INTERNATIONAL JOURNAL
OF TECHO-ENGINEERING
ISSN : 2057-5688

Cluster based classification model for trend prediction in Time series data

T.Rajesh
Research scholor,JNTUH
Asst professor, CSE,GNITS
rajesht531@gmail.com

Dr. K.VG Rao
Professor
GNITS
kvgrao1234@gmail.com

ABSTRACT

Real-time stock market trend prediction plays a essential role in the technical analysis for trend prediction process. Traditional historical technical indicators are difficult to trace and predict the trend due to noise or uncertainty in the training dataset. Since, most of the traditional homogeneous or heterogeneous stocks are used to predict the overall market sentiment for the institutions or retailors. Also, traditional approaches are difficult to find the outliers in the technical indicators data for the clustering process. These models only consider the static metrics or static outlier meaures due to variation in data distribution according to time basis. In this work, a hybrid knn classifier is used to predict the trend of the stock.

Keywords: Stock market technical data, outlier analysis, clustering approach , stock market.

1.Introduction

Machine learning (ML) is a data analytics method that acquires knowledge from previous knowledge and provides information directly to the systems[1]. Due to the complex interaction between the features, it is a difficult task to select the most relevant features from the existing ones. There are many types of algorithms for machine learning. Among them, supervised and unsupervised learning are the two main ones. Other machine learning algorithms, except for the above two, are reinforcement learning, recommendation systems, etc. Two main steps, i.e. training time and testing process, are undertaken by a supervised machine learning task. The model is built during the training period and is also tested during the testing phase. The selected data is then split into a dataset for training and testing. Finally, training begins and enters an iterative process of optimizing parameters and a phase of post-processing. The performance is measured during this post processing phase and the best model is selected.

The final model is built when the termination criteria are reached. Online stock market trading today is a highly beneficial and demanding field for investors. Therefore, stock market predictions provide investors with a challenging role in deciding when and where they can invest their money[2]. Since the financial market is highly risky because it depends on many of the country's social and economic factors, stock market anticipation can be achieved through different technological implementations. Investors use different traditional techniques for better investment decisions, including a complete analysis of the company's revenues, growth rate and market situation, etc., while the technical inspection decides when to buy and sell the market share. In conventional methods, neural network models and statistical models are used for stock price estimation[3]. During ancient times, the word "stock" was used to collect agricultural products after harvesting and to open them for sale. Then the terms severely used by traders were collecting farm products from farmers and trading with other traders. The company originated during the sale and purchase of stock creates a market known as the stock market . The word "stock" from the agricultural sector to the financial sector is then inherited. This way, many businesses dump their share for sale at different agencies. Through stock exchange, a company can enlist the name of their company to sell their share and the public can buy according to their choice and investment ability[4].

Technical analysis, which can be used for market timing decisions and price target forecasts, can be used to seek help. Technical analysis alerts investors to the imminent risk by suggesting the likely price and time levels as if the market could fall[5]. Passive investors can capitalize on such stock market falls by exiting their existing holdings in index funds at a higher price and buying them back later at a lower price or simply short selling index futures. With the help of various technical instruments, this can be done. An index is a compilation of the prices of a number of representative assets for the purpose of capturing the overall market behavior. Charles Dow, who developed the Dow Jones Industrial Average in 1896, conceived the idea of an index. All stock exchanges around the globe have built their own indices since then. Indices have numerous uses. Investors follow an index to understand the overall performance of the daily market, economists use it to study long-term relationships with other economic factors to analyze and predict business cycle and economic growth patterns, chartists plot and analyze an index's price and volume changes to predict future market direction.

