INDIAN JOURNAL OF SCIENCE AND TECHNOLOGY

RESEARCH ARTICLE

Forecasting Business Status of Organizations by Analyzing Historic Earnings Call Transcripts with the Aid of Text Refinement Framework

Prakhyath Rai[1]*, **Bellipady Shamantha Rai**[1], **Permanki Guthu Rithesh Pakkala**[1], **R Akhila Thejaswi**[1]

1 Sahyadri College of Engineering & Management, Mangaluru, Karnataka and Affiliated to Visvesvaraya Technological University, Belagavi, Karnataka, India

Check for updates

OPEN ACCESS

Received: 17-02-2023

Accepted: 23-05-2024

Published: 10-06-2024

Citation: Rai P, Rai BS, Pakkala PGR, Thejaswi RA (2024) Forecasting Business Status of Organizations by Analyzing Historic Earnings Call Transcripts with the Aid of Text Refinement Framework. Indian Journal of Science and Technology 17(24): 2469-2477. https://doi.org/10.17485/IJST/v17i24.347

* **Corresponding author**.

prakhyathrai9@gmail.com

Funding: None

Competing Interests: None

Published By Indian Society for Education and Environment (iSee)

ISSN
Print: 0974-6846
Electronic: 0974-5645

Abstract

Objectives: The research work focuses on providing an effective framework for automated text refinement that aids in financial condition projections for the company based on prior transcripts of earnings calls. The proposed framework captures the ad hoc advancements of the organizations described in the earnings call as sentiments and computes a score based on the captured sentiments. The sentiment score is then used as prime parameter to predict the stock values of the organizations. **Methods:** The framework is equipped with sentiment analysis or opinion mining to identify and extract the subjective content using text mining and Natural Language Processing (NLP). The extracted sentiments help in yielding a sentiment score to aid in the process of stock projection. The research also illustrates how the sentiment score-based stock prediction enhances in projections of stock compared to existing ML frameworks like LSTM, Random Forest, ARIMA and Regression models. **Findings:** The proposed work has an accuracy score of 93%, precision 96% and recall 95% which is comparatively better than existing ML frameworks framed on LSTM, Random Forest, ARIMA and Regression models. **Novelty:** The research framework overcomes the influence of regular features and test data in stock prediction by using the computed sentiment prediction score from the extracted sentiment phase to aid in prediction stock values and determine the financial status of organizations. The existing frameworks project the stock price based on trained model from previous stock price repository, which tend to fail capturing ad hoc changes incurring in the organization such as change of management or any economic disaster which can poses a high impact on stock projections, the proposed research work captures these organizational changes from the earnings call transcripts as sentiments and build a score to yield the stock projection framework.

Keywords: Natural Language Processing; Text Refinement; Loughran McDonald Sentiment Classifier; Term Frequency Inverse Document Frequency; Stock Price

1 Introduction

Each fiscal quarter, a company's management notifies investors and analysts on the condition of the company through an earnings call. Financial analysts and business experts carefully evaluate these conversations in pursuit of fresh information. The transcripts of these conversations may be used to get the most recent facts and information about a firm's revenue and strategic direction from corporate leadership. Because of this, the analysis report of these transcripts is extremely significant. Sentiment based features play a vital role in projection of organizational financial strengths[1]. Ad Hoc organizational changes like management change in case of takeovers based on state legislation play a vital role in measuring corporate integrity, which in turn can affect the socio-economic conditions of any organizations[2]. Investor sentiments is also a vital factor contributing towards the projections of organization's financial status[3]. Data acquisition is a significant process in any forecasting scenario, verbal communications or interviewed contents can contribute towards formation of rich data repository[4]. Earnings call is a teleconference held primarily by a publicly traded firm to discuss financial results online. Accurately documenting such events might have a consequence on the stock and its shareholders[5]. For investors and equity analysts, earnings calls are one of the most important sources of information. A company's fundamental analysis can be accomplished using the information supplied during earnings calls. In fundamental analysis, analysts might merge the facts and information gathered from the conference with the data offered in the management, discussions, and analysis portion of the company's reports. Earnings calls are necessary since investors typically plan their transactions close to the start of an upcoming conference. Equity analysts revise their financial targets using the data presented at such gatherings. The resulting earnings call transcripts will be extremely data-heavy, with each call averaging 20–30 pages of textual information. One of the main problems is coverage; it is extremely difficult for an analyst to physically pay heed to every important conference call throughout earnings season, and identifying, analyzing, and understanding investment relationships from these calls is a difficult task whether it is carried out manually or using simple algorithms. Knowledge graph provides state-of art prototype for representing text models[6].

