STOCK PREDICTION MODEL

A Project Work

Submitted in the partial fulfillment for the award of the degree of

BACHELOR OF ENGINEERING

IN

COMPUTER SCIENCE

ENGINEERING (H) IN BIG DATA

ANALYTICS

Submitted by:

DAKSH RAWAL (19BCS3849)

ARYAN GUPTA (19BCS3828)

PRAJWAL SRIVASTAV (19BCS3831)

Under the Supervision of:

DR. RAHUL RASTOGI

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING
APEX INSTITUE OF TECHNOLOGY

CHANDIGARH UNIVERSITY, GHARUAN, MOHALI - 140413,

PUNJAB

MAY & 202

DECLARATION

We, students of **'Bachelor of Engineering in Computer Science Engineering in Big Data Analytics'2023,** Department of Computer Science and Engineering, Apex Institute of Technology, Chandigarh University, Punjab, hereby declare that the work presented in this Project Work entitled '**Stock Prediction Model'** is the outcome of our own bona fide work and is correct to the best of our knowledge and this work has been undertaken taking care of Engineering Ethics. It contains no material previously published or written by another person nor material which has been accepted for the award of any other degree or diploma of the university or other institute of higher learning, except where due acknowledgment has been made in the text.

Daksh Rawal(19BCS3849)

Aryan Gupta(19BCS3828)

Prajwal(19BCS3831)

Date: 18-05-2022

Table of Contents

ABSTRACT

In this project we attempt to implement machine learning approach to predict stock prices. Machine learning is effectively implemented in forecasting stock prices. The objective is to predict the stock prices in order to make more informed and accurate investment decisions. We propose a stock price prediction system that integrates mathematical functions, machine learning, and other external factors for the purpose of achieving better stock prediction accuracy and issuing profitable trades. There are two types of stocks. You may know of intraday trading by the commonly used term "day trading." Intraday traders hold securities positions from at least one day to the next and often for several days to weeks or months. LSTMs are very powerful in sequence prediction problems because they're able to store past information. This is important in our case because the previous price of a stock is crucial in predicting its future price. While predicting the actual price of a stock is an uphill climb, we can build a model that will predict whether the price will go up or down.

ACKNOWLEDGEMENT

It is a matter of great pleasure to present this progress report on " Stock Market Prediction Model ". We are grateful to our project supervisor Dr. Rahul Rastogi, who has been of invaluable assistance to our project with his advice and suggestions. We are very much thankful to our sir for giving us the opportunity to do this project.

Group Members

Daksh Rawal (19BCS3849)

Aryan Gupta (19BCS3828)

Prajwal Srivastav(19BCS3831)

INTRODUCTION

The financial market is a dynamic and composite system were people can buy and sell currencies, stocks, equities and derivatives over virtual platforms supported by brokers. The stock market allows investors to own shares of public companies through trading either by exchange or over the counter markets.

This market has given investors the chance of gaining money and having a prosperous life through investing small initial amounts of money, low risk compared to the risk of opening new business or the need of high salary career. Stock markets are affected by many factors causing the uncertainty and high volatility in the market. Although humans can take orders 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 evaluate and control the performance of ATSs, the implementation of risk strategies and safety measures applied based on human judgements are required. Many factors are incorporated and considered when developing an ATS, for instance, trading strategy to be adopted, complex mathematical functions that reflect the state of a specific stock, machine learning algorithms that enable the prediction of the future stock value, and specific news related to the stock being analyzed.

Time-series prediction is a common technique widely used in many real-world applications such as weather forecasting and financial market prediction. It uses the continuous data in a period of time to predict the result in the next time unit. Many timeseries prediction algorithms have shown their effectiveness in practice. The most common algorithms now are based on Recurrent Neural Networks (RNN), as well as its special type - Long-short Term Memory (LSTM) and Gated Recurrent Unit (GRU). Stock market is a typical area that presents time-series data and many researchers study on it and proposed various models. In this project, LSTM model is used to predict the stock price.

Today we live and breathe data. Forecasting the stock exchange data is an important financial subject which involves an assumption that the fundamental information publicly available in the past has some predictive relationships to the future stock returns. Stock market forecasting contains uncovering the market trends, planning investment tactics, identifying the best time to purchase the stocks and which stocks to purchase. A stock exchange or equity business sector is a non-direct, non-parametric framework that is difficult to model with any sensible exactness. It is the mix of speculators who need to purchase or offer or hold a share at a specific time. Prediction will continue to be an exciting locale of research, making scientists in the analytics field always desiring to enhance the existing forecasting models. The motivation is that companies and individuals are empowered to make investment decisions to develop viable system about their future endeavors.

Stock price prediction is a heated topic in prediction study of financial area. The stock market is essentially a non-linear, nonparametric system that is extremely hard to model with any reasonable accuracy. Investors have been trying to find a way to predict stock prices and to find the right stocks and right timing to buy or sell. Most of the techniques used in technical analysis are highly subjective in nature and have been shown not to be statistically valid. Recently, data mining techniques and artificial intelligence techniques like decision trees, rough set approach, and artificial neural networks have been applied to this area. Data mining refers to extracting or mining knowledge from large data stores or sets. Some of its functionalities are the discovery of concept or class descriptions, associations and correlations, classification, prediction, clustering, trend analysis, outlier and deviation analysis, and similarity analysis. Data classification can be done in many different methods; one of those methods is the classification by using Decision

Tree. It is a graphical representation of all possible outcomes and the paths by which they may be reached.

The use of ANN in business environments has been increasing over the last few years. Excellent algorithm has been applied to predict stock price or index. Interest in neural networks has led to a considerable surge in research activities in the past decade. Artificial neural network models are based on the neural structure of the brain. The brain learns from experience and so do artificial neural networks. As a useful analytical tool, ANN is widely applied in analyzing the business data stored in database or data warehouse. Identifying customer behavior patterns and predicting stock price are emerging areas of neural network research and its application. Most of the companies have created new methods of evaluating financial data and investment decisions. Artificial Neural Networks are being used by most companies for improved forecasting capabilities in analysis of stock market. So, artificial neural network suits better than other models in predicting the stock market.

The idea of forecasting using neural network is to find an approximation of mapping between the input and output data through training. The trained neural network is then used to predict the values for the future This research work presents the use of artificial neural network as a forecasting tool for predicting the stock market price.

Mostly the approaches are in terms of fundamental approach and technical approach. For the long-term valuation fundamental approach is used. Every stock is having its own value that does not depend on the price of the stock that is known as Intrinsic value. The proposed model works through phases of data collection, feature processing, fuzzy logic mapping and stock value calculation. Fuzzy logic is used to map the quality as well as quantity valuation factors. The IF THEN rules are applied on the linguistic variable. The fuzzy model outcomes the stock value which is used to provide stock worth. The stock value is calculated by Dividend discount model. Accuracy of the system is 0.77. The results offer the backbone for the value and not the price.

Another method is DATA MINING. Decision making process for business can be risky. Corporate decision makers have to make decisions to protect company‟s benefit and lower the risk. In order to evaluate data mining approach on forecasting, a tool, called IFF, was developed for evaluating and simulating forecasts. Specifically data mining techniques‟ and simulation‟s ability to predict, evaluate and validate Port Industry forecasts is tested. Accuracy is calculated with data mining methods. Finally the probability of user‟s and simulation model‟s confidentiality is calculated. The results of the research indicate that data mining approach on forecasting and Monte Carlo method have the capability to forecast on Port industry and, if properly analyzed, can give accurate results for forecasts.

We study a multivariate Markov chain model for categorical data sequences to fuzzy time series. The proposed method gets higher average forecasting accuracy rate than some of the existing methods on temperature prediction.

The future stock returns have some predictive relationships with the publicly available information of present and historical stock market indices. ARIMA is a statistical model which is known to be efficient for time series forecasting especially for short-term prediction. In this paper, we propose a model for forecasting the stock market trends based on the technical analysis using historical stock market data and ARIMA model. This model will automate the process of direction of future stock price indices and provides assistance for financial specialists to choose the better timing for purchasing and/or selling of stocks. The results are shown in terms of visualizations using R programming language. The obtained results reveal that the ARIMA model has a strong potential for short-term prediction of stock market trend.

Hardware Requirements:

