Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 235 (2024) 820–829

Procedia Computer Science

www.elsevier.com/locate/procedia

International Conference on Machine Learning and Data Engineering (ICMLDE 2023)

Semantic Analysis Using Deep Learning for Predicting Stock Trends

Manisha Galphade[a], V. B. Nikam[b], Dhanalekshmi Yedurkar[c], Prabhishek Singh[d], Thompson Stephan[e,*]

[a]*Department of Computer Science & Engineering, School of Computing, MIT Art Design & Technology University, Pune, India*
[b]*Department of Computer Engineering & IT, Veermata Jijabai Technological Institute, Mumbai, India*
[c]*Department of Computer Science & Engineering, School of Computing, MIT Art Design & Technology University, Pune, India*
[d]*School of Computer Science Engineering and Technology, Bennett University, Greater Noida, India*
[e]*Department of Computer Science & Engineering, Graphic Era Deemed to be University, Dehradun, Uttarakhand, India*

Abstract

Company Investors and financial professionals mostly rely on quarterly reports to help them decide the ways to invest in stocks and assess the company's current performance. Quarterly company reports offer an abstracted perspective of the company's overall past performance, as well as its present situation and the market value of its market share. Financial text streams in quarterly report are unstructured naturally, but they represent cooperative expressions that are of important in any financial decision for stake holder. It will be both daunting and necessary to procedure intelligence of unstructured textual data. In this study, we address important queries related with the explosion of interest in a method to extract useful information from unstructured data and the way to work out if such insight provides any hints regarding the trends of financial markets. There is a lack of availability in the labeled dataset for financial sentiment analysis applications. The pre-trained language model employs very little labeled parameters that is used for a variety of domain specific corpora including financial sentiment analysis. In this paper, FinBERT, a model built on the BERT framework, to address linguistics challenges in the financial domain. The proposed work uses twelve transformer layers and twelve attention layers with several million parameters. The design of encoder and decoder comprises of several attention layers along with RNN. This arrangement aids to recognize instances processing the strongest relation between the words within a particular sentence. The experimentation results shows that the presented method surpass the state-of-the-art methods for financial datasets. The results are also compared with other existing models using the same financial dataset. It is observed that the FinBERT attains an accuracy of 84.77% on quarterly reports despite using a lesser training set.

Keywords: Deep learning; emotion mining; financial analysts; investors; sentiment analysis.

* Corresponding author.
E-mail address: thompsoncse@gmail.com

1877-0509
10.1016/j.procs.2024.04.078

1. Introduction

All of the data about items that are traded in a marketplace is reflected in prices in a free-trade environment. It is impossible to continually beat the markets since every economic actor updates their positions whenever fresh data becomes available, and prices change to reflect this. However, if improved techniques for retrieving it become readily accessible, the concept of ”new information” may alter, and early acceptance of such technologies may offer a short-term benefit.

Analyzing financial documents, such as news articles, expert reports, and formal corporate statements, has become an increasingly daunting task due to the vast volume of information generated daily. This sheer magnitude makes manual analysis and drawing meaningful insights practically impossible for any single entity. Consequently, the adoption of computerized sentiment analysis using natural language processing (NLP) methods by financial institutions and professionals has surged over the past decade. The business operations and financial information of a corporation are communicated through financial statements. Investors, financiers, and market professionals evaluate the company and reach the necessary judgments using financial reports and declarations. The report for the prior quarter includes language and numeric breakdowns of the financial data. About 20% of the numerical information is made up of the record, the summary of profit and loss, and the report of money flows. The remaining 80% of the quarterly reports are financial textual information of a textual nature and consist of statements, notifications from executive management, plans, information about the leadership team, information about the shareholders, and numerous reports, such as organization reports, sustainability reports, business governance statements, and reports. These low-cost textual aspects are equally important to the numerical data and give readers of quarterly reports knowledge they can use to make informed decisions. These facts provide specifics on how organizational behavior is affected by the suggestion of behavioral strategies.

