Stock Market, Sentiment Analysis and Summarization: A Literature Review

[1] Ishitva Awasthi, [2] M. Pushpalatha

[1] [2] Department of Computing Technologies, School of Computing, SRM Institute of Science and Technology, India
Corresponding Author Email: [1] ia9817@srmist.edu.in, [2] pushpalatha.m@ktr.srmuniv.ac.in

Abstract— Accurate stock market models can give investors tools for better data-based decision-making. In addition to stock market prediction, the sentiment of the investor also influences the stock prices. Furthermore, to aid stock market investors, important information must also be filtered out from the news. The combination of these models can aid traders in lowering investing risk and choosing the best stocks. In this literature review, we investigated machine learning algorithms for stock market prediction, sentiment analysis and automatic text summarization. It examines recent studies that were published as well as surveys and other reviews of the stock market.

Index Terms— sentiment analysis, stock market, summarization.

I. INTRODUCTION

The stock market has a significant impact on a country's economic performance. The unpredictable nature of stock price forecast presents a difficult risk. One of the most significant problems many stockholders encounter is the prediction of stock prices. This makes it more unstable to invest our hard-earned cash out of fear of losing it. As a result, many people are reluctant to speculate in the stock market. [21] Stock market funding produces impressive revenue for the nation. If an effective stock prediction model is created, we will develop a sense of the market's movement over time and be able to spot trends that we might not have otherwise noticed.

Machine learning is the most reasonable approach to overcome this issue given the growing computational capacity of computers. The effectiveness of machine learning models like SVM, CNN, RNN, LSTM, etc. has been compared in numerous research. The dataset used to forecast the movement of stock prices is the most crucial component of machine learning algorithms. Because even minor modifications to the data might result in significant changes in price predictions, the dataset should be as precise as possible.[22]

The most recent development in stock analysis uses sentimental analysis techniques and Natural Language Processing (NLP) to pinpoint the emotions expressed in financial news headlines. Text data, such as news stories, would significantly affect stock values in the stock market. When sentiment analysis is applied to news headlines, the goal is to determine the importance of data points on different emotions in order to comprehend how news affects the public's overall sentiments. It is frequently disputed if human sentiments run the stock market or if human sentiments drive the stock market.

By examining market data, stock market investors hope to increase their profits. Financial news, which is regarded as a key source of market information, is rapidly growing in both volume and transmission frequency. Investors were able to make better decisions because to models that could make forecasts based on news items. On the other hand, as news items are by nature free-style literary compositions, they could have sounds even while they offer valuable information. Summarization is a good candidate to be included to the preparation of financial news items since it extracts the main information that the documents aim to express and filters the sounds that can divert readers' attention.

In order to help stock investors, extract the important information and cut out the distraction of unnecessary information when reading news reports, numerous researchers have recently adopted news sentiment as an input element for stock prediction. The problem of information overload and shifts in the polarity of the news have been successfully solved by many local and international experts who have provided ways for automatic text summarization. This allows the sentiment tendency of the information to be retrieved more correctly. The study of the effect of news digest on stock fluctuation and the prediction of the trend of stock fluctuation will be of theoretical and practical significance to investors' investment decisions and improve market structure as a result of the application of existing natural language processing technology and stock prediction technology.

The most important part is the news to be selected as it will be base for the stock market prediction. The combination of summarization, sentiment analysis and stock market prediction will help in achieving better profits and insights.

II. LITERATURE SURVEY

A. Stock Market

In the paper [1], the authors have conducted a survey about various machine learning techniques for stock market prediction. Their observations show that deep learning techniques like Recurrent Neural Network (RNN) have increased interest of researchers in recent times. They also mentioned that the pre-processing steps has helped in making simpler, more interpretable, and potentially more accurate models. They did not include soft-computing methods, like regression analysis for study.

In the systematic review [2], the researchers have considered work which include technical indicator, daily stock news and social media networks as datasets. They found that technical indicators help the most in predicting the stock movement. The drawback is that social media network all news were considered.

In [3], the author proposed using Transformer Encoder-based Attention Network (TEANet) framework for stock market prediction. They have used historic stock prices of five calendar days and the related text to predict stock market. The model needs more practical implementation and taking into consideration the actual stock price value can prove more beneficial.