Current works and research methodologies referenced in this paper use the earnings call transcripts to either develop extractive summarization framework or extract the polarity of phrases extracted from earnings call transcript. Existing frameworks on stock prediction fail to capture the ad hoc changes incurring in any organization such as management change, economic failure, major investments, research funding status etc., these factors act as prime indication in fostering stock projections. The proposed work captures these ad hoc changes in the form of sentiments and computes a numerical score which therein aids in the stock projection process as a prime parameter.

Contributions: The following are the objectives of the research which contribute to the process stock projection,

- Textual Data Pre-Processing to eliminate noise and irrelevant contents from the transcript.
- Design and Develop Autonomous Text Refinement framework to capture and screen the sentiments on pre-processed data to aid the prediction.
- Compute Sentiment Prediction Score and Develop Visualizations to illustrate the financial status of units.

Using a literature study, it was possible to effectively grasp the current state of the field of text rewriting and Natural Language Processing (NLP) research and controversies. It was found that various text processing techniques had shortcomings that would prevent the proposed research work from moving forward during the survey. Following a thorough analysis, it was determined that utilizing spaCy was the most profitable text processing technique. One of the biggest problems was feature extraction, which was restricted to Word2Vec and term frequency-inverse document frequency due to the survey's identification of several methods (TF-IDF). However, the suggested research work was better suited for TF-IDF. A detailed overview of the general processes required for the research work's effective implementation is provided by the literature review as follows,

Carta et al.[7] proposed a machine learning framework to combat a binary classification issue with the idea of estimating future stock price discrepancies for certain businesses. This method may be divided into four steps: lexicon creation, feature extraction, prediction algorithm, and model explanation. When compared to other classifiers considered to be state-of-the-art classifiers, as well as the best baseline mentioned in the literature to date, this technique takes advantage of being more accurate. The lack of compassion for the semantic richness of its content, which might give extra information to increase predicting accuracy, is a downside of this technique.

Cho et al.[8] examined a study that leverages machine learning methods for binary classification in conjunction with natural language processing on corporate research papers issued in the Korean financial market to investigate and affect an investing strategy. Using KoNLPy and Mecab, they inferred the portion of speech from the report. They adopted the k-NN technique, which is dependent on the idea that nearby data points with comparable characteristics will be found. They used gradient boosting, an ensemble model that combines weak classifiers, commonly decision trees, to build a robust classifier.

Cheng et al.[9] developed a deep learning-based framework for predicting whether enhancement reports will be acknowledged that considers the user's opinions expressed in the text. To eliminate the noise, they first preprocess the textual data in all improvement reports. The sentiment of each enhancement report is then calculated using Senti4SD. They learn the new deep representation of preprocessed text by coupling the bag-of-words (BOW) representation and conventional word2vec based representation, which they consolidate in the third step. Fourth, they produce a deep learning-based classifier for the acceptability prediction of enhancement reports based on sentiment and deep representation. This strategy has the benefit of estimating duplicate SE-reports, inaccurate categorization in SE-reports, and priority in SE-reports for upcoming analysis. This strategy's drawback is main obstacles with utilizing CNN for training involve overfitting, exploding gradients, and class variance.

Askerov et al.[10] enforced a model to see any discernible communication types of technology firms. This analysis targeted on 3 text analytics strategies ordinarily employed in text mining research: word2vec, bag-of-words and nostalgic analysis and uses Naive Bayes Classifier, SVM (Support Vector Machines), TF-IDF (Term Frequency-Inverse Document Frequency) classification. The advantage of this model is that it achieves higher prediction accuracy by victimization KNN classification. The disadvantage of this model is that the SVM formula isn't appropriate for big knowledge sets and data set having a lot of noise.

Jin et al.[11] proposed an LSTM-based model for stock price prediction using the Empirical Mode Decomposition (EMD) Model and the Convolutional Neural Network (CNN) Model. They first recommend using investor sentiments into stock prediction. Second, a platform that combines EMD to extract the trend term from the stock price sequence. Third, LSTM is used to forecast stock closing prices. The drawback of LSTM is that it takes a while to train and get ready for usage in real-world applications.

Mishev et al.[12] proposed a model Based on a combination of text representation techniques and machine-learning classifiers that we employ to examine the efficacy and performance of various sentiment analysis algorithms. The examination began with lexicons for sentiment analysis in finance; as a result, the study expanded to include word and sentence encoders. As compared to lexicons and fixed word and phrase encoders, the results experience an increase in the effectiveness of contextual embeddings in sentiment analysis. The benefit of this is that, despite the dataset's small size, we were able to get effective findings, indicating that this method is suitable for fields where huge, annotated data is not readily available. The disadvantage of this technique is that it determines document similarity solely in the word-count space, which may become stagnant for large vocabularies.