• RAM: 4 GB

• Storage: 500 GB

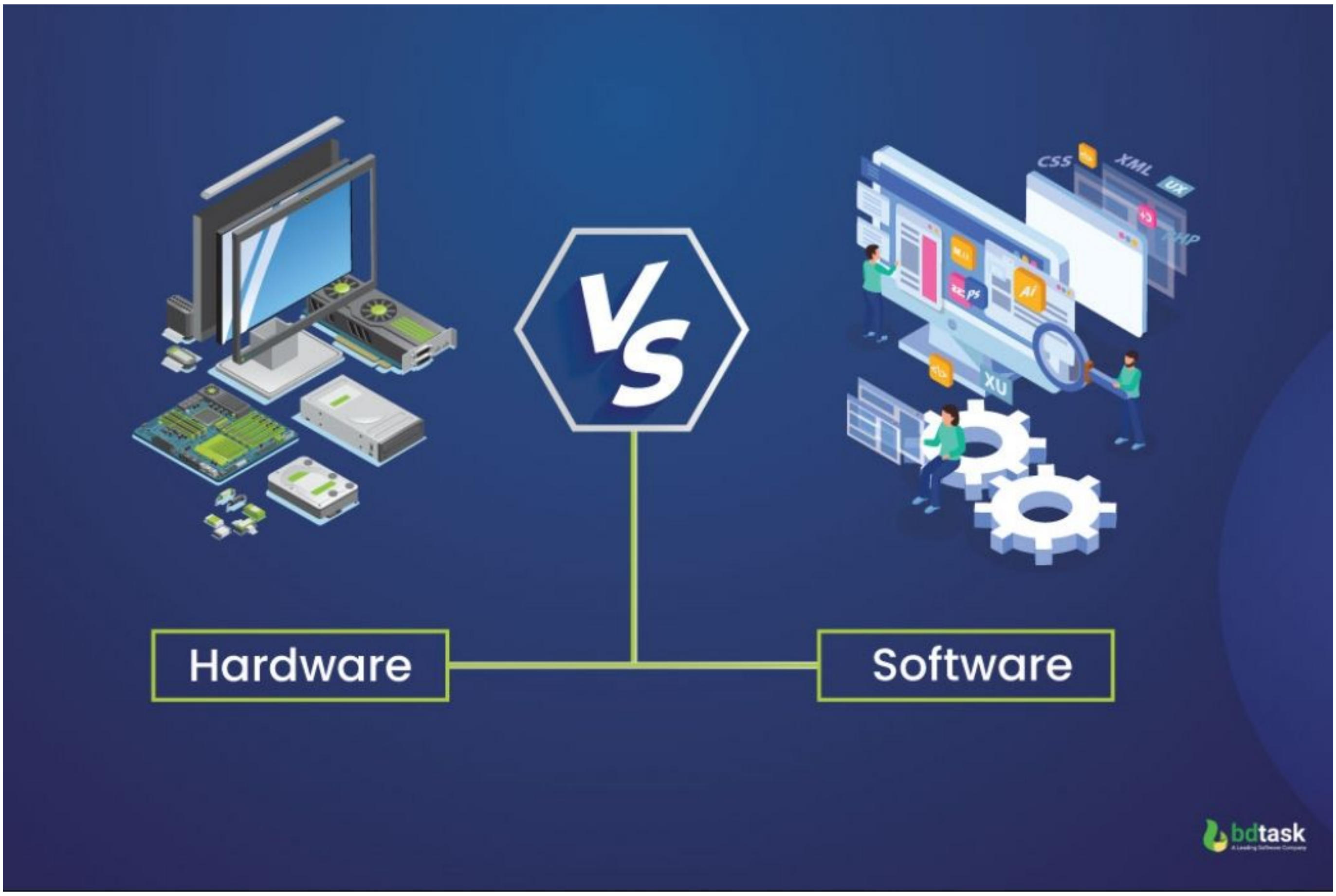
• CPU: 2 GHz or faster

• Architecture: 32-bit or 64-bit

Software Requirements:

• Python 3.5 in Google Colab is used for data pre-processing, model training and prediction.

• Operating System: windows 7 and above or Linux based OS or MAC OS



LITERATURE REVIEW

"What other people think" has always been an important piece of information for most of us during the decision-making process. The Internet and the Web have now (among other things) made it possible to find out about the opinions and experiences of those in the vast pool of people that are neither our personal acquaintances nor well-known professional critics — that is, people we have never heard of. And conversely, more and more people are making their opinions available to strangers via the Internet.

The interest that individual users show in online opinions about products and services, and the potential influence such opinions wield, is something that is driving force for this area of interest. And there are many challenges involved in this process which needs to be walked all over in order to attain proper outcomes out of them. In this survey we analyzed basic methodology that usually happens in this process and measures that are to be taken to overcome the challenges being faced.

Over the past two decades many important changes have taken place in the environment of financial markets. The development of powerful communication and trading facilities has enlarged the scope of selection for investors.

Forecasting stock return is an important financial subject that has attracted researchers' attention for many years. It involves an assumption that fundamental information publicly available in the past has some predictive
relationships to the future stock returns. In order to be able to extract such relationships from the available data, data mining techniques are new techniques that can be used to extract the knowledge from this data. For that reason, several researchers have focused on technical analysis and using advanced math and science. Extensive attention has been dedicated to the field of artificial intelligence and data mining techniques. Some models have been proposed and implemented using the above mentioned techniques, the authors of Tsang, P.M., Kwok, P., Choy, S.O., Kwan, R., Ng, S.C., Mak, J., Tsang, J., Koong, K., and Wong, T. made an empirical study on building a stock buying/selling alert system using back propagation neural networks (BPNN), their NN was codenamed NN5. The system was trained and tested with past price data from Hong Kong and Shanghai Banking Corporation Holdings over the period from January 2004 to December 2005. The empirical results showed that the implemented system was able to predict short-term price movement directions with accuracy about 74%.

The research by Wu, M.C., Lin, S.Y., and Lin, C.H., used decision tree technique to build on the work of Lin. where Lin tried to modify the filter rule that is to buy when the stock price rises k% above its past local low and sell when it falls k% from its past local high. The proposed modification to the filter rule was by combining three decision variables associated with fundamental analysis. An empirical test, using the stocks of electronics companies in Taiwan, showed Lin's method outperformed the filter rule. According to Wu, M.C., Lin, S.Y., and Lin, C.H.,, in Lin's work, the criteria for clustering trading points involved only the past information; the future information was not considered at all. The research by Wu, M.C., Lin, S.Y., and Lin, C.H., aimed to improve the filter rule and Lin's study by considering both the past and the future information in clustering the trading points. The researchers used the data of Taiwan stock market and that of NASDAQ to carry out empirical tests. Test results showed that the proposed method outperformed both Lin's method and the filter rule in the two stock markets.

The model of Wang, J.L., Chan, S.H. (2006) “Stock market trading rule discovery using two-layer bias decision tree”, applied the concept of serial topology and designed a new decision system, namely the two layer
bias decision tree, for stock price prediction. The methodology developed by the authors differs from other studies in two respects;
first, to reduce the classification error, the decision model was modified into a bias decision model.
Second, a two-layer bias decision tree is used to improve purchasing accuracy. The empirical results indicated that the presented decision model produced excellent purchasing accuracy, and it significantly outperformed than random purchase.

The authors Enke, D., Thawornwong, S. presented an approach that used data mining methods and neural networks for forecasting stock market returns. An attempt has been made in this study to investigate the predictive power of financial and economic variables by adopting the variable relevance analysis technique in machine learning for data mining. The authors examined the effectiveness of the neural network models used for level estimation and classification. The results showed that the trading strategies guided by the neural network classification models generate higher profits under the same risk exposure than those suggested by other strategies.

The research by Cao, Q., Leggio, K.B., and Schniederjans, M.J., was basically a comparison between the work of Fama and French’s model and the artificial neural networks in order to try to predict the stock prices in the Chinese market. The purpose of this study is to demonstrate the accuracy of ANN in predicting stock price movement for firms traded on the Shanghai Stock Exchange. In order to demonstrate the accuracy of ANN, the authors made a comparative analysis between Fama and French’s model and the predictive power of the univariate and multivariate neural network models. The results from this study indicated that artificial neural networks offer an opportunity for investors to improve their predictive power in selecting stocks, and more importantly, a simple univariate model appears to be more successful at predicting returns than a multivariate model.

Al-Haddad et al., presented a study that aimed to provide evidence of whether or not the corporate governance & performance indicators of the Jordanian industrial companies listed at Amman Stock Exchange (ASE) are affected by variables that were proposed and to provide the important indicators of the relationship of corporate governance & firms’ performance that can be used by the Jordanian industrial firms to solve the agency problem. The study random sample consists of (44) Jordanian industrial firms. The study founds a positive direct relationship between corporate governance and corporate performance.

Hajizadeh et al. provided an overview of application of data mining techniques such as decision tree, neural
network, association rules, and factor analysis and in stock markets. Prediction stock price or financial markets
has been one of the biggest challenges to the AI community. Various technical, fundamental, and statistical indicators have been proposed and used with varying results. Soni surveyed some recent literature in the
domain of machine learning techniques and artificial intelligence used to predict stock market movements.
Artificial Neural Networks (ANNs) are identified to be the dominant machine learning technique in stock market prediction area.

El-Baky et al., proposed a new approach for fast forecasting of stock market prices. The proposed approach
uses new high speed time delay neural networks (HSTDNNs). The authors used the MATLAB

tool to simulate results to confirm the theoretical computations of the approach.

V. Vamitha, M. Jeyanthi, S. Rajaram and T. Revathi,s research about Multivariate Markov Chain also gave a new approach in the stock market prediction systems. Since 1993 researchers proposed many methods for forecasting enrollments, Temperature prediction, stock price etc in time variant and time invariant first order, higher order, two factor and dual variables. In this paper, we propose a model to temperature prediction from correlated categorical data sequence obtained from similar source. We study a multivariate Markov chain model for categorical data sequences to fuzzy time series. The proposed method gets higher average forecasting accuracy rate than some of the existing methods on temperature prediction.

Anass Nahil prposed a new method on stock market prediction which will help many investors to invest their money in right time by which they will get more benefit in near future. Their proposed method was about support vector machine (SVM). It is a popular tool in time series forecasting for the capital investment industry. This machine learning technique which is based on a discriminative classifier algorithm, forecasts more accurately the financial data. By examining the stock price of 5 Moroccan banks, the experiment shows that the SVM can perform better when we add the global evolution of the market to the independent variables. To express the global evolution of the market, three indices of the Casablanca Stock Exchange are used : MASI, MADEX and Banks Sector Index.

Narendra Pahuja, Abhishek Oturkar, Kailash Sharma, Jatin Shrivastava, Dimple Bohra’s ARIMA model made a huge change in stock market prediction. Over the years it is observed that stock market data is nonlinear, chaotic & dynamic. This paper is going to present a predictive model for prices of the stocks with the help of ARIMA model. The stock data which is published from the Bombay Stock Exchange (BSE) & National Stock Exchange (NSE) has been used with the model developed for the prediction of stock price. From the results which are obtained, we come to the conclusion that for short-term prediction the ARIMA model has a great potential & also it shows competence with the already present methods for stock price prediction

