**DEPARTMENT OF MECHATRONICS ENGINEERING**

**Stock Prices Prediction** **Using Machine Learning**

Synopsis

*Submitted by*

**STUDENT NAME REG. NO. E-MAIL ID CONTACT NO.**

SURAMPALLI RAKESH. 160929044. [rakeshsurampalli@gmail.com](mailto:rakeshsurampalli@gmail.com) 9652529111

*Under the guidance of*

**External Guide : Sanjay Singh, PhD**

**Professor**

**Department of Information & Communication Technology**

**Internal Guide : Mahesh Anil Inamdar**

**Assistant Professor**

**Department of Mechatronics Engineering**

**MANIPAL INSTITUTE OF TECHNOLOGY, MANIPAL**

**January 2019**

**1. Introduction**

In recent years, Artificial Intelligence (AI) has been a significant technology that is being applied in making of driverless cars, intelligent robots, image and speech recognition, automatic translations, and medical assistants (6). Hence, forecasting stock market in the light of AI and by utilizing different machine learning methods has been an important issue in financial and economic fields. Consequently, it has made the researchers think to come up with reliable predictive models over the decade (7). The urge to predict stock prices more accurately is making the researchers to work for the betterment of the current predictive machine learning models. The reason is that shareholders and investors have the freedom to make plans and strategical approaches towards taking decision about investments and future activities. This leads the organizations and individuals to get any predictive method that ensures more income from the stock market easily along with minimum investment risk. In finance, forecasting stock market is considered to be one of most difficult tasks to do till now because of the stochastic behaviours and complex dependencies of stock market (8). Because of the unpredictive nature of stock market there exists no certain models of machine learning that can precisely forecast about stock market and there is more work to perform in this sector which is the inspiring factor for us to research and build a better predictive system. Various methods of machine learning had been applied for forecasting stock market throughout the recent years. Among them models such as Support Vector Regression (SVR), Artificial Neural Networks (ANNs), Bayesian Neural Network (BNN) and so on had been exploited to improve time series forecasting (8). Moreover, different hybrid methods had also been generated to improve the efficiency of prediction. Yet, there is little evidence about their relative performance as standard forecasting models. ARIMA has been extensively used for its efficiency in financial time series forecasting especially for short-term prediction than the most used neural network techniques.

**2. Literature Review**

While using **Time Series analysis** first step is to check whether historical stock market data is stationary using Plotting Rolling Statics Dickey Full Test. Second step is to eliminate TREND and SESSIONALITY from the series to make the data stationary series. For short term predictions ARIMA model has great potential. Finally the forecasted values are converted to original scale by applying TREND and SESSONALITY(1).

**Arima model** is one method for forecasting time series, it is assumed that past value of the series plus previous error terms contain information for the purpose of forecasting. The main advantage of Arima forecasting is that it requires data on time series in question only. However, Arima model are essentially backward looking, they are generally poor at predicting turning points, unless the turning point represents a return to a long-run equilibrium Exchange rate forecasting means estimating the rate which will be any of future time(2).

By using **DATA MINING** the algorithms which are implemented complement each other i.e. if the results are analysed together better investment plans can be generated. Also the sentimental model on its own provides basis for the investment. The accuracy of the system overall is around 70-75% which is seemingly reasonable in such a volatile environment(9).

**3. Problem Definition**

The aim is to predict the adjusted closing prices of Stock Market using data from the previous N days of a company. This can be done by using time series analysis. We will split this dataset into 60% train, 20% validation, and 20% test. The model will be trained using the train set, model hyperparameters will be tuned using the validation set, and finally the performance of the model will be reported using the test set. To evaluate the effectiveness of our methods, we will use the root mean square error (RMSE) and mean absolute percentage error (MAPE) metrics. For both metrics, the lower the value, the better the prediction.

**4. Objectives**

* To predict the value, we should need variables, parameters, algorithm and past data.
  + - These can be done by two different types of financial analysis to predict stock market prices:
      * Fundamental Analysis: it is based on the health of the company and this includes qualitative and quantitative factors such as interest rate, return on assets, revenues, expenses and price to earnings among others. The aim of this analysis is to check the long- term sustainability and strength of the company for the purpose of long-term investment.
      * Technical analysis: It is based on time series data. Traders analyse historical price movements and chart patterns and consider time as a crucial parameter in the prediction. Technical analysis can rely on three main keys: stock prices movement although many times the movement seems to be random, historical trends which are assumed to repeat as time passes, and all relevant information about a stock(10).
      * These features can be considered as technical analysis features for the stock market as they are based on mathematical calculations as described below:

Log return: a finance term that represents the logarithmic difference between the close price at time t and close price at time t-1

Pseudo-log-return: the logarithmic difference between average prices of consecutive minutes

Trend Indicator: a linear model applied on 1-minute tick data to generate a linear equation with a certain slope. A negative slope implies a decrease in the price while a positive slope implies an increase and a slope close to zero implies that the price is almost stable.