Organizations utilize quarterly reports to communicate their past, present, strategy, and goals for the future. Quarterly reports include both verbal and numeric data. The vast majority of textual information in quarterly reports is intricate. Companies cover up bad performance and negative press publicity by hiding their quarterly reports using industry-related trends reports, which results in biases in reporting. Despite being unstructured, textual material in financial reports often contains combined expressions that are crucial for financial decision-making. In this study, we discuss important issues related to the attention explosion to extract insight information from unstructured data and to determine whether such insight offers any cues about the direction of the financial markets. A predicted sentiment analysis model makes use of grammars provided by linguistic research. This model broadens sentiment analysis to include financial phrases within each sentence in addition to term tokens. The joint concepts shown in the texts are abstracted using a valuation heuristic. Additionally, evaluations are done to judge how well the model is working.

Deep learning approaches are now enhancing effectiveness while offering more effective solutions considering to the vast amounts of data and graphics processing units. In diverse applications which includes healthcare, speech recognition,identifying fraud, and object detection, deep learning show-case improvement in performance when compared to machine learning techniques. To prove the effectiveness of the FinBERT model in comparison with state-of-the-art methods the experimental outcomes using various deep learning techniques are explored.[1].

The main contributions of the presented research is listed as follows

(i) A novel hybrid approach based on pre-trained language model is proposed;

(ii) The presented work highlighted improvement in performance for pre-trained language model in financial domain;

(iii) The current investigation will help traders to make decisions based on the suggestions of the presented model.

The rest of the paper is summarised as follows: Section II highlights elementary backgrounds about sentiment analysis and existing systems. Section III presents the proposed framework using deep learning model. Section IV reports validation results of the proposed method. At last Section V concludes the paper.

2. Review of Literature

In [2] the authors extracted future related sentences from financial reports written in Japanese language. The published work extracts future-expressed sentences using machine learning techniques. The study also provided evidence that all the evaluation metrics for this approach averaged around 0.9, suggesting that their model demonstrated high performance. Furthermore, the combination of two SVM models proved to be effective in efficiently extracting sentences with the desired distinctive features[2].

In this research, various feature extraction methods and classifiers are investigated for the classification of business text reviews. The study utilizes a substantial dataset, namely the Yelp challenge dataset, which encompasses more than 1.6 million reviews. Notably, the Linear SVC and SGD classifiers achieved an impressive accuracy of 94.4% when employing the initial feature extraction approach. In terms of performance, the Naïve Bayes and Logistic Regression classifiers yielded slightly lower results. [3].

This study suggests a quick text sentiment analysis technique that is backed by a bi-level attention model and doesn't rely on manual features or outside information. To lessen the effects of knowledge sparsity, our model combines a spotlight mechanism and a neural topic model. By including implicit topic knowledge into word-level semantic representation, our approach at the word level broadens the outcome of word representation. A new topic-word attention method that derives word semantics from topic attitude is proposed. The connection amongst local and global emotion expression is captured at the sequence level via an attention method. Studies on the datasets like NLPCC-ECGC-16 and the ChnSentiCorp-Htl-ba-10000 ensures the efficiency of the BAM model [4].

In the work published in [5], the researchers utilised bidirectional contextualized transformer language models like BERT and XLNet, for the classification of sentiment. The outcome of this work showcased better performance in the cross-domain sentiment classification. During analysis it is found that XLNet outperforms BERT network. It is worth-while to note that XLNet properly capture the context of the sentence by utilizing few fine-tuning training samples.

This study suggests a semantic analysis-based strategy for classifying the sentiment of internet reviews. The subject vectors of online review text documents are determined using the LDA approach, assigning a distinct topic vector to each review text document. To capture the meaning of phrases within the analyzed content and the emotional expressions from the sentiment dictionary, the Word2Vec method is employed for training a word vector model. The phrases in the text of the review document's semantic similarity to the sentimental terms in the sentimental dictionary as well as the semantic correspondence among the phrases used in the assessment of text document's subject vector are calculated. To improve the quality of selecting text features, enhance the capacity for classifying feature words, and boost the effectiveness of sentiment categorization in reviews, it is achievable to establish semantic links between words by employing the feature selection techniques based on the two semantic equivalence calculation methods mentioned earlier.

In order to increase the quality of text feature selection, advance the capacity for classification of feature words, and enhance the sentiment categorization effect of reviews, it is therefore possible to acquire the semantic link between words by implementing the techniques for choosing features based on the two semantic equivalent calculation methods mentioned above [6].