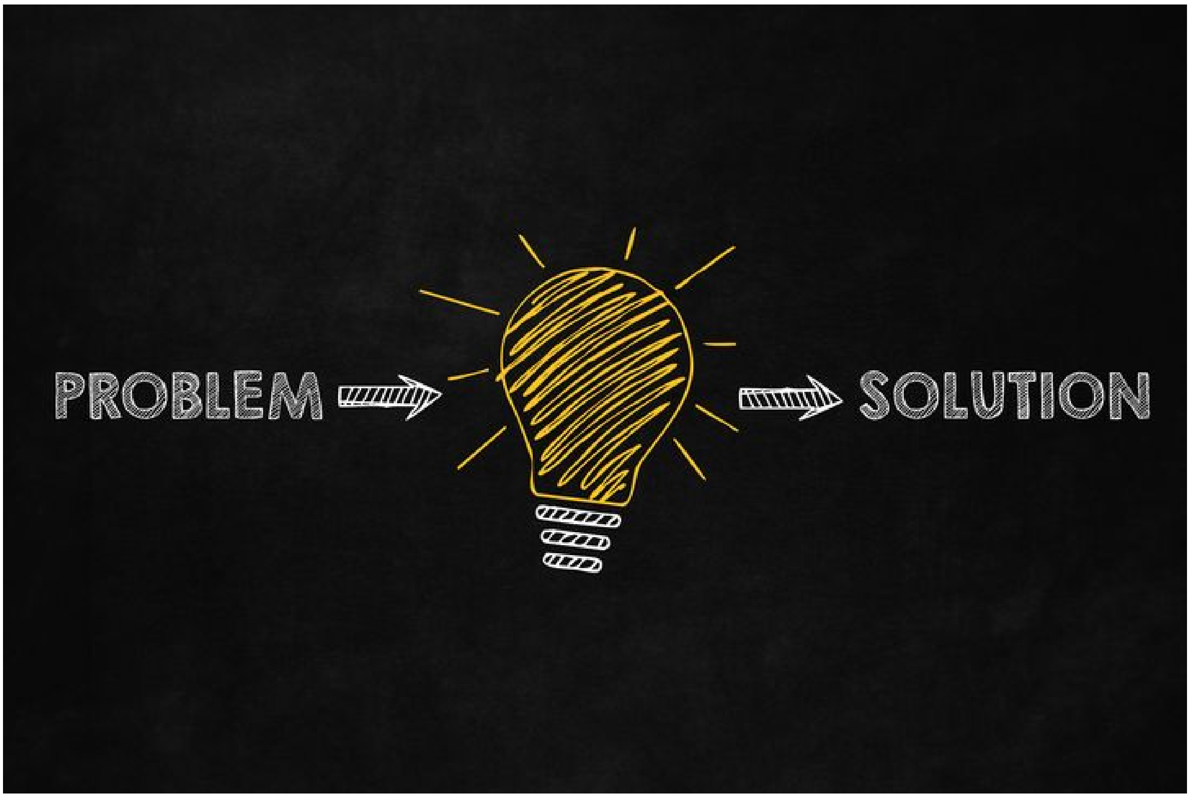
Mahantesh Angadi, Amogh Kulkarni Sai proposed the method of stock market prediction by using Data Mining Techniques with R. In this work, we explore recurrent neural networks with character-level language model pre-training for both intraday and interday stock market forecasting. In terms of predicting directional changes in the Standard & Poor’s 500 index, both for individual companies and the overall index, we show that this technique is competitive with other state-of-the-art approaches.

Nghiem Van Tinh,Nguyen Thi Thu Hien Nguyen Tien Duy proposed the method by using k-means clustering algorithm. Most of the fuzzy forecasting methods based on fuzzy time series used the static length of intervals, i.e., the same length of intervals. The drawback of the static length of intervals is that the historical data are roughly put into intervals, even if the variance of the historical data is not high. In this paper, we present a new method for forecasting enrolments based on Fuzzy Time Series and K-Mean clustering(FTS-KM). To verify the effectiveness of the proposed model, the empirical data for the enrolments of the University of Alabama are illustrated, and the experimental results show that the proposed model outperforms those of previous some forecasting models with various orders and different interval lengths.

PROBLEM FORMULATION

During software development, clones can occur in software intentionally or unintentionally. Developers tend to clone fragments of software during development to save efforts and expedite the development process of the Stock Prediction Model

From the Literature review, it is observed that studies highlight the need of efficient and scalable approach for detecting code clones having software vulnerability. The existing techniques are not able to detect all types of vulnerable code clones. Different approaches suffer from high false negative rate and not scalable to large software systems due to high time complexity. Time Series forecasting & modelling plays an important role in data analysis. Time series analysis is a specialized branch of statistics used extensively in fields such as Econometrics & Operation Research. Time Series is being widely used in analytics & data science. Stock prices are volatile in nature and price depends on various factors. The main aim of this project is to predict stock prices using Long short-term memory (LSTM)



OBJECTIVE

The main objective of this study is to study about different methodology's and get a stock market prediction tool to obtain more accurate stock prediction price and to evaluate them with some performance measures. This study can be used to reduce the error proportion in predicting the future stock prices.

It increases the chances for the investors to predict the prices more accurately by reducing error percentage and thus gain benefits in share markets.

After getting the idea about different methods of stock market forecasting techniques we can understand that by using which methods we will get more accurate results. Then we will be able to reduce the amount of error by which investors can invest their valuable money in stock market at a right time.

EXISTING METHODS

Stock Market Prediction Using Machine Learning

In the finance world stock trading is one of the most important activities. Stock market prediction is an act of trying to determine 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 time series analysis is used by the most of the stockbrokers while making the stock predictions. The programming language is used to predict the stock market using machine learning is Python. In this paper we propose a Machine Learning (ML) approach that will be trained from the available stocks data and gain intelligence and then uses the acquired knowledge for an accurate prediction. In this context this study uses a machine learning technique called Support Vector Machine (SVM) to predict stock prices for the large and small capitalizations and in the three different markets, employing prices with both daily and up-to-the-minute frequencies.

Forecasting the Stock Market Index Using Artificial Intelligence Techniques

The weak form of Efficient Market hypothesis (EMH) states that
it is impossible to forecast the future price of an asset based on the information contained in the historical prices of an asset. This means that the market behaves as a random walk and as a result makes forecasting impossible. Furthermore, financial forecasting is a difficult task due to the intrinsic complexity of the financial system. The objective of this work was to use artificial intelligence (AI) techniques to model and predict the future price of a stock market index. Three artificial intelligence techniques, namely, neural networks (NN), support vector machines and neuro-fuzzy systems are implemented in forecasting the future price of a stock market index based on its historical price information. Artificial intelligence techniques have the ability to take into consideration financial system complexities and they are used as financial time series forecasting tools.
Two techniques are used to benchmark the AI techniques, namely,
Autoregressive Moving Average (ARMA) which is linear modelling technique and random walk (RW) technique. The experimentation was performed on data obtained from the Johannesburg Stock Exchange. The data used was a series of past closing prices of the All

Share Index. The results showed that the three techniques have the ability to predict the future price of the Index with an acceptable accuracy. All three artificial intelligence techniques outperformed the linear model. However, the random walk method out performed all the other techniques. These techniques show an ability to predict the future price however, because of the transaction costs of trading in the market, it is not possible to show that the three techniques can disprove the weak form of market efficiency. The results show that the ranking of performances support vector machines, neuro-fuzzy systems, multilayer perceptron neural networks is dependent on the accuracy measure used.

Indian stock market prediction using artificial neural networks on tick data

A stock market is a platform for trading of a
company’s stocks and derivatives at an agreed price. Supply and demand of shares drive the stock markets. In any country stock market is one of the most emerging sectors. Nowadays, many people are indirectly or directly related to this sector. Therefore, it becomes essential to know about market trends. Thus, with the development of the stock market, people are interested in forecasting stock price. But, due to dynamic nature and liable to quick changes in stock price, prediction of the stock price becomes a challenging task. Stock m Prior work has proposed effective methods to learn event representations that can capture syntactic and semantic information over text corpus, demonstrating their effectiveness for downstream tasks such as script event prediction. On the other hand, events extracted from raw texts lacks of common-sense knowledge, such as 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 address 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 market prediction, show that our model obtains much 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). As the technology is increasing, stock traders are moving towards to use Intelligent Trading Systems rather than fundamental analysis for predicting prices of stocks, which helps them to take immediate investment decisions. One of the main aims of a trader is to predict the stock price such that he can sell it before its value decline, or buy the stock before the price rises. The efficient market hypothesis states that it is not possible to predict stock prices and that stock behaves in the random walk. It seems to be very difficult to replace the professionalism of an experienced trader for predicting the stock price. But because of the availability of a remarkable amount of data and technological advancements we can 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.

The Stock Market and Investment

Increasing integration of European financial markets is likely to result in even stronger correlation between equity prices in different European countries. This process can also lead to convergence in economic development across European countries if developments in stock markets influence real economic components, such as investment and consumption. Indeed, our vector autoregressive models suggest that the positive correlation between changes equity prices and investment is, in general, significant. Hence,
6
monetary authorities should monitor reactions of share prices to monetary policy and their effects on the business cycle.

Automated Stock Price Prediction Using Machine Learning

Traditionally and in order to predict market movement, investors used to analyse the stock prices and stock indicators in addition to the news related to these stocks. Hence, the importance of news on the stock price movement. Most of the previous work in 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. In this work, we propose an automated trading system that integrates mathematical functions, machine learning, and other external factors such as news' sentiments for the purpose of achieving better stock prediction accuracy and issuing profitable trades. Particularly, we aim to determine the price or the trend of a certain stock for the coming end-of-day considering the first several trading hours of the day. To achieve 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 highest accuracy (82.91%) of which was achieved using SVM for Apple Inc. (AAPL) stock.

Stock Price Correlation Coefficient Prediction with ARIMALSTM Hybrid Model

We apply LSTM
recurrent neural networks (RNN) in predicting the stock price correlation coefficient of two individual stocks. RNN's are competent in understanding temporal dependencies. The use of LSTM cells further enhances its long-term predictive properties. To encompass both linearity and nonlinearity in the model, we adopt the ARIMA model as well. The ARIMA model filters linear tendencies in 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 such as the full historical model, constant correlation model, single-index model and the multi-group model. In our empirical study, the predictive ability of the ARIMA-LSTM model turned out superior to all other financial models by a significant scale. Our work implies that it is worth considering the ARIMALSTM model to forecast correlation coefficient for portfolio optimization.

Event Representation Learning Enhanced with External Common-sense Knowledge

Prior work has proposed effective methods to learn event representations that can capture syntactic and semantic information over text corpus, demonstrating their effectiveness for downstream tasks such as script event prediction. On the other hand, events extracted from raw texts lacks of common-sense knowledge, such as 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 address 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 market prediction, show that our model obtains much 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 market.

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

We employ both random forests and LSTM networks (more precisely CuDNNLSTM) as training methodologies to analyse their effectiveness in forecasting outof-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 respect to the closing prices, but also with respect 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, prior to transaction costs, of 0.64% using LSTM networks, and 0.54% using random forests. Hence, we outperform the singlefeature setting in Fischer & Krauss (2018) and Krauss et al. (2017) consisting only of the daily returns with respect to the closing prices, having corresponding daily returns of 0 .41% and of 0 .39% with respect to LSTM and random forests, respectively. 1 Keywords: Random forest, LSTM, Forecasting, Statistical Arbitrage, Machine learning, Intraday trading.

