

# Hybrid Approaches for Stocks Prediction and Recommendation System

Vikram Sharma\*, Manik Rakhra, Gauri Mathur

Department of Computer Science and Engineering, Lovely Professional University, Phagwara, India

\*Corresponding author: vikramsh2002@gmail.com

**Abstract.** Hybrid approaches to stock prediction and recommendation are a critical area of research for individual investors and financial institutions. Traditional methods have limitations, leading to the emergence of hybrid models. This paper reviews current research on hybrid models, including GAN-based, LSTM-based, and neural network-based models, Soft Computing based, GRU based models to provide optimal results, for stock recommendation techniques include sentiment analysis, which uses natural language processing to analyze news articles and social media posts, and network analysis, which examines the relationships between stocks to identify stocks likely to move together. It also discusses evaluation metrics used to assess the performance of these models and then it provides the generalize pipelines that can be kept in mind while researching and developing a recommender engine, it also shows the future direction in order to build the hybrid recommenders as well as predictors, making it a valuable contribution to the stock prediction and recommendation field.

## 1 Introduction

Stock price prediction has been a research area for years, benefiting investors and financial institutions. Traditional methods have limitations, leading researchers to use machine learning and artificial intelligence techniques to develop more accurate models. Hybrid models combining multiple machine-learning techniques are promising solutions, and on the other hand Stock recommendation, which advises investors on stocks based on investment goals, risk tolerance, and market conditions, has become increasingly important in recent years because there is high trend making the systems smart enough so that they can act as a personal assistant or advisor.[1][25]

Techniques like fundamental analysis and machine learning have revolutionized stock recommendations. These approaches analyze large amounts of data, identifying patterns and relationships that human analysts may not detect. Hybrid approaches, which combine multiple machine-learning techniques, provide more accurate and reliable recommendations, enhancing the reliability of stock recommendations. These advanced approaches offer a more accurate and reliable approach to stock analysis. In General, the high level working of recommender system is shown below in Fig1

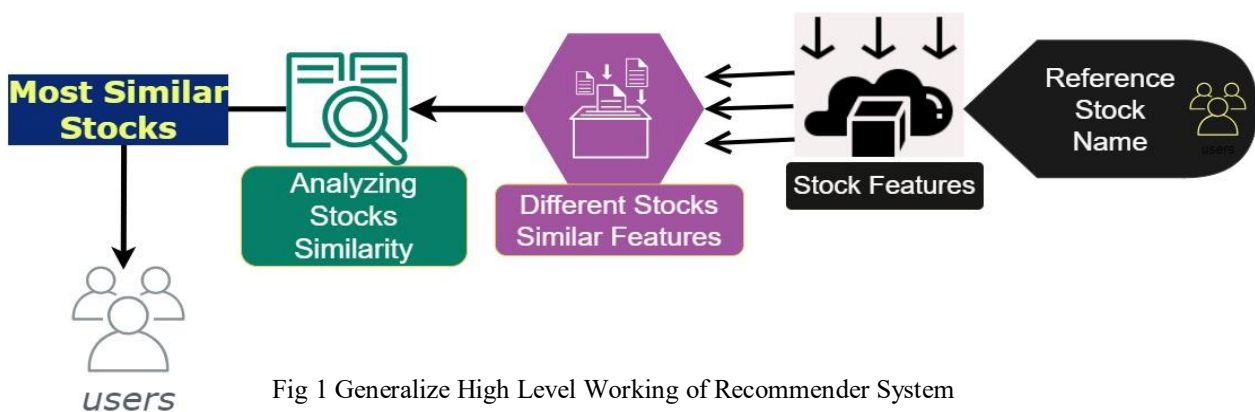


Fig 1 Generalize High Level Working of Recommender System

Other techniques that have been proposed for stock recommendation include sentiment analysis, which uses natural language processing techniques to analyse news articles, social media posts, and other sources of information to determine market sentiment, and network analysis, which examines the relationships between stocks and can identify stocks that are likely to move together.

