STOCK MARKET PREDICTION

REVIEW OF LITERATURE

MACHINE INTELLIGENCE

BACHELOR OF TECHNOLOGY

Department of Computer Science & Engineering

V Semester Section - A

SUBMITTED BY

Batch No : 05

Student name 1 : Abhishek C Hiremath SRN : PES2UG20CS011

Student name 2 : Abhishek T SRN : PES2UG20CS013

Student name 3 : Akshay C SRN : PES2UG20CS032

Paper 1: **Stock Market Forecasting Using Machine Learning Algorithms**

Shunrong Shen, Haomiao Jiang
Department of Electrical Engineering
Stanford University
{conank,hjiang36}@stanford.edu

INTRODUCTION:

Researchers from several professions have spent a great deal of time studying the prediction of stock movement. The potential of machine learning, an established method with many other uses, to forecast financial markets has been thoroughly investigated. Support vector machine (SVM) and reinforcement learning are two well-known algorithms that have been shown to be highly good in tracking the stock market and assisting in maximising the profit of stock option purchases while minimising risk [1-2].

In this research, we suggest using data from various financial products along with data on stocks from around the world as the input features for machine learning algorithms like SVM. The correlation between the closing prices of the markets that finish trading just before or at the start of US markets is of particular interest to us. Globalization has tightened the linkages between national economies, making domestic financial market disruptions more likely to come from abroad.

ALGORITHMS:

- **Feature selection**

In this project, we focus on the prediction of the trend of stock market (either increase or decrease).
Therefore, the change of feature over time is more important than the absolute value of each feature. We define xi(t), where $i \in \{1,2,..., 16\}$, to be feature i at time t. The feature matrix is given by

$F = (X1, X2, ..., Xn)^T$

where

$X_t = (x1(t),x2(t),...,x16(t))$

The new feature which is the difference between two daily prices can be calculated by

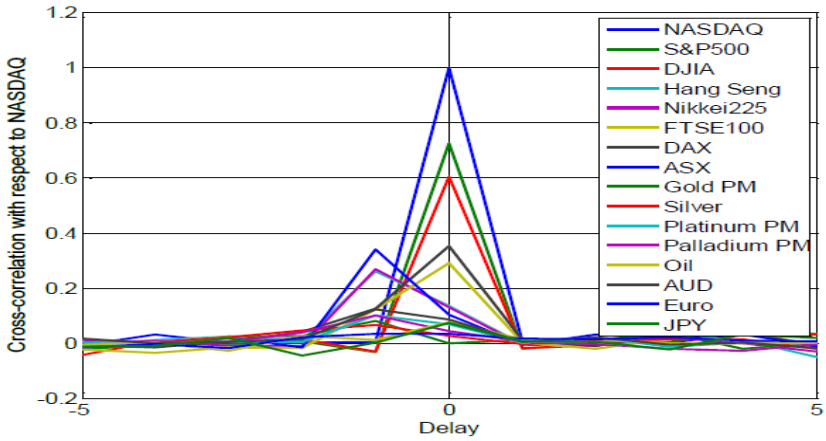
$\nabla\delta F = (\nabla\delta X(\delta + 1), \nabla\delta X(\delta + 2), ..., \nabla\delta X(n))$

The normalization can be implemented as:

$$normal(X(t)) = \frac{\mathcal{N}(\nabla_\delta X(t))}{|\mathcal{N}(\nabla_\delta X(t))|}$$

As discussed above, the performance of a stock market predictor heavily depends on the correlation between the data used for training and the current input for prediction. Intuitively, if the trend of stock price is always an extension to yesterday, the accuracy of prediction should be fairly high. To select input features with high temporal correlation, we calculated the autocorrelation and cross-correlation of different market trends (increase or decrease).

The results shown in Figure use NASDAQ as the base market.



EXPERIMENTAL RESULTS:

- **Single Feature Prediction**

We used cross-correlation to estimate the importance of each feature. To verify the information given by correlation analysis, we use individual feature to predict daily NASDAQ index trend. The prediction accuracy by each single feature is shown in Figure.

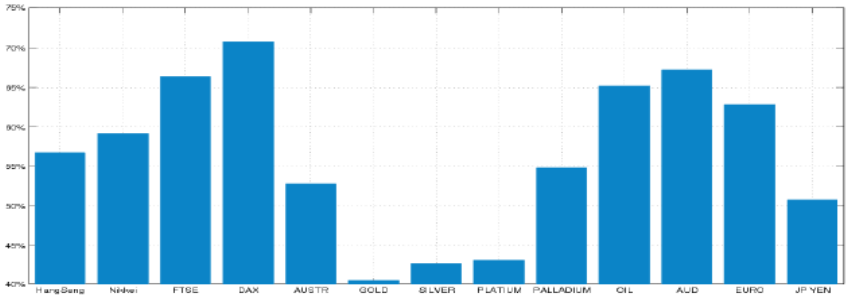
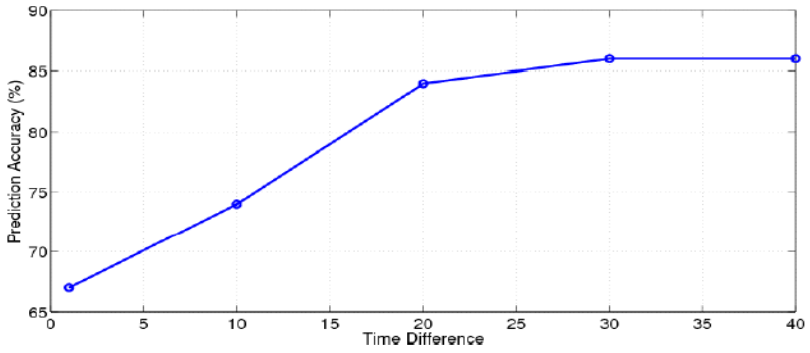


Fig. 4. Prediction accuracy by single feature

We can see that DAX yields the best results, 70.8% accuracy of prediction. Prediction accuracies of Australian dollar, FTSE and oil price are also relatively high, reaching 67.2%, 66.4% and 65.2% respectively.