* + - Different machine learning models and risk strategies have been applied to stock market prediction task trying to predict mainly the direction of the price for different time frames and using different features that would affect market prices.
    - Fusible Datasets for the algorithm has to be collected. Like technical indicators (Stochastic Oscillator, Larry William, Relative Strength index) that represent the variation of stock price over time and S&P index.

**5. Methodology**

A time series is a series of [data points](https://en.wikipedia.org/wiki/Data_point) indexed in time order. Most commonly, a time series is a [sequence](https://en.wikipedia.org/wiki/Sequence) taken at successive equally spaced points in time.

Time series analysis: It comprises methods for analyzing time series data in order to extract meaningful statistics and other characteristics of the data.

Time series forecasting: It is the use of a [model](https://en.wikipedia.org/wiki/Model_(abstract)) to predict future values based on previously observed values.

[regression analysis](https://en.wikipedia.org/wiki/Regression_analysis) is often employed in such a way as to test theories that the current values of one or more independent time series affect the current value of another time series, this type of analysis of time series is not called "time series analysis", which focuses on comparing values of a single time series or multiple dependent time series at different points in time.

A given time series is thought to consist of three systematic components including level, trend, seasonality, and one non-systematic component called noise. These components are defined as follows:

Level: The average value in the series.

Trend: The increasing or decreasing value in the series.

Seasonality: The repeating short-term cycle in the series.

Noise: The random variation in the series.

First, we need to check if a series is stationary or not because time series analysis only works with stationary data.

ADF (Augmented Dickey-Fuller) Test: The Dickey-Fuller test is one of the most popular statistical tests. It can be used to determine the presence of unit root in the series, and hence help us understand if the series is stationary or not. The null and alternate hypothesis of this test is:

Null Hypothesis: The series has a unit root

Alternate Hypothesis: The series has no unit root.

If we fail to reject the null hypothesis, we can say that the series is non-stationary. This means that the series can be linear or difference stationary. If both mean and standard deviation are flat lines, the series becomes stationary.

ARIMA models are applied in some cases where data show evidence of [non-stationarity](https://en.wikipedia.org/wiki/Stationary_process), where an initial differencing step can be applied one or more times to eliminate the non-stationarity.

**References**

*Journal / Conference Papers*

1. Stock Market Prediction Using Time Series Analysis Authors Farhan Islam Hira, Mazharul Ferdous Maruf, Afzal Hossain
2. Autoregressive Integrated Moving Average (ARIMA) Model for Exchange Rate (Naira to Dollar)Steve C Nwankwo
3. INVESTMENT IMPLICATIONS OF THE FRACTAL MARKET HYPOTHESIS ADAM KARP\* and GARY VAN VUUREN† School of Economics Department of Risk Management North-West University Potchefstroom Campus, South Africa \*
4. Group Polarization, Transition Points and Coherent Market Model Maciej Ciołek\*
5. Reinforcement Learning in Financial Markets Terry Lingze Meng and Matloob Khushi \* School of Computer Science, Building J12, University of Sydney, 1 Cleveland Street, Darlington, NSW 2006, Australia
6. Mostafa, M. M. (2010). Forecasting stock exchange movements using neural networks: Empirical evidence from kuwait. *Expert Systems with Applications*, 37(9):6302–6309.
7. Ariyo, A. A., Adewumi, A. O., and Ayo, C. K. (2014). Stock price prediction using the arima model. In *Computer Modelling and Simulation (UKSim), 2014 UKSim-AMSS 16th International Conference on*, pages 106–112. IEEE.
8. Wei, L.-Y. (2013). A hybrid model based on anfis and adaptive expectation genetic algorithm to forecast taiex. *Economic Modelling*, 33:893–899.
9. Stock Market Prediction Using Data Mining Techniques
10. Automated Stock Price Prediction Using Machine Learning,Mariam Moukalled Wassim El-Hajj Mohamad Jaber

**Name of the External Guide with contact details and email address:**

Sanjay Singh, PhD  
Professor,  
Department of Information & Communication Technology,  
Manipal Institute of Technology, Manipal-576104 INDIA  
Ph: +91-820-2925386 (O), +91-820-2572132 (R)  
Email: [sanjay.singh@ieee.org](mailto:sanjay.singh@ieee.org)

**Signature of Guide**

**Name of the Internal Guide with contact details and email address:**

Mahesh Anil Inamdar

Assistant Professor

Department of Mechatronics Engineering

[mahesh.inamdar@manipal.edu](mailto:mahesh.inamdar@manipal.edu)

**Signature of Guide**

**Date of commencement of the project:**