This paper uses ensembled LSTM to predict the KSE-100 index of the Pakistan Stock Exchange. Akima spline interpolation is used instead of cubic spline interpolation. The IMFs and single monotone residue are used to deal with noise. The proposed model is compared against preexisting popular techniques such as SVM, Random Forest, and Decision Tree. The proposed method is found to be more efficient relative to other techniques.[1] This paper analyses various machine learning algorithms for stock price prediction, including ANN (Artificial Neural Network), SVM (Support Vector Machine), and LSTM (Long Short-Term Memory). ANN is the best-predicting method due to its efficiency and accuracy rate of 90%. However, it is difficult for investors to trust these computational techniques over their own knowledge and experience, which needs to be revolutionized to ensure the future of mankind.[2] This paper reviews the 30-study done for stock predictions and the approaches used for it, focusing on finding the most used model for stocks. It also looks at the Approaches vs Yearly used, with LSTM, SVM, and ANN. ANN is the most common approach. The most popular year for stock market prediction using ML was 2013, despite the high number of papers published in 2015-2017.[3]

This paper used two popular models (ARIMA and LSTM) to predict the stock price of Apple from 2016-2020. The best-performing model was LSTM due to its low rmse (0.237) compared to ARIMA's 3.261 (3.261). The major shortcoming of both models is that they have neglected external factors such as news, revenue, etc. which play a key role in stock fluctuations. The paper suggests that further research should be done to combine LSTM with other models and consider the neglect of external factors.[4] This paper proposed a new approach for stock price prediction, i.e., IIFI (Interpretable Intuition Fuzzy Inference) based model with 6 layers. It solves the dependency of traditional fuzzy systems and creates a membership method to solve fuzzification and defuzzification, considering the CSI 300 and 500 stocks for price prediction.[5] This research paper proposes a hybrid approach for long-term stock market price trend prediction. It includes the use of Raw data and Social Media data to create a Stock index, for preprocessing the Butterfly optimization, Feature Extraction, Feature extensions, and the Brown Planthopper Optimization is applied. Deep Neural Network is also proposed, which integrates firefly-induced extreme learning with the DNN. The performance is measured using 11 global stock market indices and compared with current trendy models with accuracy, precision, and f-measure.[6]

This paper presents a deep learning neural network for predicting stock prices based on S&P 500 data. The model is compared to benchmark algorithms like SVM, linear regression, multilayer perception, and RNN. It can be improved by increasing the number of hidden layers and considering economic factors like GDP and inflammation rates. Bias Variance trade-off problem regularizations and dropouts are also added to solving the bias Variance. The mean absolute error can be reduced by choosing the shorter period and when the current price changes are high.[7] This paper has shown an approach based on a temporal graph-based algorithm, that is used in finding the user's similarity as a parameter for Hidden Markov Model. The critical features of this algorithm. is that it extracts only the important information, it is very accurate in weight estimation due to the use of similarity-based weighted average, it reduces the complexity of measurement of similarity, and this improves the performance on various metrics. The performance can be improved by considering item ratings and the feedback of the user.[8]

This paper proposes a hybrid model based on CNN and a Bidirectional Gated Recurrent unit, with the Feature selection applied. It uses CNN as a feature extractor to collect local features. The models form and compare are CNN (2.06%), LSTM (1.86%), GRU (1.83%), CNN+LSTM (4.64 mape), and FS+CNN+BGRU (1.42 mape). The future scope of development is that the model is restricted to Chinese stock data, but there are more ways of hybridization possible. [9] This paper presents 8 different LSTM and GRU-based architectures for stock market predictions. The four-block neural network architecture is constructed using a Mean absolute percentage error (mape), Root mean squared percentage error (rmpe), and Root mean dimensional percentage accuracy (rmdpe). The model has trained over 4 stocks and the most accurate model architecture is Model 1 of GRU, which performs well in all cases or as per each metric. Boxplot Whisker results in LSTM models having the least accuracy deviation compared to GRU.[10] This survey reviewed 86 papers from 2015 to 2021 on stock price prediction, forex movement prediction, and other models like CNN, LSTM, RNN, Reinforcement Learning, HAN, and NLP based. The metrics of comparison are RMSE, MAPE, MAE, and MSE. It was found that hybrid models have been studied, but there is less work done irrespective of their promising results.[11] This paper presents a three-step procedure for data preprocessing, phase space reconstruction (PSR), and data structuring partitioning with the time window. The Deep Neural LSTM-based architecture is designed and trained over the data and compared to the existing model using four metrics: RMSE, Direction Accuracy (DA), MAPE, and Correlation Coefficient. The proposed model is compared to analyze the impact of PSR and the data of stock from 2010 to 2017. It was found that the established DNN-based prediction had a front foot in comparison to others.[12] This article had done research and analysis of RNN and LSTM-based neural networks to predict stock prices. The two stocks of Nike (NKE) and Google (GOOGL) are of the same stock. The model architecture consists of 4 LSTM layers with 50X96 each having a dropout after a Dense layer. The same model is trained for different epochs and evaluated for various epochs, with GOOGL stock having a loss of 0.0011 and -16.0019 for 12 epochs on NKE and 0.5 and 0.874 respectively for 100 epochs in both stocks respectively.[13]

