STOCK MARKET PREDICTION USING RECURRENT NEURAL NETWORK

[1]Rishabh Sharma [2] Sanjit Kumar [3] Shubham Srivatsava [4] Siri A Muddebihal
[5] Dr. Saravana Kumar R

[1]B.E. 4th year, Computer Science , Dayananda Sagar Academy of Tech & Mgt, Karnataka, India
[2] B.E. 4th year, Computer Science , Dayananda Sagar Academy of Tech & Mgt, Karnataka, India
[3] B.E. 4th year, Computer Science , Dayananda Sagar Academy of Tech & Mgt, Karnataka, India
[4] B.E. 4th year, Computer Science , Dayananda Sagar Academy of Tech & Mgt, Karnataka, India
[5] Associate Professor , Computer Science , Dayananda Sagar Academy of Tech & Mgt, Karnataka, India

ABSTRACT

The prediction and analysis of stock market data have got an important role in today's economy. Our project is based on the prediction of stock values for the next 7 days that shows the variation if the stock is going to rise or fall. We will extract data from different websites using API's then we will form a dataset from the past 2 years and then implement the neural network techniques like LSTM and Linear Regression to train and test the data and finally predicting the output values and using Matplotlib tool to plot a graph for better understanding to the users which will show the stock prices for the next 7 days. This will help the investors find potentially profitable stock. It can be used to gain an insight in to the economy as a whole. Having accurate information helps make better decisions.

Keyword: *LSTM, Neural Network, Stock Prediction, Deep Learning*

1. INTRODUCTION

Stock market prediction is determining the upcoming worth of stock market trends. Stocks are bought by investors when they are cheap and sold when stocks are expensive. A stock market is where traders and brokers can buy and sell shares and bonds. The stock market is unpredictable which abstain people from investing in the stock market. Thus, the need to determine stock market value is even more crucial.

A stacked LSTM consists of various LSTM layers where every layer contains various memory cells. A stacked LSTM with high dimensionality is used. A stacked LSTM is a stable method for sequence prediction problem. Given that LSTMs work on order data, meaning that the addition of layers join levels of abstraction of input observations over time period. In effect, chunking observations over time or representing the difficulty at different time scales

2. PRIOR AND RELATED WORK

The main purpose of our project to anticipate the future outcomes of any given share based on how the value of that particular share has varied in past in simpler sense we are take historical financial time-series data and predict how it may vary in future. There have been many works done in this field and also in fields related to them.

A group of scientists and computer experts did a comparative study and compared many machines learning as well as many deep learning techniques to foresee stock values and showed which method gave the most promising result. (Reference: - Forecasting Stock Market Patterns Using Machine Learning and Deep Learning Approaches Via Continual and Duplex Data; a Comparative Analysis). Another group of scientists worked on the forecasting of stock market trends with the help of tick data. (Reference: Indian stock market value forecasting using ANN on tick data)

2.1 Behavioral Analysis

With fast paced digitalization, the amount of data which we have at our disposal is increasing exponentially. A bulk of these data comes from various social media platform where people express their feelings, thoughts, ideas, reactions etc. everyday. These data can be highly productive in knowing the general sentiment and morale of people at any given time or on any given topic. Many important decisions, which can either be of political nature or economical, could be taken by considering this into account. Yuan in 2016 ventured into this field by exploring lexicon, rule and ML-based sentiment categorization. Lakshmi et al. (2017) surveyed on the demarcation of Twitter data using NB.

2.2 Stock Market Volatility

We all know how volatile stock market is in terms of its nature. Any false news or rumours can turn the market upside down and on top of that there are other external factors as well, such as financial crisis, recession etc., which can also affect the stock market adversely. Many efforts have been made to foresee such possible future events. Kumar and Patil (2015) used time series and machine learning techniques as well, in order to forecast volatility of S&P 500 index.

Author	**Year**	**Focus**	**Approach**
Pawan Kumbhare et. al.	2017	Predicting stock market using SVM & Decision Tree	SVM & Decision Tree Algorithm
Amir Mosavi et. al.	2018	Predicting Stock Market Trends via Continuous & Binary Data	Comparative Analysis using Machine Learning & Deep Learning Algorithm
Adarsh Paul et. al	2019	Machine Learning Algorithm to predict the longevity of stock	Random Forest Algorithm was used to predict the stock market trends.
Vineet Kumar et. al.	2019	Analyzed Indian stock market prediction using artificial neural networks (ANN) on tick data	They use neural networks based on three different learning algorithms LevenbergMarquardt, Scaled Conjugate Gradient and Bayesian Regularization for stock market prediction

Table -1: Prior Work

3. EXISTING SYSTEM

The existing system works on various different approaches such as applying OCR technique on finance news, tweets, some other approaches are feature selection, using different indicators such as EMA, RSI.

