Multi-Classifier Ensemble System With Dynamic Rule Based Algorithm For Stock Prediction: A Survey

Sandeep Sharma[1], Gurpreet Kaur[2]
[1]Dept of CSE
[2]Assistant professor, Dept of CSE
[1, 2] Rayat Bhara Group of Institutions, Patiala, Punjab, India.

Abstract- *In monetary markets, it is mutually important and demanding to forecast the daily path of the stock market return. Among the some studies that focus on predicting daily stock market returns, the data mining measures utilized are either unfinished and not efficient, especially when a plethora of features are involved. This paper presents a whole and capable data mining process to anticipate the everyday direction of the S&P 500 Index ETF return based on 60 financial and economic features. To make it more accurate, I study many other techniques but SOP (Self Organizing Map) is very proficient technique. To attain an accurate stock market prediction, the identification of the effective features is essential. In other words, the representative features of the factors play a key role in prediction efficiency. Technical and fundamental analyses are two indispensable tools in financial market evaluation. Fundamental analysis can be used to estimate a firm's performance and financial status over a period of time by carefully analyzing the institute's financial statement. Technical analyses (TA), equally, evaluate securities by means of statistics such as past price and volume that are generated by market activities. The major analysis of TA is that it only reflect on the price movement and ignore the fundamental factors related to the company. The multiple classifier ensemble system (MCS), single type of machine learning technique, has newly become the focus of a new methodology for obtaining higher accuracy in predictions*

Keywords- Stock Market , Machine Learning , Self Organizing Map, Data Mining

I. INTRODUCTION

Data extracted from social networking Websites is unstructured and fuzzy in nature. In everyday life conversations as noticed on social networking Websites, people do not care about the spellings and accurate grammatical construction of a sentence that leads to different types of uncertainties, such as lexical, syntactic, and semantic. So, analyzing and extracting information patterns from such data sets are more difficult. A significant amount of research has already been carried out to categorize data/sentence into various categories of emotion. Most of the emotions that have been worked upon are either positive or negative or finding the polarity i.e. the level of emotion expressed by the author. India being the largest democracy, gives its people the right to express their emotions and sometimes some ill willed people take due advantage of such facts. As the social media is becoming a need of life today, so is the interaction/discussions of people through social media has also increased. A great impact of all this can be seen on the Indian Army Fans page on Facebook[6, 9]. The page that is followed by around 1.6 million people in the country has many thought leaders that can easily influence the thought process of several others. And it can also be said that it can be really helpful to control various difficult situations in the country if we identify those thought leaders who mainly circulate negative (anger/fear) emotions among people. This will mainly help to control situations like riots, anger among people due to roomers and false information.

So as to achieve this, an accurate understanding of how emotions are represented both in the human mind and in the computer environment is essential in the study of affect detection. The relationship between emotion and text is also essential when mapping textual information to emotion gap. In general, the study of emotions in written text is conducted from two opposite points of view. The first is the viewpoint of a writer. This is concerned with how emotions influence a writer of a text in choosing certain words and/or other linguistic elements. The second point of view is concerned with how a reader interprets the emotion in a text, and what linguistic clues are used to infer the emotion of the writer. In this thesis, the second point of view is taken into consideration because we are interested in the way people infer emotions. When an event occurs, each individual has his own perceptions and his own thought process that leads to a reaction regarding that event. As each one of us react in a different way so is the way to express our emotions is also different either verbal or textual. Now if we study only the textual expression of a group of people regarding an event, we shall have a variety of text, with different languages and ways to express. This becomes quiet difficult to mine the relevant information from such variety of text. So as to make this task much simpler and efficient we need to work onto certain areas if we are working with the dictionary based techniques.

Since the stock market is an important and active part of financial markets, the stock market prediction has achieved widespread concern from academic and business communities. There are several economic theories about future prices of stock market. The first is Efficient Market Hypothesis [9], which asserts that the price reflects all known information. Another is Random Walk Hypothesis [2], which believes that the price movement is random. These theories mean that the price of stock cannot be predicted effectively. However, recent researches suggest that stock prices could be predicted to some extent from the perspective of socioeconomic theory of finance and behavioral economics [5]. Traditionally, the stock market prediction is based on historical stock data. The various technical indicators can be extracted from historical data to predict the trend of stock market [7]. These studies focus on the past market information, but ignore the latest information. Financial news articles, known as one of the most important part of market information, are used as the latest information for the stock market prediction [1, 3]. This kind of news includes big mergers, bankruptcy of some companies or economic crisis. Financial news usually has significant impact on stock market because traders rely on them to make judgment for future trading decisions. However, financial news articles have following disadvantages. Firstly, financial news just reflects the viewpoints of editors and reporters. Secondly, because the news release is not scheduled, it is necessary to align news articles with the time of stock market. Thirdly, web pages of financial news are full of a large amount of advertisement and consist of more noise data.