The major objective of this paper is to build a recommender system for managers that create stock portfolios. Data from 2005 to 2016 is analysed using various approaches, including simple stock averages and Bayesian belief networks. The best performer approach is dimensionality reduction based on precision and recall metrics, which shows a good predictive

power on future trades and recommendation is stronger for US stocks.[14] This paper proposes a new approach to stock prediction based on a generative adversarial network with a multi-layer perceptron as the discriminator to decide between real or generated stock data. It uses S&P 500 dataset to make closing price predictions and performs well compared to other prediction algorithms like LSTM, ANN, and SVR. Future work should explore and find more influential financial features to improve accuracy.[15] This article focuses on finding the optimal model for the predictions of the 5th day and the daily buying and selling strategy for the S&P 500 index stocks. Data from 2000-2017 is considered and a complete pipeline of modelling is used to convert every feature into a relative standard score. GRU-based and LSTM-based models are compared based on self-designed metrics and evaluated for their optimism, pessimism, and return ratio. The GRU model is the best performer with an R2\_score of 0.94, but it has a better return ratio and can deal with extreme price dips and sharp price spikes. Factors like interest rate and GDP growth inflation can help improve the model's performance.[16]

In this paper, DNNs are used to predict the daily return direction of SPDR S&P 500 ETF using 60 economic and financial features. Data is pre-processed and transformed using PCA, and the DNNs show increasing classification accuracy. PCA-represented data performs better than other algorithms, with 31 components providing the best accurate results.[17] This research proposed a hybrid approach of stock prediction based on Adaline Neural Nets and a modified version of PSO. The model uses 6 swarms with 500 iterations and an average accuracy of 98.9%. The accuracy of Bayesian ANN is 98.14, mape is 1.96, and CMSPOS is 95.8 and 4.2 respectively. It predicts the open price of the Bombay stock exchange.[18] An ensemble model is proposed based on an LSTM-based network and CNN-based network stacked one after the other to analyze stock prices and generate stock investment recommendations. Historical data, tweets, and news affecting the stock are also taken into consideration while building the model, historical data has been taken from INSE for the past 20 years.[19] This paper proposed a system that can use for stock trade prediction and recommendation. This system is based on two techniques that are information granulation and fuzzy transformation on raw time series. The task of Information Granulation is to transform a raw time series into a meaningful and interpretable one. The pipeline followed in this paper is the raw time series is taken and then Fuzzy Cognitive mapping is applied in order to cluster it, then information granules are constructed, a fuzzy set is defined over the granules(fuzzifying the time series), then a fuzzy relationship is found, then it does a forecast that gives the gives a value between 0 to 1 that indicates sell or buy the stocks.[20] This paper proposes an approach to a stock recommendation based on core machine learning. The pipeline includes selecting the best indicators, then the 5 most common ML algorithms are Linear, Ridge, Stepwise, generalized boosted regression, and random forest. The model with the least rmse model is used to rank stocks, and then the given stock is verified or tested by the popular portfolio allocation methods. Metrics used for comparison are the Sharpe ratio and cumulative returns. The S&P 500 stocks data is also used.[21]