The strategy for creating a single vulnerability knowledge base that has been provided in this research allows us to integrate and reconcile diverse susceptibility information gathered from several SVDBs. In order to assist information designers in interpreting and repurposing ideas throughout system level ontologies, as well as to improve knowledge integration and reuse, the knowledge modelling approach makes use of FCA. By using the unified representation to validate the consistency of the data in these SVDBs and to improve the current SVDB knowledge assets by connecting them to other resources, we used two instances to demonstrate the practical value of our knowledge model [7].

This paper introduced an evaluation model for sentiment analysis in commodities using BERT-CNN. The problem of a single CNN model, specifically ignoring context-dependent semantic interactions in the review text, was resolved by combining features obtained by the two supplementary models. Additionally, characteristics extracted by the two supplementary algorithms were combined to successfully dig linguistic connections among brief phrases in long texts. The overall performance of the model is evaluated using matrix such as F1 value and accuracy. The dataset used is samples obtained from the mobile phone in real time on JD Mall. The experimentation show-cased improvement in performance for the BERT-CNN model [8].

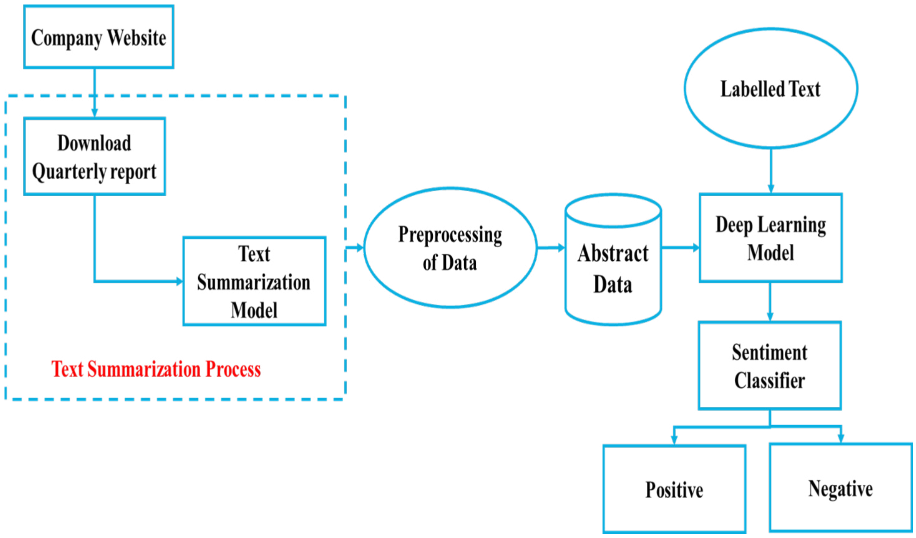

Fig. 1. System Architecture.

2.1. Existing System

Researchers have been investigating text mining and using it in the financial industry since the 1960s. Because they were created to identify structured data and produce assessment report backed data, knowledge mining techniques heavily rely on structured data. In order to qualitatively illuminate the outcomes of the text mining process, numerous academics have delved deeper into the topic of text mining and created numerous little algorithms. However, processing unstructured data will not produce the desired results, making it impossible for investors and stockholders to place their trust in this study. Additionally, data processing techniques do not forecast the results of a supported current report study [8], [9], [10].

In order to combat the issue of studying unstructured data, several new technologies and deep learning techniques have been put into the market. Quarterly reports now contain a significant amount of unstructured data. Because of the data processing technique, several crucial pieces of information lack effectively assessed [11]. It leads to incorrect analysis of the quarterly report because certain investors and stockholders don't rely on it. It has been challenging to figure out with an unstructured data in data processing, since most of the data is out there in the market [12]. However before performing data processing it requires to be in a very structured format, standard tests with economical data sets can assist in assessing distinct models and methodology. Nevertheless they will never disclose people which method performs well on them on a selected customer data set [13], [14], [15].

3. Proposed Methodology

The working of proposed system is divided into two parts first one is summarization of quarterly report which will be considered as an input to our system. The second component of this system involves the utilization of the FinBERT model. This model is designed to conduct sentiment analysis on the input data generated by the summarization model. The output produced by the FinBERT model will be used to predict the future trend of the company.