- **Long Term Prediction**

Here, we define our problem as predicting the sign of difference between tomorrow's index value with respect to that of certain days ago. We use SVM as a training model and the prediction accuracies under different time spans are shown in Figure



- Multi Term Prediction

  We compare the prediction accuracies of SVM algorithm and MART (a decision tree based boosting algorithm). The prediction results are shown in Table

	Top 4 Features	All Features
SVM	74.4%	63.1%
MART	70.3%	73.9%

  We can see that accuracies from SVM and MART learner can reach as high as 74%. This daily trend prediction accuracy is higher than most of models and the values reported on financial analysis websites.
  In addition, we note that SVM algorithm is very sensitive to the size of training data. When the size of training set is not large enough, hyper-plane found by SVM algorithm might not be able to split the data properly. Thus, feature selection is essential when using SVM. In contrast, Multiple Additive Regression Trees (MART) algorithm requires less training data and prefers high dimensional feature set.

- Regression

  The precise increase in the stock index may offer more information for an investment strategy than the stock trend. The classification problem has thus transformed into a regression problem. We utilise the square root of mean square error (RMSE) as a metric to assess the effectiveness of our model.

$$\mathrm{RMSE} = \left(\frac{1}{N}\sum_{i=1}^{N}(\tilde{y}_i - y_i)^2\right)^{\frac{1}{2}}$$

  We use linear regression, generalized linear model (GLM) and SVM algorithm to predict exact value of daily NASDAQ movement.

https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.278.6139&rep=rep1&type=pdf

Paper 2: **Stock Market Prediction Using Machine Learning**

By V Kranthi Sai Reddy

INTRODUCTION:

The prediction of a stock using machine learning is explained in this study. The majority of stockbrokers employ technical, fundamental, or time series analysis when making stock predictions. Python is the computer language used to make stock market predictions using machine learning. In this article, we suggest a Machine Learning (ML) method that will be taught using the stock market data that is currently accessible, gain intelligence, and then use the learned information to make an accurate prediction. In this regard, this study used the Support Vector Machine (SVM) machine learning technique to forecast stock prices for the large and small capitalizations and in the three separate markets, using prices with both daily and up-to-the-minute frequencies.

METHODOLOGY:

Support Vector Machines:

A discriminative classifier that is officially described by the separating hyperplane is called a Support Vector Machine (SVM). The Support Vector Machine (SVM) algorithm is regarded as one of the best ones available for time series prediction. Both regression and classification can be done using the supervised technique. The SVM includes visualising data as a point in an n-dimensional space.

The hyper-plane is defined by:

$w.x+b=0$

The SVM decision rule is :

$$L = \sum_i \alpha_i - \frac{1}{2} \sum_i \sum_j \alpha_i \alpha_j y_i y_j \overrightarrow{x_{SV}i} \cdot \overrightarrow{x_{SV}j}$$

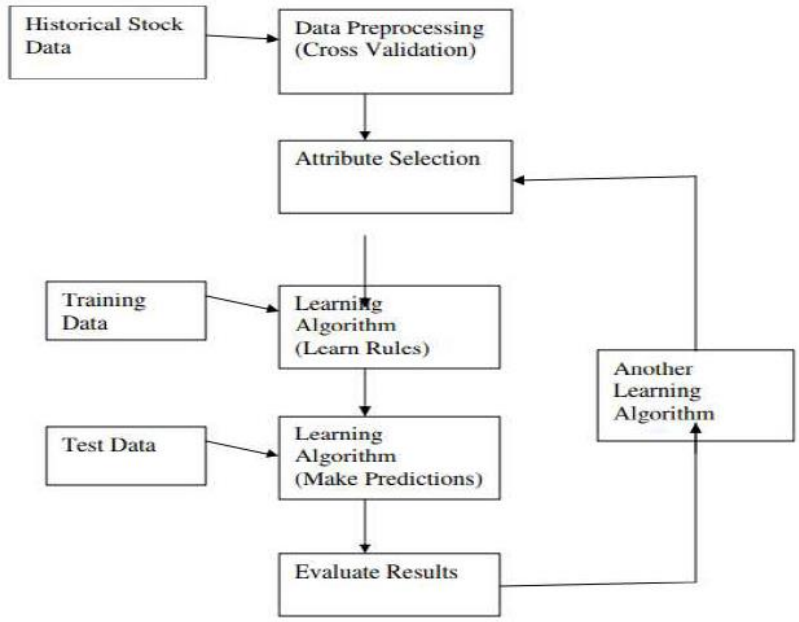
Radial Basis Function:

A radial basis function is a real-valued function whose value solely depends on how far it is from the origin or, alternately, how far it is from another point, known as the centre. A radial function is any function that satisfies the property.