This research proposes a new method based on deep learning and 2D2 PCA with radial bias-based feed forward Neural Net and with RNN. It has 4.8% more accurate than the RBFNN and is more accurate in terms of total returns and rmse. The best-fitted window size is 20 and with a component size of 10x10. Stock prices are volatile, making predicting predictions difficult.[22] This paper proposed a deep learning-based structure or formalization for stock price prediction. There was a comparison of 3 stocks TCS, Infosys, and Cipla. The model was only trained with Infosys and can predict the other two. The metrics used were error percentages, and the average error percentages over 3 stocks were for RNN 5.12, LSTM 5.31, and CNN 4.98. The best performer model is CNN based irrespective of the relatively high percentage error for TCS.[23] The study examines how changes in stock recommendations by Foreign Institutional Investors (FIIs) affect the Taiwan equity market. The results show that these changes provide valuable information for the market, and FIIs may prioritize profit over reputation. Additionally, experienced retail investors exhibit a lack of sophistication, selling upgraded stocks quickly for gains and holding on to downgraded stocks for longer periods while continuing to purchase them as prices fall. This suggests a disposition effect among experienced Retail investors in response to FII stock recommendation changes.[24] The proposed model is based on news article data and uses a distributed vector representation approach to deal with textual or new articles. It is compared to numeric approaches like BOW and Numeric data-only methods. The five industries considered are Transport Equipment, Pharmaceutical, Machinery, Electronics, and Wholesale Trade. The LSTM, SVR, MLP, and RNN are all comparable to be profitable in terms of profit gains in all industries but from them, LSTM performs best with a profit of 1211.90.[25]

## 2 Proposed Work

After analysing the previous work done in the field of stock prediction and recommendation the generalized hybrid pipeline for training the models is shown in Fig2.