A Deep Reinforcement Learning Library for Automated Stock Trading in Quantitative Finance

As deep reinforcement learning (DRL) has been recognized as an effective approach in quantitative finance, getting hands-on experiences is attractive to beginners. However, to train a practical DRL trading agent that decides where to trade, at what price, and what quantity involves error-prone and arduous development and debugging. In this paper, we introduce a DRL library FinRL that facilitates beginners to expose themselves to quantitative finance and to develop their own stock trading strategies. Along with easily-reproducible tutorials, FinRL library allows users to streamline their own developments and to compare with existing schemes easily. Within FinRL, virtual environments are configured with stock market datasets, trading agents are trained with neural networks, and extensive back testing is analysed via trading performance. Moreover, it incorporates important trading constraints such as transaction cost, market liquidity and the investor's degree of risk-aversion. FinRL is featured with completeness, hands-on tutorial and reproducibility that favors beginners: (i) at multiple levels of time granularity, FinRL simulates trading environments across various stock markets, including NASDAQ-100, DJIA, S&P 500, HSI, SSE 50, and CSI 300; (ii) organized in a layered architecture with modular structure, FinRL provides finetuned state-of-the-art DRL algorithms (DQN, DDPG, PPO, SAC, A2C, TD3, etc.), commonly used reward functions and standard evaluation baselines to alleviate the debugging workloads and promote the reproducibility, and (iii) being highly extendable, FinRL reserves a complete set of user-import interfaces. Furthermore, we incorporated three application demonstrations, namely single stock trading, multiple stock trading, and portfolio allocation.

An innovative neural network approach for stock market Prediction

To develop an innovative neural network approach to achieve better stock market predictions. Data were obtained from the live stock market for real-time and off-line analysis and results of visualizations and analytics to demonstrate Internet of Multimedia of Things for stock analysis. To study the influence of market characteristics on stock prices, traditional neural network algorithms may incorrectly predict the stock market, since the initial weight of the random selection problem can be easily prone to incorrect predictions.
Based on the development of word vector in deep learning, we demonstrate the concept of "stock vector." The input is no longer a single index or single stock index, but multi-stock high-dimensional historical data. We propose the deep long short-term memory neural network (LSTM) with embedded layer and the long short-term memory neural network with automatic encoder to predict the stock market. In these two models, we use the embedded layer and the automatic encoder, respectively, to vectorize the data, in a bid to forecast the stock via long short-term memory neural network. The experimental results show that the deep LSTM with embedded layer is better. Specifically, the accuracy of two models is 57.2 and 56.9%, respectively, for the Shanghai A-shares composite index. Furthermore, they are 52.4 and 52.5%, respectively, for individual stocks. We demonstrate research contributions in IMMT for neural network-based financial analysis.

An Intelligent Technique for Stock Market Prediction

A stock market is a loose network of economic transactions between buyers and sellers based on stocks also known as shares.

In stock markets, stocks represent the ownership claims on businesses. These may include securities listed on a stock exchange as well as those only traded privately. A stock exchange is a place where brokers can buy and/or sell stocks, bonds, and other securities. Stock market is a very vulnerable place for investment due to its volatile nature. In the near past, we faced huge financial problems due to huge drop in price of shares in stock markets worldwide. This phenomenon brought a heavy toll on the international as well as on our national financial structure. Many people lost their last savings of money on the stock market. In 2010–2011 financial year, Bangladeshi stock market faced massive collapse [1]. This phenomenon can be brought under control especially by strict monitoring and instance stock market analysis. If we can analyse stock market correctly in time, it can become a field of large profit and may become comparatively less vulnerable for the investors. Stock market is all about prediction and rapid decision making about investment, which cannot be done without thorough analysis of the market. If we can predict the stock market by analysing historical data properly, we can avoid the consequences of serious market collapse and to be able to take necessary steps to make market immune to such situations.

LSTM Fully Convolutional Networks for Time Series Classification

With the proposed models, we achieve a potent improvement in the current state-of-the-art for time series classification using deep neural networks. Our baseline models, with and without fine-tuning, are trainable end-to-end with nominal preprocessing and are able to achieve significantly improved performance.

LSTM-FCNs are able to augment FCN models, appreciably increasing their performance with a nominal increase in the number of parameters. An LSTM-FCNs provide one with the ability to visually inspect the decision process of the LSTM RNN and provide a strong baseline on their own. Fine-tuning can be applied as a general procedure to a model to further elevate its performance.

The strong increase in performance in comparison to the FCN models shows that LSTM RNNs can beneficially supplement the performance of FCN modules for time series classification. An overall analysis of the performance of our model is provided and compared to other techniques.

There is further research to be done on understanding why the attention LSTM cell is unsuccessful in matching the performance of the general LSTM cell on some of the datasets. Furthermore, an extension of the proposed models to multivariate time series is elementary but has not been explored in this work.

Learning Long term Dependencies with Gradient Descent is difficult

Recurrent networks are very powerful in their ability to represent context, often outperforming static network. But the factor off gradient descent of an error criterion may be inadequate to train them for a task involving long-term dependencies. It has been found that the system would not be robust to input noise or would not be efficiently trainable by gradient descent when the long-term context is required. The theoretical result presented in this paper holds for any error criterion and not only from mean square error.

it can also be seen that the gradient either vanishes or the system is not robust to input noise. The other imp factor to note is that the related problems of vanishing gradient may occur in deep feed-forward networks. The result presented in this paper does not mean that it is impossible to train a recurrent neural network on a particular task. It says that the gradient becomes increasingly inefficient when the temporal span of the dependencies increases. So at one point in time, it is evident that it becomes obsolete.

Improving N Calculation of the RSI Financial Indicator Using Neural Networks

There has been growing interest in Trading Decision Support Systems in recent years. In spite of its volatility, it is not entirely random, instead, it is nonlinear and dynamic or highly complicated and volatile. Stock movement is affected by the mixture of two types of factors: determinant (e.g. gradual strength change between buying side and selling side) and random (e.g. emergent affairs or daily operation variations).

There are 3 modules that are talked about in this research paper. The Neural Network Module is responsible for providing the N values that are used to calculate RSI and decide if an investor should invest in a certain company.

Trading system Module analyzes the result given by neural network module. When a query is formulated to the system, it takes the actual values of the market and builds a query to the neural network. If RSI value is higher than 70 the decision that trading system return is a sell signal. If RSI value is lower than 30 the decision that trading system return is a buy signal.

The heuristic module is in charge of managing the different formulas that provide the heuristic used to generate the optimal values for RSI indicator.

Stock Trend Prediction Using Simple Moving Average Supported by News Classification

The simple moving average is one of many times series analysis technique. Time series analysis is a method of timely structured data processing to find statistics or important characteristics for many reasons. The simple moving average shows stock trend by calculating the average value of stock prices on specific duration. The prices that are used are closing prices at the end of the day. This technique can avoid noises and therefore smooth the trend movement.

The main objective of financial news classification is to classify and calculate each news' sentiment value. The positive news is marked by sentiment value which is greater than 0, while negative news is marked by less than 0 sentiment value. If there are news having 0 sentiment value, they will be omitted as their neutralism does not affect the stock trend.

Machine learning using artificial neural network algorithm is used to predict a stock trend. The artificial neural network uses three features along with one label. The three features are a simple moving average distance which is a subtraction of long-term and short-term simple moving average, the total value of positive sentiment value for one-day news, and the total value of negative sentiment value for one-day news. Stock trend label is used and classified as uptrend and downtrend. On one hand, learning component is done by a background process. On the other hand, prediction component is foreground process which is seen and interact with the user.

VISUALIZING AND UNDERSTANDING RECURRENT NETWORKS

Character-level language models have been used as an interpretable test bed for analyzing the predictions, representations training dynamics, and error types present in Recurrent Neural Networks. In particular, the qualitative visualization experiments, cell activation statistics and comparisons to finite horizon n-gram models demonstrate that these networks learn powerfully, and often interpretable long-range interactions on real-world data.

The error analysis broke down cross entropy loss into several interpretable categories and allowed us to illuminate the sources of remaining limitations and to suggest further areas for study.

In particular, it was found that scaling up the model almost entirely eliminates errors in the n-gram category, which provides some evidence that further architectural innovations may be needed to address the remaining errors.

LSTM: A Search Space Odyssey

This paper reports the results of a large-scale study on variants of the LSTM architecture. We conclude that the most commonly used LSTM architecture (vanilla LSTM) performs reasonably well on various datasets. None of the eight investigated modifications significantly improves performance.

The forget gate and the output activation function are the most critical components of the LSTM block. Removing any of them significantly impairs performance. We hypothesize that the output activation function is needed to prevent the unbounded cell state to propagate through the network and destabilize learning. This would explain why the LSTM variant GRU can perform reasonably well without it: its cell state is bounded because of the coupling of input and forget gate.

The analysis of hyperparameter interactions revealed no apparent structure. Furthermore, even the highest measured interaction (between learning rate and network size) is quite small. This implies that for practical purposes the hyperparameters can be treated as approximately independent. In particular, the learning rate can be tuned first using a fairly small network, thus saving a lot of experimentation time.

Neural networks can be tricky to use for many practitioners compared to other methods whose properties are already well understood. This has remained a hurdle for newcomers to the field since a lot of practical choices are based on the intuitions of experts, as well as experiences gained over time. With this study, we have attempted to back some of these intuitions with experimental results. We have also presented new insights, both on architecture selection and hyperparameter tuning for LSTM networks which have emerged as the method of choice for solving complex sequence learning problems. In future work, we plan to explore more complex modifications of the LSTM architecture.

The difficulty of training recurrent neural networks

We provided different perspectives through which one can gain more insight into the exploding and vanishing gradients issue. We put forward a hypothesis stating that when gradients explode we have a cliff-like structure in the error surface and devise a simple solution based on this hypothesis, clipping the norm of the exploded gradients.

The effectiveness of our proposed solutions provides some indirect empirical evidence towards the validity of our hypothesis, though further investigations are required. In order to deal with the vanishing gradient problem, we use a regularization term that forces the error signal not to vanish as it travels back in time.

This regularization term forces the Jacobian matrices ∂xi ∂xi−1 to preserve norm only in relevant directions. In practice, we show that these solutions improve the performance of RNNs on the pathological synthetic datasets considered, polyphonic music prediction and language modeling.

Deep Sparse Rectifier Neural Networks

Sparsity and neurons operating mostly in a linear regime can be brought together in more biologically plausible deep neural networks. Rectifier units help to bridge the gap between unsupervised pre-training and no pre-training, which suggests that they may help in finding better minima during training.