In the paper [5], the authors suggested using two channel attention mechanism fusion model on Convolutional Neural Network (CNN) - Long Short-Term Memory (LSTM) for prediction of stock price. The suggested method was performed on China news website and stock market. The method suggested by authors gives greater positive return on flexible stock prediction. They could have included other stock market news to get better results.

In the paper [10], the researchers have done sentiment analysis on stock market from financial news article. The methodology used is LSTM and Explainable AI. The method predicts stock market using historical stock data as well as sentiment analysis from news. The drawback of the method is that it is not automated and the method is not multi-lingual.

The authors in paper [11] proposed a method for stock market prediction for Indian market. They are using sentiment analysis and deep learning for making the predictions. The forecast was made using LSTM and Random forest. The method predicts future trends for the stock. The accuracy is only 60% because the news considered was not filtered.

In the paper [15], the author have suggested a model fo identifying stock price trends using sentiment analysis and automatic text summarization. The machine learning algorithms used are Sequence-to-Sequence (Seq2Seq) and LSTM. The model was performed the exploration on large scale chinese short text summary dataset, news from authorized websites and dictionaries for sentiment analysis. The setback was that the dataset was limited to financial information website only. Inclusion of shareholder forum would have helped in increasing accuracy.

In the paper [16], the author used transformer model for forecasting stock market index. The index considered as dataset were CSI 300, S&P 500 and Hang Seng index. The observation made in the paper was that transformer outperforms RNN, CNN and LSTM while forecasting stock market index. The only downside was that forecast for each index was done separately.

In [17], the authors have done a comprehensive study on various hybrid deep learning techniques for prediction of stocks. The deep techniques taken into consideration for the review are Auto Regressive Integrated Moving Average (ARIMA), LSTM, CNN, Hybrid LSTM and Hybrid CNN. The first observation made was that the hybrid techniques outperform the single deep learning algorithms. Secondly, it was also observed that sentiment of the people also influences the stock prices. Finally, if other hyperparameters would also be considered for prediction then, the results could have been increased.

B. Sentiment Analysis

In [4], the authors have studied sentiments and financial signals for stock movement. They are using Normalizing flow to predict the movement of stock. The normalizing flow was applied on tweets and historical stock data. It was found that the normalizing flow helps in gaining better insight. The setback was that the content of tweets affects the normalizing flow.

In the paper [6], the authors have performed a survey for evaluation of sentiment analysis. The survey is focused on finance sector primarily. They have considered all the techniques from lexicons to transformer for the survey. The study starts with a lexicon-based method that uses term and sentence encoders and ends with modern NLP transformers. They conducted all the algorithms on one dataset, that is, Financial Phrase-Bank dataset.

In the paper [12], the researcher has performed sentiment analysis using Sentence-level Sentiment Score (SSS). The dataset taken into consideration is RSS stock news feed. In the method, blends news feed emotion from the general public with Sensex data to forecast stock market behavior. The accuracy achieved was 78.75% and can be increased if more stock indicator levels were included.

In the paper [18], the authors have proposed to forecast stock market movement using investor sentiment and Support Vector machine (SVM). The additional point of consideration was days-of-the-week effect and it improves the precision of sentiment indexes. The index taken for experiment is SSE 50. The method cannot process a large amount of data in real time.

C. Summarization

The paper [7] is a comprehensive survey of automatic text summarization methods. The survey includes Extractive methods, Abstractive methods and Hybrid methods. The

summarization was performed on the CNN daily news dataset and DUC 2001-2006 dataset. The survey covers nearly all the method of summarization systematically. The few setbacks would be that it does not include summarizing of multiple documents, summarization tailored to the needs of the user, and use of text summarization.

In paper [8], the authors have proposed a survey for abstractive text summarization. The methods which are present in the survey are Sequence-to-sequence (Seq2Seq) models and Reinforcement Learning (RL) techniques. The positive observation was that the combination of sequence-to-sequence and reinforcement learning improves the results significantly. The suggested next development is the addition of multi-reward and semantic-based metrics to RL.