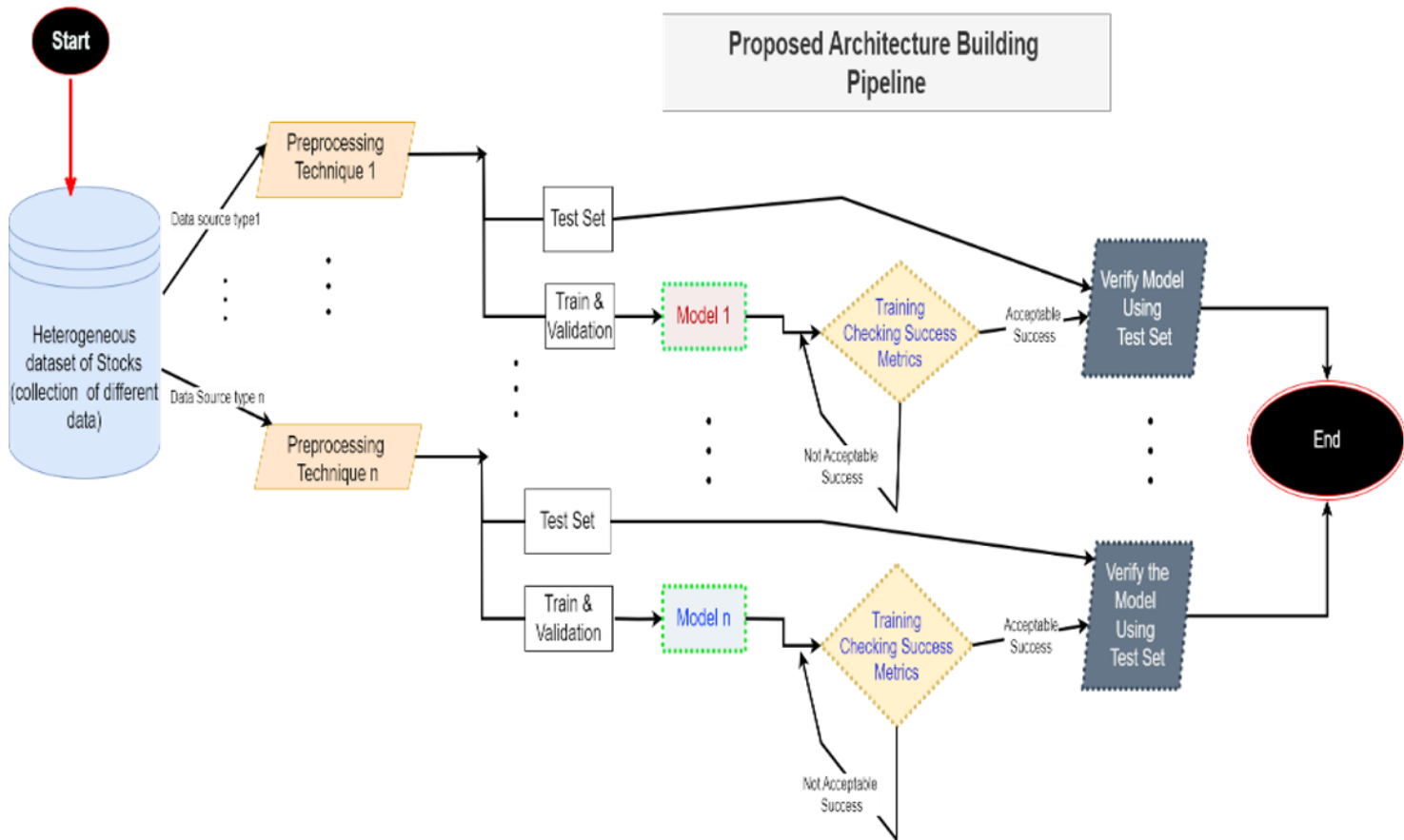


Fig 2 Generalize Architecture Building Pipeline

The process followed in the building pipeline procedure is as followed.

- Data is collected with various sources this data is heterogenous in nature because it can include the OHCL data of stocks, news, etc.
- Then as per the data the preprocessing techniques are applied for example for the OHCL there is a common approach that is moving average for removing the seasonality and remove the noises from it, similarly for each data there is different preprocessing approaches.
- Then this data is split into two parts that are Training + Validation Set and Testing Set.
- Then for Each type of data a model is trained with the Training and Validation Set respectively, this training is continued until we get an optimal success rate.
- As soon as we reach to the optimal training and validation accuracy then the Test set is used in order to verify the models separately.

Since, the above procedure in Fig 2 gives various models trained and verified now these models will be utilized in the recommendation pipeline, in-order to get the optimal stocks as per the user-need and that procedure is shown in the Fig3 below:

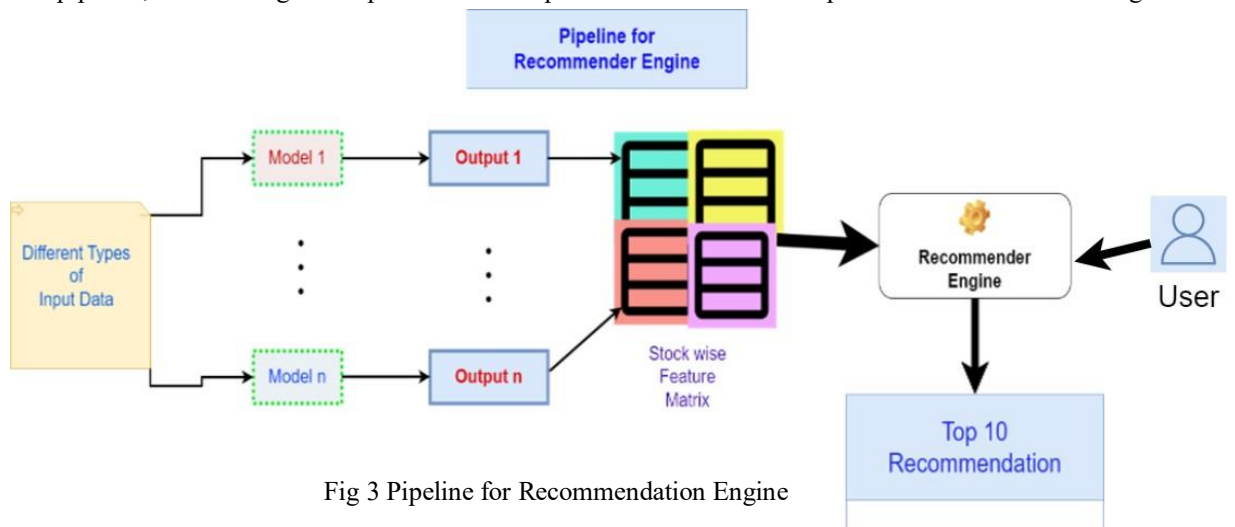


Fig 3 Pipeline for Recommendation Engine

In the above Fig 3 the procedure followed to get the recommendation is divided into two steps:

1. Stock Feature Matrix Construction
2. Stocks Comparison

Stock Feature Matrix generation involves the following steps in it

- From the Heterogenous data stack the data is feed as per the model needs as input
- Then the model processes the input and produce the result as output
- These different Output is grouped as per the stocks.

