

Analysis Of Stock Prediction Parameters and Their Impact on Effective Selection Of Stock

Er. Mausami Sawarkar Dagwar^{1*}, Dr. Sachin Agrawal²

^{1*}Research Scholar, COETA, SGB Amravati University & Assistant Professor at Priyadarshini J. L. College of Engineering, Nagpur

²Assistant Professor at College of Engineering & Technology, Akola

Abstract: Stock value prediction is a multi-disciplinary field which requires efficient knowledge about the stock's historic values, its news feeds, twitter sentiments, impact of global stock market(s) on the stock, etc. In order to effectively analyse a stock's trend for inter-day, intra-day or long term, analysts have to evaluate these values on a continuous basis. Along with these values analysts also have to analyse non-stock data like recent news about the company, management changes in the company, tweets related to the company, global news & global stock market trends which affect the company in any way possible. Each of these data sources have a different effect on the stock's value change, and it is recommended for a good stock prediction system to analyse the effects of these values before real-time deployment. Neglecting even a single parameter before deployment of the stock prediction system might result into a multitude of prediction errors. For instance, if twitter feeds for a stock are not considered during prediction of a nicely performing stock, and suddenly some news about a product fail comes online, then the stock prices might plummet, and the system will not be able to track it. In order to reduce the effect of these outlier events on predicted value of the stock, this paper analyses different parameters that affect stock prediction, and suggest the impact of these parameters on stock performance. Researchers can use this information in order to improve the accuracy of their deployed systems, and make these systems future proof.

Keywords: Stock Market, Parameters, Prediction, Accuracy, Machine Learning

1. INTRODUCTION

In the post CoVID scenario, stock markets have behaved wildly due to various global and national factors. These factors include, but are not limited to,

- Fear in the minds of people about currently high performing stocks
- Future trust on a particular stock performance
- Media channels boosting or sagging particular stocks
- Post vaccination confidence on pharma and other CoVID related companies
- Rise in global depression levels, etc.

In order to effectively predict a stock's value, it is absolutely necessary to capture these events, and develop a stock prediction model. An example of such a system can be observed from figure 1, wherein data from multiple sources is considered before prediction of the final stock value.

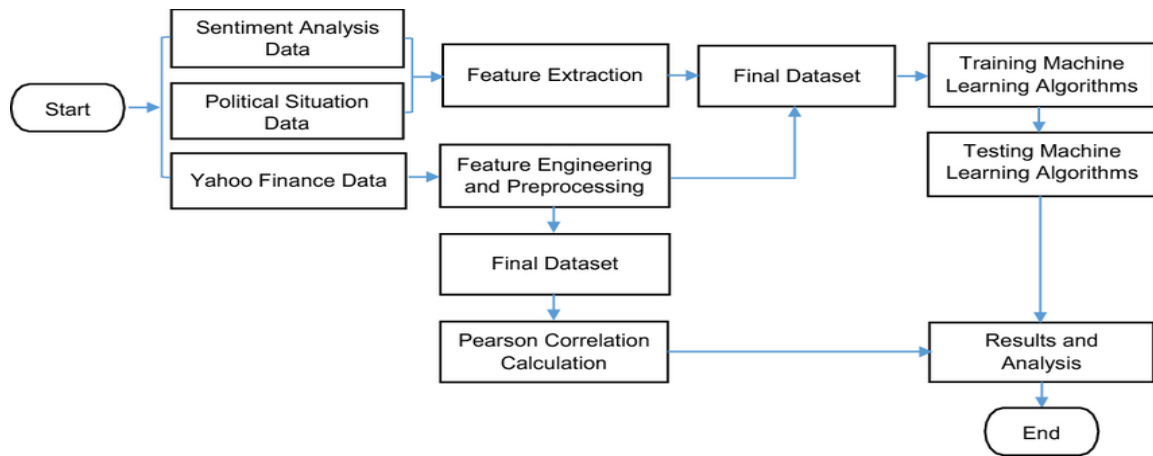


Figure 1. A stock prediction system covering multiple stock features

In this system, textual data from different sources which indicates stock sentiments and political situation is collected. This data is given to a feature extraction unit, wherein features like word2vec, parts-of-speech, etc. are evaluated. The features from financial data are also evaluated and a final dataset is created. This final dataset is given to a series of machine learning algorithms, which evaluate linkage of stock value with these features. Finally, the stock value is predicted based on this linkage. A large number of algorithms have been proposed which take a lot of different parameters as input in order to accurately predict stock values. The next section describes these algorithms in brief, and suggests methods in which these algorithms can be improved. An analysis of these methods and the used parameters is done in the following section, which will assist the researchers to select the most stock affecting features suited for their prediction application. Finally, this text concludes with some interesting observations about these algorithms, and suggests ways to cascade and fuse these algorithms for obtaining better efficiency.