Abdi et al.[13] created a deep-learning-based technique to analyze a user's viewpoint presented in reviews (called RNSA). To take benefit of sequential processing and address several shortcomings in conventional approaches, the RNSA uses the Recurrent Neural Network (RNN), which is made of Long Short-Term Memory (LSTM). In compared to other well-known techniques in the literature, this model has the benefit of providing greater performance increases. This model's drawback has to do with how comparative and caustic sentences are handled.

Onan et al.[14] developed a two-stage framework for topic extraction from scientific literature. The conceptual approach enhances word embedding systems that use word vectors generated by the word2vec, POS2vec, word-position2vec, and LDA2vec schemes. An advanced clustering ensemble architecture is being used in the clustering phase, including traditional clustering techniques through iterative voting consensus. Jaccard coefficient, Folkes & Mallows measure, and F1 score all improved because of the results produced by the suggested framework. The drawback of the system is its limitation around extracting lexical tokens from compound statements.

Pratik et al.[15] developed method that combines channel separation, topic modeling and sentence selection with punctuation restoration to generate more readable call transcript summaries to provide a better understanding of customer concerns and agent recommended solutions. The proposed extractive summarizer in the paper claims to restore full punctuation to the summaries generated from either ill-punctuated or un-punctuated original call transcripts using a novel BERT transformer-based model. The paper illustrates a summarization framework capable enough to restore the sentence grammar but does not facilitate root word extraction or synonyms handling which is significant while clustering the extracted phrases.

Marchai et al.[16] developed method based on LSTM to predict the stock price of organization, the paper illustrates the importance of regular features and nature of train data in the forecasting framework. The paper illustrates Long-Short Term Memory (LSTM) techniques used to predict the closing price from five different companies. The framework proposed adopts ML framework with pre-defined feature set and fails to capture the influence of ad hoc changes such as management change, economic failure, major investments, research funding status etc., within an organization which have a high impact on stock projection.

Shen et al.[17] developed a framework based on deep learning for stock market price trend prediction with pre-defined features set. The paper concentrates on feature reduction and enhancement to improve the training of underlying ML framework. The paper concludes by stating the importance of diverse types of information such as tweets, news, and other text-based data in the prediction system and foresees these factors in its future enhancements.

The proposed research work has a text refinement pipeline to address the limitations around phrase extraction listed on existing frameworks such as handling compound statements, root word extraction by handling synonyms, handling comparative and caustic sentences, and handling large vocabularies with noise. Also, the proposed framework aids the stock prediction process with the aid of sentiment score feature which is computed from the classified sentiment clusters from the text refinement module.

2 Methodology

The proposed model employs an NLP framework to pre-process the Earnings Call Transcript by eliminating noise and irrelevant content and extract the sentiment phrase from the pre-processed data. The sentiment phrases extracted are clustered to aid in generating a sentiment prediction score based on weighted distribution which then facilitates in stock prediction for a particular organization.

System Architecture:

The earnings call transcript is initially processed to remove any irrelevant data or noise from the transcript. The data is obtained from the company's investor relation section or from third party sources. This transcript data will be in the form of HTML, .pdf, .txt, .docx etc. The HTML information can be cleaned using BeautifulSoup. Using spaCy we can further eliminate irrelevant data by parsing the data to identify named entities, parts-of-speech etc. This parsed data can be further cleaned using Gensim, using bi-grams to identify words that do not contribute to the analysis. Then using TF-IDF we can extract the sentiments that have higher value in the transcript and these sentiments can be classified using Loughran McDonald Sentiment Word Lists. These sentiments can be then overlaid with the share price movement of the company to understand the financial status by comparative analysis.

Figure 1 displays the architecture diagram of the proposed methodology used in the research work and it contains the following modules:

1. Data Collection and Preprocessing
2. Feature Extraction
3. Sentiment Analysis and Classification
4. Compute Sentiment Prediction Score and Visualization