Since the introduction of social media, companies are gradually adopting social media technologies, using Twitter to put out to customers or YouTube to express product features. “The wisdom of crowds,” equipped with data mining rules and algorithms can automatically create estimation of future performance on a range of subjects, such as stock market performance, sports outcomes, election results and box office sales. In this work attempt has been made to study the prediction power of the collective sentiments of micro-blogging websites on the stock market. The data used in the explore are tweets collected from stocktwits.com, which is an online financial communication platform for the financial and investing community. At the time of writing, there are more than 150,000 investors on the site, which can be viewed by audiences of 40 million across the financial web and social media platforms. As a sister service to Twitter, StockTwits is composed of a large user base of trading and investing professionals, who can integrate their StockTwits accounts with their Twitter accounts if they choose to. The objectives of this work can be presented in several steps: First, we examine performance of several natural language processing approaches to detect the public sentiment of users on StockTwits. Second, we analyze the correlation between the trade volume of stocks and the activity of userss discussions of the stocks on StockTwits, to determine the predictive power of the social media data on the daily stock market performance. Our ultimate goal is not to build an ideal model for stock market prediction, but to test whether the feature of social media sentiments contributes to the stock market analysis, and to assess its predictive power.

II. RELATED WORK

The future status of companies offering stocks is of great importance to stock market practitioners. According to the efficient market theory, it is impossible to predict prices based on historical stock data. This theory also states that the prediction of the classical criteria of risk and return cannot bring advantages to share- holders. There is abundant evidence in the literature, however, that argues against the efficient nature of the market. A precise prediction of companies’ future financial status provides investors with the security to make a confident and profitable investment. To achieve an accurate stock market prediction, the identification of the effective features is crucial. In other words, the representative features of the factors play a key role in prediction efficiency. Technical and fundamental analyses are two essential tools in financial market evaluation. Fundamental analysis can be used to evaluate a firm’s performance and financial status over a period of time by carefully analyzing the institute’s financial statement. Technical analysis (TA), conversely, evaluates securities by means of statistics such as past price and volume that are generated by market activities. The major criticism of TA is that it only considers the price movement and ignores the fundamental factors related to the company. Moreover, TA takes a comparatively short-term approach to analyzing the market. Fundamental analysis seeks to find the essential features of stock and market movements. In fact, the logic behind fundamental analysis is that if a company has a proper fundamental strength, then long term stock investment in the company will be more se- cure and stable. Thus, the stocks of these fundamentally strong companies, which are making money, gaining profit and growing their businesses, represent an opportunity for a successful investment. For this reason, in this paper, fundamental analysis is applied in order to determine the fundamental features that decide which company is a good bet for a secure investment. Stock return forecasting is a fascinating endeavour with a long history. From the standpoint of finance practitioners, asset allocation requires real-time forecasts of stock returns and an improved stock return forecast holds the promise of enhancing the investment performance. For an efficient investment, the return consideration is not sufficient. In fact, the risk and return must be considered simultaneously to create an accurate portfolio evaluation. In this paper, the prediction of stock return and risk are implied concurrently based on fundamental features in order to build a more comprehensive model for stock market analysis. Although the statistical approaches such as logistic regression and regression analysis are widely applied to forecast the return and risk of stocks, the results of machine learning approaches are generally superior in comparison to statistical methods. The multiple classifier ensemble system (MCS), one type of machine learning technique, has recently become the focus of a new methodology for obtaining higher accuracy in predictions. The rationale is that the optimization of a combination of relatively simpler predictors appears more convenient than optimizing the design of a single complex prediction. In fact, three fundamental issues are effective for establishing a successful MCS model: accuracy of individual classifiers, diversity among classifiers, and the choice of the fusion methods that will be used. The aim of this combination scheme is to gain increased precision with proper single classifiers and eliminate the uncorrelated individual classifier errors, which are the errors made by individual classifiers on various parts of input space