2. LITERATURE REVIEW

Stock prediction methods are divided into 4 different categories, which can either be value-based prediction methods, sentiment-based prediction, hybrid-prediction methods, and movement-based prediction methods. Some of these methods can accurately identify the value of stock, while others assist in identifying the direction of movement for the given stock. Nature of future predicted value from all these methods depends on the nature of historical data given to these methods. For example, if a method is trained using inter-day data separated by hourly stock values, then it will predict inter-day hourly values only, it cannot be used for prediction of intra-day or long-term predictions. For instance, the work in [1] uses long-term temporal data along with social media data and web-based news in order to predict next value of the stock. Architecture for this system can be observed in figure 2, wherein a multi-source multiple-instance classifier is used for gathering data from these sources and predicting the final stock value.

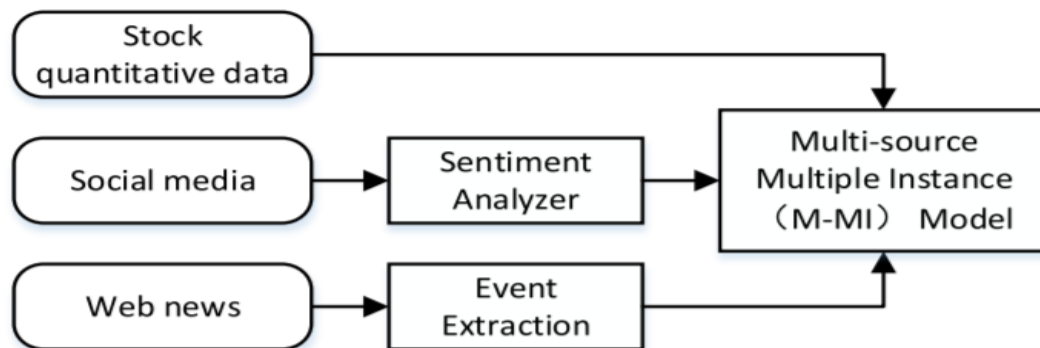


Figure 2. Multi-source stock value prediction [1]

This multi-source method is able to produce more than 59% accuracy of stock prediction, which is low considering that based on this prediction customers might shift their buying patterns. The accuracy of this system can be improved with the help of deep-learning models. Such a model is described in [2], wherein stock market trends are given to a deep learning model with feature engineering. Feature engineering includes calculation of indices like simple moving average (SMA), double moving average (DMA), stochastic oscillator, relative strength, etc. followed by feature selection. The feature selection module basically reduces redundant features and uses highly weighted features for effective stock prediction. Post this a dimensionality reduction technique combined with principal component analysis (PCA) is used. The PCA method reduces features such that accuracy of prediction can be improved. Finally, a long short-term memory (LSTM) based classifier is used for prediction of stock values. Architecture for this system can be observed from figure 3, wherein the final prediction result can be seen to have been obtained from raw data.

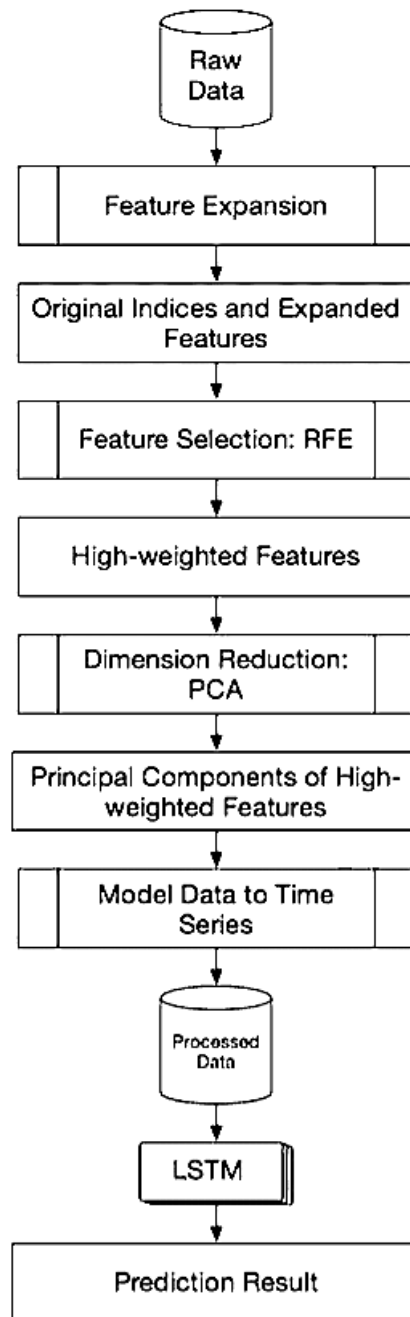


Figure 3. A deep learning-based prediction model [2]