After Stock Feature Matrix Generation Stock Comparison is done by the following steps

- The Grouped Matrix of Stocks Features is taken then based on that the similarity value is calculated, this creates a stock-by-stock pivot table.
- As the user provide a stock from the pivot it selects those stocks whose value is most for a given stock.

### 3 Results and Discussion

After analysing the papers, It has been observed that various authors used various different approaches which leads to heterogenous metric of success which are RMSE (root mean squared error), MAPE (mean absolute percentage error), accuracy, MAE (Mean Absolute Error), R2 Score of Accuracy, F1 score. Some of the authors have used multiple approaches that are performing very well so it has also been considered in the comparison and it is shown in the Table 1

Table 1 Comparison of Different Methods for Stock Price Prediction

Reference	Year	Methods	Metric
[1]	2023	Hybrid-Akima-EMD-LSTM	RMSE: 299.541 MAPE: 0.578
[2]	2023	Multilayer Feedforward Network with Session Training by Backpropagation	Accuracy: 89
		RNN and LSTM based model	Accuracy: 90
		Data mining technique with ANN	Accuracy: 93.89
		Resilient Backpropagation Algorithm	Accuracy: 94.55
		LSTM based model	Accuracy: 94.57
		Relative Difference Method	Accuracy: 95
		Soft Computing with ANN	Accuracy: 99.9
[4]	2023	LSTM based Network	RMSE: 0.237
		ARIMA based model	RMSE: 3.261
[6]	2022	FEL-DNN	Avg-Accuracy: 83.66 Avg-F1-Score: 84.38
[7]	2022	DFNN	MAE: 0.78
[9]	2022	FS+CNN+BGRU	R <sup>2</sup> score: 0.983808
[10]	2022	LSTM based Model	MAPE: 96.98%
		GRU based Model	MAPE: 96.84%
[11]	2021	LSTM based Model	MAE: 6.85E-4
[15]	2019	Gan Based Model	MAPE: 0.0137
[16]	2019	LSTM based Model	RMSE: 0.000428 R <sup>2</sup> Score: 0.94
		GRU based Model	RMSE: 0.000511 R <sup>2</sup> Score: 0.93
[17]	2019	ANN (with PC's =31)	MSE: 0.3011
[18]	2018	Proposed Hybrid Model	Accuracy: 98% MAPE: 1.1
[22]	2017	RBFNN (Radial Bias Function NN)	MAPE: 0.080059 RMSE: 0.00674

[23]	2017	CNN based Proposed Model	Accuracy: 95.09
[25]	2016	Proposed LSTM Network	Trade-Gains: 1211.90(Ten Thousand of Yen)

The best model of the year 2023 is Hybrid Model of ANN and Soft Computing, for the year 2022 is LSTM based model proposed in [10]. The best model based on the accuracy over the past years (2016-2023) is a hybrid model based on ANN and Soft Computing proposed in [2], The best model with the least RMSE over the years(2016-2023) is LSTM based model proposed in [16].

In the field of recommender engines based on the research and work done so far they can be classified into the 5 renowned classes showed in the Fig 4.