III. THE PROPOSED STOCK PREDICTION ALGORITHM

The proposed model is designed for the prediction on the time-series data obtained from the live stock API (application programming interface) from GOOGE, which can be targeted for the selective stocks in one API call. The regression model has been applied over the time series data in the specific hierarchy, which mandates the preprocessing, data cleaning, missing value fixation and application of the support vector regression model for stock value prediction. The following algorithm describes the overall working of the proposed model using SVR:

Algorithm 1: Support Vector Model based Predictive Analysis on Selective Stocks

1. *Initialize the tickers containing the stock names, {'AAPL', MSFT', 'SPY']*
2. *Initialize the data source ['GOOGLE']*
3. *Initialize the start and finish date to acquire the time series data ['START DATE' 'FINISH DATE']*
4. *Acquire the data from the API, API_CALL [tickers, data source, start date, finish date]*
5. *Obtain the adjacent close values for each stock from the acquired data for all weekdays*
6. *Extract the specific stock data from the stock data matrix, ['MSFT']*
7. *Fix the "Not a number" values in the extracted stock market vector*
    a. *If iteration index is lower than 3*
        i. *If current value is "NaN"*
            1. *Fill zero in place of NaN*
    b. *Otherwise*
        i. *Extract the three value starting from current value, initial counter at (current-3) to initial*
        ii. *Run the iteration over the mini vector acquired on step 7(b)(ii)*
            1. *If current value is NaN*
                a. *Save the value id in the mini vector*
            2. *Otherwise*
                a. *Add the current value total sum of vector, sumVec = sumVec + currentValue*
        iii. *Compute the average value of mini vector, mvAvg = sumVec / NoOfValues*
        iv. *Fill all mini vector values marked on 7(b)(ii)(1) with average value (mvAvg) on step 7(b)(iii)*
        v. *Restore the mini vector to the output vector*
8. *Apply the support vector regression (SVR) over the MSFT vector after fixation*
9. *Return the prediction from the regression model*
10. *Save the regression result*

The self organizing maps (SOM) have been utilized for the purpose of stock market prediction, where the neural network has been designed with the multilayer preceptron based formation for the purpose of stock prediction. The data of variable length (15 days and 30 days) windows has been used to train the SOM based classifier, which utilizes the group feature vectors based upon sliding window of 30 day, and the original variations in the stocks labeled as "UP", "DOWN" or "NEUTRAL". During the initialization of the weights computational model, the term weights are computed and normalized before being clustered or structured under the entity relationship library analysis. This algorithm produces the relationship matrix between the entities, which further undergoes the final analysis to predict the stock prices over the given data. The algorithm is described as following:

Algorithm 2: Weighted Stock Feature Categorization Algorithm

1. Acquire the data matrix containing the historical data of stocks
2. Analyze the number of classes or categories in the acquired data matrix
3. Group the data in the time-based segments, which may involve the month-on-month or quarter-on-quarter analysis
4. Compute the difference between last value of segment and the next value for each of the group under the data analysis
5. Label the data according to the different of two values on step 4
    a. Label "UP" if next value is higher than last value of segment
    b. Label "DOWN" if next value is lower than last value of segment
    c. Label "LEVEL" if next value equals last value of segment
6. Then prepare the prediction vector and training matrix accordingly
7. Repeat the steps from 2 to 6 for all segments according to upper bound limit
8. Return the training data matrix and training prediction (observation) vector.

The final algorithm involves the application of self organizing maps (SOM) based multi-layered perceptron based classification for the prediction of the stock prices over the historical data obtained from the Google Finance API. The SOM-model is utilized to describe the features from the input data matrix, which involves the averaging factors, dimensionality reduction and similar techniques over data to describe the final features from the data matrix. The SOM classification model is utilized to match and classify the matching patterns in order to predict the stock market variations. The algorithm is explained in the following section:

Algorithm 3: SOM based Multilayer Perceptron based Stock Price Prediction Model

1. The input data acquisition is done on the Google Finance API for the target stocks.
2. Input dataset with stock price values for given dates
3. Number of categories is a pre-defined number
4. Algorithm determines the pre-defined data points equal to the segment number
5. The algorithm evaluates the distance of each data point from all of the pre-defined initial data points.
6. The point is added to the segment between the lower and upper bound limits
7. Return the segmented & classified (labeled) training data for the SOM based predictive model.
8. The stock market data contains the variable and volatile patterns about the various stocks or stock market sectors
9. The historical stock market data matrix is further evaluated for the Self Organizing Maps (SOM) based predictions
    a. The individual entries for multiple stocks are individually inspected and classified according to their inter-feature similarity.
    b. The similarity analysis depicts the mode of operation to divide the feature descriptors in the smaller segments for the efficient analysis.
    c. After the classification the intensity of the individual feature descriptors on the training data is inquired and calculated, which prepares and classifies the future trends for the selective stocks.
10. The SOM is the operation performed on the given data matrix (test matrix) obtained after the processing of selective stock.

IV. EXPERIMENT ON VALUES OF PREDICTION

The proposed model has been analyzed for the accuracy of the predictions among the proposed prediction model. The proposed model has been designed by combining the SVM with feature descriptor extraction, missing value fixing and data cleaning in order to acquire the balanced and best value data as the training matrix. The results of the proposed model has been analyzed in the form of predicted values of the stocks in the next day of trading, which has undergone the proposed model's SVR model as well as polynomial and linear regression models. The results are shown in detail in the following table:

Table 5.1.: Table of Predictions and Actual Values

Index	UpperBound Limit (Data Size: 100 days)	Original Value	Support vector regression	Polynomial Regression	Linear Regression
0	100	26.84	26.55	27.21	29.51
1	101	26.27	27.26	26.70	29.42
2	102	26.07	25.17	26.13	29.30
3	103	25.01	26.17	25.60	29.18
4	104	26.00	23.77	24.91	29.03
5	105	25.80	29.31	24.56	28.92
6	106	17.27	23.28	24.22	28.80
7	107	25.89	**2.70**	21.96	28.37
8	108	26.46	**38.23**	22.02	28.27
9	109	26.86	24.21	22.22	28.20
10	110	25.79	27.36	22.52	28.14
11	111	25.29	24.93	22.58	28.04
12	112	25.11	25.35	22.54	27.94
13	113	24.79	25.62	22.49	27.82
14	114	25.00	24.72	22.40	27.71
15	115	25.66	25.44	22.40	27.60
16	116	25.50	26.68	22.56	27.51
17	117	26.58	24.96	22.70	27.43
18	118	26.32	29.41	23.07	27.38
19	119	26.37	24.23	23.36	27.33
20	120	26.44	27.16	23.65	27.28
21	121	25.95	26.95	23.94	27.23
22	122	25.77	25.00	24.12	27.17
23	123	25.31	26.12	24.25	27.11
24	124	25.00	**25.05**	24.30	27.03
25	125	24.53	**24.63**	24.29	26.94
26	126	24.31	24.64	24.20	26.85
27	127	23.31	25.00	24.09	26.74
28	128	23.01	22.37	23.81	26.61
29	129	23.16	23.67	23.52	26.48
30	130	23.27	23.89	23.30	26.35
31	131	15.48	23.03	23.14	26.23
32	132	23.82	2.82	21.54	25.88
33	133	24.30	39.45	21.68	25.78
34	134	24.41	22.73	21.92	25.71
35	135	24.27	24.60	22.17	25.63
36	136	24.83	25.34	22.39	25.56
37	137	25.13	25.70	22.70	25.51
38	138	25.44	**25.24**	23.05	25.46
39	139	25.51	**25.47**	23.43	25.43
40	140	24.89	25.89	23.80	25.39
41	141	25.23	24.30	24.04	25.35
42	142	25.48	26.42	24.32	25.31
43	143	25.12	**25.07**	24.62	25.28
44	144	25.84	23.68	24.84	25.24
45	145	25.81	28.03	25.17	25.22
46	146	26.10	23.93	25.46	25.21
47	147	26.16	26.88	25.79	25.20
48	148	25.95	26.61	26.09	25.19
49	149	26.03	25.38	26.34	25.18