In the paper [9], the researchers have recommended a method for Multi-Document Summarization (MDS) using fuzzy logic. It also includes improved sentence scoring and redundancy removal techniques. The method was performed using DUC 2004 dataset. The challenge in MDS is dealing with redundant information, which is resolved using cosine similarity and fuzzy logic weights.

In the paper [13], the authors have suggested automatic summarization for financial reports. The summarization is performed using sentence-based clustering as well as section-based clustering. The increase the F1 score is achieved through assigning weight to each section. The efficiency can be improved by better identification of section.

In the paper [14], the authors wanted to measure the effect of summarization on stock market prediction. They wanted to do the analysis of news impact via summarization. They used Self-Present Sentence Relevance (SPSR) for summarization of news articles. The stock market prediction was done using Support Vector Machine (SVM). It was found that the summarized article help in predicting the stock better as compared to full length article. They also mentioned that the

accuracy can be increased if the various other machine learning algorithms were used for stock market prediction along with summarization.

In the paper [19], the authors have proposed summarization of financial reports using TIBER (Tf Idf Bert dEpendency pRediction). TIBER is an extractive summarization technique. The machine learning algorithms used for summarization are Bi-LSTM, n-gram and Bidirectional Encoder Representations from Transformers (BERT). For choosing summary sentences, they have considered semantic information which is more important than syntactic information. It's crucial that the generated summaries contain any financial terms, which is represented by n-grams having $1 \leq n \leq 3$. The dataset taken is Financial Narrative Summarization (FNS 2021) shared report which is extremely long. The future enhancement also suggest using other transformer techniques for better results.

In the paper [20], the authors have taken tweet summarization of news articles. The summarization was done by considering the tweet similarity graph. The tweets considered were from New York Post. For better identification, community detection methodology aids in locating the tweets expressing these various viewpoints. The number of tweets considered for summarization were less (actual ~ 36) and if more tweets were a better result could have been obtained.

III. EVALUATION PARAMETERS

In this paper, we have reviewed various machine learning and deep learning algorithms for stock market prediction, summarization and sentiment analysis and each of which have different evaluation parameters for identifying the accuracy of the model. The table 1 mentions the parameter, the formula and the value range of the parameter.

Table 1. Evaluation Parameters

Areas	Parameter	Formula	Value Ranges
Stock Market Prediction and Sentiment Analysis	Hit Ratio	HR = (cache hit)/(cache hits + cache misses)	0% to 100%
	Accuracy	Accuracy = (tp+tn)/(tp+tn+fp+fn)	0% to 100%
	RMSE	$RMSE = \sqrt{\frac{1}{N}\sum_{t=1}^{N}(observed_t - predicted_t)^2}$	0 to1
	MSE	$MSE = \frac{1}{N}\sum_{t=1}^{N}(observed_t - predicted_t)^2$	0 to 1
	MAE	$MAE = \frac{1}{N}\sum_{i=1}^{N}|observed_t - predicted_t|$	0 to1
	MCC	$MCC = \frac{tp \times tn - fp \times fn}{\sqrt{(tp+fp)(tp+fn)(tn+fp)(tn+fn)}}$	-1 to 1
Summarization	Precision	Precision = tp/ (tp + fp)	0 to 1
	Recall	Recall = tp/ (tp + fn)	0 to 1

Areas	Parameter	Formula	Value Ranges
	F1 score	$F1\ score = 2 * \dfrac{Precision\ *\ Recall}{Precision\ +\ Recall}$	0 to 1
	Rouge Score	Rouge score is calculated based on the number of consequent words chosen for consideration from the candidate summary to the reference summary.	0 to 1

IV. CONCLUSION AND FUTURE SCOPE

In this paper, we have reviewed various machine learning and deep learning algorithms for stock market prediction, summarization and sentiment analysis. We discovered that even in recent years, the adoption of deep learning algorithms has remained on the rise quickly. We have also provided the future research for each paper reviewed. The paper also discusses the various performance metrics for stock market prediction, sentiment analysis and summarization. The future enhancement would be to analyze and implement method for combining the three technologies together and gaining better insights as well as profits from the stock market.

REFERENCES

[1] Kumbure, Mahinda Mailagaha, et al. "Machine learning techniques and data for stock market forecasting: a literature review." Expert Systems with Applications (2022): 116659.