As depicted in Figure 1, the functioning of our system begins with the input data, which comprises quarterly reports gathered from company websites [16]. These reports are stored within a designated folder. When a specific company name is selected from our list, the system proceeds to select and process the corresponding quarterly report associated with that chosen company [17]. Once the quarterly report has been successfully obtained, it is then passed through a summarization model. This model utilizes various Python modules and incorporates transformer layers to generate an extractive summary of the PDF file. This summary serves as a concise representation of the report's contents, providing a brief overview of its key information. For each BERT encoder, there exists a preprocessing

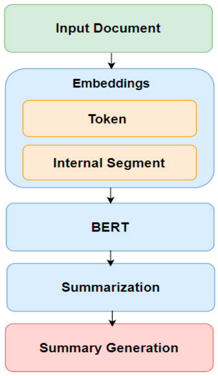

Fig. 2. Summarizing model.

model. These models leverage TensorFlow operators from the TF.text package to transform raw text into numeric input tensors that are compatible with the respective encoder. Unlike conventional Python preprocessing, these operations can seamlessly integrate into a TensorFlow model. This allows for direct processing of text inputs and subsequent model output generation. The preprocessed data is fed into the FinBERT model [18], a more advanced iteration of the BERT model. This specialized model has been pre-trained on financial datasets, which enhances its understanding of financial language and context. FinBERT employs a labeling technique, utilizing labeled phrases during its training process. These phrases are annotated with sentiments like negative, positive, or neutral as shown in Table 1. Following its training, when input data is provided to the FinBERT model, it conducts sentiment analysis. Based on its analysis, it generates an output predicting whether the stock's trend is likely to go up or down [19], [20]. Detailed explanations of these two steps are provided in the following section.

3.1. Summerization

Extractive summarization is most difficult task as it create the abstract view of pdf without replacing or missing any word from pdf file. This work uses the BERT model normalization layer to create the extractive summarization model. This aligns with our understanding that a proficient summarizer possesses the capability to comprehend meaning within the text. It can then selectively pick sentences that correspond to the constructed abstract representation of the PDF. This representation is constructed solely based on the report's internal structure. To practice BERT model for extractive summarization, we need it to produce the illustration for each sentence on every page of pdf. The resulting vectors of BERT are arranged inline with the respective tokens in spite of the network being trained as a masked language model. Although BERT uses segmentation embedding to represent many phrases, it only uses two labels per segmentation (sentence A or sentence B), as opposed to the multiple sentences used in extractive summarization. Therefore, in order to make it feasible to extract summaries from a quarterly report PDF file, we adjust the input sequence and embedding of BERT. Figure 2 shows the summarizing model utilized in this work. For a numerous set of sentences sentence1, sentence2 . . . ., sentence n where two probability values such as 0 and 1 are considered. The probability value 1 specifies the belongingness of a particular sentence to be selected, where as probability value 0 denotes for unselected category. In comparison to other existing models, the presented enhanced embedding model generates text summarizations for PDF files containing multiple sentences.

- **Token embeddings** encompass pre-trained embeddings for a diverse range of words. These pre-trained token embeddings are fashioned using a method known as WordPiece tokenization for text tokenization. WordPiece tokenization is a data-driven approach designed to find an optimal balance between vocabulary size and the ability to handle out-of-vocabulary words. WordPiece functions as a subword segmentation algorithm, starting with individual characters and expanding the vocabulary by combining these individual elements.
- **Segment Embedding -** This technique is used to distinguish input data with diverse characteristics by utilising binary coding [21], , [22]. Say if input1 is "I love books" and input2 is "I love sports", the token embed-

ding technique generates the outcomes as "[CLS],I,love,books,[SEP],I,love,sports" and segment embedding produces results as "[0, 0, 0, 0, 0, 1, 1, 1]" and Input1=0 , Input2=1.[23].