Due to the use of deep learning, accuracy of prediction improves to nearly 93% with 29 selected features. This prediction of stock is based on short term data, due to which the accuracy is very high. This accuracy must be tested for long-ranged data prediction systems. Such a long-ranged data prediction system can be observed in [3], wherein one-day ahead movement planning is predicted with the help of disparate data sources. These data sources include stock market data, technical indicators, Wikipedia hits for the given stock and news data for the given company. All these parameters are given to a feature generation and feature selection unit, wherein different features for the given data are extracted. Using this feature extraction engine, a machine learning-based classifier is trained. The classifier is a combination of decision tree (DT), support vector machine (SVM) and artificial neural networks (ANNs). Parameters like accuracy, sensitivity, specificity, precision, f-measure, etc. are evaluated. All these parameters are showcased on a user interface for effective visualization. The architecture for this model can be observed from figure 4, wherein connectivity of different models can be seen. Due to use of these disparate data sources the accuracy of prediction is improved to 85%, which is very high considering that the system is able to track next-day movements with high level of effectiveness.

Another machine learning framework for stock prediction can be observed from [4], wherein a general adversarial network (GAN) is used, along with auto-encoder for improved stock value prediction. The system uses a hyper-optimizer to optimize the features to be used for prediction. Due to the combination of GAN and hyper-optimizer the accuracy of prediction for next day values is more than 90%. This is mainly due to the error correction nature of GAN, that learns from its previous mistakes, and minimizes error based on high level confidence scores. This system can be used for long-term predictions of non-volatile stocks. But when stocks of new Initial Public Offering (IPO) companies are offered to public, then due to frequent changes in buying patterns, the stock values are incorrectly predicted using inter-day predictors. In order to improve the efficiency of such stocks, the work in [5] can be used. This work is able to predict stock values for maximum returns in short-term periods for volatile stock data.

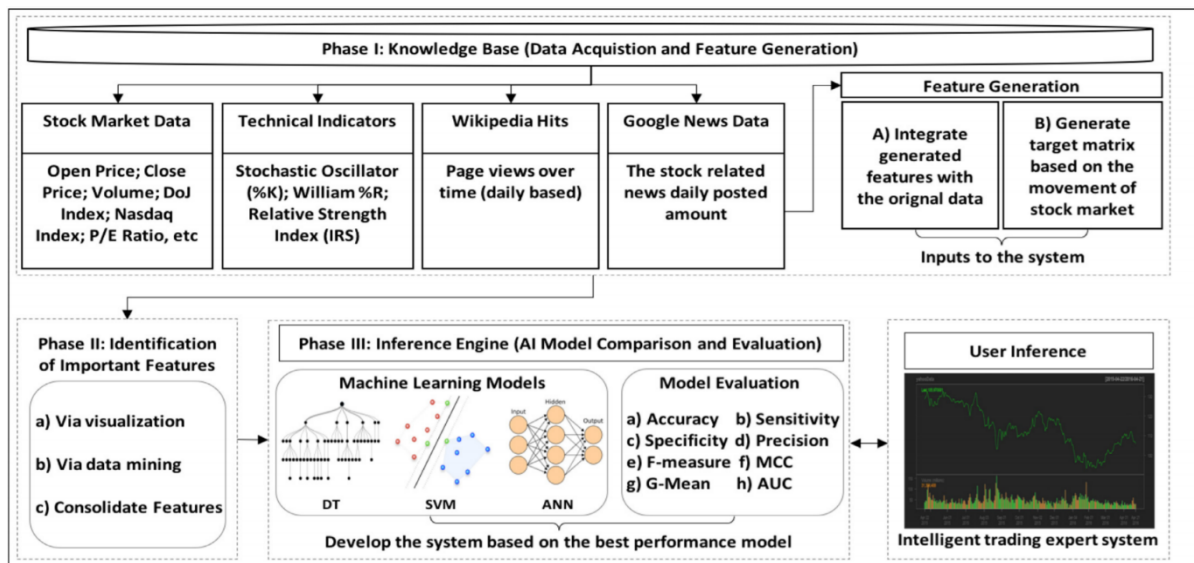


Figure 4. Disparate data sources for prediction of stock values [3]

