A Modern Approach Towards Stock Prediction Over Traditional Methods

[1st] Mr. Omkar Kadam
School of Computer Science
MIT World Peace University
Pune, India

[2nd] Mr. Krishna Gandhi
School of Computer Science
MIT World Peace University
Pune, India

[3rd] Mr. Dharamveersinh Rana
School of Computer Science
MIT World Peace University
Pune, India

***Abstract*- Prediction of stock prices have always been a buzz word when it comes to Machine Learning (ML). Precise prediction of future stock prices would help an investor to come up with a firm and confident decision regarding in which market segment to invest. ML techniques have taken over the traditional method of predicting stock market whose accuracy was not substantial, which in turn resulted in rise to the risk factor. The existing research papers focus on analysis or prediction of stock prices which do not consider political tenures, financial budgets which affects the fluctuations of stock market. Our proposed model will help an investor or any individual from a non-financial background to make secured investments. Our model will make use of Support Vector Machines (SVM) for classification, Long Short-Term Memory (LSTM) which will help us to classify and predict time-series data as the proposed model should not lag in terms of periodic factors. The proposed model will be trained with last 20 years of data making the model more robust. Data mining techniques will also be used to get the hidden insights company-wise and will be able to calculate the risk of market growth. Thus, the above proposed model will help earn trust of people in investing into stocks confidently.**

Keywords: Stock Prices, Machine Learning, Support Vector Machines (SVMs), Long Short-Term Memory (LSTM).

I. INTRODUCTION:

In Today's time Investments are considered to be most important factor in humans' life.

When it comes to investments, one of the most asked question raised "Where to invest??", "How much to invest ??" and "Will it be secured??". Also thinking about security "Investments in share market are considered to be risky". There are many financial portfolio management firms which help you take the right steps for fruitful investment, and ending up taking a chunk of your profits. With increase in your investment that chunk increases too.

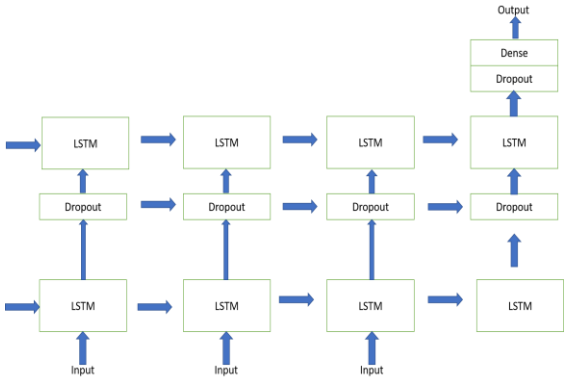
Three different models have been used to get the more accuracy out of it. LSTM have been used to cover the time forecasting factor and other regression problems have been covered using normal linear regression, Lasso regression and ridge regression, and then finally SVM to classify in what market group a company may fall.

II. METHODOLOGY:

Following are the methods used for developments proposed model:

A. LSTM (Long Short-Term Memory):

As the time factor plays an important role in prediction of stock values, therefore along with the linear regression use of neural network specifically LSTM (long short-term memory) will be used. As LSTM and Recurrent neural network will result good in working with time series also they will help in remembering the past. LSTMs are similar to small neural networks which are premeditated for memory in a bigger neural network which will be achieved through the use of a periodic node inside the LSTM cell. In the proposed model, every feed forward iteration cell will hold onto information from the previous information from the previous step along with the earlier steps. Since the looping connections weight is one, the past memories wouldn't diminish over time similar to traditional RNNs.



LSTM consists of three gates namely the update gate; forget gate and the output gate respectively. Out of which the first and second gate ensures that each element in the memory is updated and the output gate governs the amount of information to output as activation to the next level. LSTM takes a series of vectors of time series data as input X = [x1, x2, x3, ⋯, xt] and outputs a vector value figured by the neural network structures in the model's cell. The proposed model consists of 2 layers of LSTM components, first is, dropout layer in-between to avoid over-fitting.

B. Linear Regression:

Linear regression is the relation between two variables that is dependent and two or many independent variables. We are going to use Linear Regression for modelling the last "N" values which will ultimately use to predict the value of present-day date.

C. ARMA Model:

This Model, which is constantly used with Linear Model for time series where the upcoming price is assumed as a linear combination of historical error and historical values. Stock midterm prediction problem is used to set by ARMA. Let XA t be the variable based on ARMA at time t, and then we have:

XA t = f1 ({(X)t − i} pi=1) = μ +p(X)i = 1φiXt−i –q(X)i = 1ψjt−j + t

where Xt−i denotes the previous value at time t−i; t means the random error at time t; φi and ψj are the coefficients; in this constant Is μ; and integers are p and q respectively which are also frequently referred as "autoregressive and moving average polynomials".

D. SVM (Support Vector Machine):

In the proposed model SVM is used to classify whether a specific company comes under what scrip group in upcoming days using which will segregate the companies in different groups. This uses daily closing prices to estimate price unpredictability and impetus for distinct stocks and for global sector which will be used as factors to the SVM model. It will then try to forecast whether a stock rate will fluctuate or not on a given précised date. For short run we find it little predictive ability but positive predictive ability for long duration.

III. DATA DESCRIPTION:

The data used for the proposed model has being retrieved from the BSE (Bombay Stock Exchange) repository for last 20 years which will help the model to understand the economic fluctuation and political changes which tends to market change.

