Technical Report Writing

Name: Shivam Tawari

Roll no: A – 58

Topic: Stock Market Price Prediction using Machine Learning

Abstract:

Support Vector Machine (SVM) is a machine learning technology that has been used to forecast stock values in recent studies. This study calculates price volatility and momentum for individual stocks and the entire technology sector using daily closing prices for 34 technology equities. These are inputs to the SVM model as parameters. The algorithm tries to predict whether a stock price will be greater or lower in the future than it is on a current day. In the short term, we find limited forecasting capacity, but in the long run, we discover definite predictive power.

Introduction:

One of the most significant issues in finance and business is stock price prediction. The stock market, on the other hand, is a dynamic and unpredictable realm. Several studies have been conducted to forecast the market in order to benefit utilising various approaches such as statistical analysis, technical analysis, and fundamental analysis, among others, with varying outcomes.

Forecasting may be described as the analysis of historical data to anticipate some future occurrence or events. It covers a wide range of topics, including business and industry, economics, environmental science, and finance, to name a few. There are several types of forecasting issues.

The analysis of time is involved in many forecasting issues. A time series data set is a chronological series of observations for a single variable. The variable in our instance is the stock price. It might be univariate or multivariate in nature. Multivariate data comprises stock values from several companies at different points in time, whereas univariate data only provides information about one stock. Time series data analysis aids in the discovery of patterns, trends, and periods or cycles in the data. When it comes to the stock market, knowing the bullish or bearish mood early on will help you invest money properly. Pattern analysis also assists in finding the best-performing firms over a certain time span. As a result, time series analysis and forecasting are major study areas.

Literature Review:

In [1], Aditya Gupta and Bhuwan Dhingra utilised a Hidden Markov Model to forecast the following day's close price of stocks. They utilised historical stock prices from firms like Apple Inc., IBM Corporation, TATA Steel, and Dell Inc. to make their calculations. High Price, Low Price, Open Price, and Close Price were the inputs. Each stock's model was meant to be independent of the others. The model was first trained for seven months. MAPE values were used to test the model.

Authors utilized an SVM-based technique to forecast the price of stock market changes in [2]. They divided the challenge into two parts: feature selection and market trend prediction. The SVM correlation was used to determine the features that have the greatest impact on the price. To predict the direction, linear SVM is applied to the data series. They demonstrated that the algorithm can select the best feature and control overfitting when predicting stock market trends. Lin et al. employed an SVM-based technique to forecast the price of stock market changes in [2].

Methodology:

Support Vector Machines:

Support One of the finest binary classifiers is the Vector Machine. They draw a decision border with the majority of points in one category on one side and the majority of points in the other category on the other. Consider x = (X1,...,Xn), which is an n-dimensional feature vector. A linear boundary (hyperplane) has the following definition:

$$\beta_0 + \beta_1 X_1 + \ldots + \beta_n X_n = \beta_0 + \sum_{i=1}^{n} \beta_i X_i = 0$$

The sum of items in one category will be larger than 0, while the sum of elements in the other category will be less than 0.

We carry out our research by altering this parameter n to examine how patterns in volatility and momentum, both of the individual stock and the index, may be utilised to forecast future changes in the stock.

Let n1 be the index parameter and n2 be the supplied stock parameter, with n1,n2 being 5,10,20,90,270.

One week, two weeks, one month, one quarter, and one year are represented by these numbers. We supply a combination of n1,n2 in each iteration, and we utilise these parameters to compute our feature sets, train on the training data, predict on the testing data, and check the correctness of the results. We perform 25 iterations, one for each of the n1,n2 combinations.

We look at every trade day from 2007 to 2014 and compute the four characteristics on that date in order to determine the features. On that specific day, the four characteristics are combined into a single vector. We begin calculating feature vectors on the d = (max(n1,n2) +1)-th day since we average over the previous n1 days for index and n2 days for stock. If n1 = 5, and n2 = 10, then d = 11 and we begin on the 11th date. This is due to the fact that volatility and momentum are both computed using data from the previous day, yet there is no data before the first date.