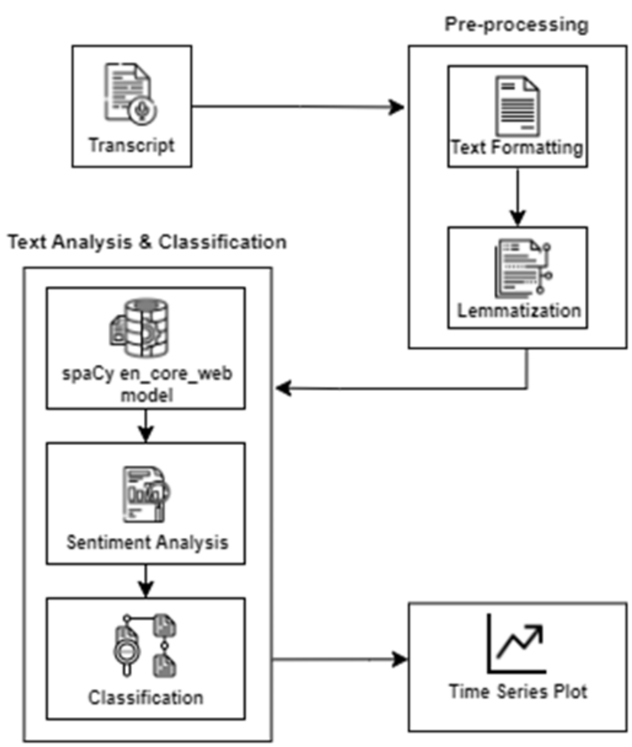

Fig 1. Architecture Diagram of Automated Text Refinement Framework

1) Data Collection and Preprocessing:

The information is available in HTML, text, PDF, and DOC forms. Using BeautifulSoup, one may sanitize the HTML data. For extracting data from HTML and XML files, use the Python module Beautiful Soup. The parse tree may be searched, navigated, and modified in idiomatic ways thanks to it and the parser. It does so by parsing the page's source code to produce a parse tree, which can then be used to extract data in a hierarchical and legible way. To identify named entities, parts of speech, etc., the data from the earnings call transcripts is processed. SpaCy, an open-source software library for sophisticated NLP, is used for this. Part-of-speech tagging, named entity identification, and other tasks are made easier by SpaCy's statistical models. POS tagging is the process of automatically giving each word in a phrase a specific tag. Additionally, SpaCy offers named entity recognition, a type of natural language processing that enables automatic article scanning, the extraction of some key textual components, and the classification of those components into predetermined categories. Stop-words, dates, and other irrelevant information are eliminated because they don't help with the search for new feelings. Stop words are frequently used words that can be eliminated from the text since they offer nothing to the analysis. These phrases have little or no significance.

2) Feature Extraction:

Further, the Gensim Library can be used to identify bi-grams and remove non-relevant words. Gensim is a Python package for big corpus similarity retrieval, document indexing, and topic modelling. The information retrieval (IR) and natural language processing (NLP) communities are the intended audiences. Gensim is designed to extract semantic topics from documents. Gensim also provides efficient implementations for increase processing speed. Using TF-IDF, pertinent features are retrieved that support the analysis. Term Frequency Inverse Document Frequency of records is referred to as TF-IDF. It may be summed up as determining how pertinent a word is to a corpus or sequence of words in a text. The way word frequency works is by counting how often a certain phrase appears in the document. How frequently (or infrequently) a term appears in the corpus is examined via inverse document frequency. With the help of the TF-IDF weighting scheme, each word in a document is given a weight based on its term frequency (tf) and the reciprocal document frequency (tf) (idf). The words with higher weight ratings are seen to be more important.

3) Sentiment Analysis and Classification:

The sentiments are then classified from the transcript, this can be done using Loughran McDonald Sentiment Word Lists. It is a dictionary used to determine which tokens (collections of characters) are classified as words. For maintaining consistency in word counts, the dictionary provides a method of identifying which tokens (collections of letters) are genuine words. Loughran - McDonald is a dictionary approach, which is used to determine whether data is positive, negative, or neutral. The proposed framework fosters extraction of 6 categories of sentiment phases (positive, negative, interesting, litigious, constraining, uncertainty).

4)Compute Sentiment Prediction Score and Visualization:

The sentiment phrases extracted and classified aids in generating a prediction score which facilitates in stock prediction. The sentiments score from the earnings call can be overlaid with the related share price movement to understand the financial status of that organization. Prediction score is calculated using the related ratio of classified sentiment phrases as illustrated in Equation (1).

Calculation of Sentiment Prediction Score:

V1 = count(positive)

V2 = count(negative)

V3 = count(interesting)

V4 = count(litigious)

V5 = count(constraining)

V6 = count(uncertainty)

T = total phases extracted

$$sentiment\ score = ((V1 + 0.5 * V3) - (V2 + 0.5 * V4 + 0.5 * V5))/(T - V6) \quad (1)$$

Positive and Negative phrases impose a high impact on stock prediction hence have higher weights when computing the sentiment prediction score.

Matplotlib is used to visualize the approach's results. A tool for visualizing data, Matplotlib is a low-level graph charting framework written in Python. There are several plot types available in Matplotlib. Plots aid in making connections between trends, patterns, and the financial health of an organization.