The work uses a combination of Auto Regressive Integrated Moving Average (ARIMA) model in order to predict the stock value using quantitative and qualitative stock data. Initially the model estimates whether a stock is stationary or not, if the stock is non-stationary then Auto-correlation function (ACF) and partial ACF values are evaluated. These values are given to an ARIMA model for final forecasting. An accuracy of more than 89% is observed for stock value prediction, while an accuracy of less than 50% is observed for trend prediction of the stock. This is majorly due to volatility of the stock values. This accuracy can be improved using ensemble classification techniques like data boosting, value bagging, temporal blending, using super learners like DTs, SVMs and Neural Network (NN). Such a work is implemented in [6], wherein an array of such classification methods is implemented for final value prediction

of stock. It is observed that stacking different classifiers improves the accuracy of prediction to nearly 90%, while blending results for different classifiers improves the accuracy to nearly 95%. Bagging and boosting of data are not recommended for stock prediction, as it introduces a lot of false positive values into the stock data. The accuracy is tested on limited volatile stock data, but it must be evaluated on large-scale long-term values for better performance evaluation. In order to do this the work in [7] can be referred, wherein large-scale heterogeneous data from different sources is taken and fused in order to effectively predict stock values. The data is similar to [1] and [3] but uses a tensor-based processing unit in order to improve effectiveness of stock prediction. Along with this tensor processing unit, a stock correlation unit is also designed which takes care of the stock's value-to-sentiment correlation. This correlation unit utilizes intra-dependency features of a stock value with the attributes of the stock like web-news, tweets, etc. Due to this correlation unit, the accuracy of stock movement prediction is improved. Based on this stock movement value, the predicted final stock value has less error, which results in an 89% accuracy. This accuracy can be further enhanced using advanced deep-learning methods like convolutional neural networks, deep-nets, LSTMs, etc. Moreover, social media mining aspects can also be further improved with the help of opinion mining techniques. Such a technique is mentioned in [8], that utilizes deep-nets for opinion mining, and combines this information with the technical indicators in order to improve the stock prediction performance to more than 90%.

The performance of prediction can also be improved by selecting proper stock value indicators, for instance the work in [9] suggests the use of Simple Moving Average, Weighted Moving Average, Relative Strength Index, Accumulation/Distribution Oscillator and Average True Range indicators, and combines them with support-vector regression (SVR) model in order to effectively analyse long-term stock values. This results in a 91% accuracy for long-term stock value prediction, which is sufficiently high for making a guess of both the stock direction and the final stock price in a given period of time. While SVR is an efficient technique for stock value analysis, the work in [10, 11] suggests that artificial neural networks (ANNs), ARIMA models and random forest (RF) classifiers can also be considered for effective stock prediction. They also suggest that Root Mean Square Error, Mean Absolute Percentage Error, Mean Bias Error and Box-Jenkins methods can be used as effective indicators to evaluate error in stock price. An accuracy of more than 90% can be achieved when these parameters are optimized for stock value prediction. While this accuracy is high enough, it can be further improved with the help of deep-learning models like the ones mentioned in [12]. These models include recurrent neural networks (RNNs) and LSTM-based predictors. Due to the use of RNNs and LSTM the accuracy of prediction can go to as high as 95%, for both long-term and short-term prediction. Another deep learning-based method namely multi-filters neural network (MFNN) is described in [13], wherein feature extraction is optimized using the neural network, and an improved feature vector is selected for stock prediction. The MFNN model outperforms RNN, CNN, LSTM and SVM models by almost 6% in terms of prediction accuracy, thus making it the most suitable model for value-based stock prediction. But this accuracy is based on value-based (candle-stick) prediction and is only 70% effective [14]. In order to further improve the model's effectiveness, it is recommended that news and web-feeds be incorporated into final stock value prediction. These web-feeds can be in the form of twitter-based sentiments, and can be used to analyse stock movements as discussed in [15]. This can give an accurate measure of the direction of the stock price, and based on this final stock value can be estimated. Some deep learning models are also suggested in [16] for detection of this movement. They suggest the use of a 3D CNN model combined with Pearson product-moment correlation coefficient for improved movement prediction accuracy. The model can be observed from figure 5, wherein a combination of 3D CNNs with LSTMs is done for final value and trend prediction of the stock.