The Radial basis function kernel, is also called as the RBF kernel, or Gaussian kernel, is a kernel that is in the form of a radial basis function (more specifically, a Gaussian function). The RBF kernel is defined as:

$$K_{RBF}(x, x') = \exp[-y ||x - x'||^2]$$

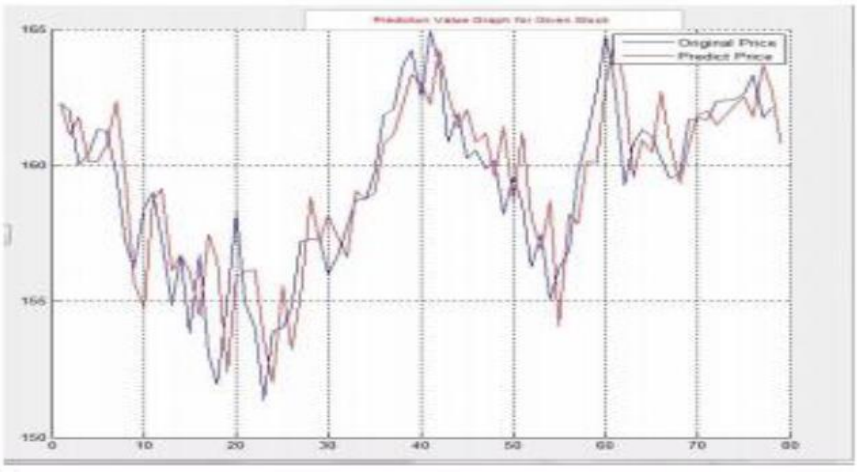
The Learning Environment:



MODEL CREATION AND EVALUATION METHODS:

In this paper we focus on predicting the Stock Market using Machine Learning model i.e., Support Vector Machine (SVM) by RBF kernel.

RESULTS:



https://www.academia.edu/download/57764282/IRJET-V5I10193.pdf

Paper 3: **Stock Market Prediction using Machine Learning: A Systematic Literature Review**

By: Siddhartha Vadlamudi

INTRODUCTION:

The application of machine learning techniques for stock value prediction is described in this study. It is determined which attributes can be used to train the algorithm in this manner. There is also discussion of several additional variables that may impact stock value.

The purpose of this literature review is to provide a thorough overview of the many machine learning algorithms that have been employed by various researchers. In this literature review, the work that various scholars have given will be discussed. The research question will be formulated and responded to with the intention of improving understanding. The papers from various sizable databases serve as the foundation for the literature review.

METHODOLOGY:

- **Inclusion and Exclusion**
  This criterion ensures that the papers are written in the English language, completely available to study and are relevant to stock market prediction. All the papers that do not fulfill this inclusion criteria were excluded.
- **Quality Assessment**
  The quality of the papers was assessed before including them in this literature review.

RESEARCH & DISCUSSION:

- **Necessity for Stock Prediction**
  The stock market is generating growing interest among people. To reap the rewards, they wish to invest their money. However, there is also a danger of losing the money invested or receiving a loss rather than a benefit. If a person invests in a stock whose future value will decline, they will suffer a loss. Predictions concerning the stock's future worth are necessary.

- **Attributes & Factors affecting the prediction**
  A prior data set is required to train the machine learning algorithm. The data in this data set may take the shape of different variables or properties. The cost of fuel, currency exchange, commodity exchange, and news are a few of these elements. The trend of the stock might also be influenced by many macroeconomic factors.

ALGORITHMS:

- **Linear Regression**

This algorithm for machine learning is within the supervised category. Instead of guessing the category, this algorithm predicts the values that fall inside a range. The prediction of daily stock value, according to the researchers, is a significant difficulty that must be overcome. The investors can invest their money with greater confidence thanks to this algorithm's prediction.

- **Neural Network**
  This approach for unsupervised machine learning consists of a number of nodes, each of which does a different computation. This method has been applied by several researchers to predict stock prices.

- **SVM**
  The SVM algorithm's objective is to establish the best line or decision boundary that can divide n-dimensional space into classes, allowing us to quickly classify fresh data points in the future. A hyperplane is the name given to this optimal decision boundary. SVM selects the extreme vectors and points that aid in the creation of the hyperplane. Support vectors, which are used to represent these extreme instances, form the basis for the SVM method.

- **K-Nearest Neighbour**
  By foreseeing the stock value at the outset, this algorithm can also assist investors in making the appropriate financial decisions. In the financial industry, the KNN can be used to forecast stock values and determine when it is ideal to buy and sell equities.

- **Random Forest**
  When investing in the stock market, this method can help to reduce risk. The accuracy this classifier offers is 83%, making it the best one for this kind of prediction.

CONCLUSION:

For investors, stock value forecasting is crucial. The investors can increase their profits by using an accurate prognosis. Investors have a difficult time deciding whether to buy and when to sell. Making this choice may benefit greatly from the algorithm provided by the machine learning algorithms. When investors are aware of the potential future worth of the stocks, they may take the risk of investing more easily. Various researchers have sought to forecast the stock's future values. In order to save the reader from having to read all the articles, this work is addressed in this paper.