This finding has been verified for four image classification datasets of different scales and all this in spite of their inherent problems, such as zeros in the gradient, or ill-conditioning of the parameterization. Rather sparse networks are obtained (from 50 to 80% sparsity for the best generalizing models, whereas the brain is hypothesized to have 95% to 99% sparsity), which may explain some of the benefits of using rectifiers.

Rectifier activation functions have shown to be remarkably adapted to sentiment analysis, a text-based task with a very large degree of data sparsity. This promising result tends to indicate that deep sparse rectifier networks are not only beneficial to image classification tasks and might yield powerful text mining tools in the future.

Furthermore, rectifier activation functions have shown to be remarkably adapted to sentiment analysis, a text-based task with a very large degree of data sparsity. This promising result tends to indicate that deep sparse rectifier networks are not only beneficial to image classification tasks and might yield powerful text mining tools in the future.

Stock Market Trends Prediction after Earning Release

As known to the public, the stock market is known as a chaotic system and it has been proved that even model built with empirical key features could still result in low accuracy. In our work, we tried to limit our scope to earnings release day, and it turned out that we could build models achieving around 70% prediction accuracy.

To build the model, We take financial statistics collected from company's quarterly earnings report, market surprise due to consensus expectations in terms of digital data, and sentiment analysis of relevant articles from mainstream media of financial professionals as two sets of input features, and make stock market movements prediction in after-hour period and trend in the day after the release day. SVM and LWLR model outperforms other models as shown by experiments, as they control the correlation among data, which was discussed in Section VI.

However, due to the limited number of company choices, we have small data size (300 samples) which could lead to high bias and overfitting. The stock price is not only affected by certain financial features, consensus news, but also company direction and future business guidance, which are difficult to be digitized.

Predicting Stock Trends through Technical Analysis and Nearest Neighbor Classification

Tech Examination is built on the philosophies of the Dow Theory and practices the past of prices to forecast upcoming actions. The method used in tech examination can be enclosed as an outline credit problem, where the ideas are resulting from the history of values and the output is an estimate of the price or an estimate of the prices trend.

The most significant evidence of this type of examination is that the marketplace action reductions everything. It means the specialist believes that anything that can perhaps affect the marketplace is already reflected in the prices, as well as that all the new evidence will be directly reflected in those prices. As an import, all the technician needs is to analyze the past of prices.

The main gears of the tech examination are the capacity and price charts. Based on the data of values and size the tech pointers are built. Tech pointers are math formulations that are applied to the price or volume statistics of a safekeeping for demonstrating some aspect of the association of those amounts.

METHODOLOGY

The following methodology will be followed to achieve the objectives defined for proposed research work:

SYSTEM DESIGN AND ARCHITECTURE

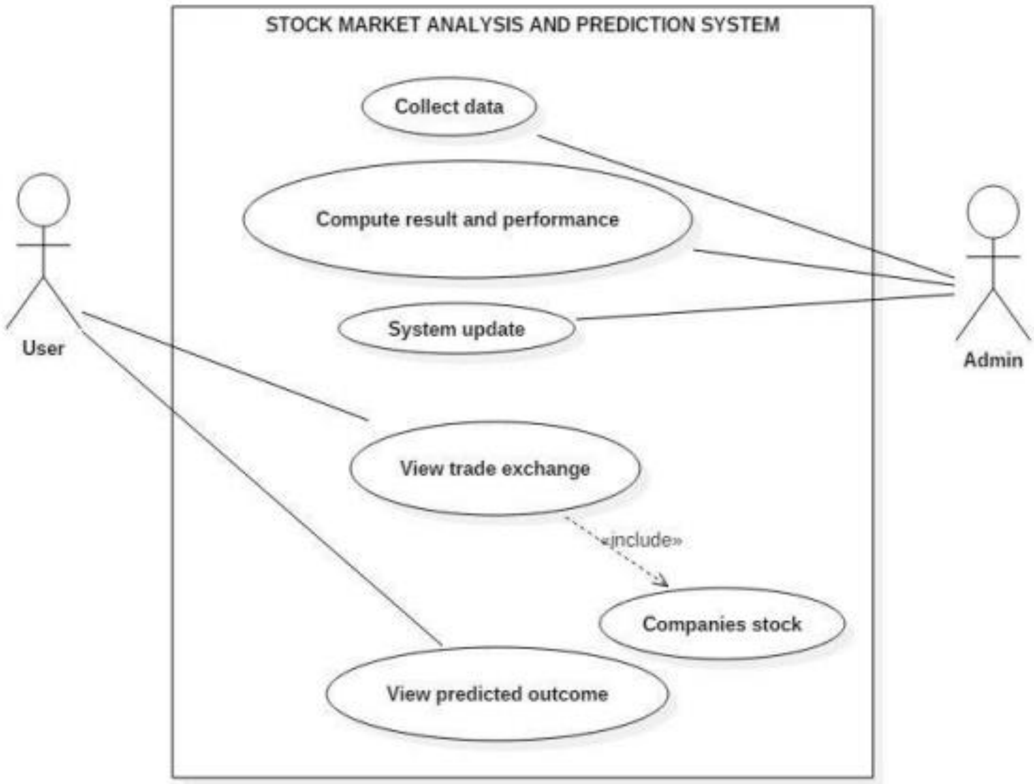


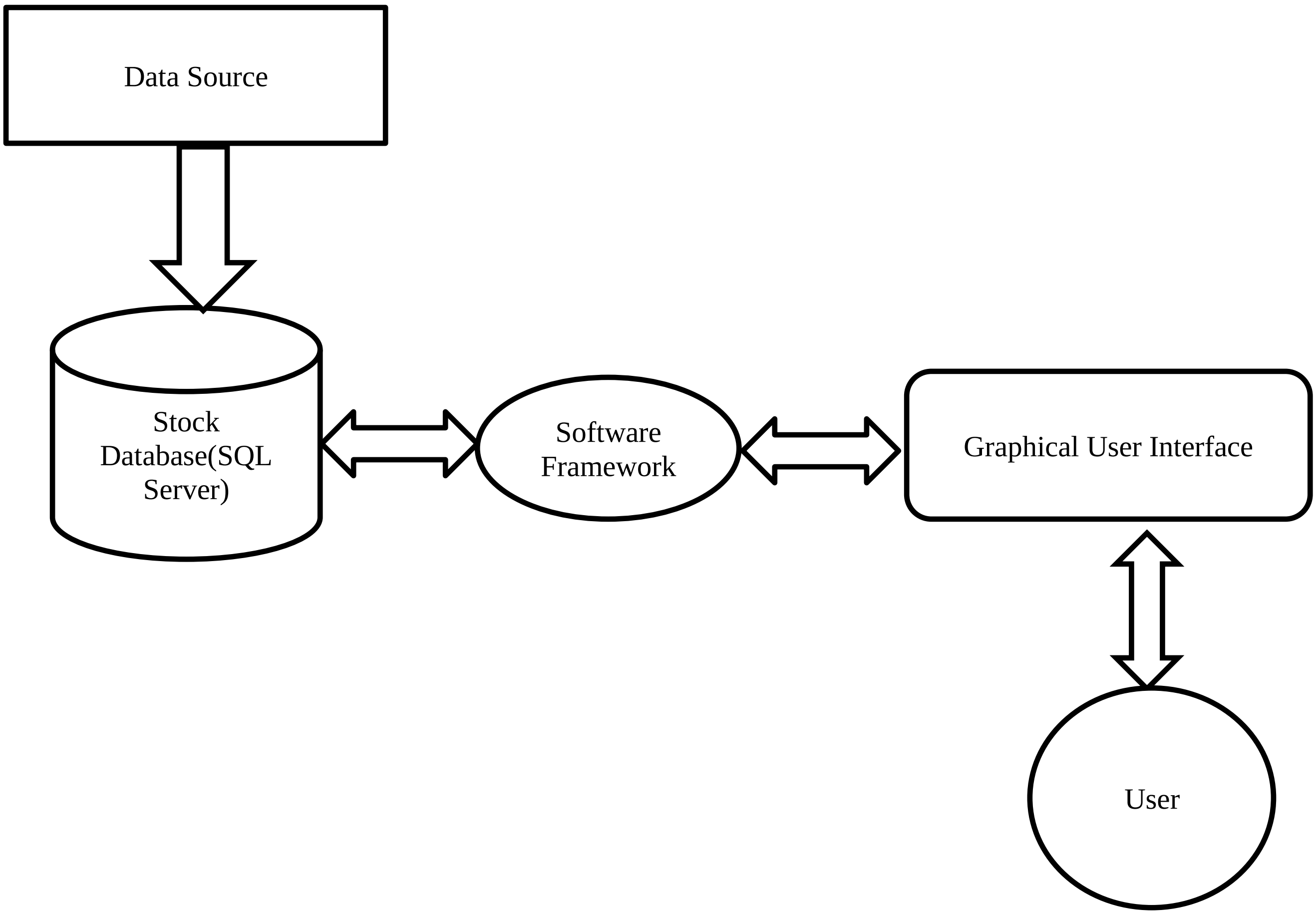
Figure 4.1: Use Case Diagram for the system

Use case index

Use case ID	Use case name	Primary actor	scope	complexity	priority
1	Collect data	admin	in	high	1
2	Compute result and prepare	admin	in	high	1
3	System update	admin	in	high	1
4	View trade exchange	user	in	medium	2
5	Company stock	user	in	medium	2
6	View predicted outcome	user	in	high	1

Use case description:
Use case ID:1
Use case name: Collect data
Description: Every required data will be available in Nepal stock exchange. Admin will be able to collect the data for system.
Use case ID:2
Use case name: Compute result and performance
Description: Prediction result will be handled and generated by admin. The system will be built, through which the result of prediction and system performance will be analyzed.
Use case ID: 3
Use case name: System update
Description: With the change of market and technology regular update of system is required. Beside there the predict result of stock exchange and their actual price will be updated by admin in regular basis.
Use case ID: 4
Use case name: View traded exchange
Description: Company trading which is held at NEPSE can be viewed by user.
Use Case ID: 5
Use Case Name: Company Stock
Description: It is extended feature of view traded exchange. This includes the stock value of particular company.
Use Case ID: 6
Use Case Name: View predicted outcome
Description: This use case is must important in whole project. The key feature of this project is to predict the stock value of hydropower companies. Thus, this will be available in user interface and viewer can observe them.