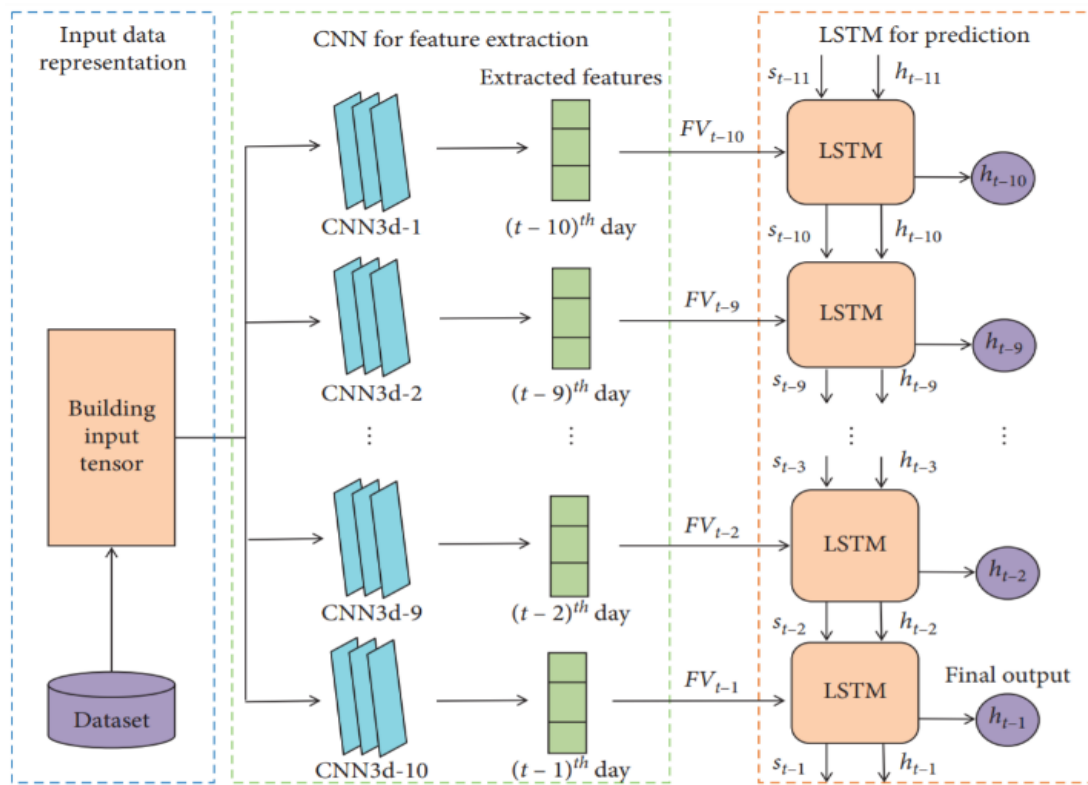


Figure 5. A 3D CNN with LSTM model for stock value and direction prediction [16]

This model also has an accuracy performance similar to MFNN, but can be extended to support multiple indicator types, and multiple input types for better prediction performance. This prediction performance can be further improved using sliding-window analysis of stock values as mentioned in [17]. Using this sliding window analysis, stock values are predicted in chunks, and accuracy of each chunk is optimized to optimize the final model accuracy. Example of such an architecture can be observed from figure 6, wherein different parameters are combined with a sliding-window validator for better prediction performance. Accuracies in the range of 91% to 97% can be achieved for long-term prediction, which makes the system similar in performance to MFNN and 3D CNNs, but facilitates the prediction using chunk-based analysis.