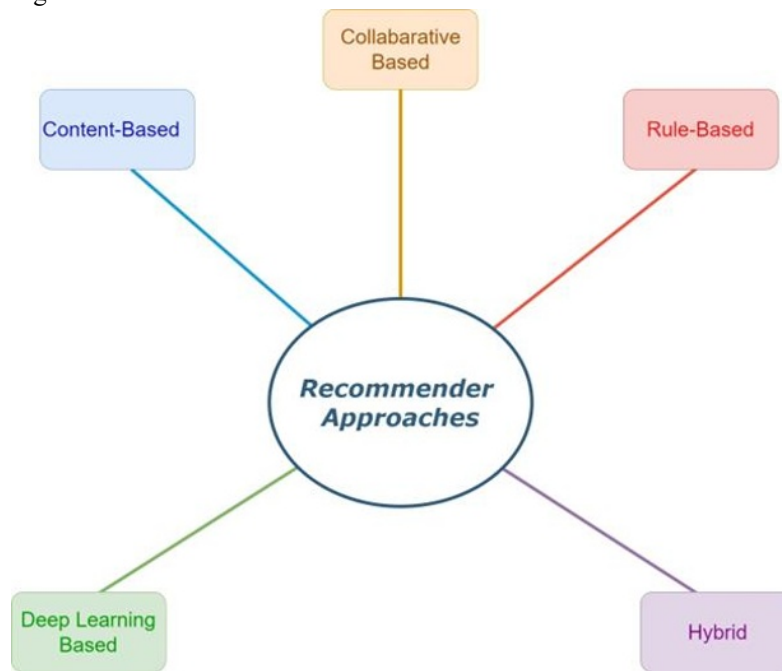


Fig 4 Classification of Recommenders Based on Approaches

## 4 Conclusion And Future Scope

Recent research has led to the development of a hybrid methodology for stock prediction and recommendation. These methodologies take the best from various approaches, such as ANN, RNN, GAN's, LSTM, GRU, ARIMA, Bidirectional GRU, and Neural Networks with Radial bias. The hybrid models provide accurate results based on various factors such as different price types (OHCL), textual and visual data. There improvements can be measured using metric like accuracy, precision, root mean squared error(rmse), mean absolute percentage error(mape), mean absolute error(mae) or by their reliability, performance in the real world. This improvement is clearly visible by the results shown in this paper where a hybrid model of Artificial Neural Net and Soft Computing has given the best accuracy i.e.,99.9%. This paper also gives a future direction towards the various hybridization that can be done in order to create a robust recommender based on the volatile, and heterogenous data available.

## References

- [1] M. Ali, D. M. Khan, H. M. Alshanbari, and A. A.-A. H. El-Bagoury, "Prediction of Complex Stock Market Data Using an Improved Hybrid EMD-LSTM Model," *Applied Sciences*, vol. 13, no. 3, p. 1429, Jan. 2023, doi: 10.3390/app13031429.
- [2] D. Sheth and M. Shah, "Predicting stock market using machine learning: best and accurate way to know future stock prices," *International Journal of System Assurance Engineering and Management*. Springer, Feb. 01, 2023. doi: 10.1007/s13198-022-01811-1.