Pseudocode

Step 1: Pre-Processing

Extract Data: Convert HTML and XML earnings call input data into text data using python's Beautiful Soup package

Deduce to Text Format: Convert rtf, docx, pdf files to text format

Step 2: Feature Extraction

Lemmatization: Generate root work using lemmatization technique

Handling Bi-Grams: The bi-grams are expelled from the data

De-Noising: Elimination of stop words and noisy characters from data

Step 3: Sentiment Classification

Classification: Loughran McDonald Sentiment Word Lists based classification into classes as Negative, Positive, Uncertainty, Litigious, Constraining, and Interesting

Step 4: Compute Sentiment Score and Visualization

Compute Sentiment Score: Sentiment prediction score is calculated using the related ratio of classified sentiment phrases

Visualize the Results: Sentiments are then charted in matplotlib to see the variation across various calls, these recognized sentiments are overlaid with stock prices to observe stock price movement.

The proposed framework is evaluated against the existing stock prediction frameworks build on ML models like LSTM, Random Forest, ARIMA and Regression models which are trained based on historic prediction datasets. The proposed work yields stock projections aided with the sentiment score feature thereby experiencing the enhanced results. The earnings call transcripts used in testing and training the model are acquired from SPGISpeech corpus. The proposed framework computes the sentiment score based on cluster derived using Loughran McDonald Sentiment Word Lists, such as Negative, Positive, Uncertainty, Litigious, Constraining, and Interesting, which foster in computation of sentiment score in a weighted distribution. The sentiment score then aids in the process of stock projections as a prime feature.

3 Results and Discussion

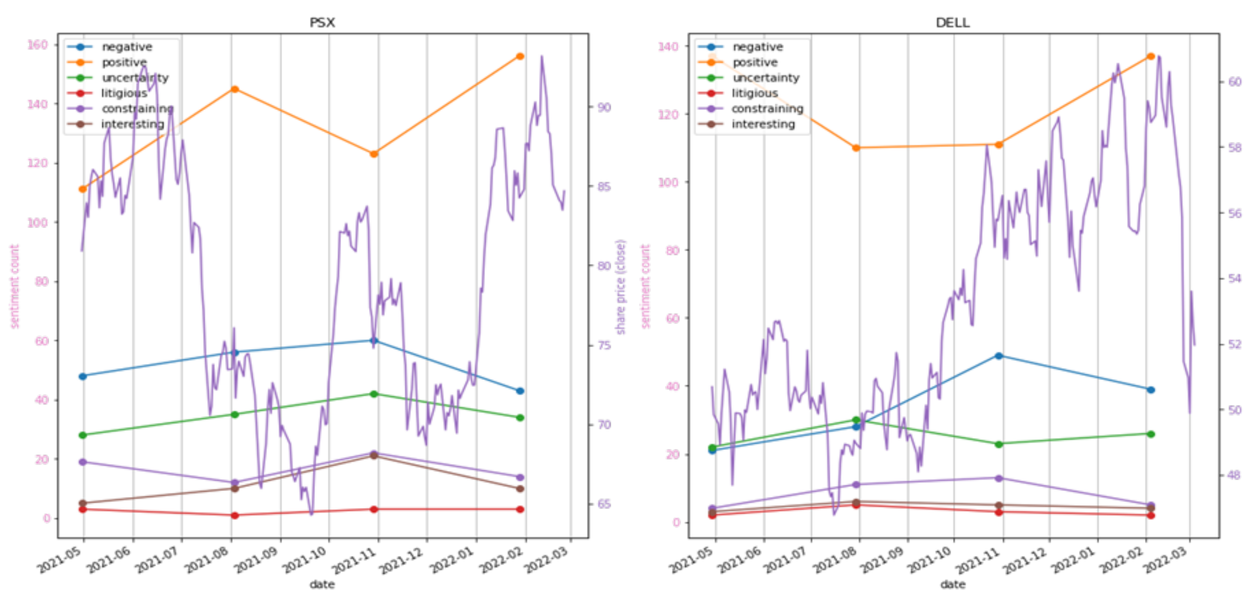

Fig 2. Sentiment graph imposed with the stock prices of Phillips and Dell

To evaluate the organization's financial situation, the sentiments expressed during the earnings call can be merged with the relevant share price movement. Matplotlib is used to visualize the method's findings. A tool for visualizing the data, Python-based Matplotlib is a low-level framework for graph visualization. There are many different graphs available with Matplotlib. Plots aid in making connections between trends, patterns, and the financial well being of an organization.

The graphs that forecast the sentiment classification from the transcript data are shown in Figures 2 and 3, which demonstrates that for both businesses, the stock prices that have been overlaid indicate a progressive rise over time as the proportion of positive feelings rises and the proportion of negative sentiments falls.

The proposed methodology has an accuracy sore of >88% as shown in Figure 4.