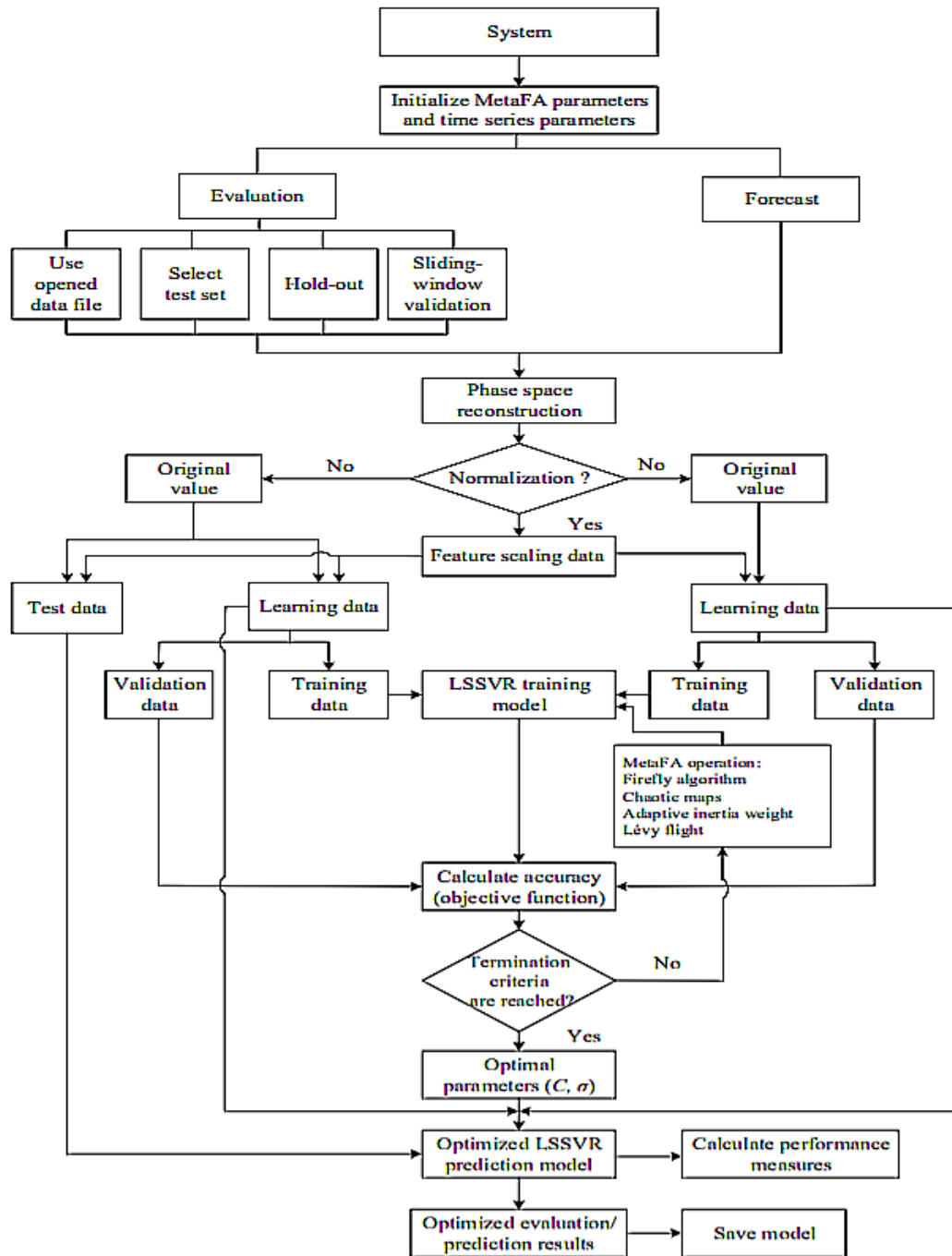


Figure 6. A sliding-window based stock value prediction classifier [17]

Different technical indicators have different impact on the final predicted value of stock. A comprehensive analysis of this impact can be observed from [18], wherein indicators like Simple Moving Average (SMA), Weighted Moving Average (WMA), Double Exponential Moving Average (DEMA), Average Directional Index (ADX), Range of Change (RoC), Relative Strength Index (RSI), Kaufman Adaptive Moving Average (KAMA), Triple Exponential Moving Average (TEMA), Minus Directional Indicator (MDI) and Williams %R (WR) are used for analysis of stock price variation. They suggest that each of these indicators are of lagging type, and can only be used to predict values of stock when sufficient training data is available. Leading indicators do not require large amount of data for final prediction, and are not available for stock value prediction purposes. From their analysis it is observed that the combination of WMA, KAMA, Momentum, EMA and DEMA is most effective for minimum error performance in stock

value prediction. A combination of these techniques can be applied to other models like boosted DT and logistic regression (LR) given in [19] in order to improve its prediction accuracy from 70%. Ensemble classifiers and neural network-based classifiers like the ones given in [20, 21] can also be used for further improvement in prediction accuracy. A detailed review of these methods can be found in [22], which indicates that CNN and RNN models are the most widely used models for effective stock value prediction when combined with stock sentiments. Moreover, techniques like multi-variate independent component analysis (MV-ICA) and Canonical Correlation Analysis (CCA) mentioned in [23] can be used for further improving correlation of stock value with market sentiments. An accuracy of more than 95% can be achieved for some selected stocks if these models are properly selected and applied. This accuracy can be further boosted to nearly 98% with the help of tick-data based stock analysis as mentioned in [24], wherein a combination of Scale Conjugate Gradient (SGD) with Bayesian Regularization (BR) and Levenberg-Marquardt (LM) optimization are used for error reduction. From this review it can be analysed that CNN-based models when combined with effective feature selection and web-feed analysis can be a suitable model for stock value prediction. Analysis of these techniques can be observed in the next section.