ISSN : 2057-5688

The Index also serves as a benchmark for evaluating the periodic performance of mutual funds[6]. Studies have found that the respective benchmark index is not beaten by mutual fund managers. Technical analysis metrics are simply a way to explain and measure the stock price trend. Filter benefits include quick computational times (usually much faster than wrapper selection methods) and simple scalability. Filter advantages include fast computational times (usually much faster than wrapper feature selection methods). For high-speed stock data, scalability is of particular importance where selection is required quickly and data dimensionality is high[7]. They further confirmed empirically that the sequential price jumps in equity prices were statistically and economically significant and were autocorrelated positively. Trading volume, however was found to be 60 percent higher and bid-ask spreads on pattern formation were lower. They examined the profitability of the candlestick pattern based trading strategy in U.S. markets. Following the completion of the pattern, the strategy involved buying a stock and holding it for ten trading days. The trading strategy suggested initiating trades on the day after completion of the pattern at the opening price. By holding the trading position for one to ten days, profitability was tested. A new trading position on the candlestick pattern formation was initiated and held for 10 days. The actual returns were compared using the bootstrap methodology to those obtained. Different trends have been implemented in different ways in different markets. So far, sufficient research on w.r.t candlestick patterns in Indian stock markets has not been undertaken to the best knowledge of the author. The Indian stock markets have become a hot destination for venture capital funds, mutual funds, hedge funds, PMS (portfolio management services), private equity funds, etc as a rapidly growing economy and an attractive destination for foreign portfolio investors coupled with mass domestic participation. In addition, increased financial awareness has motivated individuals to invest directly and embrace stock trading as a full-time profession. Thus in the Indian context, there is a need to test this oldest commercial technical school of thought. In addition, the study attempts to evaluate its profitability over various holding time periods using separate trading strategies. These are categorized broadly into patterns of reversal and continuation. A pattern of reversal means a change in the previous trend, and a pattern of continuation means that the previous trend will continue. So in order to use the candlestick for prediction, it is important to identify the trend. Such patterns require either a downtrend or an uptrend to be the present trend.

Reversal patterns are classified as bullish and bearish reversal patterns, depending on the nature of the previous trend. The downtrend is paused by a bullish reversal pattern, which indicates that the stock price may not fall further and it begins to go either up or sideways. Hammer, inverted hammer, bullish engulfing, piercing pattern, morning star, bullish harami, three inside up three outside up, tweezer bottom, three white soldiers, etc, are some of the popular bullish reversal patterns. Likewise, a Bearish reversal pattern stops the upward trend, indicating that the stock price may not increase further and that it begins to go sideways or downwards. This study focuses exclusively on the patterns of reversal. Their profitability is evaluated by assuming that on the day following the pattern, a trading position is initiated at the opening price and held for one to ten days. In order to achieve better prediction efficiency, many researchers have applied artificial intelligence (AI) techniques to financial markets. ANN, GA, fuzzy logic, SVM and optimization models such as particle swarm optimization, ant colony optimization and teaching-based optimization are some of the popular techniques.

2. Related works

Most of the outlier and clustering models are difficult to find the essential outlier detection measures due to the variance in ranges of each stock prices in the realtime market data[8]. Traditional outlier detecton models are used to find the outlier based on the static average range of all stocks and it is independent of each stock technical feature. Also, most of the clustering models are difficult to group the trend based or extreme outlier based techical stock details on realtime data. used adaptive fuzzy-GARC H and PSO to plan the model for stock index prediction. They selected an RBFNN model for data set training and predicted the SSE index.

3. Proposed model

In the section, a hybrid knn model is proposed to find the trend stocks on the realtime market data in a new filtered based clustering model on technical data. The continuous technical data types for trend prediction are evaluated for this model. The flow chart of the proposed stock market trend prediction model is listed in figure . Initially, data from stock exchange sites such as tradingview or wallmine are taken from the real time market. The

ISSN : 2057-5688

training data is used for stock-related technical factors like mark, price, ADX, ADR, RSI, MACD, news sentiment score, etc. Technical dat a preprocessing and clustering operations are performed on stock technical data.