Some system uses SVM and Decision Tree approach but they are good only when the dataset is of smalltime duration when the datasets become large it overfits with big noise.

Some studies show that deep learning methods such as RNN and LSTM works good when they use duplex datasets instead of unbroken one.

4. PROPOSED APPROACH

4.1 MODULES IDENTIFIED

4.1.1 DATA EXTRACTION MODULE

First thing needed to train a model is Data. Datasets of various stock companies can be extracted from the various financial websites such as Yahoo Finance that provides API's services which can be implemented to get the up-to date data of the company. For our venture, we are thinking about S&P 500 Companies. The rundown of organizations in S&P 500 can be gotten from Wikipedia. Utilize stock's ticker image from stage a to get information from Yahoo Finance.

Then we can save that Datasets into a csv file and can train the model on the particular dataset.

We have split the dataset into two sections where 65% of the information will be utilized for preparing and 35% of the information will be utilized for testing.

4.1.2 BUILDING THE PREDICTING MODEL

As human don't start thinking from scratch every second and understand each word based on your understanding of previous words while reading an essay or in general, in the same way LSTM also worklist based RNN models are able to connect precious information to the present task that is why they are used in time-series data for categorization, converting and building prediction.

LSTM recurrent neural network belongs to the group of machine and deep learning formulas, approaches and methods. It is a repeated network because of the response connections in its constructing design which allows it to operation the entire order of facts. LSTM are explicitly made to avoid the long-term decencies problem. Remembering data for a long period of time is practically their default behavior something they struggle to learn.

All RNN has a configuration of a string of repeating modules of neural network. In normal RNNs, this reproduces section will have an easy design, such as a one tanh surface. But LSTM structure comprises of 4 parts: the cell, input, output and forget gateway. The unit part of the LSTM recall values over random time period, and the three gateways check the run of in and out of the unit. The unit of the prototype is used for storing path of the reliance between the element in the input order. The input gateway is responsible for amount of data that will move in the unit cell, the forget gate manages the size to which a data last in the unit, and the output gateway responsible the expanse to which the worth is used to calculate the LSTM unit.

4.1.3 DATA PREPROCESSING MODULE

Information Pre-processing is that progression where the information gets changed, or encoded, to carry it's anything but an express that now the machine can undoubtedly parse it. All in all, the highlights of the information would now be able to be effectively deciphered by the calculation. Information objects are portrayed by various highlights, that catch the essential qualities of an item, for example, the mass of an actual article or the time at which an occasion happened, and so forth Highlights are regularly called as factors, qualities, fields, traits, or measurements. Preprocessing of data involves various steps such as dealing with missing values, encoding categorical data, splitting dataset into training and test set, feature scaling, etc. If our dataset contains some missing data, then it may create a huge problem for our machine learning model. Hence it is necessary to handle missing values present in the dataset, there are various methods to deal with the missing value like mean substitution, maximum likelihood, etc.

4.2 TOOLS USED

4.2.1 PROGRAMMING LANGUAGE

Python is an open-source programming language that is utilized in web programming, information science, man-made consciousness, and numerous logical applications. Learning Python permits the software engineer to zero in on taking care of issues, instead of zeroing in on sentence structure. Its overall size and worked on linguistic structure give it an edge over dialects like Java and C++, yet the plenitude of libraries gives it the force expected to get incredible things done.

Python is a broadly utilized universally useful, significant level programming language. It was made by Guido van Rossum in 1991 and further created by the Python Software Foundation. It was planned with an accentuation on code lucidness, and its punctuation permits developers to communicate their ideas in less lines of code. Python is a programming language that allows you to work rapidly and incorporate frameworks all the more proficiently. There are two significant Python renditions: Python 2 and Python 3. Both are very unique.

4.2.2 SOFTWARES USED

Visual Studio Code is a source-code editorial manager made by Microsoft for Windows, Linux and macOS. Features incorporate help for troubleshooting, linguistic structure featuring, keen code fulfillment, scraps, code refactoring, and inserted Git. Clients can change the topic, console alternate ways, inclinations, and introduce expansions that add extra usefulness. Microsoft has delivered the greater part of Visual Studio Code's source code on the Microsoft/VS Code vault of GitHub utilizing the name, under the lenient MIT License, while the deliveries by Microsoft are restrictive freeware.