3. ANALYSIS OF REVIEWED ALGORITHMS

In order to analyse the reviewed algorithms and their respective technical indicators, this section segregates the algorithms w.r.t. the algorithm used, list of features used, impact of these features and the obtained accuracy. Impact of these features is divided into L, M and H values, which showcases the effect of including the said feature in improving accuracy of the predictor. A summary of this study can be observed in table 1, wherein all this information is aggregated.

Table 1. Analysis of different stock indicators and methods

Method	Features Used	Impact of these features	Accuracy (%)
Multi-source indicators [1]	Twitter feeds, closing stock value, news data	H, M, M Short term	59
PCA with LSTM [2]	Selective Technical indicators	H Short Term	93
DT, SVM and ANN [3]	Selective Technical indicators	H Long term	85
GAN [4]	Selective Technical indicators	H Long term	90
ARIMA [5]	Technical indicators	L Short term	89
Blending DT, SVM and NN [6]	Technical indicators	M Short term	95
Stacking DT, SVM and NN [6]	Technical indicators	M Short term	90
Tensor-based CNN [7]	Twitter feeds, closing stock value, news data	H, M, M Short term	89
LSTM [8]	Twitter feeds, closing stock value, news data	H, M, M Short term	90
SVR [9]	Technical indicators	M Long term	91
ARIMA and RF [10, 11]	RMSE, MAE	L, L Short term	90
RNN with LSTM [12]	Technical indicators	M Short term	95
MFNN [13]	Technical indicators	M Short term	98
3D CNN with LSTM [16]	Technical indicators, global indices	M, M Long term	98
Sliding window with LSTM [17]	Chunk-based indicators	H Short-term	91
SVR [18]	WMA, KAMA, Momentum, EMA and DEMA	H, H, H, H, H Short term	85
MV-ICA [23]	Technical indicators, global indices	M, M Short term	95

Based on this analysis, it can be observed that CNN, RNN, LSTM and 3D CNNs are the most suitable methods for stock prediction. Moreover, a combination of tweet data, web-news, global indices and technical indicators will result in higher accuracy when compared to using only technical indicators for prediction.

CONCLUSION & FUTURE WORK

Based on the result analysis, it can be observed that in order to predict short-term stock values, the technical indicators are usually sufficient. But due to the limited applicability of these indicators they are only 70% effective in terms of real-time application. This indicates that the predicted values although are 95% accurate theoretically, but when applied to real time systems will be only 63% accurate. In order to improve this accuracy, it is recommended that other features like twitter feeds, web-news, global stock indices and other social feeds be considered while predicting real-time stock value. Out of these, twitter feeds have a very high impact as they are updated in real-time and thus evaluate the impact of any negative or positive news on the stock's performance very quickly. Web-news and global indicators affect the accuracy moderately, while other social feeds have been found to have low impact on stock value change. Technical indicators WMA, KAMA, Momentum, EMA and DEMA are observed to have huge impact on stock value prediction accuracy, and thus must always be used for effective stock selection. In future, researchers are suggested to work on machine learning models that incorporate global indices, other social data, web-scraping and data from local micro-blogging sites in order to further improve the accuracy and reliability of stock selection.