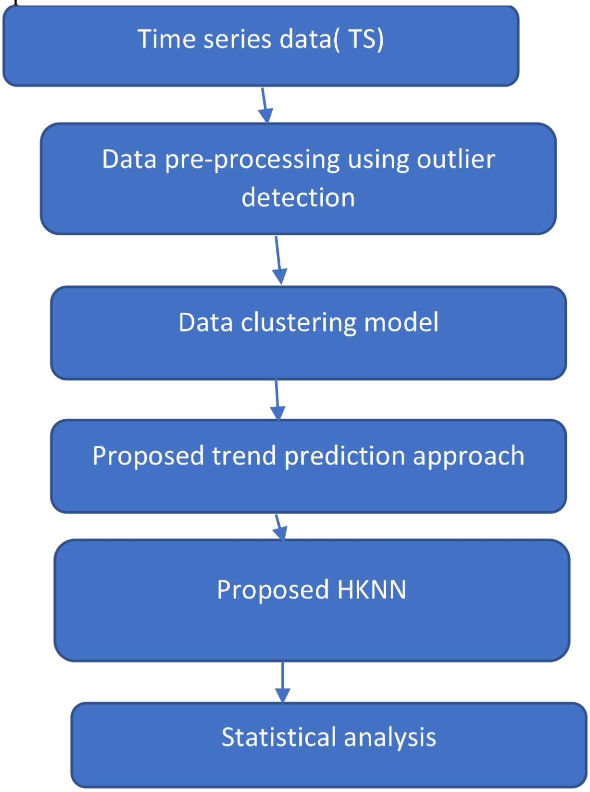


Figure : Proposed framework

In the proposed frame work initially time series data is collected then performs outlier detection using proposed appraoch then filtered data is send to the proposed hybrid clustering approach to form the clusters then the data is send to the hybrid classifier for trend prediction .

Hybrid KNN we used the log normal distance as newly introduced distance measure then we also used distance based probability mechanism to find out the nearest neighbors.

Algorithm 1: Proposed HKNN

Input : Clustered training data D, test samples T, k value and classes C.
Output: Trend class prediction
Procedure: To each instance p in D
do
Compute lognormaldistance$(D(t,p)/t \in D, p \in D, t \,!= p) = \log(\sum (\| t\,i \| - \| pj \|)^2)$;
done
Sort k neighbors according to their distances.
sort(k, D(t,p))
Compute probability distributions using the neighbor distances.
To each instance t(i) in k-neighbors(sort(k, D(t,p)))
do
$\text{DistProb[]} = \frac{1}{\sqrt{2.\pi}} \int e^{-D(t(i),p)^2} \, dD(t(i),p) / |N|; N = \text{Total attributes}$
done
To each test sample t in k-neighbors(sort(k, D(t,p)))
do
Compute class membership probabilities of each test sample t
assign class to t sample using classifier.
done

Algorithm 1, describes the proposed HKNN model in order to find the trend prediction for the stock time series data. In this algorithm. Log normal values are calculated then neighbors identified finally class membership is predicted based on newly introduced measures.

4. Experimental results

Experimental results are simulated with java and third party data in real time. In this model, different stocks data and it's technical data are taken into account for trend prediction. Also, by using various statistical features are used to find outliers. Statistical measures are evaluated using java libraries from third parties.

Table 1: Sample technical data collected from the tradingview website.

```
@relation data
@feature MFI numeric
@feature Volatility numeric
@feature STO numeric
@feature ATR numeric
@feature RSI numeric
@feature PPO numeric
@feature ADX numeric
@feature MACD numeric
@feature StochRSI numeric
@feature Price numeric
@feature 'Performance today' numeric
@feature Sentiment {Buy_Predict,Sell}
@data
FORTIS,69.39,1.76,57.21,2.46,58.8,0.23,35.8,0.96,33.33,139.5,-
0.035832,Sell
LUMAXIND,42.85,2.81,39.01,49.8,47.86,-1.09,10.15,10.79,24.35,1752.55,-
1.052953,Sell
BINDALAGRO,20.45,5.72,11.3,0.79,32.72,-3.23,20.93,-0.38,0,13.75,-
0.36232,Sell
CENTEXT,25.61,4.63,25.24,0.25,23.57,-3.37,24.08,-0.18,0,5.5,1.85185,Buy
MARUTI,89.18,1.85,93.49,135.29,61.76,3.32,25.99,127.96,58.37,7048.9,-
3.719995,Sell
MAHSEAMLES,23.18,2.02,20.72,9.56,44.72,-0.5,10.99,-
0.31,0,474.3,0.327866,Buy
GPPL,37.63,3.45,35.64,3.24,46.07,-0.19,22.14,1.31,0,93.45,-0.373141,Sell
JPINFRATEC,45.77,6.67,37.75,0.16,44.27,-14.14,26.49,-
0.2,66.67,2.5,4.166663,Buy
FIEMIND,47.32,3.12,33.17,15.6,51.18,-0.17,14.13,3.58,0,496.55,-
0.769386,Sell
LITL,0,0,0,0,19.67,0,61.58,0,0,0.3,0,Buy
AUTOLITIND,63.77,5.61,31.92,2.11,51.85,-0.94,50.21,0.43,34.11,37.15,-
1.196801,Sell
TORNTPOWER,61.41,2.35,51.43,6.09,53.4,0.38,24.44,1.87,40.08,257.85,-
0.559191,Sell
DCM,77.09,5.01,61.19,2.94,41.87,-2.46,17.31,-0.94,33.33,58.6,-
0.170362,Sell
KEI,54.82,3.81,63.73,15.62,53.72,-0.3,27.43,7.41,66.67,413.6,0.890351,Buy
GREENPLY,69.2,3.36,79.08,5.61,62.85,5.22,41.84,6.05,0,166.9,-
```