In the table 5.1, the result analysis shows the predicted values by the different testing models. The squared difference of the predicted and original values has been recorded in the table 5.2 between the original with SVR, Linear regression and Polynomial regression models. The original values in table 5.1 are explained with the help of table 5.2 in the following discussion. The testing models in the experimental results include the support vector regression (SVR), polynomial regression and linear regression models. In the case of SVR model, the indices on the positions 7, 8, 31, 32 and 33 shows the anomalies with higher differences between the original observations and predicted values by the proposed SVR model over the data of 100 days arranged in the linear formation with slider window function. The predicted values on the indices of 24, 25, 38, 39 and 43 are predicted with the minimum difference, which does not exceed the maximum difference of 0.04 as squared difference. The polynomial linear prediction model has been recorded with five anomalies on the indices (6, 7, 8, 9 and 31), which is equal to the anomalies counted for SVR model. The nearest prediction by linear model are recorded on the indices 2, 25, 26, 30, and 48, out of which the maximum difference has been recorded at 0.06 as squared difference, which is significantly higher than the SVR model. The linear prediction model has been found most reliable in our case, which has been recorded in two values on 28th and 31^{st} indices in the case of anomalies. But the nearest matching values are obtained in three indices (38, 39 and 41), where maximum difference has been found at 0.01 as squared difference The linear regression model can be declared as the most significant model, if the assessment is based upon the count of anomalies, whereas the highest number of matching models elaborates the support vector regression (SVR) model, which can be declared as most significant predictive model.

Table 5.2: Squared Sum of Prediction v/s Actual value of next day stock value

Index	SVR	Polynomial Regression	Linear Regression
1	0.09	0.14	7.14
2	0.97	0.19	9.9
3	0.8	0	10.44
4	1.35	0.35	17.42
5	4.98	1.2	9.17
6	12.32	1.55	9.71
7	36.14	48.41	133
8	537.86	15.41	6.14
9	138.62	19.75	3.28
10	7.04	21.51	1.78
11	2.46	10.67	5.52
12	0.13	7.37	7.59
13	0.06	6.62	7.99
14	0.69	5.3	9.21
15	0.08	6.75	7.32
16	0.05	10.64	3.75
17	1.39	8.64	4.06
18	2.62	15.09	0.72
19	9.57	10.59	1.13
20	4.57	9.06	0.92

21	0.53	7.77	0.71
22	1	4.03	1.65
23	0.6	2.73	1.97
24	0.66	1.12	3.23
25	0	0.5	4.12
26	0.01	0.06	5.83
27	0.11	0.01	6.43
28	2.85	0.61	11.8
29	0.41	0.64	12.99
30	0.26	0.13	11.01
31	0.38	0	9.49
32	57.11	58.73	115.62
33	440.92	5.22	4.22
34	229.63	6.85	2.2
35	2.84	6.19	1.68
36	0.11	4.39	1.86
37	0.26	5.95	0.53
38	0.32	5.89	0.14
39	0.04	5.7	0
40	0	4.31	0.01
41	0.99	1.18	0.25
42	0.86	1.42	0.01
43	0.89	1.36	0.03
44	0	0.25	0.03
45	4.67	1	0.36
46	4.92	0.41	0.34
47	4.72	0.4	0.8
48	0.52	0.14	0.93
49	0.43	0.02	0.58
50	0.43	0.1	0.73

The table 5.2 shows the count of anomalies as well as the count of most matching events in the predictive model with the squared distances. The squared difference values have been acquired by computing the squared distances between the original values on the next day of the stock window (100 days), and the estimated values by support vector regression (SVR), polynomial regression and linear regression models.

The proposed model has been also analyzed for the performance of the proposed SOM based model with the existing models of Naïve bayes, Decision tree, C45 and Maximum Entropy. The proposed model has been tested with the different sizes windows for the prediction of the stock prices, which includes 7 days, 15 days, 21 days and 30 days.The proposed model has been also analyzed for the performance of the proposed SOM based model with the existing models of Naïve bayes, Decision tree, C45 and Maximum Entropy. The proposed model has been tested with the different sizes windows for the prediction of the stock prices, which includes 7 days, 15 days, 21 days and 30 days.

