

A Cluster based Non-Linear Regression Framework for Periodic Multi-Stock Trend Prediction on Real Time Stock Market Data

Lakshmana Phaneendra Maguluri¹, R. Ragupathy²

Department of Computer Science and Engineering, Annamalai University
Annamalai Nagar, Chidambaram, Tamil Nadu 608002, India

Abstract—Trend prediction is and has been one of the very important tasks in the stock market since day one. For a sophisticated trend prediction using real time stock market data, stock sentiment news and technical analysis plays a vital role. While predicting the trend in the conventional way, technical indicators are delayed due to temporal data and less historic data. All the conventional stock trend predicting methods sustained without sentiment scores, technical scores and time periods for trend prediction. Considering the fact that all the previous conventional methods of stock trend predictions are bound to take single stock for trend prediction due to high computational memory and time, this prototype of highly functioning algorithms focus on trend prediction with multi stock data breaking all the conventional rules. This multi stock trend prediction model commissions and implements the effectively programmed algorithms on real time stock market data set. In this multi-stock trend prediction model, a new stock technical indicator and new stock sentiment score are proposed in order to improve the stock feature selection for trend prediction. In order to find the best real time feature selection model, a technical feature selection measure and stock news sentiment score are developed and incorporated. We used integrated stock market data to make a hybrid clustered model to find the relational multi stocks. Giving a final verdict, this is a cluster based nonlinear regression multi stock framework in order to predict the time-based trend prediction. The multi stock trend regression accuracy is bettered by 12% and recall by 11% while we cross check the experimental outcomes, henceforth making this model more accurate and precision furnished.

Keywords—Multi-stock trend prediction; stock market; clustering; nonlinear regression

I. INTRODUCTION

Stock markets provide investors with the most profitable avenue to spend their money. The investors can't find another way to make a high rate of return growth from anywhere else. They will have to bear the loss if things go south. This implies that investment is quite a risky thing to be involved. There is every opportunity to make returns bigger and bigger, and to lose everything [1]. When looked into theories, any change in share prices is associated with change in fundamental variable relevant to share price assessment. For example, Stock earnings, size, dividend pay-out ratio, various economic variables etc. A moving average can be calculated using a predetermined period by using the mathematical analysis of the stock's average price value. Every time the price of the stock

changes, the average price either goes up or down [2]. Simple (arithmetic), triangular, exponential, and variable and weighted moving averages, calculated using open, close, low, high and stock price volumes are the different type of moving average indicators. The trend of the price maybe hiked or inflated at times at any point. These common outlines the flow of market trading. The investors gain profit when they buy a stock that shows uptrend. The uptrend stock here is value-appreciating stock. In uptrend, the value of the stock appreciates consistently over a period, even if there are consolidated. For example, a company's particular stock price began to say at ₹ 275, its price reached say ₹ 380, we say there is an uptrend for that stock, even though there were brief dips in stock prices between them. An uptrend can span hours, months, and years [3]. A downtrend is pretty much contrary of an uptrend. In downtrend, the stock price depreciates steadily over a period that may include some brief rises. If he follows these signals the investor can be benefited. Moving average trading is profitable if the price level shifts between selling and buying signals are adequate, otherwise it will result in losses. Bollinger Bands can be a great aid for dealers in dualistic decisions [4]. New openings to the trade can be opened by them. The market will likely get to seize around when the market approaches a Bollinger band. This information alone suffices for a dualistic decision to be won. Bollinger bands need a simple indication of how much they should make an attempt on the markets. Types of dualistic options with high outputs such as hierarchy options or one touch choices need this prediction, which is Bollinger Bands can turn a normal strategy into one that is highly profitable. Bollinger bands form essential levels of conflict and trend prediction. The relative strength index [5] is an indicator of momentum that measures the level of current price changes in order to assess the over-bought or over-sold conditions of a stock price or other quality. The RSI compares the momentum of stock price predictions [6].

When we select the appropriate technique of pre-processing data the sentiment analysis can be improved. This very fact makes pre-processing of data a crucial step in the process. Aside from the usual pre-processing techniques, some of the news articles require different pre-processing techniques because the content produced by the user community, for example, received messages from Twitter [7]. The views expressed in the news are either positive or negative or neutral opinion which plays an important role in the trend prediction. Analyzing sentiment is the task of selecting the sentiment label

for a given news article. It may also be considered a task to classify. As a result of this news reporting on a company, the opinions are formed among the investors, so they can make informed decisions about their share in that company's stock. All inputs are deemed independent from each other to predict the test class labels. The financial media reporters gather the information through reliable sources and the same would be disseminated in news article format [8]. The news articles which are published must be checked for trustworthiness. There are different methods of media where these news articles can be disseminated, and source from which the sentiment is to be derived from the news articles published in that source must be decided with utmost care. A Research Paper's result accuracy is based on the credible sources. Yahoo finance, the money control is few official websites we can say where the news articles are trustworthy. In the text classification approach, each word in the article is weighted with the frequency and this is classified within the specified group. Considering the importance of trading volume in understanding stock market microstructure, comprehensive empirical studies were conducted to research the relation between price, volatility and volume of trading. Market investors are often inclined to look for better investment options providing higher returns as the investment decision is made to gain better returns than other avenues available, or to expect a higher return than others. The probability of not achieving the anticipated or targeted return is commonly known as risk, but risk estimation is a difficult activity. Volatility is usually taken as the indicator of risk. Simple words volatility is a standard deviation in returns. Volatility can be actual volatility, historical volatility, the volatility implied and the volatility forward. Although considering the reasons for volatility, the economists argue that the market is moving according to the information provided to the market; others argue that volatility has little to do with the economic or external factors, and it is the reaction of the investors that exerts greater market impact. Investors are generally averse to risk. At the same time investment with volatile assets has to be made. The investment in security usually has varying purposes. Some buy stock and keep long to have the privilege of owning these capital assets. But some others are buying stock to sell and have the price differences. The return on equities varies with shifts in stock prices. Stock rates rarely remain the same. It is unpredictable. On the one hand, price volatility [26] is an opportunity, on the other, a threat to the investors. Price stability will reduce the risk stemming from price volatility. Yet stock prices cannot stay steady over time, because they are more prone to shifts in environmental factors. There are no bounds or barriers to the flow of funds in a globalized environment. The major players now in the Indian Stock Market are the FII (Foreign institutional investors). With the incurring losses, the risk of the stock increases, this is in fact measured by standard deviation statistic. This is the dispersion from what is required of the real. The larger the dispersion the greater the perceived security risk will be. The risk of a stock is viewed in relation to the market as well. Each safety is susceptible to market influence. Market influence may be greater or smaller. But the fact is, to a greater extent, the fortune of the individual stock is governed by the market. This part of the risk to the stock is called the systematic risk. All stocks on the market must share

that class of risk. Such risks are therefore also known as non-diversifiable risk, because they cannot be eliminated through diversification. The statistics used to measure this portion of overall risk are beta. A security beta tells how far the security is market related. Operation on the stock market became popular nowadays. Present investors don't find investing in the stock market as pointless. They find investment in stocks to be more remunerative than other opportunities. Formerly stock investment has not received due respect and it has been treated as somewhat speculative that even today some discounts for its social acceptability are considerable. Stocks give not just the institutional investors but also the small retail investors a better opportunity. But that doesn't mean everyone knows the surgery. Market operation transparency is still in jeopardy. The SEBI is trying hard to get things working.

Looking back at our previous contributions, we have developed a single stock trend prediction using the technical and news data in an intraday process. Now, we propose a single stock trend prediction model using the technical and news data in different periodic time intervals. In this contribution an advanced multi-stock trend prediction model is designed and implemented on real time stock market data in different periodic time intervals. In this paper, new multi-stock technical and sentimental scores are developed to improve the stock selection process. A multi-stock clustering algorithm and classification models are developed in order to predict the periodic multi-stock trend.

II. RELATED WORK

Jeon et al. [9] demonstrated behavior next day by using a random subsample of collected tweets for stock market. They've gathered the NASDAQ, S&P 500 and DJIA tweet posts. For each day, they considered the factor of combined fear and hope, and analyzed the relationship between market indicators and these factors. They reported that the above mentioned stocks had been negatively associated with emotional tweets. Their findings have proved that stock market reaction on the very next day can be predicted by collecting emotional data [10].