- **Position Embeddings-** One of the widely used position embedding approach is BERT. It can process input sequences up to 512 tokens which results in vector dimensions of (512, 768). Positional embedding is employed to account for the fact that a word's position within a sentence can influence the sentence's contextual meaning and should not share the same vector representation. As an example, "We didn't play, however we were spectating" [24]. In this sentence "we" must not have same vector representations [23], [25].
- **BERT-** The BERT base model comprises of twelve transformer layers, twelve attention layers and several million parameters. A Transformer layer is a fusion of a encoder and decoder layers in which encoder is equipped with RNN and attention layers, and the decoder consists of additional attention layer between them, same as sentence-to-sentence model. This design aids in identifying and emphasizing the significance of important words.
- **Summerization layer -** Self-attention layer is the crucial architectural change that differentiates between the BERT model and the RNN. In the summarization model, the objective is to identify and emphasize the most robust connections among words, thereby aiding in representation. To capture specific summarization features, the BERT model output is combined with an LSTM model layer, where each LSTM cell is normalized accordingly.

3.2. FinBERT model

BERT proved to be an ideal choice for our financial sentiment analysis task. Even with a relatively limited dataset, we could leverage the capabilities of state-of-the-art NLP models. However, given the distinct nature of our domain – finance, which significantly differs from the general-purpose corpus on which BERT was originally trained, we deemed it necessary to introduce an additional step before embarking on sentiment analysis. Our approach involved taking the pre-trained BERT model and subjecting it to further training on a specialized financial corpus, specifically Reuters TRC2. This corpus is available upon request. The primary objective was to enhance domain adaptation by acquainting the model with financial jargon before fine-tuning it for the specific task at hand. We conducted this additional pre-training of BERT by utilizing the outstanding transformers library from Hugging Face, which was referred to as PyTorch-pretrained-BERT at the time. Subsequently, we fine-tuned the language model for our task by adding one or more task-specific layers [26], [27]. The merit of this approach is it requires no fine tuning for an extensive dataset since the model inculcates the basic understanding of a language obtained during the initial training process.

Once we had the pre-trained and domain-adapted language model in place, we proceeded with the subsequent steps to fine-tune it using labeled financial statement data for the financial sentiment classification process. For this purpose, we utilized the Economic Phrase Bank dataset, which is a meticulously curated and extensively labeled resource, albeit somewhat limited in size. This dataset was created by extracting 700,000 sentences from a variety of news articles containing financial terms with associated labels. The process of labeling was a collaborative effort, involving financial experts and master's students with backgrounds in finance. They not only assigned labels to the sentences but also determined the level of inter-annotator agreement for each sentence. This agreement level indicates the extent to which experts agreed on categorizing sentences as positive, neutral, or negative in sentiment.

FinBERT could also be a language model supported BERT. It additional trains the BERT model for financial data. And further training corpus can be a group of 1.8M Reuters' news articles and Economic PhraseBank. The prominent dataset for sentiment analysis is the Financial PhraseBank, encompassing 700,000 English sentences randomly sourced from financial news on the Kaggle database. Subsequently, these sentences were evaluated by a panel of 16 individuals with expertise in finance and business. FinBERT sentiment analysis model is accessible on Hugging Face model hub. The architecture is described in Figure 3 and Figure 4. The overall flow of the FinBERT model is shown in Figure 5.

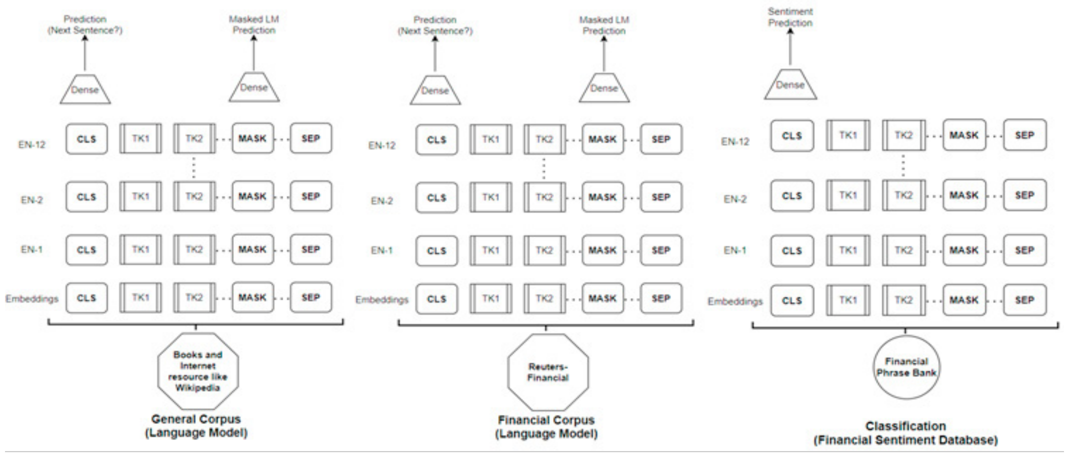

Fig. 3. Architecture of FinBERT model.