[2] Bustos, Oscar, and Alexandra Pomares-Quimbaya. "Stock market movement forecast: A systematic review." Expert Systems with Applications 156 (2020): 113464.

[3] Zhang, Qiuyue, et al. "Transformer-based attention network for stock movement prediction." Expert Systems with Applications 202 (2022): 117239.

[4] Tai, Wenxin, et al. "Learning Sentimental and Financial Signals With Normalizing Flows for Stock Movement Prediction." IEEE Signal Processing Letters 29 (2021): 414-418.

[5] Sun, Lin, Wenzheng Xu, and Jimin Liu. "Two-channel Attention Mechanism Fusion Model of Stock Price Prediction Based on CNN-LSTM." Transactions on Asian and Low-Resource Language Information Processing 20.5 (2021): 1-12.

[6] Mishev, Kostadin, et al. "Evaluation of sentiment analysis in finance: from lexicons to transformers." *IEEE access* 8 (2020): 131662-131682.

[7] El-Kassas, Wafaa S., et al. "Automatic text summarization: A comprehensive survey." Expert Systems with Applications 165 (2021): 113679..

[8] Alomari, Ayham, et al. "Deep reinforcement and transfer learning for abstractive text summarization: A review." Computer Speech & Language 71 (2022): 101276..

[9] Patel, Darshna, Saurabh Shah, and Hitesh Chhinkaniwala. "Fuzzy logic based multi document summarization with improved sentence scoring and redundancy removal technique." *Expert Systems with Applications* 134 (2019): 167-177.

[10] Gite, Shilpa, et al. "Explainable stock prices prediction from financial news articles using sentiment analysis." PeerJ Computer Science 7 (2021): e340.

[11] Darapaneni, Narayana, et al. "Stock Price Prediction using Sentiment Analysis and Deep Learning for Indian Markets." arXiv preprint arXiv:2204.05783 (2022).

[12] Bharathi, Shri, and Angelina Geetha. "Sentiment analysis for effective stock market prediction." International Journal of Intelligent Engineering and Systems 10.3 (2017): 146-154.

[13] Abdaljalil, Samir, and Houda Bouamor. "An Exploration of Automatic Text Summarization of Financial Reports." Proceedings of the Third Workshop on Financial Technology and Natural Language Processing. 2021.

[14] Li, Xiaodong, et al. "Does summarization help stock prediction? A news impact analysis." IEEE intelligent systems 30.3 (2015): 26-34.

[15] Tianfang, Qi, and Jiang Hongxun. "Exploring stock price trend using Seq2Seq based automatic text summarization and sentiment mining." Management Review 33.5 (2021): 257

[16] Wang, Chaojie, Yuanyuan Chen, Shuqi Zhang, and Qiuhui Zhang. "Stock market index prediction using deep Transformer model." Expert Systems with Applications 208 (2022): 118128.

[17] Shah, Jaimin, Darsh Vaidya, and Manan Shah. "A Comprehensive Review on Multiple Hybrid Deep Learning Approaches for Stock Prediction." Intelligent Systems with Applications (2022): 200111.

[18] Ren, Rui, Desheng Dash Wu, and Tianxiang Liu. "Forecasting stock market movement direction using sentiment analysis and support vector machine." IEEE Systems Journal 13, no. 1 (2018): 760-770.

[19] Vanetik, Natalia, Marina Litvak, and Sophie Krimberg. "Summarization of financial reports with TIBER." Machine Learning with Applications 9 (2022): 100324.

[20] Chakraborty, Roshni, Maitry Bhavsar, Sourav Kumar Dandapat, and Joydeep Chandra. "Tweet summarization of news articles: An objective ordering-based perspective." IEEE Transactions on Computational Social Systems 6, no. 4 (2019): 761-777.

[21] Lekshmi S, Deepa Rajan, and Sam G Benjamin. "Survey Paper On Stock Prediction Using Machine Learning Algorithms". International Journal of Creative Research Thoughts. , no. 9 (2021): c446-c450.

[22] SINHA, ANKIT, YASH AGRAWAL, VIDHAN KUMAR, and CHANDAN KUMAR. "Survey of Stock Price Prediction Using Sentiment Analysis." no. May (2020): 1140-1143.