References

- [1] X. Zhang, S. Qu, J. Huang, B. Fang and P. Yu, "Stock Market Prediction via Multi-Source Multiple Instance Learning," in *IEEE Access*, vol. 6, pp. 50720-50728, 2018, doi: 10.1109/ACCESS.2018.2869735.
- [2] Shen, J., Shafiq, M.O. Short-term stock market price trend prediction using a comprehensive deep learning system. *J Big Data* 7, 66 (2020). <https://doi.org/10.1186/s40537-020-00333-6>
- [3] Stock market one-day ahead movement prediction using disparate data sources, <https://www.sciencedirect.com/science/article/abs/pii/S0957417417301331>
- [4] Bi Q., Yan H., Chen C., Su Q. (2020) An Integrated Machine Learning Framework for Stock Price Prediction. In: Dou Z., Miao Q., Lu W., Mao J., Jia G. (eds) *Information Retrieval. CCIR 2020. Lecture Notes in Computer Science*, vol 12285. Springer, Cham. https://doi.org/10.1007/978-3-030-56725-5_8
- [5] S. M. Idrees, M. A. Alam and P. Agarwal, "A Prediction Approach for Stock Market Volatility Based on Time Series Data," in *IEEE Access*, vol. 7, pp. 17287-17298, 2019, doi: 10.1109/ACCESS.2019.2895252.
- [6] Nti, I.K., Adekoya, A.F. & Weyori, B.A. A comprehensive evaluation of ensemble learning for stock-market prediction. *J Big Data* 7, 20 (2020). <https://doi.org/10.1186/s40537-020-00299-5>
- [7] Improving Stock Market Prediction via Heterogeneous Information Fusion, <https://arxiv.org/abs/1801.00588>
- [8] Huang, J-Y, Liu, J-H. Using social media mining technology to improve stock price forecast accuracy. *Journal of Forecasting*. 2020; 39: 104– 116. <https://doi.org/10.1002/for.2616>
- [9] Stock price prediction using support vector regression on daily and up to the minute prices, <https://www.sciencedirect.com/science/article/pii/S2405918818300060>
- [10] Stock Closing Price Prediction using Machine Learning Techniques, <https://www.sciencedirect.com/science/article/pii/S1877050920307924>
- [11] S. M. Idrees, M. A. Alam and P. Agarwal, "A Prediction Approach for Stock Market Volatility Based on Time Series Data," in *IEEE Access*, vol. 7, pp. 17287-17298, 2019, doi: 10.1109/ACCESS.2019.2895252.
- [12] Nikou, M, Mansourfar, G, Bagherzadeh, J. Stock price prediction using DEEP learning algorithm and its comparison with machine learning algorithms. *Intell Sys Acc Fin Mgmt*. 2019 ; 26: 164– 174. <https://doi.org/10.1002/isaf.1459>
- [13] Deep learning-based feature engineering for stock price movement prediction, <https://www.sciencedirect.com/science/article/abs/pii/S0950705118305264>
- [14] H. Qiu and F. Liu, "Candlestick Analysis in Forecasting U.S. Stock Market: Are They Informative and Effective," 2019 IEEE 4th International Conference on Big Data Analytics (ICBDA), Suzhou, China, 2019, pp. 325-328, doi: 10.1109/ICBDA.2019.8713248.
- [15] Real-Time Sentiment Analysis of Twitter Streaming data for Stock Prediction, <https://www.sciencedirect.com/science/article/pii/S1877050918308433>
- [16] Can Yang, Junjie Zhai, Guihua Tao, "Deep Learning for Price Movement Prediction Using Convolutional Neural Network and Long Short-Term Memory", *Mathematical Problems in Engineering*, vol. 2020, Article ID 2746845, 13 pages, 2020. <https://doi.org/10.1155/2020/2746845>
- [17] J. Chou and T. Nguyen, "Forward Forecast of Stock Price Using Sliding-Window Metaheuristic-Optimized Machine-Learning Regression," in *IEEE Transactions on Industrial Informatics*, vol. 14, no. 7, pp. 3132-3142, July 2018, doi: 10.1109/TII.2018.2794389.
- [18] F. B. Oriani and G. P. Coelho, "Evaluating the impact of technical indicators on stock forecasting," 2016 IEEE Symposium Series on Computational Intelligence (SSCI), Athens, Greece, 2016, pp. 1-8, doi: 10.1109/SSCI.2016.7850017.
- [19] Prediction Models for Indian Stock Market, <https://www.sciencedirect.com/science/article/pii/S1877050916311619>

- [20] A. P. Ratto, S. Merello, L. Oneto, Y. Ma, L. Malandri and E. Cambria, "Ensemble of Technical Analysis and Machine Learning for Market Trend Prediction," 2018 IEEE Symposium Series on Computational Intelligence (SSCI), Bangalore, India, 2018, pp. 2090-2096, doi: 10.1109/SSCI.2018.8628795.
- [21] Stock market index prediction using artificial neural network Predicción del índice del mercado bursátil utilizando una red neuronal artificial, <https://www.sciencedirect.com/science/article/pii/S2077188616300245>
- [22] Kumar, G., Jain, S. & Singh, U.P. Stock Market Forecasting Using Computational Intelligence: A Survey. *Arch Computat Methods Eng* (2020). <https://doi.org/10.1007/s11831-020-09413-5>
- [23] Guo Z, Wang H, Liu Q, Yang J (2014) A Feature Fusion Based Forecasting Model for Financial Time Series. *PLoS ONE* 9(6): e101113. <https://doi.org/10.1371/journal.pone.0101113>
- [24] Selvamuthu, D., Kumar, V. & Mishra, A. Indian stock market prediction using artificial neural networks on tick data. *Financ Innov* 5, 16 (2019). <https://doi.org/10.1186/s40854-019-0131-7>

DOI: <https://doi.org/10.15379/ijmst.v10i4.3444>

This is an open access article licensed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/3.0/>), which permits unrestricted, non-commercial use, distribution and reproduction in any medium, provided the work is properly cited.