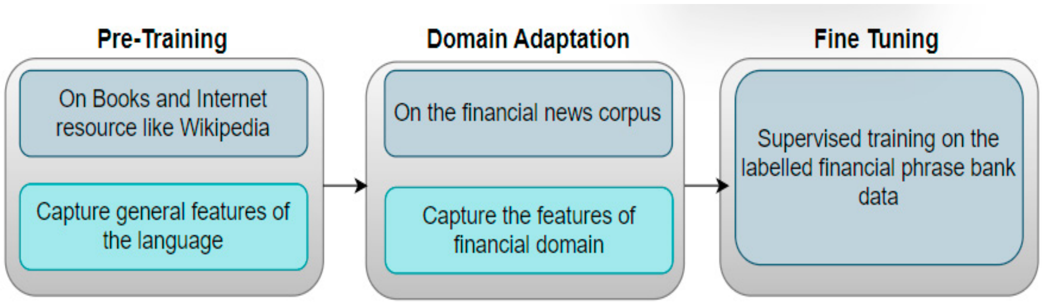

Fig. 4. Training Steps in FinBERT.

4. Results and Analysis

4.1. Dataset

We will be working with two datasets: the first comprises company quarterly reports, which serve as input for our summarization model. The output generated by the summarization model is then utilized as input for our FinBERT model. Due to the substantial volume of unstructured data typically found in quarterly reports, and the varying sizes of PDF files across different companies, it is not feasible to directly input these PDF files into the model for sentiment analysis. Therefore, it is imperative to preprocess the quarterly report data, summarizing it to achieve a well-structured format suitable for conducting sentiment analysis with FinBERT. This processed data is saved as $FinBERT_{c}lean.csv$ for subsequent Sentiment Analysis. The dataset consists of 7 Lakh sentences, each labeled with a sentiment score. To enhance the accuracy of the FinBERT model, it is imperative to train it with an appropriate dataset. Therefore, we have divided the dataset into three segments: training, testing, and validation. Approximately 80% of the dataset is allocated for training the model, while the remaining 10% is dedicated to testing its performance. An additional 10% of the data is reserved for the validation of the model. This division allows for comprehensive training, evaluation, and validation of the FinBERT model's accuracy.

4.2. Validation And Accuracy

In the presented work specific hyperparameters are used which includes dropout probability ($p = 0.1$), warm-up percentage (0.2), maximum sequence length (64), learning rate ($2e^5$) and a minibatch (size = 64). In order to choose the best model, the presented network trained for 5 epochs there by evaluating the performance periodically. The

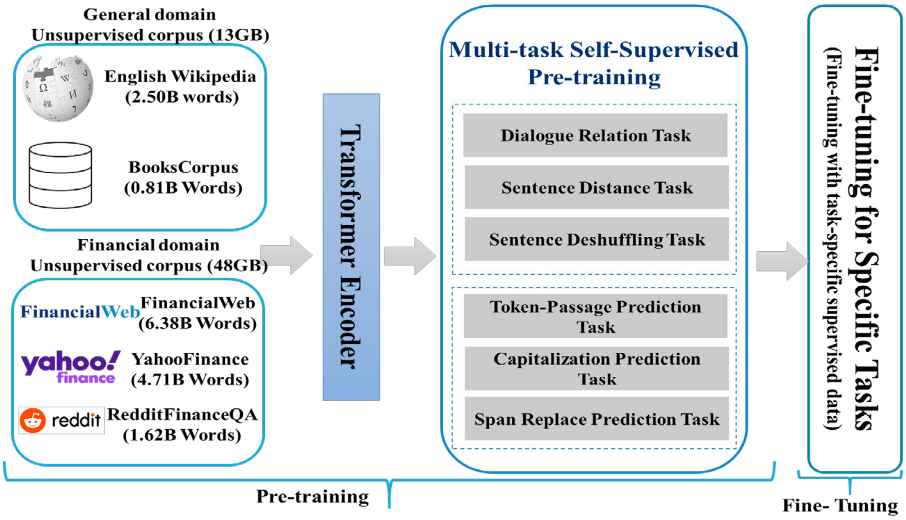

Fig. 5. Flow of FinBERT model.