Vu [11] proposed a new machine learning system by integrating features, consumer assurance and last 3 days data into the products. The cross-validation method has been adopted in a Decision Tree classifier for integrating all of the filtered features. Pre-processing steps include extracting noisy data, normalizing tweets and selecting data. The model was tested with NER (Named Entity Recognition Task) and without NER for Google, Apple, Amazon, and Microsoft companies stock and yielded 80.49 percent, 82.93 percent, 75.00 percent, and 75.61 percent for up and down NER (Named Entity Recognition Task) labels, respectively [11].

Vijh et al. [12] created a thorough study of stock prediction from data collection (how to collect it from twitter and tweet description), cloud storage, and then the process of opinion analysis (software and techniques) and finally the phase of prediction. Over time they examined the correlation between financial markets and social media data. They built a cloud-based system in JSON format to store various dimensions of public emotions contained in fetched tweets. The program was assessed for four companies listed under the UK Stock

Exchange, and the data checked were collected for 30 days. Their finding will help the firms assess the concerns of stakeholders and establish a new market strategy. Overall, the research enhanced the efficiency in the forecasting phase with emotional analysis and synthesizing.

Zhang et al. [13] used the twitter and survey index sentiments and attentive indicators, volatility and trading volume of S&P index 500 to forecast returns. Various supervised learning techniques and the Diebold-Mariano test were conducted and compared with autoregressive baseline model to confirm the significance of sentiments and attention-based predictions. They noted that tweet volume and sentiments were relevant to predicting lower-market capitalization portfolios. In addition, they show that Kalman Filter indicators and Twitter sentiment were helpful in forecasting some sentiment labels based on surveys.

Chen et al. [14] analyzed the data obtained from various networking networks called chat rooms, web forums, and micro blogs and found different characteristics to be present. They believed that chat room posts at the activity level are strongly correlated with the trend in stock and assumption that is true. For chat room post sentiments the same performance was achieved with short posts reported from previous studies. The result indicated that post sentiments improved stock price return forecasting as compared to using only historical prices. They also developed a trading strategy and reported a return of 21 per cent over seven months. Proposed Model The overall process of predicting stock market direction consists of different steps that include data collection, pre-processing of text and selection of features. The programming is required for the overall work to be carried out. R language, python, was used with Java. The packages of those tools have been used to implement the algorithms proposed. The data our problem requires are of two types [15]. The historic stock values and the news stories from which the emotions are to be derived. Unlike the other systems which used the static data, our system is based on both the streaming data and the static data. The crawler crawls on the specified website and extracts the specified company's news articles for which the future direction of the stock is to be predicted. Since the stock prices must be correlated with the news articles, the news articles must be extracted along with the time stamps. Such news articles then act as the input to the module for the study of sentiments. The researchers have been pursuing paths of sentiment analysis for many years, and have come up with many different algorithms to characterize the text's feeling. Every algorithm has some advantages and disadvantages. Choosing the algorithm for sentiment analysis [25] may depend on the available datasets, domain and prior experience. One approach is to be chosen among approaches, linguistic-based, lexicon-based, and machine learning. If the approach is selected, the correct algorithm must be determined in that approach [16]. It is very crucial to decide what data set is being used for the research. Our framework does not dispose of readily accessible data sets. Most of the data is data processing that is being processed and stored in the database. Our system has to have two types of data. One relates to historical stock values and the other set of data containing news articles which are published online. The data used is from a combination of

two different sources to study the correlation between news articles and stock prices: a dataset of historical data and a corpus of news articles. The initial source of data used to extract news articles is the website of money control, which has a large reservoir of critical news for the individual stocks. Money control is India's premier source of financial information. They derived historical values for 2012 from <http://ichart.finance.yahoo.com> for the Infosys stock in NIFTY. This data is then loaded into a table of databases that can then be queried and processed. The moneycontrol.com Website was used for the news articles. To predict future prices based on the sequence of events, historical data are extracted from the moneycontrol.com web site for all the companies listed in BSE (around 3000) [17]. The code (scraper) is written to extract the open price of each company, close prices for the years 2007 to 2014. For this system, events from disclosure records and pieces of content collected from indiatimes.com, moneycontrol.com, sebi.com, watchoutinvestors.com, courts.gov.in, cibil.com are used as corpus.

Data pre-processing greatly decreases word space but there are still incentives for knowledge loss. Kang [18] measured the volatility of Indian stock market day-to-day returns. The study period was 1961-2005, and data were gathered together from the Economic Times Index and S&P CNX Nifty. The series observed volatility clustering quiet intervals of big returns were interspersed with cycles of volatility of great returns. The GARCH model was used to check the volatility effect asymmetry, and the result indicated a volatility asymmetry. It was known that high price movements started in response to strong economic fundamentals, and that the real reason for sudden movement was market imperfection [19].

Smruti et al. [20] proposed an extreme learning based PCA approach to predict the stock market data on limited training dataset. Shangkun et al. [21] proposed a gradient boosting approach to detect the trend in the china market. Shanoli et al. [22], proposed a novel time series model to predict the stock market data using the rule based approach on the training data.

III. FILTER BASEDD STOCK TECHNICAL PREDICTION MODEL

In the paper [23], we have proposed a novel filtered based classification model on the technical and stock news datasets in order to predict the trend of the to find the bullish trend stocks on the real-time market data. This model is tested on the continuous type of technical data for trend prediction. In the proposed framework, a correlated multi-stock trend prediction model is designed and implemented on the real-time market data. In the initial phase, a real-time stock technical data and its related news are extracted from the money control and zerodha websites. These technical data and news data are pre-processed using the novel approaches developed in the papers [24]. In this work, an improved version of technical indicator and sentiment scores are defined based on the contextual information of the stock data. These scores and technical data are integrated to form the training data. A novel clustering measure is used to form the clusters based on the integrated data features. This clustering model is implemented to form the clusters based on the technical data and scores of an integrated dataset. Finally, this clustered data is given to classification

model to predict the trend of the multiple stocks based on the input test sample as shown in the Fig. 1.

A. Stock Technical and Comments Data Collection and Pre-Processing

In this phase, all the stock related technical and news data are extracted from the zerodha or trade view or money control websites for data collection. All these collected data are pre-processed using the models in the papers [20][21]. Text pre-processor is applied on the stock news data as text filtering. In this work, a modified version of stock technical score and sentiment score are developed on the technical and stock news datasets.

B. Hybrid Stock News Score

To each comment in the stock corpus S, we construct a dictionary of words that contains bullish and bearish words. In the stock news training data S, each input stock news is represented as sn[i] and term frequencies of the sn[i] is presented as tsn[i][j], where i, j represent the jth term of the ith stock news s. Here, the term frequency and normalized term frequencies are used to find the news score of the stock. T represents the total tokens in the ith stock news. This normalized data is scaled by using inverse document frequency (idf) and multi-stock scaling factor (mssf) is represented in Eq. (1)

$$tfidf(sn(i), S, mssf) = tf(tsn(i), S) \times idf(tsn(i), mssf)$$

$$idf(sn(i), S, mssf) = \log \frac{mssf}{1 + \max\{tsn[i][j]\}; j = 1..T.}$$

$$mssf = \max\{tf(S[i], tf(S[j]))\} / N * Prob(tf(S[i]/tf(S[j])); i \neq j \tag{1}$$

C. Proposed New Stock Technical Indicator

In the paper, a novel mutual information (MI) is proposed to find the contextual relationship of the bullish and bearish stocks using the technical indicators. Hybrid technical mutual information is represented in terms of bullish and bearish cases as shown in Eq. (2).

$$Technical\ Indicator = TI = stgv(l, bu) \log \frac{strv(l, be)}{rsi(bu) * rsi(be)} \tag{2}$$

Where bu represents the bullish and be represents the bearish stock type.

$$T1 = P(\beta) * CP_s - MA \left(\frac{n}{2} + 1 \right)$$

$$Bu: T1 \geq 0; n: time\ interval$$

$$Be: T1 < 0$$

$$T2 = P(\beta) * \frac{S_{ma(3)} + S_{ma(6)} + S_{ma(12)} + S_{ma(24)}}{4};$$

$$Bu: CP_s \geq T2$$

$$Be: CP_s < T2$$

$$Final\ Multi-Stock\ technical\ indicator = MT = (T1 + T2) / TI$$

where, S_{ma(i)} is the ith stock moving average. Strv: Super trend red value, Stgv: Super trend green value. MA: moving average. CPs: Closed price of a stock s.