Results and Discussion:

When it comes to predicting price direction one day ahead (m = 1), the SVM model is only tenths of a percentage point better than plain random guessing. This has a number of significant ramifications. For starters, it bolsters the Efficient Markets Hypothesis. If there is a model that integrates a variety of historical data formats, as well as some properties such as momentum that economists have shown are present in stock price data, is unable to perform better than a toss a coin This is really solid evidence when it comes to predicting the price direction for the next day, that prices move in a random manner. Prices that already reflect available data will only alter in response to new data, so the price direction for tomorrow will be determined solely by new data that arrives tomorrow. Because all prior data should already be included into the price, a model like ours that analyses solely historical data should not be able to anticipate price direction.

As a result, EMH is supported by our model's difficulties in predicting the next day's stock price.

As the parameter m is increased, certain trends arise. The most noticeable feature is that when m = 5,10,20, the mean and median returns rise, but then fall somewhat when m = 90 and m = 270. Figure 1 depicts the mean prediction accuracy as a function of the parameter m, where mean prediction accuracy is defined as the average of the mean accuracies for all 25 combinations of n1,n2 with a fixed m, where each combination gives the mean accuracy across the 34 stocks.

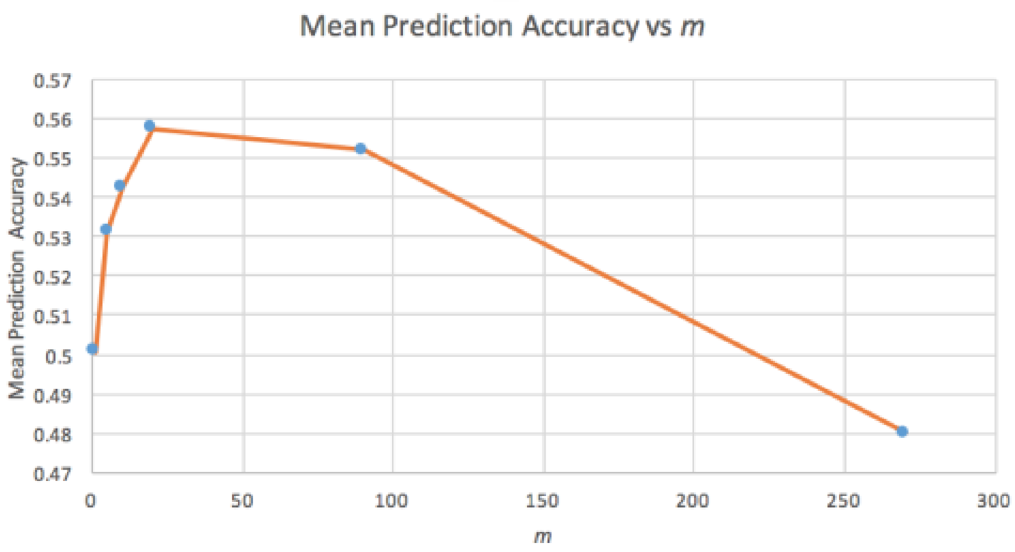

Figure 1

Conclusion:

This research offers an SVM-based stock market trend prediction system for selecting a good feature subset, evaluating stock indicators, and controlling over-fitting in stock market trend prediction. There are three different advantages to the proposed system.

One is that a good feature subset is chosen, one that includes features that are substantially associated with the output but not with one another. The second is that specific features are assessed and ranked. The correlation-based SVM filter approach is used to choose features and evaluate them at the same time.

Finally, the suggested solution is resilient since the piecewise linear principle is used to implement it, and the feature weights are used to construct the best separation hyperplane. Our approach outperforms existing prediction systems in terms of robustness, according to simulation data. It eliminates the over-fitting flaws that plague traditional stock market trend forecasting systems.

References:

[1] Lin, Y., Guo, H., & Hu, J. (2013). An SVM-based approach for stock market trend prediction. In Neural Networks (IJCNN), The 2013 International Joint Conference on (pp. 1–7). IEEE.

[2] Gupta, A., & Dhingra, B. (2012). Stock market prediction using hidden markov models. In Engineering and Systems (SCES), 2012 Students Conference on (pp. 1–4). IEEE.