REFERNCE:

https://www.abc.us.org/ojs/index.php/ajtp/article/view/521

PAPER – 4 : Machine learning techniques for stock prediction

BY : Vatsal H. Shah

https://bigquant.com/community/uploads/default/original/1X/5c6d3b9959a8556a533a58e0ac4568dfc63d6ff4.pdf

Introduction to Stock Market Prediction :

Recently, a lot of interesting work has been done in the area of applying Machine Learning Algorithms for analyzing price patterns and predicting stock prices and index changes. Most stock traders nowadays depend on Intelligent Trading Systems which help them in predicting prices based on various situations and conditions, thereby helping them in making instantaneous investment decisions.

Stock Prices are considered to be very dynamic and susceptible to quick changes because of the underlying nature of the financial domain and in part because of the mix of known parameters (Previous Days Closing Price, P/E Ratio etc.) and unknown factors (like Election Results, Rumors etc.)

An intelligent trader would predict the stock price and buy a stock before the price rises, or sell it before its value declines. Though it is very hard to replace the expertise that an experienced trader has gained, an accurate prediction algorithm can directly result into high profits for investment firms, indicating a direct relationship between the accuracy of the prediction algorithm and the profit made from using the algorithm.

Motivation behind the Project :

In this paper, we discuss the Machine Learning techniques which have been applied for stock trading to predict the rise and fall of stock prices before the actual event of an increase or decrease in the stock price occurs. In particular the paper discusses the application of Support Vector Machines, Linear Regression, Prediction using Decision Stumps, Expert Weighting and Online Learning in detail along with the benefits and pitfalls of each method.

The paper introduces the parameters and variables that can be used in order to recognize the patterns in stock prices which can be helpful in the future prediction of stocks and how Boosting can be combined with other learning algorithms to improve the accuracy of such prediction systems.

Stock Prediction in Detail :

Fundamental Analysis: Performed by the Fundamental Analysts, this method is concerned more with the company rather than the actual stock. The analysts make their decisions based on the past performance of the company, the earnings forecast etc.

Technical Analysis: Performed by the Technical Analysts, this method deals with the determination of the stock price based on the past patterns of the stock.

When applying Machine Learning to Stock Data, we are more interested in doing a Technical Analysis to see if our algorithm can accurately learn the underlying patterns in the stock time series.

The Efficient Market Hypothesis (EMH) :

The EMH hypothesizes that the future stock price is completely unpredictable given the past trading history of the stock. There are 3 types of EMH's: strong, semi-strong, and weak form. In the weak EMH, any information acquired from examining the stock's history is immediately reflected in the price of the stock.

The Random Walk Hypothesis :

The Random Walk Hypothesis claims that stock prices do not depend on past stock prices, so patterns cannot be exploited since trends to not exist.

Indicator Functions :

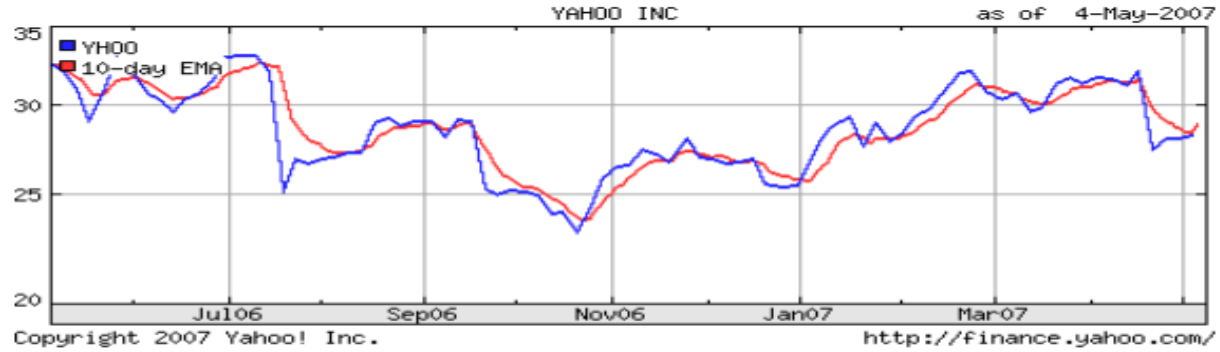
Indicators can be any of the following:
Moving Average (MA) : The average of the past n values till today.
Exponential Moving Average (EMA) : Gives more weightage to the most recent values while not discarding the older observation entirely.
Rate of Change (ROC) : The ratio of the current price to the price n quotes earlier. n is generally 5 to 10 days.
Relative Strength Index (RSI): Measures the relative size of recent upward trends against the size of downward trends within the specified time interval (usually 9 – 14 days).

For this Project, the EMA was considered as the primary indicator because of its ability to handle an almost infinite amount of past data, a trait that is very valuable in time series prediction (It is worth noting that the application of other indicators might result in better prediction accuracies for the stocks under consideration).

EMA (t) = EMA (t-1) + alpha * (Price (t) - EMA (t-1))
Where, alpha = 2/ (N+1), Thus, for N=9, alpha = 0.20