SYSTEM FLOW DIAGRAM

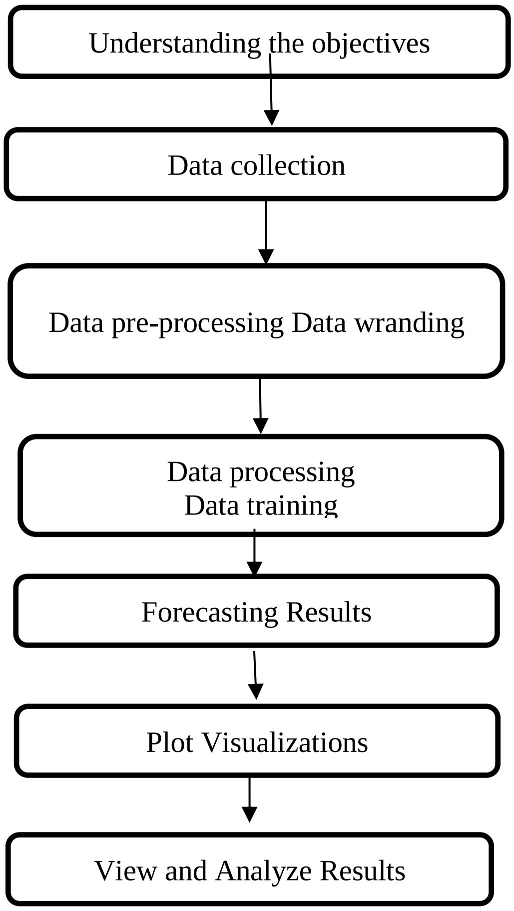


IMPLEMENTATION STEPS

System architecture is a model that defines the behavior of a system in the conceptual model. The huge systems are decomposed into subordinate systems to provide similar set of services. The beginning layout strategy of perceiving these sub-systems and building up a structure for sub-systems control and cooperation is called architecture design. As shown above, Fig. 4.1 includes seven major steps to implement the system and each step is explained below.

A. Understanding the Objective

The first step in developing a project is to understand the objective which involves an understanding of the intent and essentials of a system. This comprehension is used as a problem description and a preparatory system to accomplish the expectations. The objective of our project is neither to build a system that makes billions nor to waste billions too. But the objective is to develop a system that finds the direction of change of stock price indices based on the co-relations between stock prices and help the investors in the stock market in taking a decision whether to buy/sell/hold a stock by providing the results in-terms of visualizations.



B. Data Collection

Once the understanding of the objective is over, the next step is to collect the data. Data collection involves the understanding of initial observations of the data to identify the useful subsets from hypotheses of the hidden information. We use the data from Google finance.

C. Data Pre-processing: Data Wrangling

The data pre-processing stage involves all the activities to prepare the final dataset from the preparatory raw information. The data preparation tasks can be performed several times as there is no specific order. These tasks include the selection of a record, table, attribute and cleaning of data for modeling tools.

D. Data Processing: Data Training

In technical analysis investors use the auto regressive and moving average models to forecast the stock trends. Major steps involved here are **identification, parameter estimation and forecasting**. These steps are repeated until an appropriate model is identified for prediction.

E. Forecasting Results

The process of making predictions of the future by relying upon the past and present data is known as forecasting. Various prediction techniques are used by the stock analysts to evaluate the future stock trends value. Prediction also offers a significant standard for organizations that have a long-term perception of actions. We use ‘forecast’ package for predicting the future stock trends based on the analysis of past trends. This ‘forecast’ package provides a number of forecasting functions for displaying the time series predictions along with exponential smoothing and space models.

F. Plot Visualizations

Data visualization is a graphical representation of the numerical data. After forecasting the stock market trends we visualize the results for short-term investment assistance in-terms of line charts, candlesticks charts, bar charts, and histograms.

G. View and Analyse Results

Once after plotting the results in-terms of visualizations we can find out the correlations to get the short-term predictions. In the next section we provide some of the screen shots by which the investor can analyze and predict the future stock trends of a particular company at a specific time period. So the investors in the stock market can use this as assistance to sell/buy/hold a share.

Surveyed Stock Markets And Related Data Sets

The list of stock markets authors have obtained their data for training and testing of their perspective models is shown in Table 1. Surveyed articles focus on forecasting returns of a single stock market index or of multiple stock market indexes. However, several studies concentrate in forecasting returns of a single stock or multiple stocks (Ajith, Baikunth, & Mahanti, 2003a; Atsalakis & Valavanis, 2006a, 2006b).

Articles in Table 1 may be classified in three categories.

The first category includes articles that use as input data indexes from well developed markets in Western Europe, North America and other solid economy countries.

Ettes (2000), Setnes and Van Drempt (1999) model the Amsterdam stock exchange. Brownstone (1996), Kanas and Yannopoulos (2001) model the FTSE stock index. Lendasse, De Bodt, Wertz, and Verleysen (2000) studies the Belgian market. The Madrid stock exchange is examined by Fernandez-Rodriguez, Gonzalez-Martel, and Sosvilla-Rivebo (2000), Perez-Rodriguez, Torrab, and Andrada-Felixa (2004). The German DAX is forecasted by Rast (1999), Schumann and Lohrbach (1993), Siekmann, Gebhardt, and Kruse (1999), Steiner and Wittkemper (1997). These markets belong to well developed European markets.

Baek and Cho (2002), Chun and Park (2005), Kim and Chun (1998), Kim (1998), Oh and Kim (2002) model the Korean stock index. Cao, Leggio, and Schniederjans (2005), Yiwen, Guizhong, and Zongping (2000), Zhang, Jiang, and Li (2004), Zhongxing and Liting (1993) examine the Shangai stock market. Lam (2001) forecasts the Hong Kong stock exchange, Chen, Leung, and Daouk (2003), Kuo (1998), Wang (2002), Wang and Leu (1996) study the Taiwan stock index. Baba and Kozaki (1992), Huang, Nakamori, and Wang (2005), Jaruszewicz and Mandziuk (2004), Kimoto, Asakawa, Yoda, and Takeoka (1990), Mizuno, Kosaka, and Yajima (1998) forecast the Japanese Stock.

Ajith et al. (2003a), Ajith, Sajith, and Sarathchandran (2003b), Chen, Abraham, Yang, and Yang (2005a) attempt to forecast the NASDAQ stock exchange, and Chaturvedi and Chandra (2004), Halliday (2004), Leigh, Paz, and Purvis (2002) try to forecast the NYSE. The S&P 500 has the highest percentage of preference among studies as in Armano, Marchesi, and Murru (2004), Casas (2001), Malliaris and Salchenberger (1993), Tsaih, Hsu, and Lai (1998). Olson and Mossman (2003) studies the Toronto stock exchange index. Surveyed markets are the North America well developed markets. This survey includes also studies from the Australian Stock Index, surveyed by Barnes, Rimmer, and Ting (2000), Pan, Tilakarante, and Yearwood (2005), Vanstone, Finnie, and Tan (2005).

The second category focuses on studies that use indexes to forecast emerging markets. Studies from ex-Eastern Europe include Zorin and Borisov (2002) for the Latvian Riga stock exchange index, Walczak (1999), Wikowska (1995) for the Polish stock exchange index. Egeli, Ozturan, and Badur (2003), Yumlu, Gurgen, and Okay (2004, 2005) forecast the Istanbul Stock Exchange market. From Western European emerging markets, studies include Koulouriotis Koulouriotis (2004, 2001, 2002, 2005) for the Athens stock exchange, Andreou, Neocleous, Schizas, and Toumpouris (2000), Constantinou, Georgiades, Kazandjian, and Kouretas (2006) for the Cyprus stock exchange. The Singapore stock exchange is the most popular emerging market, forecasted by Ayob, Nasrudin, Omar, and Surip (2001), Hui, Yap, and Prakash (2000), Kim (1998), Phua, Hoh, Daohua, and Weiding (2001).

The third category includes articles that do not focus on a particular stock exchange market index, but use independent stocks or portfolio of stocks, instead. A typical example of this category is the study by Pantazopoulos, Tsoukalas, Bourbakis, Bruen, and Houstis (1998) that uses as input the price of the IBM stock, the study of Steiner and Wittkemper (1997), who selected as inputs the 16 top/bottom stocks from the DAX and the research by Atsalakis and Valavanis (2006a, 2006b) applied to five stocks of the Athens Stock Exchange and the NYSE.

TECHNOLOGIES

PYTHON

Python was the language of choice for this project. This was an easy decision for the multiple reasons.

1. Python as a language has an enormous community behind it. Any problems that might be encountered can be easily solved with a trip to Stack Overflow. Python is among the most popular languages on the site which makes it very likely there will be a direct answer to any query.

2. Python has an abundance of powerful tools ready for scientific computing. Packages such as Numpy, Pandas, and SciPy are freely available and well documented. Packages such as these can dramatically reduce, and simplify the code needed to write a given program. This makes iteration quick.

3. Python as a language is forgiving and allows for programs that look like pseudo code. This is useful when pseudocode given in academic papers needs to be implemented and tested. Using Python, this step is usually reasonably trivial.

However, Python is not without its flaws. The language is dynamically typed and packages are notorious for Duck Typing. This can be frustrating when a package method returns something that, for example, looks like an array rather than being an actual array. Coupled with the fact that standard Python documentation does not explicitly state the return type of a method, this can lead to a lot of trials and error testing that would not otherwise happen in a strongly typed language. This is an issue that makes learning to use a new Python package or library more difficult than it otherwise could be.

NUMPY

Numpy is python modules which provide scientific and higher level mathematical abstractions wrapped in python. In most of the programming languages, we can't use mathematical abstractions such as f(x) as it would affect the semantics and the syntax of the code. But by using Numpy we can exploit such functions in our code.

Numpy's array type augments the Python language with an efficient data structure used for numerical work, e.g., manipulating matrices. Numpy also provides basic numerical routines, such as tools for finding Eigenvectors.

SCIKIT LEARN

Scikit-learn is a free software machine learning library for the Python programming language. It features various classification, regression and clustering algorithms including support vector machine, random forest, gradient boosting, k-means etc. It is mainly designed to interoperate with the Python numerical and scientific libraries NumPy and SciPy.

