IOSR Journal of Engineering (IOSRJEN) **www.iosrjen.org**
ISSN (e): 2250-3021, ISSN (p): 2278-8719
Volume 13, PP 06-10

Data Adaptive Stock Recommendation

Mayank H. Mehta[1], Kamakshi P. Banavalikar[2], Jigar M. Parekh[3,] Jeet N. Khetan[4], Siddharth Maheshwari[5]

[1,2,3]*Student, Department of Computer Engineering, Atharva College of Engineering, Mumbai, India*
[4,5]*Bachelors of Engineering, Mumbai, India*

Abstract : *The volatile nature of the stock market makes it difficult to guess the movement of stocks. These fluctuations make forecasting complicated. Results of the company in a financial period(quarterly, half-yearly or annual), as well as the trend in stocks of peer companies of the sector and the market in general, contribute to these movements. This paper proposes using historical data to understand movements in stocks using various technical indicators. With the help of these indicators, Technical Analysis is done using Support Vector Machine (SVM) to analyze the relationship of these factors and predicting the stock performance.*

Keywords : *Scikit, Stock prediction, Stock technical indicator, Support vector machine.*

I. Introduction

A stock is a share in the ownership of a company, which represents a claim on the company's earnings and assets. As people acquire more stock, their ownership stake in the company becomes greater. The stock exchange is a market for the exchange of securities. Stock prices change every day as a result of market forces (demand and supply). If more people want to buy a stock (demand) than selling it (supply), then the price moves up. If more people want to supply (sell) than demanding it (buying it), then the price moves down.

One of the most widely studied and challenging problem is stock price prediction, attracting researchers from many fields including economics, history, finance, mathematics, and computer science. The volatile nature of the stock market makes it difficult to apply simple time-series or regression techniques. Traders and financial institutions have created various proprietary models to try and beat the market for themselves or their clients, but rarely has anyone achieved consistently higher-than-average returns on investment. Additionally, the challenge of stock market prediction is fascinating, as with increase of just a few percentage the profit increases by millions of dollars for such institutes[1].The challenge of stock forecasting is appealing because an increase of a few basis points can bring about a profit of millions of dollars. Forecasting provides concrete data for investment decisions. There is an investment theory, efficient market hypothesis (EMH) that states it is impossible to "beat the market" because stock market efficiency causes existing prices of the share to always incorporate and reflect all relevant information. According to the EMH, stocks always trade at their fair value on stock exchanges, making it impossible for investors to either purchase undervalued stocks or sell stocks for inflated prices.[2] . In this paper, we focus on a specific machine learning technique known as Support Vector Machines (SVM). Our goal is to use SVM to predict whether a given stock's price is higher or lower on the day. These required parameters are calculated using daily closing prices for each stock from the years 2004 through 2017. We analyze whether this historical data can help us predict price direction.

Support Vector Machines are one of the best binary classifiers. They create a decision boundary such that most points in one category fall on one side of the boundary while most points in the other category fall on the other side of the boundary. We finally conclude that the accuracy of our prediction range from 70-80%.

II. Literature Survey

2.1.Stock Trend Prediction with Technical Indicators using SVM[3]

This paper proposes forecasting in the short term. According to this paper the short term is defined as 1 to 10 days. The technical indicators are analyzed using features. The prediction yields 60% accuracy for next day predication and greater than 70% for 3-5 days. This paper helps us in getting the technical indicators.]

2.2.Predicting Stock Price Direction using Support Vector Machine [4]

This paper proposes forecasting for the long term. It uses economic fundamentals such as inflation in its computation. It also takes into account the Federal Funds rate that incorporates monetary stance. The Dow Jones Industrial Average (DJIA) and it's movement is taken into account. It gets an accuracy of about 70%. This paper helped us in understanding the SVM algorithm.

2.3.Stocks Market Prediction Using Support Vector Machine[5]

This paper forecasts for the long term as well but instead of economic parameters uses factors such as closing prices. It also uses indicators such as price volatility and momentum. This paper is able to achieve accuracy in the range of 55 to 60%. This paper helps us in understanding SVM and its classification in detail.

2.4.Stock Price Prediction Based on Financial Statements Using SVM [6]

This paper forecasts the price movement using financial statements. It uses parameters such Earnings Per Share (EPS), Book value Per Share (BPS) and Net Profit Growth Rate (NPGR). On a 1 month basis it achieves an accuracy of 57.5% while that for 2 months is 50.1%. This paper helps us in understanding the Data set and the Financial statements.

III. Methodology

3.1.Support Vector Machines(SVM)

Vladimir N. Vapnik and Alexey Ya. Chervonenkis proposed the original SVM algorithm in 1963. SVM is a supervised learning algorithm that can be used for solving classification and regression problems. The data set is split into two parts. The first part of the data is used for training and the other part for testing. Each data item is plotted as a point in a N-Dimensional plane where the value of each feature is the co-ordinates of that point. Points are then distributed in this space. SVM model is mapped so that the examples of the separate categories are divided by a clear gap that is as wide as possible.