ISSN : 2057-5688

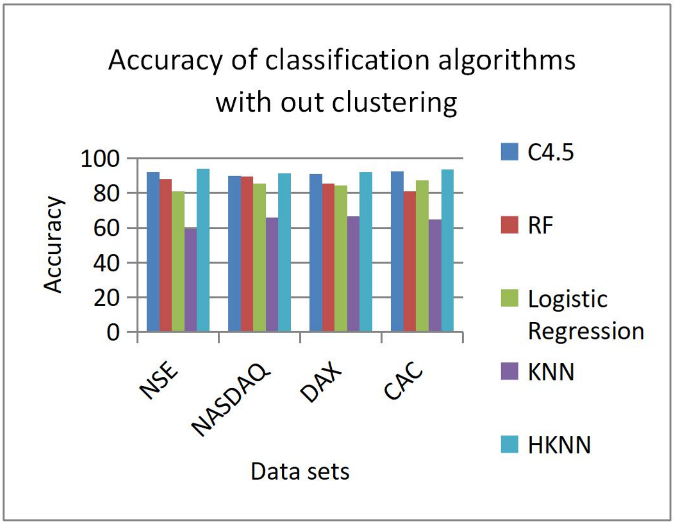
0.029951,Sell
HBLPOWER,43.77,4.37,43.87,1.1,46.7,-3.01,14.27,-0.21,33.36,24.85,-1.19284,Sell
COSMOFILMS,78.9,3.77,67.13,7.84,59.93,3.64,25.28,5.41,33.33,204.2,-1.85052,Sell
BIOCON,57.31,1.97,59.84,11.95,42.11,-0.12,12.62,-1.32,0.71,613.7,0.979023,Buy
INSECTICID,68.19,3.03,48.86,19.69,51.25,-0.45,10.07,3.19,62.35,638.6,-1.579718,Sell
PRIMESECU,39.34,3.81,25.44,1.57,54.51,-1,12.64,-0.2,66.67,40.15,1.517073,Buy
GARDENSILK,50.4,5.23,8.74,0.98,34.07,-2.99,18.98,-0.32,0,18.6,-0.799998,Sell
VLSFINANCE,70.84,4.41,45.03,2.61,52.56,0.13,24.06,0.93,66.67,60,1.35135,Buy
BALMLAWRIE,40.13,2.45,16.98,4.33,36.59,-1.84,15.77,-1.62,33.33,176.25,-0.39559,Sell
NOIDATOLL,33.34,7.17,19.76,0.38,41.03,-2.47,23.87,-0.15,51.01,5.35,0.943391,Buy
GODREJPROP,83.42,4.47,70.86,40.95,58.7,9.72,38.45,48.08,49.34,889.95,-1.51062,Sell
M&MFIN,38.64,2.66,45.97,11,47.12,-0.94,10.36,-0.22,24.54,413.2,0.1697,Buy
SKMEGGPROD,47.95,4,49.2,2.09,48.91,-2.93,12.55,-0.71,62.7,51.15,-2.198849,Sell
USHAMART,45.94,6.25,8.54,2.12,38.65,0.88,23.66,-0.02,21.68,33.3,-0.149923,Sell
WIPRO,77.76,2.75,84.19,7.94,61.97,3.43,14.62,4.88,46.28,291.1,0.988727,Buy
PATSPINLTD,56.97,7.8,29.31,0.78,41.14,-2.04,11.6,-0.2,0,9.6,-3.999996,Sell
IITL,47.05,6.2,60.31,4.96,59.96,5.02,19.45,2.59,42.12,80,-0.062465,Sell
BAJAJCORP,52.36,2.92,66.02,10.81,46.97,2.47,29.13,5.52,33.33,385.95,-0.168131,Sell
CUMMINSIND,41.41,3,62.13,21.61,44.31,-0.68,12.87,-0.55,10.72,741.6,2.828614,Buy
MOTHERSUMI,69.7,3.49,74.22,5.49,55.76,-1.66,16.36,0.15,88.06,150.1,-1.63827,Sell
GAIL,44.69,2.6,55.46,8.96,46.98,-1.28,16.1,0.69,4.46,339.45,-1.565896,Sell
COROMANDEL,31.16,3.41,21.42,15,38.8,-3.81,26.19,-8.17,60.99,426.15,-1.251302,Sell
ICIL,60.54,6.48,19.25,3.01,43.41,-4.95,20.54,0.05,0,44.15,-4.951559,Sell