Figure 5.1: F1-measure based comparative analysis

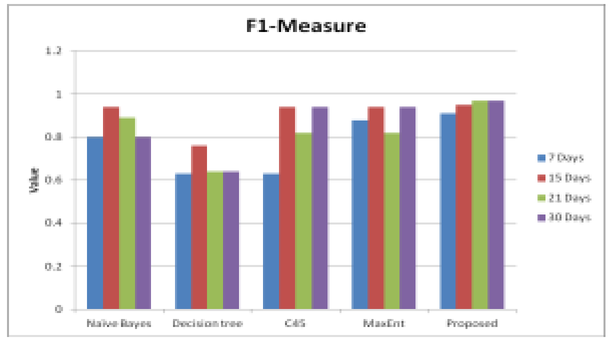


Table 5.3: F1-measure based Comparative Analysis

Test Cases	Naïve Bayes	Decision tree	C45	MaxEnt	Proposed
7 Days	0.8	0.63	0.63	0.88	0.91
15 Days	0.94	0.76	0.94	0.94	0.95
21 Days	0.89	0.64	0.82	0.82	0.97
30 Days	0.8	0.64	0.94	0.94	0.97

The proposed model has been recorded with 91% of F1-measure in SOM model against the Naïve bayes, decision tree, C45 and maximum entropy based classification models with F1-error of 80%, 63%, 63% and 88% respectively for the windows of 7 days. The proposed model has been recorded with 97% of F1-measure in SOM model against the Naïve bayes, decision tree, C45 and maximum entropy based classification models with F1-error of 89%, 64%, 82% and 82% respectively for the windows of 21 days.

Figure 5.2: Average Predictive Accuracy based comparative analysis

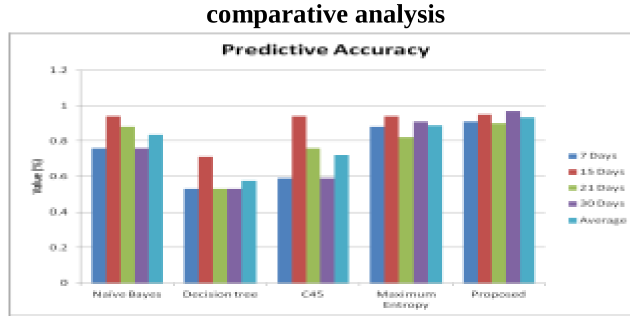


Table 5.4: Predictive Accuracy based Comparative Analysis

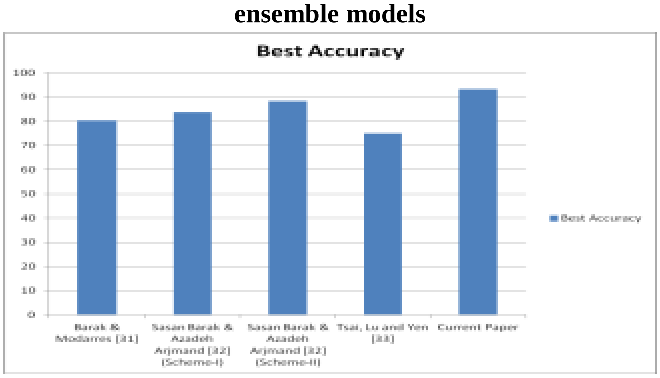
Test Cases	Naïve Bayes	Decision tree	C45	Maximum Entropy	Proposed
7 Days	0.76	0.53	0.59	0.88	0.91
15 Days	0.94	0.71	0.94	0.94	0.95
21 Days	0.88	0.53	0.76	0.82	0.901
30 Days	0.76	0.53	0.59	0.91	0.97
Average	0.835	0.575	0.72	0.8875	0.93275

The predictive accuracy of 91% in the paradigm of 7 days using the SOM model, which has been compared with Naïve bayes, Decision tree, C45 and Maximum entropy accuracy of 76%, 53%, 59% and 88% respectively, whereas the domination of the proposed model remains consistently higher for the proposed model in the prediction window sizes of 15-30 days. The average accuracy of 93.275% in the proposed model has been found higher than average accuracy of Naïve bayes, decision tree, C45 and maximum entropy based stock price prediction models with accuracy values of 83.5%, 57.5%, 72% and 88.75% respectively. These average accuracy values are obtained over the all rounds of performance testing with variable window sizes of 7, 15, 21 and 30 days.