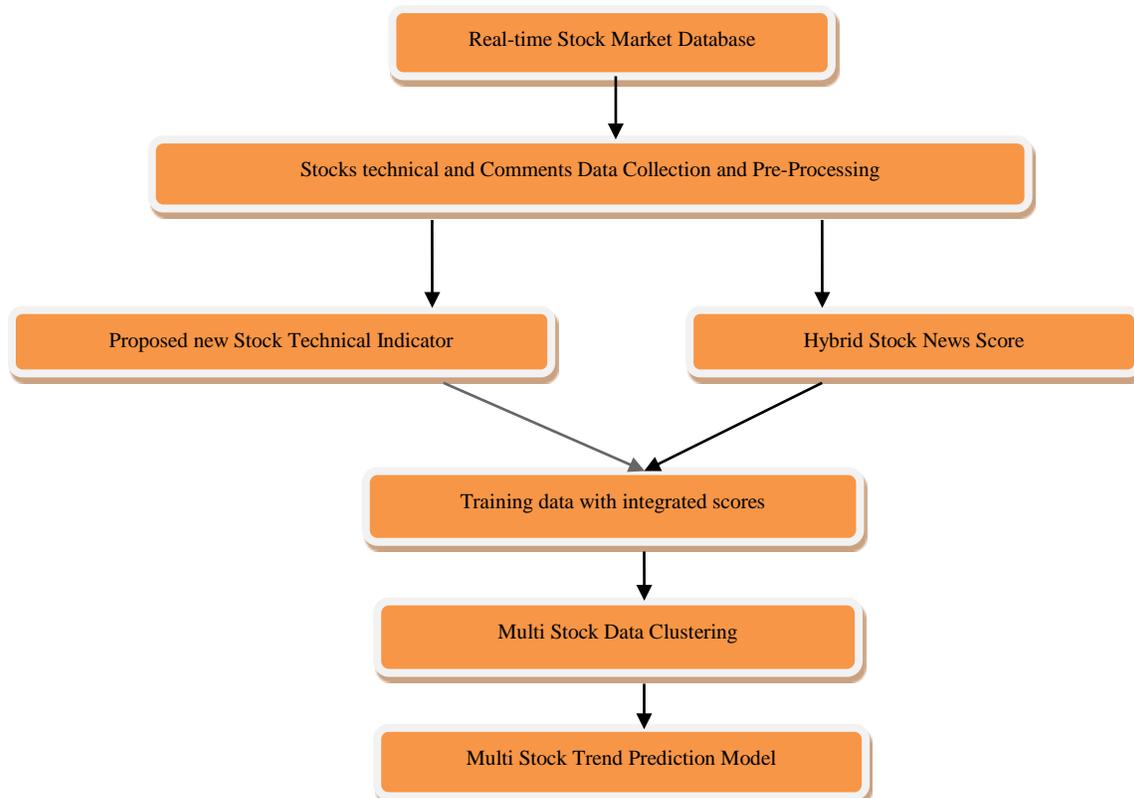


Fig. 1. Proposed Multi Stock Trend Prediction Framework.

D. Bullish Dictionary Words

Escalated, gain, enjoy, expansion, aggrandize, elevated, increment, rise, prefer, hallow, expand, supersize, idolize, positive, appreciate, plus, relish, accelerate, augment, raise, more, amplify, soar, adore, appreciative, approbatory, desire, esteem, approving, raised, swell, extend, addition, worship, climb, add, commendatory, venerate, augmentation, fancy, revere, friendly, proliferate, addendum, increased, escalate, proliferation, accumulate, love, stoke, complimentary, heightened, hype, uprising, accrual, boost, up, applauding, enlarge, admire, admiring, good, multiply, accretion.

E. Bearish Dictionary Words

Descend, recede, depreciative, abhor, drop, diminution, depreciatory, uncomplimentary, adverse, deplore, slide, detest, lower, plunge, lessen, unappreciative, depletion, dislike, dive, reduce, decrease, depressed, decreased, under, diminish, dip, low, derogatory, disapprove, unfavorable, negative, lowering, loathe, disfavor, unflattering, sink, receded, disdain, hate, decrement, unfriendly, subtract, loss, abate, decline, despise, fall, diminishment, lessening, downsize, abominate, minify, execrate, deprecate, inappreciative, dropped, shrinkage, reduction, wane, abatement, disapproving, dwindle, down.

F. Training Data with Integrated Scores

In this work, technical data is integrated with the newly computed technical score and the stocks news score as unlabeled data for data clustering. Here, multiple stock's technical data and scores are integrated to find the score-based data clustering on multiple stocks.

G. Multi-Stock Data Clustering

Input parameters: SF: stock features, SC: stock clusters

1: for each feature f_i $i = 1, \dots, N$ in the stock feature space SF

2: for each instance all points do

3: for all stock clusters SC_j $j = 1$ to k

4: do

5: Compute

$$C_i = \frac{\frac{1}{n} \sum_{i=1}^n (f_k - \mu_f)^3}{\sigma_k^3} \parallel \max\{f_k\} - \mu_f \parallel^2$$

6: Find $C_i^*(f_n) = \arg \min_i \left(\frac{\frac{1}{n} \sum_{i=1}^n (f_k - \bar{X})^3}{\sigma_k^3} \parallel f_k - \mu_f \parallel^2 \right)$

7: Update multi-stock cluster $C_i = \{f_n \mid C_i^*(f_n) = i\}$

8: repeat until k clusters

9: end for //number of features

10: end for //number of clusters

H. Multi-Stock Trend Prediction Model

In the proposed multi-stock trend prediction model, a hybrid multi-linear regression model is designed and implemented to predict the trend of the multiple stocks. In this model, a new probability estimation based non-linear regression model is designed and implemented on the training clustered dataset. A Non-linear regression estimation using the time wise trend prediction is given as.

Parameters:

$f(s(\alpha))$: input query stock.

$f(s(\beta))$: Related list of multiple stocks.

$r_{s(\alpha)}, r_{s(\beta)}$: RSI values of the input and multiple stocks.

$\varphi_{s(\alpha)}, \phi_{s(\beta)}$: BB value of input and its related multiple stocks.

1. Read the clustered training dataset.
2. Compute the non-linear regression based estimation to the time based trend estimation as

Let $m_{s(\alpha)}, m_{s(\beta)}$ are the test stock and its associated multi-stocks

$$f(m_s) \sim E(m(s), k(m_{s(\alpha)}, m_{s(\beta)}))$$

where

$$\upsilon = \text{Variance} = V(\text{RP}(\varphi_{s(\alpha)}, \gamma_{s(\beta)})) = \frac{\varphi_{s(\alpha)} \cdot \phi_{s(\beta)}}{(\varphi_{s(\alpha)} + \phi_{s(\beta)})^2 (\phi_{s(\beta)} + \gamma + 1)}$$

$$3. \quad \sigma = \text{SD} = \text{SD}(\text{RP}(\varphi_{s(\alpha)}, \phi_{s(\beta)})) = \sqrt{\frac{\varphi_{s(\alpha)} \cdot \phi_{s(\beta)}}{(\varphi_{s(\alpha)} + \phi_{s(\beta)})^2 (\varphi_{s(\alpha)} + \phi_{s(\beta)} + 1)}}$$

$ED(\chi) := f(\chi) = k(r_{s(\alpha)}, r_{s(\beta)}) \cdot \chi e^{-\chi} / \sigma$ for $x \geq 0$; // rsi based test stock and its correlated stocks data distribution estimation.