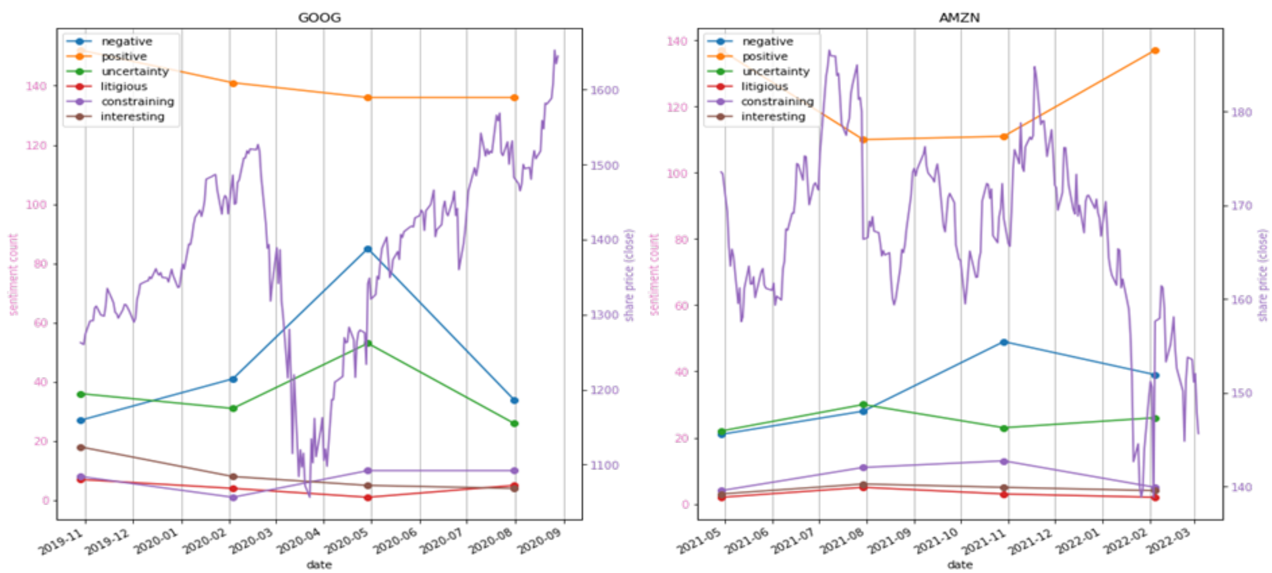

Fig 3. Sentiment graph imposed with the stock prices of Google and Amazon

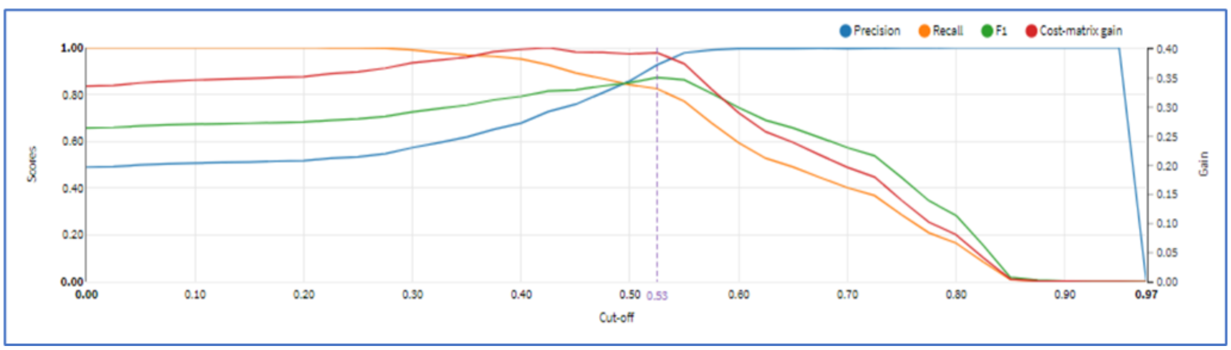

Fig 4. Evaluation Metrics of Proposed methodology

Comparative Analysis of Stock Prediction frameworks is obtained by comparing the prediction of stock for the organizations using LSTM framework regular features and the proposed frameworks sentiment prediction score in this paper,

For Model Training and Testing the following configuration is employed.