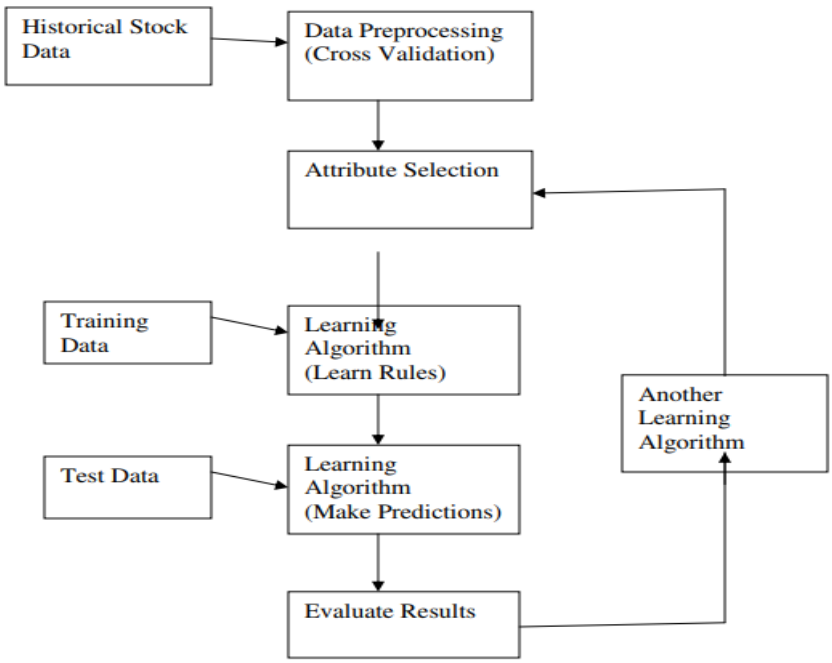
In theory, the Stock Prediction Problem can be considered as evaluating a function F at time T based on the previous values of F at times t-1,t-2,t-n while assigning corresponding weight function w at each point to F.

The technical analysis charts below show how the EMA models the actual Stock Price.



The Learning Environment :

The Weka and YALE Data Mining Environments were used for carrying out the experiments. The genral setup used is as follows:



==========X==========

PAPER – 5 : Machine learning techniques and use of event information for stock marketpredictions.

BY : Paul D. Yoo

https://ieeexplore.ieee.org/abstract/document/1631572/

Introduction:

Since the stock market was firstly introduced, many have attempted to predict the stock markets using various computational tools such as Linear Regression (LR), Neural Networks (NNs), Genetic Algorithms (GAs), Support Vector Machines (SVMs), Case-based Reasoning (CR) and others. Over the last decade, NNs have been most widely used and shown better performance over other approaches in many cases.

Neural Networks :

In the literature, it has been shown that NNs offer the ability to predict market directions more accurately than other existing techniques. The ability of NNs to discover non-linear relationships between the training input/output pairs makes them ideal for modeling nonlinear dynamic systems such as stock markets**.**

Recent researches tend to hybridize several AI techniques. Hiemstra [11] proposed fuzzy expert systems to predict stock market returns. He suggested that combining NNs and fuzzy logic capture the complexities of functional mapping and do not require the specification of the function to approximate. Tsaih et al. [41] integrated the rule based technique and NNs to predict the direction of change of the S&P 500 stock index futures on a daily basis.

Conclusion and future work :

In this paper, we examined recent developments in stock market prediction models. By comparing various prediction models, we found that NNs offer the ability to predict market directions more accurately than other existing techniques. The ability of NNs to learn nonlinear relationships from the training input/output pairs enables them to model non-linear dynamic systems such as stock markets more precisely**.**

For further research, firstly, prior-knowledge database should be built by analyzing historical events on stock markets. Based on the prior-knowledge, the event weighting schema will be developed and each event should be weighted accordingly. Finally, a proposed model which incorporates the weighed events into the numeric time series data should be compared empirically with other models.

==========X==========

PAPER – 6 : A Machine Learning Approach for Stock Price Prediction

BY : Richard Kyle MacKinnon

http://www.cs.umanitoba.ca/~ywang/papers/ideas14.pdf

Abstract :

Data mining and machine learning approaches can be incorporated into business intelligence (BI) systems to help users for decision support in many real-life applications. Here, in this paper, we propose a machine learning approach for BI applications. Specifically, we apply structural support vector machines (SSVMs) to perform classification on complex inputs such as the nodes of a graph structure. We connect collaborating companies in the information technology sector in a graph structure and use an SSVM to predict positive or negative movement in their stock prices. The complexity of the SSVM cutting plane optimization problem is determined by the complexity of the separation oracle. It is shown that (i) the separation oracle performs a task equivalent to maximum a posteriori (MAP) inference and (ii) a minimum graph cutting algorithm can solve this problem in the stock price case in polynomial time. Experimental results show the practicability of our proposed machine learning approach in predicting stock prices.

Categories and Subject Descriptors :

[Database Management]: Database Applications—data mining.
[Artificial Intelligence]: Learning—parameter learning.
[Computer Applications]: Administrative data processing—financial.

Training Labels :

The final pieces of data required to train an SSVM are the training labels. In this case, the labels are binary (i.e., consisting of a positive class and a negative class). Nodes whose companies saw increases in stock price in the months after the release of financial data were labelled with the positive class while nodes whose companies saw decreases in stock price were labelled with the negative class. Historical stock price data was obtained for all companies from Yahoo! Finance.

Since the maximum time span between historical data points obtained from Yahoo! Finance is monthly while the Quandl feature vector data was calculated for annual data points, additional processing was done on the Yahoo! Finance data to remove the extra data points. In addition, some of the csv files inexplicably had Windows line endings while others had UNIX line endings so processing was also done to standardize this aspect across all csv files.

CONCLUSIONS :

Data mining and machine learning approaches can be incorporated into business intelligence (BI) systems to help users for decision support in many real-life applications. In this paper, we focused on an interesting application of stock price prediction. Specifically, we used minimum graph cuts as parts of a cutting plane algorithm to solve the optimization problem of the structural support vector machine (SSVM).