$$m(s) = ED[f(s(\alpha)), f(s(\beta))]$$

$$\text{Pr ob}(r_{s(\alpha)}, r_{s(\beta)}) = k(r_{s(\alpha)}, r_{s(\beta)}) / \upsilon$$

$$\begin{aligned}
 k(r_{s(\alpha)}, r_{s(\beta)}) &:= \frac{r_{s(\beta)}}{\pi[(x - r_{s(\alpha)})^2 + r_{s(\beta)}^2]} \\
 \text{ProbEsti} &= D\left(\frac{r_{s(\beta)}}{\pi[(x - r_{s(\alpha)})^2 + r_{s(\beta)}^2]} * \chi * e^{(-\chi x)} / \left(\sqrt{\frac{\varphi_{s(\alpha)} * \phi_{s(\beta)}}{(\varphi_{s(\alpha)} + \phi_{s(\beta)})^2 * (\varphi_{s(\alpha)} + \phi_{s(\beta)} + 1)}}\right)\right) \\
 &= \frac{\partial}{\partial x} \left(\frac{r_{s(\beta)}}{\pi[(x - r_{s(\alpha)})^2 + r_{s(\beta)}^2]} * \chi * e^{(-\chi x)} / \left(\sqrt{\frac{\varphi_{s(\alpha)} * \phi_{s(\beta)}}{(\varphi_{s(\alpha)} + \phi_{s(\beta)})^2 * (\varphi_{s(\alpha)} + \phi_{s(\beta)} + 1)}}\right)\right) \\
 &= \left(\frac{r_{s(\beta)}}{\pi[(x - r_{s(\alpha)})^2 + r_{s(\beta)}^2]}\right) e^{-\chi x} \frac{\partial}{\partial x} (-\chi x) \cdot \sqrt{\frac{\varphi_{s(\alpha)} * \phi_{s(\beta)}}{(\varphi_{s(\alpha)} + \phi_{s(\beta)})^2 * (\varphi_{s(\alpha)} + \phi_{s(\beta)} + 1)}} \\
 &= \left(\frac{r_{s(\beta)}}{\pi[(x - r_{s(\alpha)})^2 + r_{s(\beta)}^2]}\right) e^{-\chi x} - \chi \cdot \sqrt{\frac{\varphi_{s(\alpha)} * \phi_{s(\beta)}}{(\varphi_{s(\alpha)} + \phi_{s(\beta)})^2 * (\varphi_{s(\alpha)} + \phi_{s(\beta)} + 1)}} \\
 &= -\chi \cdot e^{-\chi x} \left(\frac{r_{s(\beta)}}{\pi[(x - r_{s(\alpha)})^2 + r_{s(\beta)}^2]}\right) \cdot \left(\sqrt{\frac{\varphi_{s(\alpha)} * \phi_{s(\beta)}}{(\varphi_{s(\alpha)} + \phi_{s(\beta)})^2 * (\varphi_{s(\alpha)} + \phi_{s(\beta)} + 1)}}\right)
 \end{aligned}$$

Estimate non-linear least square equation for

$$\begin{aligned}
 \text{Trend } T^* &= \text{Max} \left\{ -\chi \cdot e^{-\chi x} \left(\frac{r_{s(\beta)}}{\pi[(x - r_{s(\alpha)})^2 + r_{s(\beta)}^2]}\right) \cdot \left(\sqrt{\frac{\varphi * \phi}{(\varphi + \phi)^2 * (\varphi + \phi + 1)}}\right) * \max\{b(s(\alpha)), b(s(\beta))\} \right\} \\
 b(s(\alpha)) &= \frac{\sum_{i=1}^n (p(s(\alpha))_i - \frac{\sum_{i=1}^n p(s(\alpha))}{n}) (q(s(\beta))_i - \frac{\sum_{i=1}^n q(s(\beta))}{n})}{\sum_{i=1}^n (p(s(\alpha))_i - \frac{\sum_{i=1}^n p(s(\alpha))}{n})^2} \\
 b(s(\beta)) &= \frac{1}{n} \left(\sum_{i=1}^n q(s(\beta))_i - b(s(\alpha)) \cdot \sum_{i=1}^n p(s(\alpha))_i \right)
 \end{aligned}$$

4. If $(T^* \geq 1M(\text{MBB}) \& T^* > 1M(\text{MACDS}))$
5. then
6. stock s has Trend Positive.
7. Get stocks S* with similar T* value which satisfies the given condition on 1 minute time frame.
8. Else
9. stock s has Trend Negative.
10. Get stocks S* with similar T* value which satisfies the given condition on 1 minute time frame.
11. If $(T^* \geq 3M(\text{MBB}) \& T^* > 3M(\text{MACDS}))$
12. then
13. stock s has Trend Positive.
14. Get stocks S* with similar T* value which satisfies the given condition on 3 minute time frame.
15. Else
16. stock s has Trend Negative.
17. Get stocks S* with similar T* value which satisfies the given condition on 3 minute time frame.
18. If $(T^* \geq 5M(\text{MBB}) \& T^* > 5M(\text{MACDS}))$
19. then
20. stock s has Trend Positive.
21. Get stocks S* with similar T* value which satisfies the given condition on 5 minute time frame.
22. Else
23. stock s has Trend Negative.
24. Get stocks S* with similar T* value which satisfies the given condition on 5 minute time frame.

25. If $(T^* \geq 10M(MBB) \& \& T^* > 10M(MACDS))$
 26. then
 27. stock s has Trend Positive.
 28. Get stocks S^* with similar T^* value which satisfies the given condition on 10 minute time frame.
 29. Else
 30. stock s has Trend Negative.
 31. Get stocks S^* with similar T^* value which satisfies the given condition on 10 minute time frame.
 32. If $(T^* \geq 15M(MBB) \& \& T^* > 15M(MACDS))$
 33. then
 34. stock s has Trend Positive.
 35. Get stocks S^* with similar T^* value which satisfies the given condition on 15 minute time frame.
 36. Else
 37. stock s has Trend Negative.
 38. Get stocks S^* with similar T^* value which satisfies the given condition on 15 minute time frame.
 39. If $(T^* \geq 30M(MBB) \& \& T^* > 30M(MACDS))$
 40. then
 41. stock s has Trend Positive.
 42. Get stocks S^* with similar T^* value which satisfies the given condition on 30 minute time frame.
 43. Else
 44. stock s has Trend Negative.
 45. Get stocks S^* with similar T^* value which satisfies the given condition on 30 minute time frame.
 46. If $(T^* \geq 1H(MBB) \& \& T^* > 1H(MACDS))$
 47. then
 48. stock s has Trend Positive.
 49. Get stocks S^* with similar T^* value which satisfies the given condition on 1 hour time frame.
 50. Else
 51. stock s has Trend Negative.
 52. Get stocks S^* with similar T^* value which satisfies the given condition on 1 hour time frame.
 53. If $(T^* \geq 1D(MBB) \& \& T^* > 1D(MACDS))$
 54. then
 55. stock s has Trend Positive.
 56. Get stocks S^* with similar T^* value which satisfies the given condition on 1 day time frame.
 57. Else
 58. stock s has Trend Negative.
 59. Get stocks S^* with similar T^* value which satisfies the given condition on 1 day time frame.
 60. If $(T^* \geq 1W(MBB) \& \& T^* > 1W(MACDS))$
 61. then
 62. stock s has Trend Positive.
 63. Get stocks S^* with similar T^* value which satisfies the given condition on 1 week time frame.
 64. Else
 65. stock s has Trend Negative.
 66. Get stocks S^* with similar T^* value which satisfies the given condition on 1 week time frame.
 67. If $(T^* \geq 1Mt(MBB) \& \& T^* > 1Mt(MACDS))$
 68. then
 69. stock s has Trend Positive.
 70. Get stocks S^* with similar T^* value which satisfies the given condition on 1 month time frame.
 71. Else
 72. stock s has Trend Negative.
 73. Get stocks S^* with similar T^* value which satisfies the given condition on 1 month time frame.
-

In the proposed multi-stock trend prediction algorithm, a non-linear regression model is used to predict the trend of the input stock with different time frames. In this work, we have used 1m,3m,5m,10m,15m,30m,1H,1D,1W,1Mt time frames in order to predict the trend of the given stock based on the clustered stock market dataset. Here, the computed non-linear regression estimator value is tested against the MACD signal value and middle line Bollinger line values to predict the similar type of trends in the real-time market.