Table 1. Training Dataset.

Sentences	Labelled	Score
ari network reports q2 eps 001 vs 003 est sales 132m vs 134m est.	Negative	-0.623
ubs maintains neutral on agilent technologies raises price target to 90.	Positive	0.438
agilent collaborates on study of performance enhancing spinach extract.	Netural	0.127

discrimination rate is set as 0.85 so as to obtain a fine tuned network. Table 2 shows training and validation accuracy of 5 epochs. Table 3 shows final accuracy of the presented model.

Table 2. Training and Validation Accuracy.

Epoch	Training loss	Validation loss	Accuracy
1	No log	0.421006	0.843333
2	0..405100	0.422754	0.841111
3	0.405100	0.441444	0.844444
4	0.150400	0.532759	0.838889
5	0.150400	0.560726	0.847778

Table 3. Final Accuracy.

Classes	Precision	Recall	F1-score	Support
0	0.87	0.81	0.84	309
1	0.85	0.91	0.88	359
2	0.90	0.89	0.89	332
Accuracy			0.87	1000
Macro avg	0.87	0.87	0.87	1000
Weighted avg	0.74	0.87	0.87	1000

4.3. Comparison with other models

We presented numerous experiments on the newspaper headline and company dataset using several Machine Learning based models with our presented technique. The methods are skilled for all sets of abilities to grasped outcomes. All the experimentation which were performed were trained with three classification techniques. The presented worked employed various deep learning models like XLNET and BERT on economical dataset. The experimentation results are as shown in Table 4. From the analysis it is observed that the proposed approach is robust for future trends prediction. More precisely, the precision values for XLNET AND BERT models are 69% and 30%.

Table 4. Comparative Analysis.

Model	Training loss	validation loss	Accuracy
XLNET	0.864	0.595	0.787
BERT	1.083	1.075	0.0511
FinBERT	0.405	0.441	0.844

Overfitting when the model has been trained with an extensive quantity of dataset serves as one of the primary problems while adjusting the models on an economic dataset. The over fitting issue occurs when a deep learning model is trained using a greater number of epochs because the training model heavily incorporates the trends unique to the sample details. In addition, once we prepared it using greater sparse phrases than it can handle, it might end up being overfitted. Nonetheless, if it did not contain correctly trained on a particular group of epochs, it ought to provide high precision on the set used for training while failing on the actual test set. In other words, by excessively fitting to the data used for training, the model loses its ability to generalize. For this reason, this work utilized only 5 epochs to keep from over fitting.

In the work reported in [28] the authors have directly utilized the entire report of a company. Here, in the pre-processing stage, positive and negative instances are derived from each of the sentences present in the whole report making it a tedious process and time-consuming process. Whereas, in the presented work, we have used the quarterly report of a company. Initially the report data has been summarised to extract the content of prime importance. It is followed by the extraction of the positive and/or negative instances effectively. This process will be comparatively easier and faster. This information is then given as an input to the deep learning model (FinBERT).

5. Conclusion and Future Scope

In contrast to other works on estimating stock market way, in this work we focused more on financial sentiment analysis, information collected from business blogs, social media platforms, and websites. We looked into the relationship between quarterly reports and the trajectory of particular company's stock, and we used a novel method to do fundamental analysis using a deep learning model for a particular financial domain, achieving an accuracy of about 84.77%. Since we recently employed three sources for the current work, numerous strategies can be used in the future to increase the model's accuracy. As part of our ongoing research, more news sentiment elements will be added and examine the extent to which the algorithm may be applied in other circumstances.

References

[1] Gautam, R., and Sharma, M. Stress Diagnosis Using Deep Learning Techniques. In International Conference on Innovative Computing and Communications: Proceedings of ICICC 2022, Vol. 2, 645-654.

[2] Tanaka, Y., Kodera, S., Sato, F., Sakaji, H., and Izumi, K., 2019. The Extraction of the Future Oriented Sentences from Annual Reports. 8th International Congress on Advanced Applied Informatics (IIAI-AAI), 2019, 679-684.