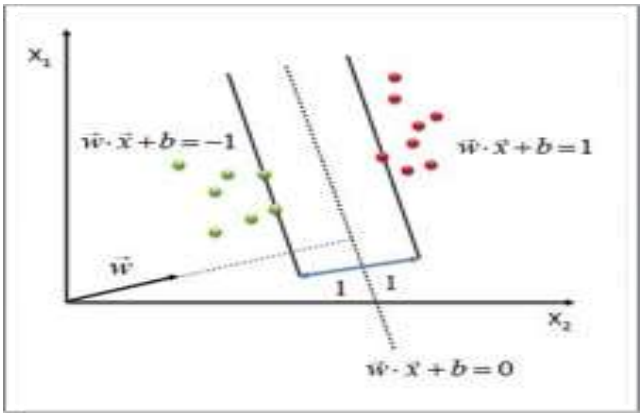


Fig. 1. An example of Support Vector Machines

Considering a 2-Dimensional feature space, the hyperplane is a straight line. The line equation is given by
: y = ax + b
The equation of hyperplane is given by dot product of two vectors
: wt = (-b -a 1) and : xt = (1 x y)
: g(x)= **w**t.**x** + w**o**
where wo is *-b* , which is intersection of line with axes.

Given a data set with n vectors xi (p-dimension) and value yi indicating if element belongs to class(+1) or not(-1).
: D = {(**x i , y i) | x i** ∈ **Rp , y i** ∈ **{-1,1}** } **i=1n**

Find two hyperplanes separating the data points with no data points between them.
: (**w** . **x** + b) = 1, for all x of class 1-HyperPlane 1
: (**w** . **x** + b) = -1, for all x of class 2-HyperPlane 2

Maximise the distance between two hyperplanes such that for each vector **x i**
: (w.x + b) >= 1
: (w.x + b) <= -1

Generalising the problem becomes
: **y i(w.x i +b) >= 1** for all 1<=i<=n

The goal of a support vector machine is to find the optimal separating hyperplane which maximizes the margin of the training data. Support vector machine constructs a hyperplane which can be used for classification,

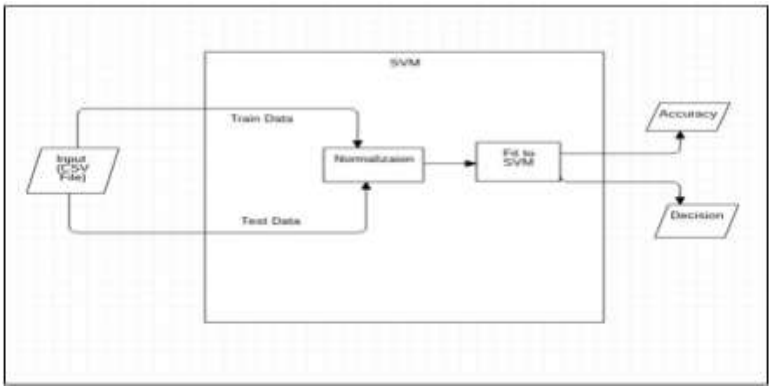
regression, or other tasks. Intuitively, a good separation is achieved by the hyperplane that has the largest distance to the nearest training-data point of any class, since in general the larger the margin the lower the error of the classifier.[8]

3.2. DataSet

The dataset for the project was pulled from yahoo finance. The stocks taken in the project are (TCS, WIPRO, INFOSYS, HCL Tech, Tech Mahs) having a time span from 25-08-2004 to 01-02-2017. The data included the following attributed:[7]

- Date
- Open
- Close
- High
- Low
- Volume
- Adjusted Close

3.3. Block Diagram



3.4. Approach

Based on the EMH, the price of the stock day before is going to have the most impact on the price of the stock on the present day. Thus as we go along the timeline, data-points which are nearer to today's price point are going to have a greater impact on today's price. Similarly the historical data-points will have lesser impact on today's price.Indicators are independent trading systems introduced to the world by successful traders. They are built on preset logic using which traders can supplement their technical study to arrive at a trading decision. Indicators help in confirming trends, sometimes predicting trends and even in buying and selling. The following technical indicators along with the directional movement of BSE index and USD-INR index were obtained and taken as the attributes of the data_frame. We have taken Exponential moving average to label each tuple of data-frame[9]. The raw dataset so obtained was used to generate a data frame. This was done using technical formulae of the indicators mentioned in Table 1. [2]

Sr. Number	Parameter	Explanation
1	Volume	Represents the number of shares that changed hands during a day's trade.
2	Momentum	Acceleration in stocks price. It may move up or down.
3	BSE Momentum	Acceleration of stocks in the Bombay Stock Exchange.
4	ROCR	Momentum oscillator that measures percent change in price
5	SMA	Computed by adding the closing price for a number of time periods and then dividing this total by the number of time periods.
6	RSI	Speed of change of price movements