IV. EXPERIMENTAL RESULTS

Experimental results are simulated using java environment and real-time market data. Proposed model is compared to the traditional stock market classification models to verify the performance of the hybrid feature selection-based clustering and classification model to the traditional models. Also, proposed model is compared to the traditional techniques by using various statistical performance measures such as accuracy, true positive rate, recall, precision, false positive rate, runtime etc. These performance metrics are analyzed and compared by using third party java libraries. Different types of statistical metrics such as recall, precision, accuracy, F-measure are evaluated on the stock market sentiment data along with the technical data. These statistical measures are evaluated based on the confusion matrix as described in Table I.

Accuracy: It is the ratio of correctly labelled stock predictions class labels to the entire stock class labels as shown in Eq. (3):

$$Stok\ Accurecy(SA) = \frac{(STP+STN)}{(STP+SFP+SFN+STN)} \quad (3)$$

Precision: It is the ratio of correctly classified positive stock classes to the all actual positive and negative labelled stock classes as shown in Eq. (4):

$$Stock\ Precision(SP) = \frac{STP}{(STP+SFP)} \quad (4)$$

Recall: It is the ratio of correctly classified positive stock classes' labels to the all predicted positive and negative labelled stock classes as shown in Eq. (5):

$$Stock\ Recall(SR) = \frac{STP}{(STP+SFN)} \quad (5)$$

F-Measure: It is the harmonic average of recall and precision as shown in Eq. (6):

$$Stock\ F - Measure(SF) = \frac{2*SR*SP}{SR+SP} \quad (6)$$

TABLE I. STOCK STATISTICAL MEASURES

Model Predicted values			
Stock positive		Stock negative	
Actual stock values	Stock positive	Stock true positive prediction(STP)	Stock false positive prediction(SFP)
	Stock negative	Stock false negative prediction(SFN)	Stock true negative prediction(STN)

Fig. 2 illustrates the 5 min candlestick pattern graph in the ZERODHA brokerage website. As shown in the Fig. 2, it is noted that the reliance industries stock is uptrend in the afternoon session.

Fig. 3 illustrates the 10 min candlestick pattern graph in the ZERODHA brokerage website. As shown in the Fig. 3, it is noted that the reliance industries stock is downtrend in the morning session and slightly uptrend in the afternoon session.

Fig. 4 illustrates the 15 min candlestick pattern graph in the ZERODHA brokerage website. As shown in the Fig. 4, it is noted that the reliance industries stock is downtrend in the morning session and slightly uptrend in the afternoon session.

Fig. 5 illustrates the 5 min candlestick pattern graph in the ZERODHA brokerage website. As shown in the Fig. 5, it is noted that the HDFC bank stock has uptrend in the afternoon session.



Fig. 2. Zerodha: 5 Min Reliance Industries Candlestick Chart.



Fig. 3. Zerodha: 10 Min Reliance Industries Candlestick Chart.



Fig. 4. Zerodha: 15 Min Reliance Industries Candlestick Chart.



Fig. 5. Zerodha: 5 Min HDFC Bank Candlestick Chart.

Fig. 6 illustrates the 10 min candlestick pattern graph in the ZERODHA brokerage website. As shown in the Fig. 6, it is noted that the HDFC bank stock is downtrend in the morning session and slightly uptrend in the afternoon session.

Fig. 7 illustrates the 15 min candlestick pattern graph in the ZERODHA brokerage website. As shown in the Fig. 7, it is noted that the HDFC bank stock is downtrend in the morning session and slightly uptrend in the afternoon session.

Fig. 8 illustrates the 5 min candlestick pattern graph in the ZERODHA brokerage website. As shown in the Fig. 8, it is noted that the nifty index is uptrend in the entire session.

Fig. 9 illustrates the 10 min candlestick pattern graph in the ZERODHA brokerage website. As shown in the Fig. 9, it is noted that the Nifty index is downtrend in the morning session and slightly uptrend in the afternoon session.

Fig. 10 illustrates the 15 min candlestick pattern graph in the ZERODHA brokerage website. As shown in the Fig. 10, it is noted that the Nifty index is downtrend in the morning session and slightly uptrend in the afternoon session.



Fig. 10. Zerodha: 15 Min NIFTY Index Candlestick Chart.

Table II describes the performance of computational runtime (ms) of stock trend feature extraction using the proposed approach on large datasets. From the Table II, it is clearly shown that the present feature extraction procedure has low computation runtime as compared to the conventional approaches.

Table III illustrates the proposed multi-stock feature selection measures on the input data. From the Table III, it is observed that the proposed multi-stock feature selection has better filtering than the conventional feature selection measures.

Table IV describes the historical data of the multiple stocks from the real-time market. From the Table IV, periodic levels of stock such as stock volume, stock MACD, stock volume change, low and high details are taken from the real-time market.



Fig. 6. Zerodha: 10 Min HDFC Bank Candlestick Chart.



Fig. 7. Zerodha: 15 Min HDFC Bank Candlestick Chart.



Fig. 8. Zerodha: 5 Min NIFTY Index Candlestick Chart.



Fig. 9. Zerodha: 10 Min NIFTY Index Candlestick Chart.

TABLE II. RUNTIME OF THE STOCK FEATURE COMPUTATION MEASURES ON THE REAL-TIME DATASET

Stocks	Roughset	PCA	PSO	MultiStockFS
#1	6238	6637	6030	3777
#2	6868	6354	6086	3682
#3	7780	7637	6301	3858
#4	6321	5105	7771	3679
#5	6641	6633	6813	3995
#6	6290	6609	5357	3756
#7	6362	7051	7573	4742
#8	6939	5876	7640	4292
#9	5349	6132	5701	4231
#10	6942	6153	7816	4780
#11	6860	5317	5191	3575
#12	6552	6384	5442	3722
#13	7709	7463	6909	4518
#14	7457	5957	6097	3666
#15	7755	6264	5348	3688
#16	6157	6746	5625	3659
#17	5796	5386	5113	4546
#18	6833	7465	6554	3369
#19	5379	5538	5078	4593
#20	6142	6056	6448	3358

TABLE III. STOCK TREND FEATURES EXTRACTION USING THE PROPOSED MODEL

Stocks	Roughset	PCA	PSO	MultiStockFS
#1	62	66	58	47
#2	58	67	64	55
#3	65	65	62	52
#4	63	61	60	54
#5	67	66	62	48
#6	67	59	63	53
#7	66	63	63	53
#8	64	67	67	48
#9	59	63	60	50
#10	64	64	64	50
#11	67	61	62	53
#12	62	61	64	55
#13	65	64	67	54
#14	62	61	58	48
#15	68	67	65	48
#16	65	63	66	49
#17	66	61	62	54
#18	68	60	64	48
#19	67	68	62	53
#20	68	61	64	54

Table V describes the historical data of the multi-stock technical data with MACD less than zero for clustering and classification. From the Table V, periodic levels of stock such as stock high, stock low, stock change, and stock volume, stock MACD are taken from the real-time market.

Table VI describes the performance of F1-measure of multi-stock trend classification using the proposed framework on large datasets. From the Table VI, it is clearly shown that the present framework has better efficiency F1-measure as compared to the conventional approaches.

Fig. 11 describes the performance of recall of multi-stock trend classification using the proposed learning framework on large datasets. As shown in the Fig. 11, it is clearly shown that the present framework has better efficiency recall as compared to the conventional frameworks.

Table VII describes the performance of precision of multi-stock trend classification using the proposed framework on large datasets. From the Table VII, it is clearly shown that the present framework has better efficiency precision measure as compared to the conventional approaches.

Fig. 12 describes the performance of accuracy of stock trend classification using the proposed multi-stock trend prediction framework on large datasets. As shown in the Fig. 12, it is clearly shown that the present framework has better efficiency accuracy as compared to the conventional frameworks.