4.2.3 FRAMEWORK/LIBRARIES USED

- **Flask**- Flask is a miniature web system written in Python. It is named a microframework on the grounds that it doesn't need specific devices or libraries. It has no information base reflection layer, structure approval, or whatever other segments where prior outsider libraries give normal capacities. In any case, Flask upholds expansions that can add application includes as though they were executed in Flask itself. Expansions exist for object-social mappers, structure approval, transfer dealing with, different open validation advances and a few normal system related devices. To access the Yahoo Finance API, we have used the Flask Services to get the data from the financial website and store it on our server.

- **Matplotlib**- Matplotlib is a plotting library for the Python programming language and its mathematical arithmetic augmentation NumPy. It's anything but an article situated API for installing plots into applications utilizing universally useful GUI tool. Matplotlib is a cross-stage, information representation and graphical plotting library for Python and its mathematical augmentation NumPy. Accordingly, it's anything but a reasonable opensource option in contrast to MATLAB. Engineers can likewise utilize matplotlib's APIs to insert plots in GUI applications. A Python matplotlib script is organized so a couple of lines of code are everything necessary in many occasions to create a visual information plot. The matplotlib prearranging layer overlays two APIs: The pyplot API is a progressive system of Python code objects beat by matplotlib.pyplot, An Object-Oriented API assortment of items that can be amassed with more prominent adaptability than pyplot. This API gives direct admittance to Matplotlib's backend layers.

- **Keras-** Keras is an open-source programming library that gives a Python interface to counterfeit neural organizations. Keras goes about as an interface for the TensorFlow library. Keras contains various executions of normally utilized neural-network building squares like layers, targets, actuation capacities, analyzers, and a large group of devices to make working with picture and text information simpler to improve on the coding important for composing profound neural organization code. The code is facilitated on GitHub, and local area support gatherings incorporate the GitHub issues page, and a Slack channel. Notwithstanding standard neural organizations, Keras has support for convolutional and intermittent neural organizations. It upholds other normal utility layers like dropout, group standardization, and pooling. Keras permits clients to productize profound models on cell phones on the web, or on the Java Virtual Machine. It additionally permits utilization of circulated preparing of profound learning models on groups of Graphics handling units and tensor handling units.

4.3 OUTPUT

Perform following steps to get the output: a) Using stock's ticker image, it gets the most recent two years information. b) Provide the information to the system. c) Train the system using the RNN algorithm. d) System will foresee the yield. When we procure a dataset, we plan to partition it into two subsets: Training set is a subset of the dataset used to assemble prescient models. Test set or inconspicuous models is a subset of the dataset to survey the probable future presentation of a model. On the off chance that a model fit to the preparation set far superior to it fits the test set, over fitting is presumably the reason. After the pre-processing is done the output graph is printed for the actual and predicted values and for the last two years to show the pattern of stock prices of that company and the next 7 days stock values are displayed.