This allowed the SSVM to learn a prediction model for a complex graph input with multiple edges per node (representing complex relationships between companies that affect the stock price by using feature vectors that contain fundamental financial information). The resulting model was applied to the problem of stock price prediction, where the positive and negative class labels corresponded to increasing and decreasing stock prices respectively.

We used 3-fold cross validation to determine that an appropriate value for the SSVM parameter C was 1000. Accuracy of the SSVM on the training samples—in terms of correctly classified nodes—was higher than 78%, suggesting the model was learnt successfully without over-fitting. This shows the effectiveness of our machine learning approach—using SSVM classification augmented with minimum graph cuts—in a practical BI application of stock price prediction.

==========X==========

PAPER – 7 : **Stock Closing Price Prediction using Machine Learning Techniques**

By Mehar Vijha

- **INTRODUCTION :**

Accurate prediction of stock market is very challenging due to volatile and non linear nature of financial stock markets but its possible with help of artificial intelligence and increased computational capabilities. In this Article they have used Artificial Neural Network and Random Forest techniques for predicting the next day closing price.

The financial data:

Open, High, Low and Close prices of stock are used for creating new variables which are used as inputs to the model. The models are evaluated using standard strategic indicators: RMSE and MAPE. The low values of these two indicators show that the models are efficient in predicting stock closing price.

Different classical algorithms like linear regression , Random Walk Theory (RWT), Moving Average Convergence / Divergence (MACD) and also using some linear models like

Autoreggressive Moving Average (ARMA), Autoregressive Integrated Moving Average (ARIMA), for predicting stock prices.

- **METHODOLOGY :**

- **ANN** is capable for finding hidden features through a self learning process. These are good approximators and are able to find the input and output relationship of a very large complex dataset.

The historical data for the five companies is collected from Yahoo Finance . The dataset includes 10 year data from 4/5/2009 to 4/5/2019 of Nike, Goldman Sachs, Johnson and Johnson, Pfizer and JP Morgan Chase and Co. The data contains information about the stock such as High, Low, Open, Close, Adjacent close and Volume.

Six new variables is created for the prediction of stock closing price and used to train the model.

1. Stock High minus Low price (H-L)

2. Stock Close minus Open price (O-C)

3. Stock price's seven days' moving average (7 DAYS MA)

4. Stock price's fourteen days' moving average (14 DAYS MA)

5. Stock price's twenty one days' moving average (21 DAYS MA)

6. Stock price's standard deviation for the past seven days (7 DAYS STD DEV)

The model uses the basic structure of Neural Network having neurons with different layers :

- Input layer - consists of new variables.
- Hidden layer - consists of these neurons.
- Output layer - consists of only one neuron which will give the predicted value in terms of closing price of the stock.

The weights on each input load is multiplied and added and sent to the neurons. The hidden layer or the activation layer. The total weight is calculated and is moved to the third layer which is the output layer

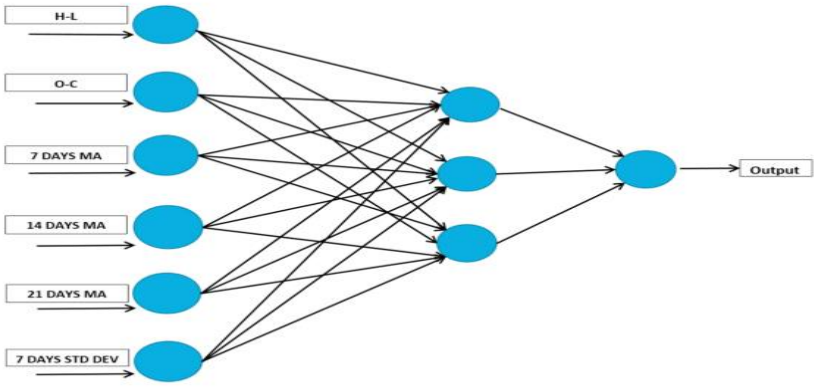


Fig. 1. Detailed architecture of Artificial Neural Network (ANN) for stock price prediction

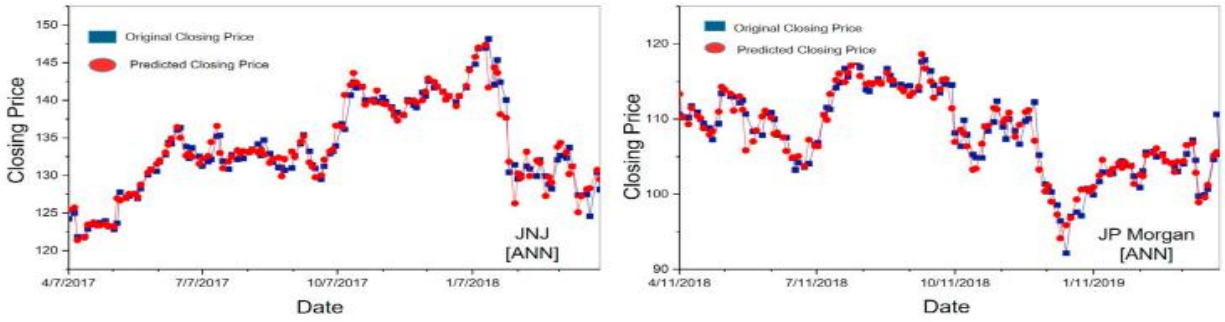
- **Random Forest (RF)** - an ensemble machine learning technique. It is capable of performing both regression and classification tasks. The idea is to combine multiple decision trees in order to determine the final output instead of relying on individual decision trees.