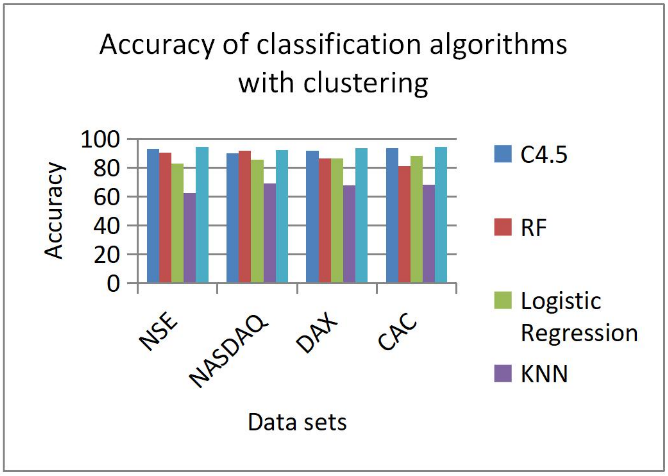
PANACEABIO,80.32,5.46,71.29,11.03,57.49,1.08,38.89,4.55,40.64,193.8,-2.905813,Sell
PBAINFRA,0,1.5,18.53,0.06,19.02,-10.47,52.56,-0.31,66.67,4,0,Sell
NITESHEST,67.43,6.39,19.58,0.39,48.88,-1.27,26.98,0.02,0,5.95,-2.459018,Sell
,24.35,47.49,41,-0.74,19.11,3.12,33.33,1465.35,-0.778684,Sell
LGBBROSLTD,45.21,2.46,21.8,9.5,46,-1.43,10.3,-3.8,66.67,393.85,2.099807,Buy
WELCORP,65.55,4.21,51.13,5.66,52.57,6.15,32.28,4.65,33.33,132.7,-1.301599,Sell
VTL,71.85,2.52,82.78,28.36,61.7,1.29,23.01,20.08,15.03,1126.55,0.244705,Buy
FDC,87.87,3.52,32.28,5.76,41.66,-2.61,13.54,-1.46,29.52,162.35,-0.794371,Sell
GAL,66.27,9.32,26.93,0.41,35.01,-3.83,28.14,-0.22,0,4.65,5.681818,Buy
BALAJITELE,39.25,3.28,12.12,2.56,37.1,-2.28,8.36,-1.41,0,76.95,-1.535515,Sell
HDIL,68.08,3.96,34.81,0.97,38.65,1.16,25.59,0.23,0,23.5,-4.081633,Sell
BALRAMCHIN,60.31,3.87,72.73,5.55,59.75,2.16,27.32,3.45,68.57,141.15,-1.534714,Sell
FINPIPE,40.3,2.96,27.7,13.93,36.7,-3.52,26.24,-7.23,33.33,466.15,-0.829698,Sell
MANGCHEFER,77.6,4.03,70.07,1.68,54.55,4.58,35.64,1.29,66.67,41.25,-0.960388,Sell
AIAENG,77.94,2.62,61.22,47.25,56.94,0.33,24.79,17.71,57.95,1809.5,0.522195,Buy
SASKEN,89.38,3.17,75.79,22.58,75.21,0.93,27.79,12.37,100,705.9,-1.023554,Sell
CARBORUNIV,18.91,2.65,20.43,9.9,45.02,0.68,17.85,1.2,1.3,372.85,0.067096,Buy
OMAXE,40.87,1.22,48.47,2.56,51.56,0.93,33.1,0.84,66.67,210.35,-0.047513,Sell
SUNPHARMA,39.94,2.51,22.89,11.41,44.85,-0.58,16.16,1.51,54.13,468.55,3.091306,Buy
UNITY,45.58,12.73,31.96,0.07,34.5,-18.91,27.4,-0.09,66.67,0.55,0,Buy
SADBHAV,14.68,3.32,19.16,7.97,52.96,-0.55,27.07,5.21,0,229.65,-2.172527,Sell
JKPAPER,66.63,3.15,75.92,4.74,57.4,1.8,21.69,2.94,9.26,145.65,0.379041,Buy
PVR,59.98,2.52,79.88,42.67,59.29,1.86,24.58,33.02,42.75,1670.55,-1.387208,Sell
WILLAMAGOR,24.85,5.15,12.2,2.39,34.35,-2.91,16.06,-0.97,23.55,45.2,-

