KLE Dr. M S Sheshgiri college of engineering and technology

Belagavi, Karnataka.

PROJECT ON

MACHINE LEARNING BASED STOCK PRICE PREDICTION

SUBMITTED BY

MANOJKUMAR B BAGEWADI	2KL17CS113
SARVESH S HATTI	2KL17SC074
K L NITEESH	2KL17SC034
NANDESH GANIGER	2KL17CS046

ABSTRACT

The process of forecasting the stock prices has been a difficult task for many of the researchers, analysts and new investors. In fact, investors are highly interested in the research area of stock price prediction. For a good and successful investment, many investors are keen in knowing the future situation of the stock market. Good and effective prediction systems for stock market help traders, investors, and analyst by providing supportive information like the future direction of the stock market. Predicting stock market price is a complex task that traditionally involves extensive human-computer interaction. There are multiple prediction methodologies for share price forecasting. Time Series Forecasting is basic for share price forecasting and other financial model forecast. As share price are more nonlinear, more intelligent time series prediction systems are required. Existing systems accuracy are not efficient enough in predicting. In this paper, we propose to use LSTM Machine Learning Algorithm for efficient forecasting of stock price. This will provide more accurate results when compared to existing stock price prediction algorithms.

INTRODUCTION

The stock market is a vast array of investors and traders who buy and sell stock, pushing the price up or down. The prices of stocks are governed by the principles of demand and supply, and the ultimate goal of buying shares is to make money by buying stocks in companies whose perceived value (i.e., share price) is expected to rise. Stock markets are closely linked with the world of economics — the rise and fall of share prices can be traced back to some Key Performance Indicators (KPI's). The five most commonly used KPI's are the opening stock price (`Open'), end-of-day price (`Close'), intra-day low price (`Low'), intra-day peak price (`High'), and total volume of stocks traded during the day (`Volume'). Economics and stock prices are mainly reliant upon subjective perceptions about the stock market. It is near-impossible to predict stock prices to the T, owing to the volatility of factors that play a major role in the movement of prices. However, it is possible to make an educated estimate of prices. Stock prices never vary in isolation: the movement of one tends to have an avalanche effect on several other stocks as well. This aspect of stock price movement can be used as an important tool to predict the prices of many stocks at once. Due to the sheer volume of money involved and number of transactions that take place every minute, there comes a trade-off between the accuracy and the volume of predictions made; as such, most stock prediction systems are implemented in a distributed, parallelized fashion. These are some of the considerations and challenges faced in stock market analysis.

LITERATURE SURVEY

From the research paper “Machine Learning in Stock Price Trend Forecasting” written by Y. Dai and Y. Zhang in Stanford University, they used features like PE ratio, PX volume, PX EBITDA, 10-day volatility, 50-day moving average, etc. to predict the next-day stock price and a long-term stock price [2]. The machine learning algorithms used in the research are Logistic Regression, Gaussian Discriminant Analysis, Quadratic Discriminant Analysis, and SVM. The accuracy ratio is defined as the number of days that the model correctly classified the testing data over the total number of testing days. With the short-term model predicting the next day stock price, it has very low accuracy, the Quadratic Discriminant Analysis is the best among all models, it scored a 58.2% accuracy. With the long term model predicting the next n days stock prices, the longer the time frame, the better in the accuracy for SVM. With a time window of 44 days, the SVM model’s accuracy reached 79.3%. Apart from that, it was found that by increasing the number of features, the accuracy increased. When all of the 16 features were used, the accuracy of the model reached 79%, while it fell to 64% when only 8 features were used, and 55% if only 1 feature was used. Our project will also investigate how the timeframe would affect the accuracy of price predictions of different models. As models have to reach a certain threshold to have significance for the users to work as a reference, it is essential for us to optimize our model to figure out what the optimal parameters and model structure are for our stock price prediction purpose.