Table 5.5: Comparative analysis with existing models

Author/year	Stock Exchange	Input data	Base Classifier	Feature Selection	Hybrid Model	The Best Accuracy
Barak & Modarres [31]	Return Forecasting in TSE-Iran	Financial ratios and Fundamental Index	Cart, Rep Tree, LAD Tree	Function based Clustering	Hybrid	80.24
Sasan Barak & Azadeh Arjmand [32]	Return Forecasting in TSE-Iran	Financial ratios and Fundamental Index	Cart, Rep Tree, LAD Tree	Function based clustering with diversity	Fusion	83.65
Sasan Barak & Azadeh Arjmand [32]	Return Forecasting in TSE-Iran	Financial ratios and Fundamental Index	DTNB, BF Tree, LAD Tree,	Function based clustering with diversity	Fusion	88.25
Tsai, Lu and Yen [33]	Taiwan	Intangible assets value variable	MLP	PCA- Stepwise Regression-decision trees-association rules- GA	MLP	75
Current Paper	S&P, NYSE, NASDAQ	Weekdays adjacent close data for acquired data	Self Organizing Maps, Multilayer Perceptron	Moving Average, Overlapping adjacent close segments, variance	Fusion	93.28

Figure 5.3: Best Accuracy based comparison with existing ensemble models



REFERENCES

[1] R.Cervello-Royo, F. Gujiarro and K. Michniuk, "Stock market trading rule based on pattern recognition and technical analysis: forecasting DJIA index with intraday data," *Expert System Application*, vol. 42, 2015, pp. 5963-5975.

[2] C.-F. Huang, "A hybrid stock selection model using genetic algorithms and support vector regression," *Appl. Soft Comput.*, vol. 12, 2012, pp. 807-818.

[3] S. Barak, J. H. Dahooie and T. Tichy, "Wrapper ANFIS-ICA method to do stock market timing and feature selection on the basis of japanese candlestick," *Expert System Application*, vol. 42, 2015, pp. 9221-9235.

[4] D.E Rapach and G. Zhou, "Forecasting Stock returns," *Handbook of Economic Forecasting*, vol. 2, 2013, pp. 328-383.

[5] S. Barak, M. Abessi and M. Modarres, "Fuzzy turnover rate chance constraints portfolio model," *European Journal of Operational Research*, vol. 228, 2013, pp. 141-147.

[6] S. H. Cheng, "A hybrid predicting stock return model based on bayesian network and decision tree," in *Modern Advances in Applied Intelligence, Springer International Publishing*, Taiwan, 2014.

[7] S.H. Cheng, "A hybrid predicting stock return model based on logistic stepwise regression and CART algorithm," in *Intelligent Information and Database Systems, Springer International Publishing*, Taiwan, 2015.

[8] S. Barak and M. Modarres, "Developing an approach to evaluate stocks by forecasting effective features with data mining methods," *Expert System Application*, vol. 42, 2015, pp. 1325-1339.

[9] S. Barak and S.S. Sadegh, "Forecasting energy consumption using ensemble ARIMA-ANFIS algorithm," *International Journal of Electrical Power Energy System*, vol. 82, 2016, pp. 92-104.

[10] C.-F. Tsai, Y.-F. Hsu and D.C. Yen, "A comparative study of classifier ensembles for bankruptcy prediction," *Appl. Soft Computing*, vol. 24, 2014, pp. 977-984.

[11] Sasan Barak, Azadeh Arjmand and Sergio Ortobelli, "Fusion of multiple diverse predictors in stock market," *Springer*, vol. 36, 2016, pp. 90-102.

[12] Chen Chen, Wu Dongxing, Hou Chunyan and Yuan Xiajie, "Exploiting Social Media For Stock Market Prediction With Factorization Machine," in *Web Intelligence (WT) and Intelligent Agent Technologies (IAT)*, Poland, 2014.

[13] Xu Feifei and Vlado Keelj, "Collective Sentiment Mining of Microblogs in 24-Hour Stock Price Movement Prediction," in *Business Informatics (CBI)*, Geneva, Switzerland, 2014.

[14] Soujanya Poria, Erik Cambria, Gregoire Winterstein and Guang-Bin Huang, "Sentic patterns: Dependency based rules for concept level sentiment analysis," *Knowledge Based Systems*, vol. 69, 2014, pp. 45-63.

[15] Yassine, Mohamed and Hazem Hajj, "A Framework for Emotion Mining from Text in Online Social Networks,"

in *Data Mining Workshops (ICDMW)*, Sydney, Austraila, 2010.

[16] Vivek Narayanan, Ishan Arora and Arjun Bhatia, "Fast and accurate sentiment classification using an enhanced Naive Bayes model," *Springer Berlin Heidelberg*, vol. 8206, 2013, pp. 194-201.