Session 1: Trained with 2019 year Data and Evaluated on 2020 year Data

Session 2: Trained with 2019 and 2020 year Data and Evaluated on 2021 year Data

Session 3: Trained with 2019, 2020 and 2021 year Data and Evaluated on 2022 year Data

Table 1. Evaluation Metrics LSTM vs Proposed Framework

Accuracy Score for Google			
Model	**Session 1**	**Session 2**	**Session 3**
LSTM Framework	0.735	0.747	0.746
Proposed Framework	0.781	0.782	0.735
Accuracy Score for Phillips			
Model	**Session 1**	**Session 2**	**Session 3**
LSTM Framework	0.725	0.737	0.736
Proposed Framework	0.751	0.721	0.725
Accuracy Score for Dell			
Model	**Session 1**	**Session 2**	**Session 3**

Continued on next page

Table 1 continued

LSTM Framework	0.701	0.712	0.718
Proposed Framework	0.743	0.741	0.754
Accuracy Score for Amazon			
Model	**Session 1**	**Session 2**	**Session 3**
LSTM Framework	0.735	0.733	0.714
Proposed Framework	0.744	0.764	0.791

As illustrated in Table 1, the accuracy of proposed framework is improved as compared to existing LSTM framework[16], The difference can be seen with respect to organization "Amazon" in the session 3 where proposed framework is relatively more accurate than the LSTM which indicates the importance of weighted score of earnings call transcript in forecasting the stock price.

Table 2. Evaluation Metrics ML Model with and without Sentiment Score from Proposed Framework

Model	Accuracy	Precision	Recall
Random Forest	0.83	0.82	0.81
ARIMA	0.91	0.92	0.93
Multiple Regression	0.94	0.95	0.93
LSTM	0.92	0.91	0.93
ML model aided with Sentiment Score Feature from Proposed Framework	0.93	0.96	0.95

As illustrated in Table 2, the metrics fostered by proposed framework is improved as compared to other ML frameworks such as Random Forest, ARIMA Models, Regression frameworks and LSTM framework. Sentiment score generated from the proposed framework has high impact in yielding the results of prediction framework. The proposed framework sees a significant improvement in accuracy, precision, and recall compared to the existing frameworks as shown in Table 2.

Table 3. Proposed Model Comparison with existing LSTM Model for Stock Prediction

Metrics	Existing LSTM Framework	ML Model with Proposed Framework
Loss	0.0848	0.087
F1 score	0.9194	0.93
Binary Accuracy	0.9193	0.93
MSE	0.0772	0.0792
MAE	0.0848	0.0878
TPR	0.92	0.91
TNR	0.91	0.92
FPR	0.08	0.10
FNR	0.09	0.07

The accuracy of ML and Deep Learning models is aided and dependent on the features and size of train data fed to the model to greater extent as compared to sentiment score-based model which takes the weighted score of extracted sentiments to predict the stock price. Table 3 illustrates the comparison of proposed work with existing LSTM ML framework[16], The evaluation metrics illustrated in Table 3 clearly indicate the significant improvements of the proposed framework which is implement with the aid of sentiment score compared to the existing LSTM ML framework trained using only pre-defined feature set.

4 Conclusion

Text Refinement framework can be built using several techniques, the proposed research work employs a text refinement framework to cleanse the textual data by eliminating stop words and special characters and fosters extraction of sentiment phrases on the intermediate cleansed content, the phrases extracted are clustered using Loughran McDonald Sentiment Word Lists into 6 sentiment tokens, Negative, Positive, Uncertainty, Litigious, Constraining, and Interesting. A numeric score is computed on the extracted sentiment phrases based on weighted distribution. The sentiment score then aids the process of stock projection as a prime feature. The proposed framework sees significant enhancement in the evaluation metrics because of its ability to capture and extract the ad hoc changes in the organization such as management change, investments, research capabilities and future goals as sentiments from the earnings call transcripts and liable them as sentiment score fostering the stock projection.

The proposed work has an accuracy score of 93%, precision 96% and recall 95% which is comparatively better than existing ML frameworks framed on LSTM, Random Forest, ARIMA and Regression models. Also, the underlying text refinement framework employed for sentiment extraction adopts Term Frequency-Inverse Document Frequency (TF-IDF) model from Scikit-Learn which is best suited to handle enormous set of data volume as compared to Word2Vec which is employed in most of the existing frameworks of phrase extraction. The ML and Deep Neural Network based models are influenced by the feature scores and size of train data, which would fail in capturing the immediate socio-economic effects fostered on the organization, and these socio-economic factors discussed in earnings call transcript and extracted as sentiment phase can aid in highly accurate prediction model.

As an enhancement to the use cases around stock price prediction, a framework can be experimented which takes both the ML as well as sentiment score-based framework parameters with micro weighted clusters to design and develop a model to forecast the stock price. Such models can effectively capture the regular feature importance score and the sentiment score to aid in stock prediction with enhanced accuracy and precision. The framework can be further enhanced by improving the underlying training process by employing feature reduction on the collective feature set. The pre-processing stage can be enhanced by employing container-based execution strategies to parallelly process multiple input documents, thereby reducing the generation time of intermediate information rich content.