TABLE IV. HISTORICAL SAMPLE MULTIPLE STOCKS AND ITS TECHNICAL DATA IN THE TRAINING DATA

Symbol	% change	price	volume	High	low	MACD(>0)
UNICHEMLAB	19.99	282.1	1284742	282.1	239.1	17.86
AMBER	16.69	1730	465661	1749.95	1490	25.69
VIMTALABS	14.63	119.5	1574820	122.85	101.55	5.36
KREBSBIO	12.74	101.75	178157	105	85.65	2.55
SOLARA	11.2	846.2	989205	897	778.3	31.99
MASTEK	10.95	635	942715	649.5	573.9	52.34
PGEL	9.95	45.85	106109	45.85	43.15	-0.28
KOPRAN	9.92	52.65	908150	52.65	47.4	3.55
ONMOBILE	9.91	32.15	255665	32.15	28.85	0.437
JUBILANT	9.81	873	1649341	895	791	45.33
SMSPHARMA	9.71	82.45	2135488	83.9	74	5
RATNAMANI	9.56	1144.95	40209	1192.85	1043.2	16.2
NEOGEN	9.39	590.15	109135	593.45	538.35	12.35
BLISSGVS	8.9	116.85	1111067	118	106.85	2.4
CAPLIPPOINT	8.31	467.85	922628	472	435	21.79
NEULANLAB	8.2	845.1	449593	850	765	63.7
ERIS	8.09	531.05	778165	550	481.05	4.84
INFOBEAN	8.06	117.35	75037	119.3	107.4	3.01
WOCKPHARMA	8.01	298.1	2887437	301.9	276	3.49
SMSLIFE	7.97	362.9	63598	368	331	10.01
GRANULES	7.85	294.75	8874157	304.3	274.4	20.76

RKFORGE	7.8	156.25	243706	160	143	-5.32
SASKEN	7.79	584.95	114039	595.6	536	26.77
ALPA	7.77	27.05	666647	27.6	24.35	1
LAURUSLABS	7.63	1004.5	10940287	1075.15	955	91.21
HIMATSEIDE	7.48	66.1	385132	66.65	61.5	0.93
BASF	7.45	1432.15	156248	1447.3	1332	46.21
INDOCO	7.42	243.35	558803	254.4	229.7	3.57
DIXON	7.4	8198.9	138471	8358	7720.15	519.66
LINCOLN	7.3	206.6	565273	211.8	191.7	8.03
BALPHARMA	7.23	46	85579	46.45	42.5	0.276
VISAKAIND	6.98	292.1	142707	295	268.85	12.82
TATAMOTORS	6.88	111.85	1.17E+08	113.5	102.9	1.34
TEXINFRA	6.64	37.75	625822	38.8	36.55	-0.327
ALKEM	6.49	2835	443912	2836	2680.35	65.93
GESHIP	6.41	240.85	727951	243	228.4	3.11
POLYMED	6.36	415	258325	421	383.1	24.5
SHILPAMED	6.16	584	462587	597	560	15.65
JUBLINDS	6.1	113.1	77728	116.8	106.5	0.826
SASTASUNDR	5.93	85.8	9005	85.85	80.1	1.22
GRPLTD	5.71	740	919	751.05	682.35	15.78
SANOFI	5.67	8208.45	59322	8287	7849	14.37
ZYDUSWELL	5.67	1702.1	194025	1739.95	1615	87.75
FDC	5.65	316.05	3007566	324.4	300.5	9.79
PANACEABIO	5.5	227.25	436753	231	217.4	5.27
ADVENZYMES	5.1	199.8	904277	202.4	190.2	5.3
FELDVR	5	17.85	262859	17.85	17.15	0.578
PSL	5	1.05	18633	1.05	1.05	0.132
TATACOMM	5	797.7	169136	797.7	770.25	42.36
CREATIVE	4.99	94.65	9760	94.65	93.7	3.93

TABLE V. HISTORICAL MULTI-STOCK TECHNICAL DATA WITH MACD LESS THAN ZERO

symbol	% change	price	volume	High	Low	MACD(<0)
MAANALU	18.3	60.45	178577	61.3	48	0.23
KGL	14.29	0.4	197734	0.4	0.35	-0.022
UVSL	11.11	0.5	18958673	0.5	0.45	-0.013
NTL	11.11	0.5	4300	0.5	0.5	-0.057
FCSOFT	11.11	0.5	2335985	0.5	0.45	-0.004
METKORE	10	0.55	7913	0.55	0.55	-0.066
CUPID	9.64	227.55	447482	231.8	209.15	3.9
ALICON	8.85	270.1	11552	283.5	250	0.403
GAMNINFRA	7.69	0.7	559919	0.7	0.65	-0.017
EROSMEDIA	7.65	19.7	753262	20	17.6	0.324
INDNIPPON	7.1	286.5	158200	300.1	268.95	-1.39
TATAMTRDVR	6.67	40	13128562	40.3	37.25	-0.752
TASTYBITE	6.63	12568.95	3575	12589.8	11601	14.29

NECLIFE	6.01	22.05	1142308	22.45	20.85	0.132
AVTNPL	5.96	40	512506	40.85	37.75	0.663
NBVENTURES	5.95	48.95	436139	49.5	46.35	0.012
MERCATOR	5.88	0.9	723310	0.9	0.8	-0.088
EUROMULTI	5.88	0.9	13420	0.9	0.8	0.028
BALAMINES	5.79	603.5	480666	614.75	566.45	20.17
MAYURUNIQ	5.74	231.05	172540	246.9	217.2	6.96
LOTUSEYE	5.71	27.75	2274	29	27	-0.761
INSPIRISYS	5.59	26.45	9846	26.75	25.1	0.346
BILENERGY	5.56	0.95	619895	0.95	0.9	-0.057
BFINVEST	5.53	313.9	40009	315	298	2.82
LUMAXIND	5.49	1309	85835	1319.6	1211	32.92
GSCLCEMENT	5.46	30.9	221319	31.35	28.85	0.268
IGARASHI	5.29	271.7	132410	275.65	258.05	-0.776
JMCPROJECT	5.24	47.2	72962	48.5	44.85	-0.804
GEOJITFSL	5.17	36.6	261160	36.8	34.3	0.929
ASTEC	5.11	987	190376	1032.9	936.1	47.17
BOMDYEING	5.08	62	1591922	62.4	58.3	-2.2
GUJAPOLLO	5.05	171.65	4630	175.95	160	-1.31
OPTIEMUS	5	21	56611	21	21	-0.677
NATHBIOGEN	5	350.7	10431	350.7	340	3.9
UNITECH	5	2.1	2111540	2.1	2.05	0.024
SHIRPUR-G	5	7.35	39968	7.35	6.8	-0.09
DQE	5	1.05	1150	1.05	1.05	-0.151
AFFLE	5	1772.95	58185	1772.95	1681.8	39.91
BFUTILITIE	4.99	229.35	133305	229.35	217.8	7.73
ESTER	4.98	59	233721	59	56.5	1.72
TRIVENI	4.98	56.9	148572	56.9	56.9	0.735
WABAG	4.98	120.15	40710	120.15	120	2.93
PREMEXPLN	4.97	115	2873	115	108.5	2.46
DWARKESH	4.97	25.35	1119087	25.35	24.3	0.229
SKIPPER	4.96	37	74287	37.3	35.25	-0.528
AURIONPRO	4.96	48.65	17269	48.65	45.1	-0.829
CIMMCO	4.96	20.1	60846	20.1	19.2	0.417
BODALCHEM	4.96	77.25	1070456	78.9	72.4	3.5
BCG	4.96	6.35	1785857	6.35	6.05	-0.588
VAKRANGEE	4.96	29.65	342576	29.65	29.65	-0.822

TABLE VI. PERFORMANCE ANALYSIS OF F1-MEASURE USING DIFFERENT TRADITIONAL CLASSIFICATION LEARNING FRAMEWORKS

StockName	SVM	RF	NB	NN	CNN	MultiStockFS
LT	0.91	0.95	0.9	0.91	0.89	0.97
ASIAN_PAINT	0.94	0.89	0.94	0.9	0.9	0.97
AXIS_BANK	0.92	0.92	0.92	0.92	0.93	0.97
BAJAJ_AUTO	0.93	0.87	0.94	0.91	0.87	0.98
BAJFINANCE	0.89	0.91	0.91	0.94	0.89	0.98