7	CCI	An oscillator used in technical analysis to help determine when an investment vehicle has been overbought and oversold.
8	EMA	moving average that is similar to a simple moving average, except that more weight is given to the latest data.
9	ATR	Is an indicator that measures volatility
10	MFI	an oscillator that uses both price and volume to measure buying and selling pressure.
11	WILLR	normalizes price as a percentage between 0 and 100.
12	OBV	momentum indicator that uses volume flow to predict changes in stock price.
13	USD_INR	The conversion value of Dollar to Rupee

Table 1: List of technical parameters

The above technical indicators cover different type of features:

1) Price change – ROCR, Momentum
2) Stock trend discovery – MFI
3) Buy&Sell signals – WILLR, RSI, CCI
4) Volatility signal – ATR
5) Volume weights – OBV

The formulas for the different technical indicators are :

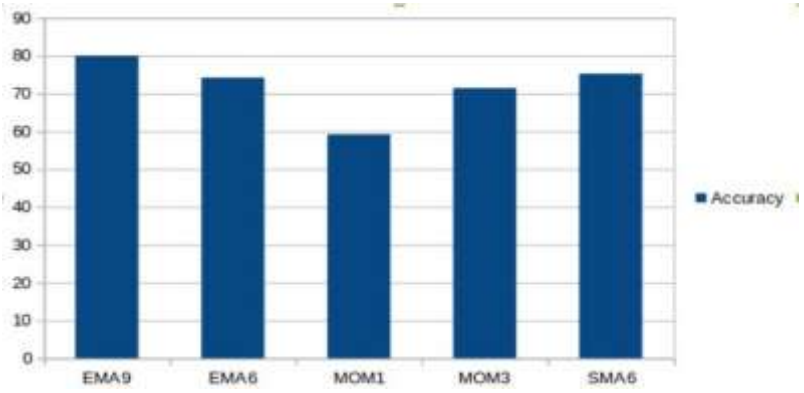
Momentum = Price(t) - Price (t-n)
ROCR = (Price(t)/Price(t-n))*100
RSI = Avg(PriceUp) / (Avg(PriceUP) + Avg(PriceDown) * 100
Where : PriceUp(t) = 1 * (Price(t)-Price(t-1))
PriceDown(t)= 1 * (Price(t-1)-Price(t))
ATR(t) = ((n-1) * ATR(t-1) + Tr(t)) / n
where Tr(t)=Max(Abs(High-Low), Abs(High-Close(t-1)), Abs(Low-Close(t-1));
WILLR = (highest-closed)/(highest-lowest)*100. [2]

IV. Result And Analysis

The data frame was pre processed and scaled before training and testing. To validate the prediction accuracy, we trained on 65% of the historical data and predicted on the 35% of the recent data.The prediction was carried on certain technical stocks and the accuracy for the respective stock are as follow :

- TCS - 74.78%
- WIPRO - 79.48%
- INFOSYS - 78.98%
- HCL Tech - 74%
- Tech Mah - 77.37%

Considering the stock of Infosys and testing it with different labels we get Exponential moving average as the optimum label for prediction.



Based on the the predicted result and available through the label the following confusion matrix is constructed: A total of 1008 data set is considered for the matrix.

	TRUE	FALSE
+VE	423	185
-VE	26	374

V. Future Scope

This project may be furthered by incorporating more parameters and more complex parameters that produce better accuracy. An index based on the probability of success and expected upside may be devised. In future one may use Kelley's formula for allocation (number of shares to purchase) of any stock and recommend to the user.

VI. Conclusion

The choice of the indicator function can drastically improve/reduce the accuracy of the prediction system. Also a particular Machine Learning Algorithm might be better suited to a particular sector of stock, say IT Stocks, whereas the same algorithm might give lower accuracies while predicting some other types of Stocks, say Energy Stocks. Moreover, while applying, we should also note the Machine Learning Algorithm for Technical Analysis. We assumed that the effect of the Unknown Factors (Political Effects, Election Results, Rumors etc.) was already embedded into the historical stock pattern. By learning from the past data we are able to get above 70% accurate prediction on the next couple days trend. For future work, it is worth adding sentiment data as feature to the technical feature. The challenge is how to eliminate as much noise in sentiment data and quantify them.

References

[1] Stocks Basics: What Are Stocks? | Investopedia http://www.investopedia.com/university/stocks/stocks1.asp#ixzz4cUmf5 lUa
[2] EfficientMarketHypothesis–EMH http://www.investopedia.com/terms/e/efficientmarkethypothesis.asp
[3] Stock Trend Prediction with Technical Indicators using SVM, Xinjie Di,SCPD, Apple Inc
[4] Predicting Stock Price Direction Using Support Vector Machine, Saahil Madge, Independent Work Report Spring 2015
[5] Stock Market Forecasting Using Machine Learning Algorithms, Hu, Zhu, Tse, 2013 6th International Conference on Information Management, Innovation Management and Industrial Engineering
[6] Stock Price Prediction Based on Financial Statements Using SVM, Heo and Yang, International Journal of Hybrid Information Technology Vol.9, No.2 (2016), pp. 57-66
[7] https://in.finance.yahoo.com/
[8] http://www.svm-tutorial.com/2014/11/svm-understanding-math-part-1/
[9] http://zerodha.com/varsity/module/technical-analysis/