References

1) Bouktif S, Fiaz A, Awad M. Augmented textual features-based stock market prediction. *IEEE Access*. 2020;8:40269–40282. Available from: https://doi.org/10.1109/ACCESS.2020.2976725.
2) Ongsakul V, Chatjuthamard P, Jiraporn P, Chaivisuttangkun S. Corporate integrity and hostile takeover threats: Evidence from machine learning and "CEO luck". *Journal of Behavioral and Experimental Finance*. 2021;32. Available from: https://doi.org/10.1016/j.jbef.2021.100579.
3) Renault T. Sentiment analysis and machine learning in finance: a comparison of methods and models on one million messages. *Digital Finance*. 2020;2(1-2):1–13. Available from: https://doi.org/10.1007/s42521-019-00014-x.
4) Pakkala R, Rai P, Bellipady SR. Impact of Syntactical and Statistical Pattern Recognition on Prognostic Reasoning. In: Handbook of Research on Machine Learning Techniques for Pattern Recognition and Information Security. IGI Global. 2021;p. 38–55. Available from: https://doi.org/10.4018/978-1-7998-3299-7.ch003.
5) Ma Z, Bang G, Wang C, Liu X. Towards Earnings Call and Stock Price Movement. . Available from: https://doi.org/10.48550/arXiv.2009.01317.
6) Wang Y, Zhang H, Shi G, Liu Z, Zhou Q. A model of text-enhanced knowledge graph representation learning with mutual attention. *IEEE Access*. 2020;8:52895–52905. Available from: https://doi.org/10.1109/ACCESS.2020.2981212.
7) Carta SM, Consoli S, Piras L, Podda AS, Recupero DR. Explainable machine learning exploiting news and domain-specific lexicon for stock market forecasting. *IEEE Access*. 2021;9:30193–30205. Available from: https://doi.org/10.1109/ACCESS.2021.3059960.
8) Cho P, Park JH, Song JW. Equity research report-driven investment strategy in Korea using binary classification on stock price direction. *IEEE Access*. 2021;9:46364–46373. Available from: https://doi.org/10.1109/ACCESS.2021.3067691.
9) Cheng J, Sadiq M, Kalugina OA, Nafees SA, Umer Q. Convolutional Neural Network Based Approval Prediction of Enhancement Reports. *IEEE Access*. 2021;9:122412–122424. Available from: https://doi.org/10.1109/ACCESS.2021.3108624.
10) Askerov R, Kwon E, Song LM, Weber D, Schaer O, Dadgostari F, et al. Natural Language Processing for Company Financial Communication Style. In: 2020 Systems and Information Engineering Design Symposium (SIEDS). IEEE. 2020;p. 1–6. Available from: https://doi.org/10.1109/SIEDS49339.2020.9106636.
11) Jin Z, Yang Y, Liu Y. Stock closing price prediction based on sentiment analysis and LSTM. *Neural Computing and Applications*. 2020;32:9713–9729. Available from: https://doi.org/10.1007/s00521-019-04504-2.
12) Mishev K, Gjorgjevikj A, Vodenska I, Chitkushev LT, Trajanov D. Evaluation of sentiment analysis in finance: from lexicons to transformers. *IEEE access*. 2020;8:131662–131682. Available from: https://doi.org/10.1109/ACCESS.2020.3009626.
13) Abdi A, Shamsuddin SM, Hasan S, Piran J. Deep learning-based sentiment classification of evaluative text based on Multi-feature fusion. *Information Processing & Management*. 2019;56(4):1245–1259. Available from: https://doi.org/10.1016/j.ipm.2019.02.018.
14) Onan A. Two-stage topic extraction model for bibliometric data analysis based on word embeddings and clustering. *IEEE Access*. 2019;7:145614–145633. Available from: https://doi.org/10.1109/ACCESS.2019.2945911.
15) Biswas PK, Iakubovich A. Extractive Summarization of Call Transcripts. *IEEE Access*. 2022;10:119826–119840. Available from: https://doi.org/10.1109/ACCESS.2022.3221404.
16) Marchai FL, Martin W, Suhartono D. Stock Prices Prediction Using Machine Learning. In: 2021 8th International Conference on Information Technology, Computer and Electrical Engineering (ICITACEE). IEEE. 2021;p. 79–84. Available from: https://doi.org/10.1109/ICITACEE53184.2021.9617222.
17) Shen J, Shafiq MO. Short-term stock market price trend prediction using a comprehensive deep learning system. *Journal of Big Data*. 2020;7:1–33. Available from: https://doi.org/10.1186/s40537-020-00333-6.