- [3] L. N. Mintarya, J. N. M. Halim, C. Angie, S. Achmad, and A. Kurniawan, "Machine learning approaches in stock market prediction: A systematic literature review," *Procedia Comput Sci*, vol. 216, pp. 96–102, 2023, doi: 10.1016/j.procs.2022.12.115.
- [4] Haoran Wu, Shuqi Chen, and Yicheng Ding, "Comparison of ARIMA and LSTM for Stock Price Prediction," *Financial Engineering and Risk Management*, vol. 6, no. 1, 2023, doi: 10.23977/ferm.2023.060101.
- [5] W. Wang et al., "An interpretable intuitionistic fuzzy inference model for stock prediction," *Expert Syst Appl*, vol. 213, p. 118908, Mar. 2023, doi: 10.1016/J.ESWA.2022.118908.
- [6] K. Venkateswararao and B. V. R. Reddy, "LT-SMF: long term stock market price trend prediction using optimal hybrid machine learning technique," *Artif Intell Rev*, 2022, doi: 10.1007/s10462-022-10291-5.
- [7] T. Al-Sulaiman, "Predicting reactions to anomalies in stock movements using a feed-forward deep learning network," *International Journal of Information Management Data Insights*, vol. 2, no. 1, Apr. 2022, doi: 10.1016/j.jjime.2022.100071.
- [8] A. Gupta and P. Srinath, "A recommender system based on collaborative filtering, graph theory using HMM based similarity measures," *International Journal of System Assurance Engineering and Management*, vol. 13, pp. 533–545, Mar. 2022, doi: 10.1007/s13198-021-01537-6.
- [9] Q. Zhou, C. Zhou, and X. Wang, "Stock prediction based on bidirectional gated recurrent unit with convolutional neural network and feature selection," *PLoS One*, vol. 17, no. 2 February, Feb. 2022, doi: 10.1371/journal.pone.0262501.
- [10] A. Lawi, H. Mesra, and S. Amir, "Implementation of Long Short-Term Memory and Gated Recurrent Units on grouped time-series data to predict stock prices accurately," *J Big Data*, vol. 9, no. 1, Dec. 2022, doi: 10.1186/s40537-022-00597-0.
- [11] Z. Hu, Y. Zhao, and M. Khushi, "A survey of forex and stock price prediction using deep learning," *Applied System Innovation*, vol. 4, no. 1. MDPI AG, pp. 1–30, Mar. 01, 2021. doi: 10.3390/ASI4010009.
- [12] P. Yu and X. Yan, "Stock price prediction based on deep neural networks," *Neural Comput Appl*, vol. 32, no. 6, pp. 1609–1628, Mar. 2020, doi: 10.1007/s00521-019-04212-x.
- [13] A. Moghar and M. Hamiche, "Stock Market Prediction Using LSTM Recurrent Neural Network," in *Procedia Computer Science*, Elsevier B.V., 2020, pp. 1168–1173. doi: 10.1016/j.procs.2020.03.049.
- [14] G. De Rossi, J. Kolodziej, and G. Brar, "A recommender system for active stock selection," *Computational Management Science*, vol. 17, no. 4, pp. 517–547, Dec. 2020, doi: 10.1007/s10287-018-0342-9.
- [15] K. Zhang, G. Zhong, J. Dong, S. Wang, and Y. Wang, "Stock Market Prediction Based on Generative Adversarial Network," in *Procedia Computer Science*, Elsevier B.V., 2019, pp. 400–406. doi: 10.1016/j.procs.2019.01.256.
- [16] A. Sethia and P. Raut, "Application of LSTM, GRU and ICA for stock price prediction," in *Smart Innovation, Systems and Technologies*, Springer Science and Business Media Deutschland GmbH, 2019, pp. 479–487. doi: 10.1007/978-981-13-1747-7\_46.
- [17] X. Zhong and D. Enke, "Predicting the daily return direction of the stock market using hybrid machine learning algorithms," *Financial Innovation*, vol. 5, no. 1, Dec. 2019, doi: 10.1186/s40854-019-0138-0.
- [18] M. R. Senapati, S. Das, and S. Mishra, "A Novel Model for Stock Price Prediction Using Hybrid Neural Network," *Journal of The Institution of Engineers (India): Series B*, vol. 99, no. 6, pp. 555–563, Dec. 2018, doi: 10.1007/s40031-018-0343-7.
- [19] M. S. Hegde, G. Krishna, and R. Srinath, "An Ensemble Stock Predictor and Recommender System," in *2018 International Conference on Advances in Computing, Communications and Informatics (ICACCI)*, IEEE, Sep. 2018, pp. 1981–1985. doi: 10.1109/ICACCI.2018.8554424.
- [20] W. Wang and K. K. Mishra, "A novel stock trading prediction and recommendation system," *Multimed Tools Appl*, vol. 77, no. 4, pp. 4203–4215, Feb. 2018, doi: 10.1007/s11042-017-4587-z.

- [21] H. Yang, X. Y. Liu, and Q. Wu, “A Practical Machine Learning Approach for Dynamic Stock Recommendation,” in Proceedings - 17th IEEE International Conference on Trust, Security and Privacy in Computing and Communications and 12th IEEE International Conference on Big Data Science and Engineering, Trustcom/BigDataSE 2018, Institute of Electrical and Electronics Engineers Inc., Sep. 2018, pp. 1693–1697. doi: 10.1109/TrustCom/BigDataSE.2018.00253.
- [22] R. Singh and S. Srivastava, “Stock prediction using deep learning,” *Multimed Tools Appl*, vol. 76, no. 18, pp. 18569–18584, Sep. 2017, doi: 10.1007/s11042-016-4159-7.
- [23] S. Selvin, R. Vinayakumar, E. A. Gopalakrishnan, V. K. Menon, and K. P. Soman, “Stock price prediction using LSTM, RNN and CNN-sliding window model,” in 2017 International Conference on Advances in Computing, Communications and Informatics (ICACCI), IEEE, Sep. 2017, pp. 1643–1647. doi: 10.1109/ICACCI.2017.8126078.
- [24] C.-H. Chang, “Exploring stock recommenders’ behavior and recommendation receivers’ sophistication,” *Journal of Economics and Finance*, vol. 41, no. 1, pp. 1–26, Jan. 2017, doi: 10.1007/s12197-015-9330-x.
- [25] R. Akita, A. Yoshihara, T. Matsubara, and K. Uehara, “Deep learning for stock prediction using numerical and textual information,” in 2016 IEEE/ACIS 15th International Conference on Computer and Information Science, ICIS 2016 - Proceedings, Institute of Electrical and Electronics Engineers Inc., Aug. 2016. doi: 10.1109/ICIS.2016.7550882.