To evaluate the effectiveness of the models, a comparison is made between the two techniques :

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(O_i - F_i)^2}{n}} \qquad MAPE = \frac{1}{n}\sum_{i=1}^{n}\frac{(O_i - F_i)}{O_i} * 100 \qquad MBE = \frac{1}{n}\sum_{i=1}^{n}(O_i - F_i)$$

'Oi' → original closing price 'Fi' → the predicted closing price

'n' → the total window size



- **CONCLUSION :**

The historical dataset available on company's website consists of only few features which are not sufficient enough. To obtain higher accuracy we create new variables using existing

variable . ANN is used for predicting the next day closing price of the stock and for a comparative analysis, RF is also implemented.

https://www.sciencedirect.com/science/article/pii/S1877050920307924

PAPER – 8 : **A Machine Learning Model for Stock Market Prediction:**

By Osman Hegazy

- **INTRODUCTION :**

This paper proposes a machine learning model to predict stock market price. The proposed algorithm integrates Particle swarm optimization (PSO) and least square support vector machine (LS-SVM).

Proposed model is based on the study of stocks historical data and technical indicators. PSO algorithm selects best free parameters combination for LS-SVM to avoid over-fitting and local minima problems and improve prediction accuracy.In most cases ANNs suffer from over-fitting problem due to the large number of parameters to fix, and the little prior user knowledge about the relevance of the inputs. vector machines (SVMs) had been developed as an alternative that avoids such limitations.

➢ METHODOLOGY :

❖ LEAST SQUARE SUPPORT VECTOR MACHINE :

Its versions of support vector machines (SVM) which are a set of related supervised learning methods that analyze data and recognize patterns, and used for classification and regression analysis and which finds the solution by solving a set of linear equations . Let X is n× p input data matrix and y is n×1 output vector.Given the {xi,yi} i=1 to n → training data set, where p xi ∈ R and yi ∈ R . Goal is to construct the function f (x) = y , which represents the dependence of the output yi on the input xi.

$$f(x) = W^T \varphi(x) + b$$

The optimization problem and the equality constraints of LS-SVM are defined as follows:

$$\min_{w,e,b} j(w,e,b) = \frac{1}{2} w^T w + C \frac{1}{2} e^T$$

$$y_i = w^T \varphi(x_i) + b + e_i$$

❖ PARTICLE SWARM OPTIMIZATION ALGORITHM :

A heuristic search method which is derived from the behavior of social groups like bird flocks or fish swarms. PSO moves from a set of points to another set of points in a single iteration with likely improvement using a combination of deterministic and probabilistic rules.

Optimization is achieved by giving each individual in the search space a memory for its previous successes, information about successes of a social group and providing a way to incorporate this knowledge into the movement of the individual. The elements of the velocity vector for particle i are updated as:

$$\upsilon_{ij} \leftarrow \omega \upsilon_{ij} + c_1 q(x_{ij}^{pb} - x_{ij}) + c_2 \gamma(x_j^{sb} - x_{ij}), j = 1,..,n$$

Where w → inertia weight

Xi → best variable vector of particle i

C1 and C2 are constants, and q and r are random numbers in the range (0, 1).

Once the velocities have been updated, the variable vector of particle i is modified according to:

$$x_{ij} \leftarrow x_{ij} + \upsilon_{ij}, j = 1,...,n.$$

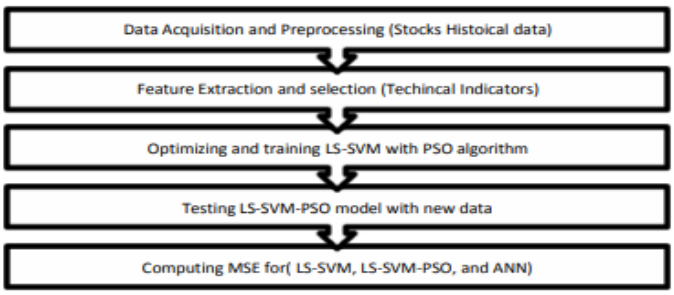


Fig. 2: The Proposed Model

Five technical indicators are calculated from the raw datasets:

1) Relative Strength Index (RSI): RSI = 100- [100 / (1+RS)]
2) Money Flow Index (MFI): MF = Typical Price * Volume
   Money Ratio (MR) = (Positive MF / Negative MF)
   MFI = 100 – (100/ (1+MR)).
3) Exponential Moving Average (EMA): EMA = [α *T Close] + [1-α* Y EMA].
4) Stochastic Oscillator (SO) =
   [(Close price – Lowest price) / (High Price – Low Price)] * 100
5) Moving Average Convergence/Divergence (MACD):
   MACD = [0.075*EMA of Closing prices] – 0.15] *EMA of closing prices]
   Signal Line = 0.2*EMA of MACD