2.691066,Sell

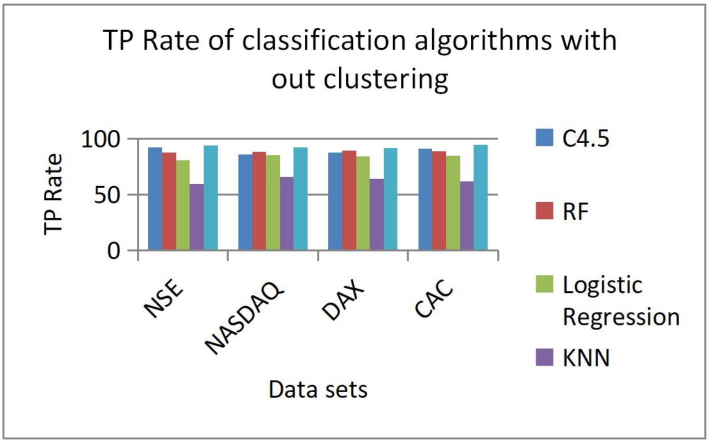
Table 1, describes the sample collected data from the tradingview website to perform the outlier detection and clustering approaches. These data contains various technical stock related features as the training data.

Following results are obtained Accuracy with out cluster based classification,



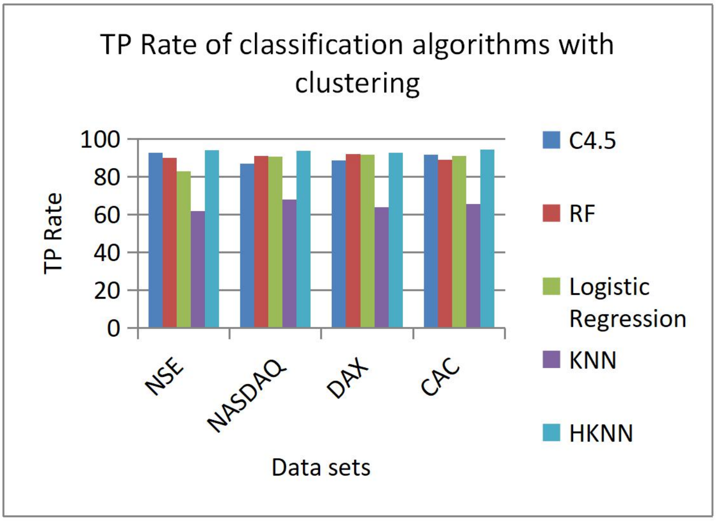


TP Rate of algorithms without cluster based classification .



.

ISSN : 2057-5688



5.Conclusion

In this work, a new cluster based classification approach is proposed to predict the individual stock trend. The bullish and bearish patterns are hard to forecast using trend indicators or realtime stock sentiment news for the most part of the current technical indicators. Experimental findings showed that the current model is computationally effective in terms of predicting trend of the stocks.

6.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

ISSN : 2057-5688

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.eswax.2019.100016.
[7]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.
[8]Y. Xu, B. Iglewicz, and I. Chervoneva, "Robust estimation of the parameters of g-and-h distributions, with applications to outlier detection," Computational Statistics & Data Analysis, vol. 75, pp. 66–80, Jul. 2014, doi: 10.1016/j.csda.2014.01.003.

[9]S. Aghabozorgi and Y. W. Teh, "Stock market co-movement assessment using a three-phase clustering method," Expert Systems with Applications, vol. 41, no. 4, Part 1, pp. 1301–1314, Mar. 2014, doi: 10.1016/j.eswa.2013.08.028.

[10]H. Esmalifalak, A. I. Ajirlou, S. P. Behrouz, and M. Esmalifalak, "(Dis)integration levels across global stock markets: A multidimensional scaling and cluster analysis," Expert Systems with Applications, vol. 42, no. 22, pp. 8393–8402, Dec. 2015, doi: 10.1016/j.eswa.2015.06.053.