Scikit-learn is largely written in Python, with some core algorithms written in Cython to achieve performance. Support vector machines are implemented by a Cython wrapper around LIBSVM .i.e., logistic regression and linear support vector machines by a similar wrapper around LIBLINEAR.

TENSORFLOW

TensorFlow is an open source software library for numerical computation using data flow graphs. Nodes in the graph represent mathematical operations, while the graph edges represent the multidimensional data arrays (tensors) communicated between them. The flexible architecture allows you to deploy computation to one or more CPUs or GPUs in a desktop, server, or mobile device with a single API. TensorFlow was originally developed by researchers and engineers working on the Google Brain Team within Google's Machine Intelligence research organization for the purposes of conducting machine learning and deep neural networks research, but the system is general enough to be applicable in a wide variety of other domains as well.

TensorFlow is Google Brain's second-generation system. While the reference implementation runs on single devices, TensorFlow can run on multiple CPUs and GPUs (with optional CUDA and SYCL extensions for general-purpose computing on graphics processing units). TensorFlow is available on 64-bit Linux, macOS, Windows, and mobile computing platforms including Android and iOS.

KERAS

Keras is a high-level neural networks API, written in Python and capable of running on top of TensorFlow, CNTK, or Theano. It was developed with a focus on enabling fast experimentation. Being able to go from idea to result with the least possible delay is key to doing good research.

Keras allows for easy and fast prototyping (through user friendliness, modularity, and extensibility). Supports both convolutional networks and recurrent networks, as well as combinations of the two. Runs seamlessly on CPU and GPU.

The library contains numerous implementations of commonly used neural network building blocks such as layers, objectives, activation functions, optimizers, and a host of tools to make working with image and text data easier. The code is hosted on GitHub, and community support forums include the GitHub issues page, a Gitter channel and a Slack channel.

COMPILER OPTION

Anaconda is a freemium open-source distribution of the Python and R programming languages for large-scale data processing, predictive analytics, and scientific computing, that aims to simplify package management and deployment. Package versions are managed by the package management system conda.

CHOOSING THE DATASET

For this project, we chose the Google stocks. The Google stocks is a large index traded on the New York stock exchange. All companies in the index are large publicly traded companies, leaders in each of their own sectors. The index covers a diverse set of sectors featuring companies such as Microsoft, Visa, Boeing, and Walt Disney. It is important to use a predefined set of companies rather than a custom selected set so that we do leave ourselves open to methodology errors or accusations of fishing expeditions. If we had selected a custom set of companies, it could be argued that the set was tailored specifically to improve our results. Since the aim of the project is to create a model of stock markets in general. Google was chosen because it is well known. The components provided a good balance between available data and computational feasibility.

GATHERING THE DATASETS

A primary dataset will be used throughout the project. The dataset will contain the daily percentage change in stock price. Luckily, daily stock price data is easy to come by. Google and Yahoo both operate websites which offer a facility to download CSV files containing a full 14 daily price history. These are useful for looking at individual companies but cumbersome when accessing large amounts of data across many stocks. For this reason, Quandl was used to gather the data instead of using Google and Yahoo directly. Quandl is a free to use website that hosts and maintains vast amounts of numerical datasets with a focus specifically on economic datasets, including stock market data which is backed by Google and Yahoo. Quandl also provides a small python library that is useful for accessing the database programmatically. The library provides a simple method for calculating the daily percentage change daily in prices.

For instance, data we gather for a Monday will be matched with, and try to predict, Tuesday's trend. This dataset was then saved in CSV format for simple retrial as needed throughout the project. This dataset containing the daily trends of companies will serve as the core dataset that will be used in most experiments later in the report.

RESULTS AND DISCUSSION

The most interesting task is to predict the market. So many methods are used for completing this task. Methods vary from very informal ways to many formal ways a lot. This tech. are categorized as:

- Prediction Methods
- Traditional Time Series
- Technical Analysis Methods
- Machine Learning Methods
- Fundamental Analysis Methods
- Deep Learning

The criteria for this category are the kind of tool and the kind of data that these methods are consuming in order to predict the market. What is mutual to the technique is that they are predicting and hence helping the market's future behavior.

Technical Analysis Methods

Technical analysis is used to attempt to forecast the price movement of virtually any tradable instrument that is generally subject to forces of supply and demand, including stocks, bonds, futures and currency pairs. In fact, technical analysis can be viewed as simply the study of supply and demand forces as reflected in the market price movements of a security. It is most commonly applied to price changes, but some analysts may additionally track numbers other than just prices, such as trading volume or open interest figures.

Over the years, numerous technical indicators have been developed by analysts in attempts to accurately forecast future price movements. Some indicators are focused primarily on identifying the current market trend, including support and resistance areas, while others are focused on determining the strength of a trend and the likelihood of its continuation. Commonly used technical indicators include trendlines, moving averages and momentum indicators such as the moving average convergence divergence (MACD) indicator.

Technical analysts apply technical indicators to charts of various timeframes. Short-term traders may use charts ranging from one-minute timeframes to hourly or four-hour timeframes, while traders analyzing longer-term price movement scrutinize daily, weekly or monthly charts.

Fundamental Analysis Techniques

Fundamental analysis uses real, public data in the evaluation a security's value. Although most analysts use fundamental analysis to value stocks, this method of valuation can be used for just about any type of security. For example, an investor can perform fundamental analysis on a bond's value by looking at economic factors such as interest rates and the overall state of the economy. He can also look at information about the bond issuer, such as potential changes in credit ratings.

For stocks and equity instruments, this method uses revenues, earnings, future growth, return on equity, profit margins, and other data to determine a company's underlying value and potential for future growth. In terms of stocks, fundamental analysis focuses on the financial statements of the company being evaluated. One of the most famous and successful fundamental analysts is the so-called "Oracle of Omaha", Warren Buffett, who is well known for successfully employing fundamental analysis to pick securities. His abilities have turned him into a billionaire.

Traditional Time Series Prediction

Time series analysis can be useful to see how a given asset, security or economic variable changes over time. It can also be used to examine how the changes associated with the chosen data point compare to shifts in other variables over the same time period.

For example, suppose you wanted to analyze a time series of daily closing stock prices for a given stock over a period of one year. You would obtain a list of all the closing prices for the stock from each day for the past year and list them in chronological order. This would be a one-year daily closing price time series for the stock.

Delving a bit deeper, you might be interested to know whether the stock's time series shows any seasonality to determine if it goes through peaks and valleys at regular times each year. The analysis in this area would require taking the observed prices and correlating them to a chosen season. This can include traditional calendar seasons, such as summer and winter, or retail seasons, such as holiday seasons.

Alternatively, you can record a stock's share price changes as it relates to an economic variable, such as the unemployment rate. By correlating the data points with information relating to the selected economic variable, you can observe patterns in situations exhibiting dependency between the data points and the chosen variable.

Machine Learning Methods

Various sectors of the economy are dealing with huge amounts of data available in different formats from disparate sources. The enormous amount of data, known as Big Data, is becoming easily available and accessible due to the progressive use of technology. Companies and governments realize the huge insights that can be gained from tapping into big data but lack the resources and time required to comb through its wealth of information. In this regard, Artificial Intelligence (AI) measures are being employed by different industries to gather, process, communicate and share useful information from data sets. One method of AI that is increasingly utilized for big data processing is Machine Learning.

The various data applications of machine learning are formed through a complex algorithm or source code built into the machine or computer. This programming code creates a model which identifies the data and builds predictions around the data it identifies. The model uses parameters built into the algorithm to form patterns for its decision-making process. When new or additional data becomes available, the algorithm automatically adjusts the parameters to check for a pattern change, if any. However, the model shouldn't change.

How machine learning works can be better explained by an illustration in the financial world. Traditionally, investment players in the securities market like financial researchers, analysts, asset managers, individual investors scour through a lot of information from different companies around the world to make profitable investment decisions. However, some pertinent information may not be widely publicized by the media and may be privy to only a select few who have the advantage of being employees of the company or residents of the country where the information stems from. In addition, there's only so much information humans can collect and process within a given time frame. This is where machine learning comes in.

An asset management firm may employ machine learning in its investment analysis and research area. Say the asset manager only invests in mining stocks. The model built into the system scans the World Wide Web and collects all types of news events from businesses, industries, cities, and countries, and this information gathered comprises the data set. All the information inputted in the data set is information that the asset managers and researchers of the firm would not have been able to get using all their human powers and intellects. The parameters built alongside the model extracts only data about mining companies, regulatory policies on the exploration sector, and political events in select countries from the data set. Say, a mining company XYZ just discovered a diamond mine in a small town in South Africa, the machine learning app would highlight this as relevant data. The model could then use an analytics tool called predictive analytics to make predictions on whether the mining industry will be profitable for a time period, or which mining stocks are likely to increase in value at a certain time. This information is relayed to the asset manager to analyze and make a decision for his portfolio. The asset manager may make a decision to invest millions of dollars into XYZ stock.

In the wake of an unfavorable event, such as South African miners going on strike, the computer algorithm adjusts its parameters automatically to create a new pattern. This way, the computational model built into the machine stays current even with changes in world events and without needing a human to tweak its code to reflect the changes. Because the asset manager received this new data on time, he is able to limit his losses by exiting the stock. Machine learning is used in different sectors for various reasons. Trading systems can be calibrated to identify new investment opportunities. Marketing and e-commerce platforms can be tuned to provide accurate and personalized recommendations to their users based on the users' internet search history or previous transactions. Lending institutions can incorporate machine learning to predict bad loans and build a credit risk model. Information hubs can use machine learning to cover huge amounts of news stories from all corners of the world. Banks can create fraud detection tools from machine learning techniques. The incorporation of machine learning in the digital-savvy era is endless as businesses and governments become more aware of the opportunities that big data presents.

DEEP LEARNING

An artificial intelligence function that imitates the workings of the human brain in processing data and creating patterns for use in decision making. Deep learning is a subset of machine learning in Artificial Intelligence (AI) that has networks which are capable of learning unsupervised from data that is unstructured or unlabeled. Also known as Deep Neural Learning or Deep Neural Network.

The digital era has brought about an explosion of data in all forms and from every region of the world. This data, known simply as Big Data, is gotten from sources like social media, internet search engines, e-commerce platforms, online cinemas, etc. This enormous amount of data is readily accessible and can be shared through fine tech applications like cloud computing. However, the data, which normally is unstructured, is so vast that it could take decades for humans to comprehend it and extract relevant information. Companies realize the incredible potential that can result from unraveling this wealth of information and are increasingly adapting to Artificial Intelligence (AI) systems for automated support.