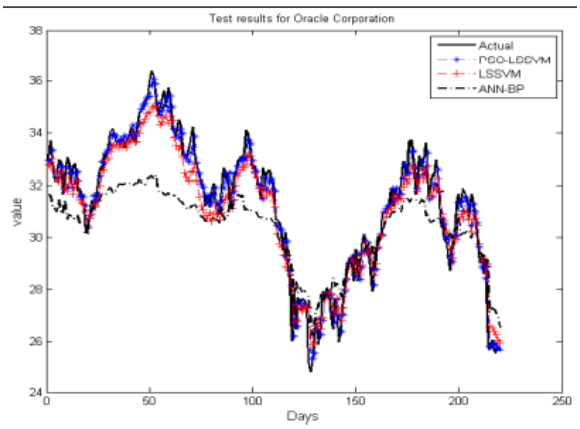


Fig. 4: Results for Oracle Company

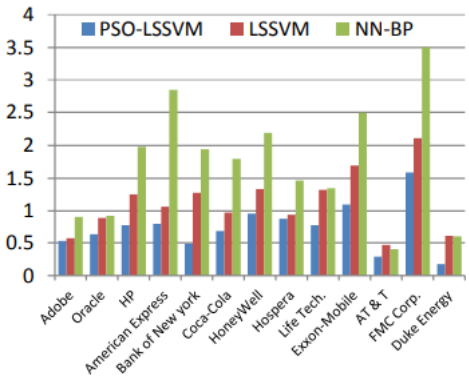


Fig. 15: MSE of proposed algorithm

➢ CONCLUSIONS:

The PSO is employed iteratively as global optimization algorithm to optimize LS-SVM for stock price prediction It is capable to overcome the over-fitting problem which found in ANN, especially in case of fluctuations in stock sector. The performance of the proposed model is better than LS-SVM.

https://arxiv.org/ftp/arxiv/papers/1402/1402.7351.pdf

PAPER - 9 : **Machine Learning Techniques for Stock Prediction**

By Vatsal H. Shah

- **Preprocessing the Historical Stock Data :**

  the historical data was downloaded from the yahoo finance section. In particular, the stock prices of two companies were studied, namely Google Inc. (GOOG) and Yahoo Inc. (YHOO) The dataset available has the following attributes:
  Date Open High Low Close Volume Adj. Close

  The price of the stock yesterday is going to have the most impact on the price of the stock today
  Thus as we go along the time-line, data-points which are nearer to today's price point are going to have a greater impact on today's price. The attribute will act as our label for predicting the movements of the stock price. For our experiments we use the EMA (Exponential Moving Average) as the indicator function.

- **Different types of Machine Learning Techniques :**

  3.1 Decision Stump
  3.2 Linear Regression (with only numeric attributes taken under consideration)
  3.3 Support Vector Machines Using C-Class Support Vector Machines which use RBF Kernels
  3.4 Boosting The AdaBoostM1 Algorithm was applied to the DataSet after applying the C-SVC Algorithm

- **Stock Prediction based on Textual Analysis of Financial News Articles:**

  Huge amount of valuable information is gathered from Financial News Articles, Company Reports and Expert Recommendations (Blogs from valid sources can also act as a source of information) . Most of them are in textual format as opposed to numerical formate problem domain can now be viewed as one that involves Mining of Text Documents and Time Series Analysis concurrently.

  One method which has been used involves defining the news impact on a particular stock: Positive, Negative, and Neutral.

  1) positive impact (or negative impact) - if the stock price rises (or drops) significantly for a period, after the news story has been broadcasted.
  2) neutral - If the stock price does not change dramatically after the news is released, then the news story.

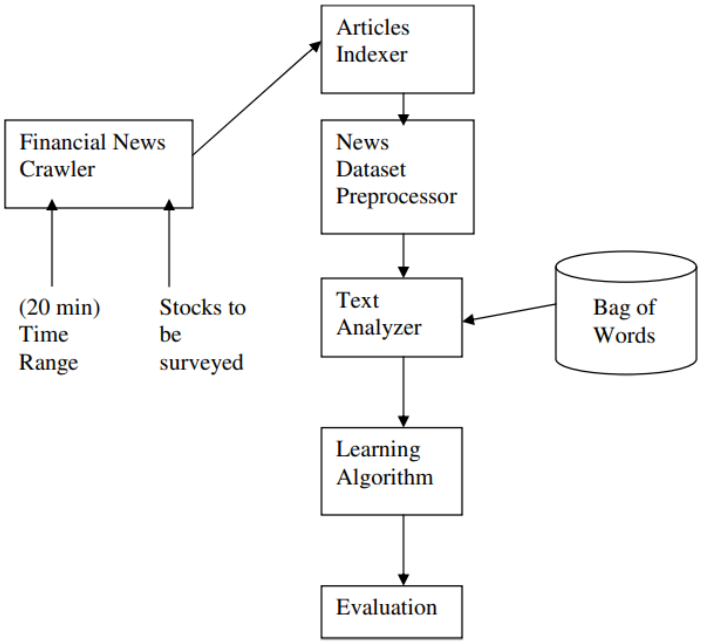
  Another method which we study in this paper relates to detecting and determining patterns in the news articles which correspond directly to a rise or fall in the stock price.

➢ METHODOLOGY :

The general architecture is as follows: A crawler continuously crawls news articles and indexes them for a particular stock portfolio. The learning environment requests the news since the last T minutes from the indexer.

The learning environment consists of several base learners which look for specific information in the text .A Bag-Of-Words consisting of Positive Prediction Terms and Negative Prediction Terms and Phrases is used by the learning environment. Each time a word/phrase from the Positive Prediction Term occurs in a particular news article, a PostiveVote is assigned to the article.

The Diagram below shows the general architecture of such a system:



https://www.researchgate.net/publication/328930285_Stock_Market_Prediction_Using_Machine_Learning

==========X==========