BAJAJFINSV	0.93	0.91	0.95	0.95	0.89	0.97
BPCL	0.91	0.9	0.9	0.89	0.9	0.96
BHARTIARTL	0.94	0.92	0.88	0.91	0.92	0.99
INFRATEL	0.88	0.88	0.92	0.92	0.92	0.96
BRITANNIA	0.94	0.92	0.94	0.92	0.91	0.97
CIPLA	0.93	0.89	0.9	0.93	0.91	0.97
COALINDIA	0.94	0.91	0.91	0.88	0.94	0.98
DRREDDY	0.9	0.88	0.87	0.88	0.93	0.97
EICHERMOT	0.89	0.95	0.91	0.95	0.95	0.96
GAIL	0.88	0.87	0.9	0.91	0.92	0.98
GRASIM	0.88	0.9	0.94	0.92	0.9	0.98
HCLTECH	0.88	0.94	0.94	0.93	0.92	0.97
HDFCBANK	0.89	0.89	0.87	0.89	0.89	0.96
HEROMOTOCO	0.94	0.88	0.89	0.89	0.89	0.97
HINDALCO	0.93	0.87	0.89	0.91	0.95	0.98

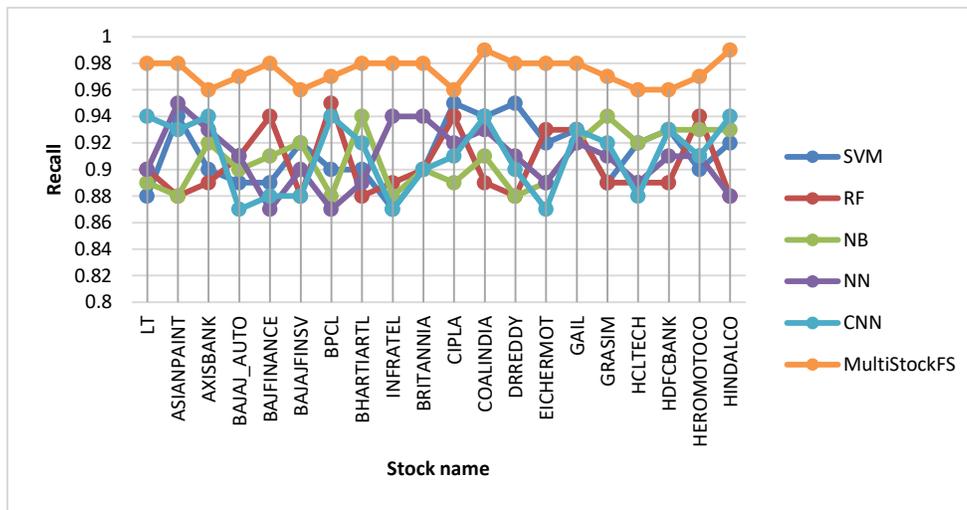


Fig. 11. Performance Analysis of Recall using different Traditional Feature Selection based Classification Frameworks.

TABLE VII. PERFORMANCE ANALYSIS OF PRECISION MEASURE USING DIFFERENT TRADITIONAL CLASSIFICATION LEARNING FRAMEWORKS

StockName	SVM	RF	NB	NN	CNN	MultiStockFS
LT	0.91	0.88	0.89	0.9	0.88	0.99
ASIANPAINT	0.87	0.92	0.88	0.93	0.88	0.99
AXISBANK	0.93	0.94	0.89	0.88	0.93	0.97
BAJAJ_AUTO	0.88	0.93	0.87	0.93	0.89	0.99
BAJFINANCE	0.9	0.92	0.89	0.88	0.95	0.98
BAJAJFINSV	0.88	0.94	0.89	0.92	0.89	0.97
BPCL	0.88	0.94	0.91	0.93	0.9	0.97
BHARTIARTL	0.88	0.92	0.91	0.88	0.95	0.98
INFRATEL	0.89	0.93	0.95	0.92	0.9	0.99
BRITANNIA	0.9	0.94	0.93	0.93	0.95	0.98
CIPLA	0.92	0.91	0.88	0.88	0.88	0.98
COALINDIA	0.93	0.95	0.87	0.89	0.89	0.96
DRREDDY	0.93	0.9	0.94	0.95	0.94	0.98

EICHERMOT	0.88	0.89	0.91	0.93	0.9	0.97
GAIL	0.95	0.94	0.91	0.92	0.94	0.99
GRASIM	0.91	0.92	0.89	0.91	0.92	0.96
HCLTECH	0.9	0.9	0.89	0.94	0.92	0.99
HDFCBANK	0.89	0.89	0.92	0.91	0.9	0.98
HEROMOTOCO	0.95	0.93	0.91	0.9	0.93	0.98
HINDALCO	0.93	0.89	0.92	0.91	0.91	0.97

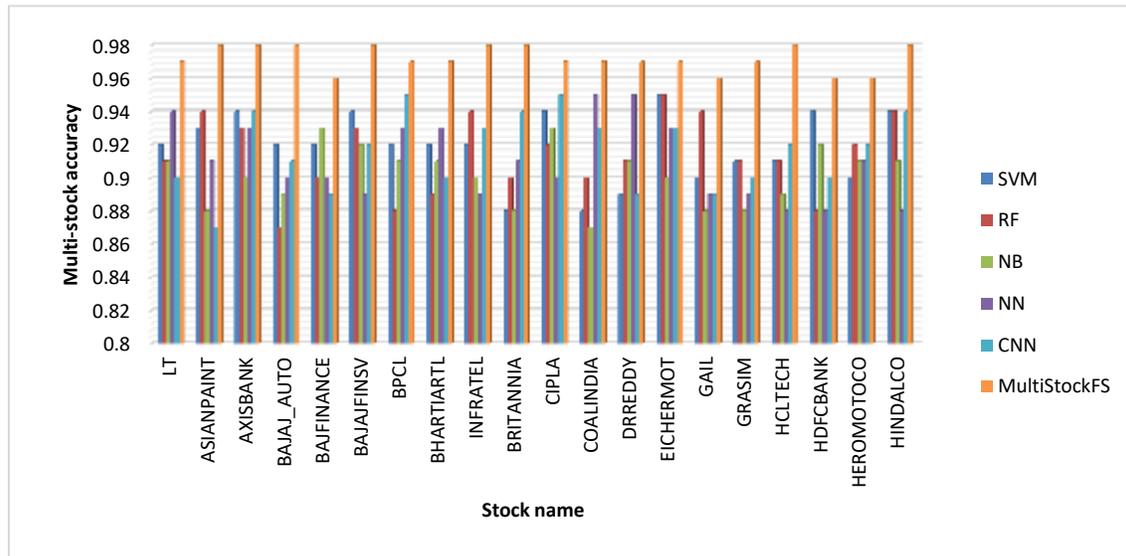


Fig. 12. Performance Analysis of Accuracy using different Traditional Classification Learning Frameworks.

V. INFERENCE

Performance of various segments such as precision, recall, F-Measure, accuracy and runtime are improved due to data filtering and feature selection in the above model as we can see in the results from the above tables. From the above tables, it is clearly identified that the proposed feature extraction and scoring approach optimizes the stock sentiment of the social media comments and its technical data. The proposed stock feature has less runtime and more efficiency in the real time stock market databases when compared with the traditional feature extraction measures. When the traditional classifiers and the proposed non-linear classifiers are compared from the above tables, it is observed that the performance of the non-linear classifiers are better than the traditional classifiers in terms of recall precision, accuracy and runtime(ms). 12% of accuracy is obtained through the proposed model when compared to the traditional stock market prediction classifiers.

VI. CONCLUSIONS

In this paper, a hybrid real-time multi-stock trend prediction model is designed and implemented on the stock market data. Since, most of the conventional single stock trend prediction models are depend on data size and limited feature space, it is difficult to find a novel feature selection measure on the stock technical data and stock news data. Also, these models are independent of temporal features for stock trend prediction. In this work, an advanced time based multi-stock trend prediction model is developed on the real-time data. In

this model, a new technical stock feature selection indicator and sentiment scores are computed for the clustering method. Finally, a cluster based non-linear regression framework for periodic multi-stock trend prediction is applied on the real time stock market data. Experimental results proved that the present model has better efficiency than the traditional technical indicators in terms of accuracy, f-measure, precision and recall. From the experimental results, it is observed that the proposed stock market trend prediction model has 9% of runtime (ms) and 12% of average classification accuracy as compared to the traditional trend prediction models on training and test dataset.