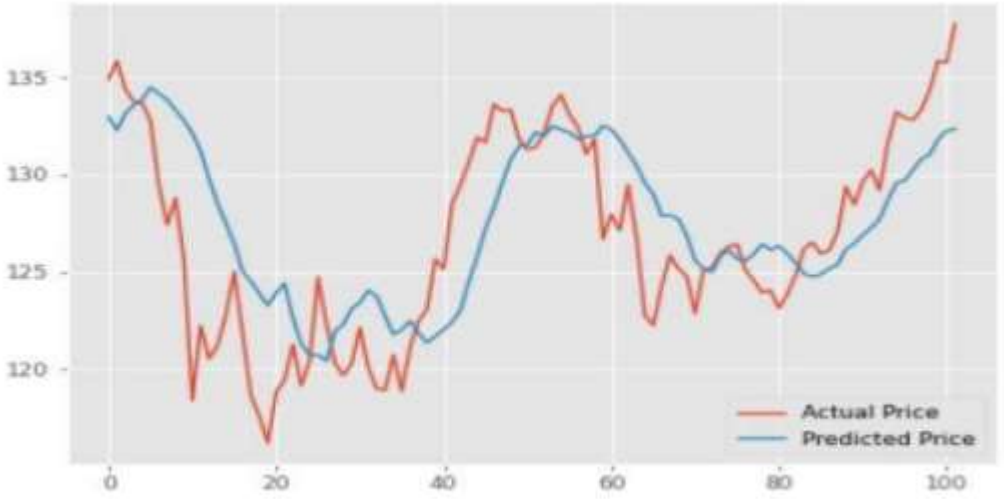

Fig -1: Test case analysis of Apple Stock

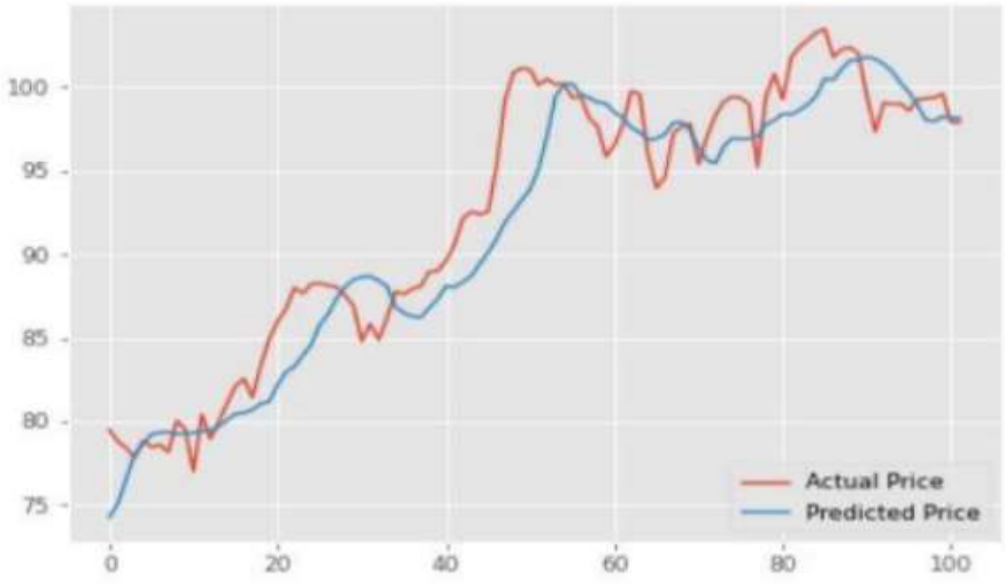

Fig -2: Test case analysis of Dell Stock

4. CONCLUSIONS

In this study, we have worked on Recurrent Neural Network (RNN) and Long Short-Term Memory (LSTM) methods to forecast the stock market trends and patterns and created an interface that user can use to select a company to see the next 7 days stock value for certain companies and can decide whether to go with buy or sell the certain stock. In future this might turn out to be a game changer in stock market. Everyone might refer to this idea before investing in the stock market. People can be sure that they will be getting something in return and they won't be at a loss. The project can be expanded to add functionalities like detecting news trends of various stocks. The financial news related to stock market can be displayed in the dashboard interface.

5. ACKNOWLEDGEMENT

Aside from the endeavors of ourselves, the accomplishment of any project relies generally upon the support and rules of numerous others. We make a move to communicate our appreciation to individuals who have been instrumental in the effective consummation of this undertaking. We might want to show our most noteworthy appreciation to Dr. Saravana Kumar R. We thank him for his colossal help and help. The direction and support got from every one of the individuals who contributed and who are adding to this venture, was fundamental for the achievement of the undertaking.

6. REFERENCES

[1] K. Hiba Sadia, Aditya Sharma, Adarsh Paul, Sarmistha Padhi, Saurav Sanyal "Stock Market Prediction Using Machine Learning Algorithms" International Journal of Engineering and Advanced Technology (IJEAT) ISSN: 2249 – 8958, Volume-8, Issue-4, April 2019.

[2] Min Wen 1, Ping Li1, Lingfei Zhang2, and Yan Chen1 "Stock Market Trend Prediction Using HighOrder Information of Time Series" School of Computer Science, Southwest Petroleum University, Japan, IEEE Access, received: January 18, 2019, accepted February 21, 2019, date of publication February 26, 2019.

[3] Yuling Lin, Haixiang Guo, Jinglu Hu "SVM based approach for stock market trend prediction" Graduate School of Information, Waseda University, JAPAN, 2017

[4] Dharmaraja Selvamuthu, Vineet Kumar and Abhishek Mishra "Indian stock market prediction using artificial neural networks on tick data" Department of Mathematics, Indian Institute of Technology Delhi, India,2019.

[5] Mojtaba Nabipour, Pooyan Nayyeri, Hamed Jabani, Shahab S. and Amir Mosavi "Predicting Stock Market Trends Using Machine Learning and Deep Learning Algorithms Via Continuous and Binary Data; a Comparative Analysis" Faculty of Mechanical Engineering, Tarbiat Modares University, Tehran, Iran, School of Mechanical Engineering, College of Engineering, University of Tehran, Iran,2018.

[6] Pawan Kumbhare; Rohit Makhija; Hitesh Raichandani "Predicting Stock Market Trends Using Decision Trees and Support Vector Machines and analysing how these algorithms performs", Santa Clara University,2017.