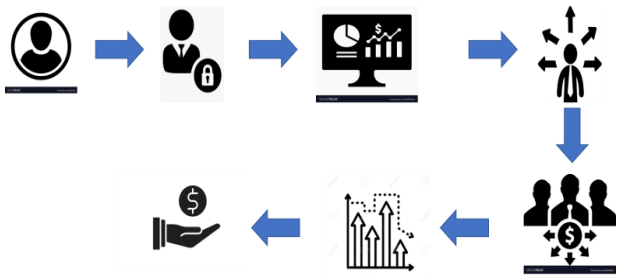
Following picture shows the overview of the dataset for single day (post market closure) used for developing the proposed model:

	A	B	C	D	E	F	G	H	I	J	K	L	M
1	SC_CODE	SC_NAME	SC_GROUP	SC_TYPE	OPEN	HIGH	LOW	CLOSE	LAST	PREVCLOSE	NO_TRADES	NO_OF_SHRS	NET_TURNOV
2	500002	ABB LTD.	A	Q	1318.2	1338.75	1312.55	1320.4	1323.55	1332	255	1816	2407362
3	500003	AEGIS LOGIS	A	Q	204.75	205.7	200.6	203.9	203.9	204.2	294	5672	1150504
4	500008	AMAR RAJA BA	A	Q	743.45	747.5	737.6	739.65	739.65	742.75	253	7307	5424260
5	500009	A.SARABHAI	X	Q	13.6	13.75	13	13.43	13.44	13.24	41	12887	172463
6	500010	HDFC	A	Q	1973.55	2018.6	1956	2009.6	2009.6	1970	2531	320697	630395875
7	500012	ANDHRA PETRO	X	Q	57	58.9	56.5	58.2	58.9	57.4	250	37525	2161436
8	500013	ANSAL INFRAS	B	Q	11.9	11.94	11.3	11.9	11.9	11.72	15	2663	31391
9	500020	BOM DYEING	B	Q	114.85	117.25	113.5	115.6	115.6	113.8	1198	117726	13634443
10	500023	ASIANHOTNR	B	Q	144	147	144	147	147	150	2	14	2037

SC stands for scrip in the following dataset

1) SC_CODE: Company Code.
2) SC_NAME: Company Name.
3) SC_GROUP: Scrip part of which group.
4) SC_TYPE: Scrip category – Equity, Preference, Debentures, Bond.
5) OPEN: Opening price of specific Company on the respective day.
6) HIGH: Highest price specific company has reached on that respective day.
7) LOW: Lowest price specific company has reached on that respective day
8) CLOSE: Closing price of specific company on that respective day.
9) LAST: The last trade price of the stock.
10) PREVCLOSE: The closing price on previous day.
11) NO_TRADES: Total Number of trades which took places on the respective day.
12) NO_OF_SHARES: Total number of shares transacted on a respective day.
13) NET_TURNOVER: The net turnover of the company after the closing bell on the specific day.

IV. WORKFLOW:



Work-Flow of Proposed Model

1) The user logs into the system.
2) Upon logging in user will be seeing his/her personal dashboard.
3) On the dashboard user may choose from investment types such as for Retirement, Marriage, Education, Property, Car, etc.
4) Now the user will be asked with his/her preferences regarding the investment which he/she is desiring to make.
5) Based on the preferences entered by the user the model will suggest a list of companies which fulfil those preferences.
6) On choosing the company from the suggested list user will be displayed with the predictions regarding the company and the returns expected.

V. ASSUMPTIONS:

Main Disadvantage of this model can be market situations. For example, a reputed company "XYZ" might have good records in last few years, the company might end up in a loss base on some inside company decisions, so such type of prediction can't be made by our model. Limitation can be internet availability too, as the prediction is based on time series fluctuation of the time may occur loss to people.

VI. FUTURE SCOPE:

Will provide one user interface via which it will look more user friendly to people and more transparency to them. As he/she will see the trends of market and can provide payment gateway for investing into specific scope.

CONCLUSION:

We here by conclude that, stock market is however good choice of investment as our model will use various algorithms to predict its price. Not only that the user will enjoy the profit as we will have 20 years of data which will involve all the factors which will or can affect the stock market prices. User will try to invest without any fear of losing money and can enjoy the benefits of stock market.

REFERENCES:

[1] G. Francis A. N. Refenes, A. Zapranis. Stock performance modeling using neural networks: A comparative study with regression models. Neural Networks, 7(2):375–388, 1994.
[2] Saahil Madge: Independent Work Report Spring 2015
[3] S. Choudhury, S. Ghosh, A. Bhattacharya, K.J. Fernandes, M.K. Tiwari, "A real time clustering and SVM based price-volatility prediction for optimal trading strategy"
[4] Zhen Hu, Jibe Zhu, and Ken Tse "Stocks Market Prediction Using Support Vector Machine", 6th International Conference

on Information Management, Innovation Management and Industrial Engineering, 2013.M.

[5] K. Hiba Sadia, Aditya Sharma, Adarrsh 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

[6] C.-L. Huang, C.-Y. Tsai,”A hybrid SOFM-SVR with a filter-based feature selection for stock market forecasting”S, Expert Syst Appl, 36 (2) (2009), pp. 1529-1539

[7] Bruno Miranda Henrique, Vinicius AmorimS obreiro, Herbert Kimura,”Stock price prediction using support vector regression on daily and up to the minute prices”, The Journal of Finance and Data Science , Volume 4, Issue 3, September 2018, Pages 183-201

[8] XinyiLi, YinchuanLi, HongyangYang, LiuqingYang1, Xiao-YangLiu,” DP-LSTM:DifferentialPrivacy-inspiredLSTMfor StockPredictionUsingFinancialNews”

[9] Hyeong Kyu Choi,"Stock Price Correlation Coefficient Prediction with ARIMA-LSTM Hybrid Model"

[10] G.P. Zhang. Time series forecasting using a hybrid arima and neural network model. Neurocomputing, 50:159–175, 2003