REFERENCES

- [1] O. Aladesanmi, F. Casalin, and H. Metcalf, "Stock market integration between the UK and the US: Evidence over eight decades," *Global Finance Journal*, vol. 41, pp. 32–43, Aug. 2019, doi: 10.1016/j.gfj.2018.11.005.
- [2] J. Bley and M. Saad, "An analysis of technical trading rules: The case of MENA markets," *Finance Research Letters*, vol. 33, p. 101182, Mar. 2020, doi: 10.1016/j.frl.2019.04.038.
- [3] A. C. Briza and P. C. Naval, "Stock trading system based on the multi-objective particle swarm optimization of technical indicators on end-of-day market data," *Applied Soft Computing*, vol. 11, no. 1, pp. 1191–1201, Jan. 2011, doi: 10.1016/j.asoc.2010.02.017.
- [4] O. Bustos and A. Pomares-Quimbaya, "Stock market movement forecast: A Systematic review," *Expert Systems with Applications*, vol. 156, p. 113464, Oct. 2020, doi: 10.1016/j.eswa.2020.113464.
- [5] Z. Dai, X. Dong, J. Kang, and L. Hong, "Forecasting stock market returns: New technical indicators and two-step economic constraint method," *The North American Journal of Economics and Finance*, vol. 53, p. 101216, Jul. 2020, doi: 10.1016/j.najef.2020.101216.

- [6] S. R. Das, D. Mishra, and M. Rout, "Stock market prediction using Firefly algorithm with evolutionary framework optimized feature reduction for OSELM method," *Expert Systems with Applications: X*, vol. 4, p. 100016, Nov. 2019, doi: 10.1016/j.eswx.2019.100016.
- [7] S. Das, R. K. Behera, M. kumar, and S. K. Rath, "Real-Time Sentiment Analysis of Twitter Streaming data for Stock Prediction," *Procedia Computer Science*, vol. 132, pp. 956–964, Jan. 2018, doi: 10.1016/j.procs.2018.05.111.
- [8] D. P. Gandhmal and K. Kumar, "Systematic analysis and review of stock market prediction techniques," *Computer Science Review*, vol. 34, p. 100190, Nov. 2019, doi: 10.1016/j.cosrev.2019.08.001.
- [9] B. N. Jeon and B.-S. Jang, "The linkage between the US and Korean stock markets: the case of NASDAQ, KOSDAQ, and the semiconductor stocks," *Research in International Business and Finance*, vol. 18, no. 3, pp. 319–340, Sep. 2004, doi: 10.1016/j.ribaf.2004.04.006.
- [10] H. M. G. E.a., V. K. Menon, and S. K.p., "NSE Stock Market Prediction Using Deep-Learning Models," *Procedia Computer Science*, vol. 132, pp. 1351–1362, Jan. 2018, doi: 10.1016/j.procs.2018.05.050.
- [11] M. S. Pagano, L. Peng, and R. A. Schwartz, "A call auction's impact on price formation and order routing: Evidence from the NASDAQ stock market," *Journal of Financial Markets*, vol. 16, no. 2, pp. 331–361, May 2013, doi: 10.1016/j.finmar.2012.11.001.
- [12] M. Vijh, D. Chandola, V. A. Tikkiwal, and A. Kumar, "Stock Closing Price Prediction using Machine Learning Techniques," *Procedia Computer Science*, vol. 167, pp. 599–606, Jan. 2020, doi: 10.1016/j.procs.2020.03.326.
- [13] Y. Zhang and L. Wu, "Stock market prediction of S&P 500 via combination of improved BCO approach and BP neural network," *Expert Systems with Applications*, vol. 36, no. 5, pp. 8849–8854, Jul. 2009, doi: 10.1016/j.eswa.2008.11.028.
- [14] M.-Y. Chen, C.-H. Liao, and R.-P. Hsieh, "Modeling public mood and emotion: Stock market trend prediction with anticipatory computing approach," *Computers in Human Behavior*, vol. 101, pp. 402–408, Dec. 2019, doi: 10.1016/j.chb.2019.03.021.
- [15] R. Jacinto, E. Reis, and J. Ferrão, "Indicators for the assessment of social resilience in flood-affected communities – A text mining-based methodology," *Science of The Total Environment*, p. 140973, Jul. 2020, doi: 10.1016/j.scitotenv.2020.140973.
- [16] M. M. Rounaghi and F. Nassir Zadeh, "Investigation of market efficiency and Financial Stability between S&P 500 and London Stock Exchange: Monthly and yearly Forecasting of Time Series Stock Returns using ARMA model," *Physica A: Statistical Mechanics and its Applications*, vol. 456, pp. 10–21, Aug. 2016, doi: 10.1016/j.physa.2016.03.006.
- [17] X. Li, P. Wu, and W. Wang, "Incorporating stock prices and news sentiments for stock market prediction: A case of Hong Kong," *Information Processing & Management*, vol. 57, no. 5, p. 102212, Sep. 2020, doi: 10.1016/j.ipm.2020.102212.
- [18] J. H. Yu, J. Kang, and S. Park, "Information availability and return volatility in the bitcoin Market: Analyzing differences of user opinion and interest," *Information Processing & Management*, vol. 56, no. 3, pp. 721–732, May 2019, doi: 10.1016/j.ipm.2018.12.002.
- [19] Z. Zhou, M. Gao, Q. Liu, and H. Xiao, "Forecasting stock price movements with multiple data sources: Evidence from stock market in China," *Physica A: Statistical Mechanics and its Applications*, vol. 542, p. 123389, Mar. 2020, doi: 10.1016/j.physa.2019.123389.
- [20] S. R. Das, D. Mishra, and M. Rout, "Stock market prediction using Firefly algorithm with evolutionary framework optimized feature reduction for OSELM method," *Expert Systems with Applications: X*, vol. 4, p. 100016, Nov. 2019, doi: 10.1016/j.eswx.2019.100016.
- [21] S. Deng, C. Wang, M. Wang, and Z. Sun, "A gradient boosting decision tree approach for insider trading identification: An empirical model evaluation of China stock market," *Applied Soft Computing*, vol. 83, p. 105652, Oct. 2019, doi: 10.1016/j.asoc.2019.105652.
- [22] S. S. Pal and S. Kar, "Time series forecasting for stock market prediction through data discretization by fuzzistics and rule generation by rough set theory," *Mathematics and Computers in Simulation*, vol. 162, pp. 18–30, Aug. 2019, doi: 10.1016/j.matcom.2019.01.001.
- [23] L.P.Maguluri, and R. Ragupathy "A New Sentiment Score Based Improved Bayesian Networks For Real-Time Intraday Stock Trend Classification". *International Journal of Advanced Trends in Computer Science and Engineering*, vol. 8, no. 4, 2019, pp. 1045-1055. *The World Academy Of Research In Science And Engineering*, doi:10.30534/ijatcse/2019/10842019. Accessed 20 July 2020.
- [24] L.P.Maguluri, and R. Ragupathy, "An Efficient Stock Market Trend Prediction Using the Real-Time Stock Technical Data and Stock Social Media Data", *International Journal of Intelligent Engineering and Systems*, Vol.13, No.4, 2020, DOI: 10.22266/ijies2020.0831.28.
- [25] M. Syamala, and N.J. Nalini, "A deep analysis on aspect based sentiment text classification approaches", *International Journal of Advanced Trends in Computer Science and Engineering*, Volume 8, No.5, September - October 2019, <https://doi.org/10.30534/ijatcse/2019/01852019>.
- [26] M. Syamala, and N.J. Nalini, "A filter based improved decision tree sentiment classification model for real-time amazon product review data", *International Journal of Intelligent Engineering and Systems*, Vol.13, No.1, 2020 DOI: 10.22266/ijies2020.0229.18.