The research paper “Predicting stock and stock price index movement using Trend Deterministic Data Preparation and machine learning techniques” written by J. Patel, S. Shah, P. Thakkar, and K. Kotecha for the “Expert Systems with Applications” international journal demonstrated a way to use trend deterministic data to predict stock price movement [3]. They conducted experiments in using 10 technical indicators’ signals as inputs, then they use prediction models to predict whether the stock will go up or down in the coming 10 days, Technical analysis indicators include SMA, EMA, Momentum, Stochastic SK, Stochastic SK, MACD, RSI, etc. The prediction models they have used include ANN, SVM, Random Forest, and Naive Bayesian models. The model outputs “up” or “down” movement signals. Experiments have shown random forest scored the highest performance with 83.56% accuracy with their inputs.

Wanjawa and L. Muchemi demonstrated the potential in predicting stock prices using ANN, as shown in the research paper “ANN Model to Predict Stock Prices at Stock Exchange Markets”. They used 70% of the training data to predict the stock prices for the next 60 days. Through optimizations, they were able to predict the actual closing prices within 0.71% mean absolute percentage error (MAPE), with the highest variance -3.2% among all of the 62 days. This demonstrated a high potential for using machine learning to accurately predict stock prices. This is one of the key components in our application where algorithms have to be designed to have high accuracy, such that the platform could be useful for retail investors.

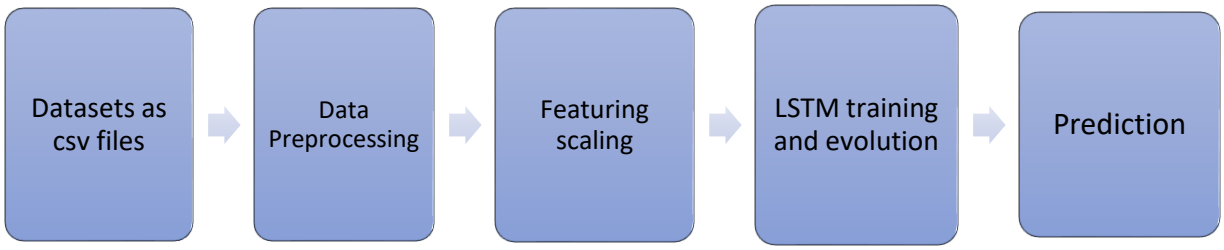
OBJECTIVES

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.

PROPOSED SYSTEM

We propose an online learning algorithm for predicting the end-of-day price of a given stock with the help of Long Short Term Memory (LSTM).

System Architecture



LSTM Training and Evaluation

Long short-term memory (LSTM) is an artificial recurrent neural network (RNN) architecture used in the field of deep learning. Unlike standard feedback neural networks, LSTM has feedback connections that make it a "general purpose computer". It can not only process single data points, but also entire sequences of data. Long Short-Term Memory (LSTM) networks are an extension for recurrent neural networks, which basically extends their memory. Therefore, it is well suited to learn from important experiences that have very long-time lags in between. The units of an LSTM are used as building units for the layers of a RNN, which is then often called an LSTM network. LSTM's enable RNN's to remember their inputs over a long period of time. This is because LSTM's contain their information in a memory, that is much like the memory of a computer because the LSTM can read, write and delete information from its memory.

REFERENCES

- Nazar, Nasrin Banu, and Radha Senthilkumar. “An online approach for feature selection for classification in big data.” Turkish Journal of Electrical Engineering & Computer Sciences 25.1 (2017): 163-171.
- Suresh, Harini, et al. “Clinical Intervention Prediction and Understanding using Networks.” arXiv preprint arXiv:1705.08498 (2017).
- Pascanu, Razvan, Tomas Mikolov, and Yoshua Bengio.“On the difficulty of training recurrent neural networks.” International Conference on Machine Learning. 2013.
- Zhu, Maohua, et al. “Training Long Short-Term Memory With Sparsified Stochastic Gradient Descent.” (2016).
- Ruder, Sebastian. “An overview of gradient descent optimization algorithms.” arXiv preprint arXiv:1609.04747.(2016).
- Recht, Benjamin, et al. “Hogwild: A lock-free approach to parallelizing stochastic gradient descent.” Advances in neural information processing systems. 2011.
- Zhao, P., Hoi, S. C., Wang, J., Li, B. “Online transfer learning”. Artificial Intelligence, 216, 76-102. (2014)