One of the most common AI techniques used for processing Big Data is Machine Learning. Machine learning is a self-adaptive algorithm that gets better and better analysis and patterns with experience or with newly added data. If a digital payments company wanted to detect the occurrence of or potential for fraud in its system, it could employ machine learning tools for this purpose. The computational algorithm built into a computer model will process all transactions happening on the digital platform, find patterns in the data set, and point out any anomaly detected by the pattern.

Deep learning, a subset of machine learning, utilizes a hierarchical level of artificial neural networks to carry out the process of machine learning. The artificial neural networks are built like the human brain, with neuron nodes connected together like a web. While traditional programs build analysis with data in a linear way, the hierarchical function of deep learning systems enables machines to process data with a non-linear approach. A traditional approach to detecting fraud or money laundering might rely on the amount of transaction that ensues, while a deep learning non-linear technique to weeding out a fraudulent transaction would include time, geographic location, IP address, type of retailer, and any other feature that is likely to make up a fraudulent activity. The first layer of the neural network processes a raw data input like the amount of the transaction and passes it on to the next layer as output. The second layer processes the previous layer’s information by including additional information like the user's IP address and passes on its result. The next layer takes the second layer’s information and includes raw data like geographic location and makes the machine’s pattern even better. This continues across all levels of the neuron network until the best and output is determined.

Using the fraud detection system mentioned above with machine learning, we can create a deep learning example. If the machine learning system created a model with parameters built around the amount of dollars a user sends or receives, the deep learning method can start building on the results offered by machine learning. Each layer of its neural network builds on its previous layer with added data like a retailer, sender, user, social media event, credit score, IP address, and a host of other features that may take years to connect together if processed by a human being. Deep learning algorithms are trained to not just create patterns from all transactions, but to also know when a pattern is signaling the need for a fraudulent investigation. The final layer relays a signal to an analyst who may freeze the user's account until all pending investigations are finalized.

Deep learning is used across all industries for a number of different tasks. Commercial apps that use image recognition, open source platforms with consumer recommendation apps, and medical research tools that explore the possibility of reusing drugs for new ailments are a few of the examples of deep learning incorporation.

ARTIFICIAL NEURAL NETWORKS (ANN)

A computing system that is designed to simulate the way the human brain analyzes and process information. Artificial Neural Networks (ANN) is the foundation of Artificial Intelligence (AI) and solves problems that would prove impossible or difficult by human or statistical standards. ANN has self-learning capabilities that enable it to produce better results as more data becomes available.

MARKET TRACK

Its stock market forecast system consists of two major parts: an extensive database and a forecast model. The forecast model reads the database and then makes a prediction of where the market is headed. From this prediction, it determines a trading position for the Dow Diamonds or the SP500 Spiders [5]. The database and forecast are updated daily at the close of trading.

It uses a neural network model in combination with a genetic algorithm to calculate the SP500 forecast. The calculations are somewhat complex but can be summarized by the following three procedural steps.

Step one: The genetic algorithm is used to find the optimum neural network structures and inputs. This calculation basically determines how the networks will be wired.

Step two: Using the information from the first step, a set of networks is initialized and then trained on about 75 percent of the market data (in-sample) in their database, which currently consists of about 7200 days of data. They use an evolutionary program to train the networks (i.e. to determine the neural network weights).

Step three: After training, the networks are rigorously tested on the remaining 25 percent of market data (out-of-sample). Networks that fail the test are discarded. Networks that pass the test are included in the library that they use to calculate the forecast. The number of neural networks currently in their library varies from day to day, but normally contains more than 400. 7 Input to the networks are technical and fundamental market data. The table below shows the types of data that are currently used by the model:

- Dow Jones Industrial Average closing value
- Dow Jones Industrial Average theoretical high value
- Dow Jones Industrial Average theoretical low value
- Dow Jones Transportation Average closing value
- Dow Jones Utility Average closing value
- New York Stock Exchange total volume

- New York Stock Exchange number of advancing stocks
- New York Stock Exchange number of declining stocks
- New York Stock Exchange number of new highs
- New York Stock Exchange number of new lows
- New York Stock Exchange advancing volume
- New York Stock Exchange declining volume
- SP500 closing value
- SP500 trailing earnings
- Yen-Dollar exchange rate
- Treasury bill discount rate
- Commodity Research Bureau index

The above data are filtered and normalized and certain functions of these data are computed. It currently computes 63 separate input variables at the close of each trading day.

Figure 2.1: Forecasting method for Markettrak 8 The 63 inputs are applied to a neural network and after some number crunching the network outputs a value between -1.0 and +1.0, with -1.0 being a very strong down market signal and +1.0 being a very strong up market signal. A value near zero would indicate a neutral market signal. They apply the inputs to each network in our library and an average of their outputs is computed. This average network output is used with position set points to determine a trading position for the Dow Diamonds or the SP500 Spiders for the next trading day.

When the computed value of the average network output is above the long position set point, a long position is indicated. When the value of the average network output is below the short position set point, a short position is indicated. When the average network output falls between these two set points, a cash position is indicated. Because of the timing of the update, trades may be made in the extended sessions or at open of the next trading day. When computing our performance, all trades are assumed to take place at the session close. The current trading position along with recent average network output values and the average network output set points are shown on our forecast page.

REQUIREMENT ANALYSIS AND FEASIBILITY STUDY

Feasibility Study

Simply put, stock market cannot be accurately predicted. The future, like any complex problem, has far too many variables to be predicted. The stock market is a place where buyers and sellers converge. When there are more buyers than sellers, the price increases. When there are more sellers than buyers, the price decreases. So, there is a factor which causes people to buy and sell. It has more to do with emotion than logic. Because emotion is unpredictable, stock market movements will be unpredictable. It's futile to try to predict where markets are going. They are designed to be unpredictable.

There are some fundamental financial indicators by which a company's stock value can be estimated. Some of the indicators and factors are: Price-to-Earning (P/E) Ratio, Price-to-Earning Growth (PEG) Ratio, Price-to-Sales (P/S) Ratio, Price/Cash Flow (P/CF) Ratio, Price-to-Book Value (P/BV) Ratio and Debt-to-Equity Ratio. Some of the parameters are available and accessible on the web but all of them aren't. So we are confined to use the variables that are available to us.

The proposed system will not always produce accurate results since it does not account for the human behaviours. Factors like change in company's leadership, internal matters, strikes, protests, natural disasters, change in the authority cannot be taken into account for relating it to the change in Stock market by the machine.\

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.

Requirement Analysis

After the extensive analysis of the problems in the system, we are familiarized with the requirement that the current system needs. The requirement that the system needs is categorized into the functional and non-functional requirements. These requirements are listed below: 3.2.1 Functional Requirements Functional requirement are the functions or features that must be included in any system to satisfy the business needs and be acceptable to the users. Based on this, the functional requirements that the system must require are as follows:

- The system should be able to generate an approximate share price.
- The system should collect accurate data from the NEPSE website in consistent manner.

Non-Functional Requirements

Non-functional requirement is a description of features, characteristics and attribute of the system as well as any constraints that may limit the boundaries of the proposed system. The non-functional requirements are essentially based on the performance, information, economy, control and security efficiency and services. Based on these the non-functional requirements are as follows:

- The system should provide better accuracy.
- The system should have simple interface for users to use.
- To perform efficiently in short amount of time

REFERENCES

[1] Ajith, A., Baikunth, N. & Mahanti, P. K. (2003a). Hybrid intelligent systems for stock market analysis. In Proceedings of International Conference on Computational Science.

[2] Ajith, A., Sajith, N., & Sarathchandran, P. P. (2003b). Modelling chaotic behaviour of stock indices using Intelligent Paradigms. Neural, Parallel & Scientific Computations Archive, 11, 143–160.

[3] Andreou, A. S., Neocleous, C. C., Schizas, C. N., & Toumpouris, C. (2000). Testing the predictability of the Cyprus Stock Exchange: The case of an emerging market. Proceedings of the International Joint Conference on Neural Networks, 360–365.

[4] Armano, G., Marchesi, M., & Murru, A. (2004). A hybrid genetic-neural architecture for stock indexes forecasting. Information Sciences, 170(1), 3–33.

[5] Atiya, A., Noha Talaat & Samir Shaheen, (1997). An efficient stock market forecasting model using neural networks. In Proceedings of the IEEE International Conference on Neural Networks.

[6] Atsalakis, G., & Valavanis, K. (2006). A neuro-fuzzy controller for stock market forecasting. Working paper, Technical University of Crete.

[7] Atsalakis, G., & Valavanis K. (2006). Neuro-fuzzy and technical analysis for stock prediction. Working paper, Technical University of Crete.

[8] Ayob, M., Nasrudin, M. F., Omar, K., & Surip, M. (2001). The effects of returns function on individual stock price (KLSE) prediction model using neural networks. In Proceedings of the International Conference on Artificial Intelligence, IC-AI 2001 (pp. 409–415).

[9] Baba, N., & Kozaki, M. (1992). An intelligent forecasting system of stock price using neural networks. Proceedings of the IEEE International Joint Conference on Neural Networks, 371–377.

[10] Baba, N., & Suto, H. (2000). Utilization of artificial neural networks and the TDLearning method for constructing intelligent decision support systems. European Journal of Operational Research, 122, 501–508.

[11] Baek, J., & Cho, S. (2002). Time to jump. Long rising pattern detection in KOSPI 200 future using an auto-associative neural network. Lecture Notes in Computer Science (2412). Springer.

[12] Barnes, M. B., Rimmer, R. J., & Ting, K. M. (2000). A study of techniques for mining data from the Australian stock exchange. In Proceedings of the Fourth World Multi-conference on Systemics, Cybernetics and Informatics, SCI 2000 (Vol. VIII, Part II, pp. 52–57).

[13] Bautista, C. C. (2001). Predicting the Philippine Stock Price Index using artificial neural networks. UPCBA Discussion Paper No. 0107.

[14] Brownstone, D. (1996). Using percentage accuracy to measure neural network predictions in stock market movements. Neurocomputing, 10, 237–250.

[15] Cao, Q., Leggio, K. B., & Schniederjans, M. J. (2005). A comparison between Fama and Frenchs model and artificial neural networks in predicting the Chinese Stock Market. Computers and Operations Research, 32, 2499–2512.