C. S. Lifna, "Stock Price Prediction Using Machine Learning Techniques," 2019 International Conference on Intelligent Sustainable Systems (ICISS), Palladam, Tamilnadu, India 2019.