[3] Lokeshwaran, K., Rajesh, A., 2018. A Study of Various Semantic Web Crawlers and Semantic Web Mining. International journal of Applied Mathematics, 2018.

[4] Aparna, N., Manohara, M. M., and Radhika, P., 2016. Prediction Models for Indian Stock Market. Procedia Computer Science. Vol. 89, 441-449.

[5] Khan, J. A., Sangroha, D., Ahmad M., and Rahman, M. T., A performance evaluation of semantic based search engines and keyword based search engines. 2014. International Conference on Medical Imaging, m-Health and Emerging Communication Systems (MedCom), 2014, 168-173.

[6] Khare, K., Darekar, O., Gupta, P., and Attar, V. Z., Short term stock price prediction using deep learning. 2017. 2nd IEEE International Conference on Recent Trends in Electronics, Information & Communication Technology (RTEICT), 2017, 482-486.

[7] Wei, L., Cao, G., and Yin, J., 2019. Bi-level attention model for sentiment analysis of short texts. IEEE Access, Vol. 7, 119813-119822.

[8] Chen, Y., Lin W., and Wang, J. Z., 2019. A Dual-Attention-Based Stock Price Trend Prediction Model With Dual Features. IEEE Access, Vol. 7, 148047-148058.

[9] Hoblos, J., Experimenting with Latent Semantic Analysis and Latent Dirichlet Allocation on Automated Essay Grading. 2020. Seventh International Conference on Social Networks Analysis, Management and Security (SNAMS), 2020, 1-7.

[10] Attanasio, G., Cagliero, L., Garza, P., and Baralis, E., Combining News Sentiment and Technical Analysis to Predict Stock Trend Reversal. 2019. International Conference on Data Mining Workshops (ICDMW), 2019, 514-521.

[11] Lien Minh, D., Sadeghi-Niaraki, A., Huy, H. D., Min, K., and Moon, H., 2018. Deep Learning Approach for Short-Term Stock Trends Prediction Based on Two-Stream Gated Recurrent Unit Network. IEEE Access, Vol. 6, 55392-55404.

[12] Zhao, F., Xiao, Z., Wang, C., Ni, D., and Li, L., 2019. External Stimuli Predict Financial Market Behavior From the Brain Perception Perspective. IEEE Access, Vol. 7, 28769-28777.

[13] Istiake, M. A., Maswood, M. M. S., and Alharbi, A. G., Deep Learning-Based Stock Price Prediction Using LSTM and Bi-Directional LSTM Model. 2020. 2nd Novel Intelligent and Leading Emerging Sciences Conference (NILES), 2020, 87-92.

[14] Xiao, Q., and Ihnaini, B. 2023. Stock trend prediction using sentiment analysis. PeerJ Computer Science, Vol. 9, e1293.

[15] Ho, T.-T., Huang, Y., 2021. Stock Price Movement Prediction Using Sentiment Analysis and CandleStick Chart Representation. Sensors, Vol. 21, 7957.

[16] Park, J., and Cho, S. 2023. Incorporation of company-related factual knowledge into pre-trained language models for stock-related spam tweet filtering. Expert Systems with Applications, Vol. 234, 121021.

[17] Rajakulanajagam, N. 2015. Stock Market Reaction to the Release of Annual Financial Statements: A Study on Selected Listed Banking Finance and Insurance Companies in Sri Lanka. Published and presented in ICCM.

[18] Wisniewski, T. P., and Yekini, L. S. 2015. Stock market returns and the content of annual report narratives. In Accounting Forum Vol. 39, No. 4, 281-294.

[19] Shi, M., and Zhao, Q. Stock market trend prediction and investment strategy by deep neural networks. 2020. 11th International Conference on Awareness Science and Technology (iCAST), 1-6.

[20] Wei, G., Zhang, W., and Zhou, L. Stock trends prediction combining the public opinion analysis. 2015. International Conference on Logistics, Informatics and Service Sciences (LISS), pp. 1-6.

[21] Yin, L., Li, B., Li, P., and Zhang, R. 2023. Research on stock trend prediction method based on optimized random forest. CAAI Transactions on Intelligence Technology, Vol. 8, No. 1, 274-284.

[22] Wang, Y., Liu, H., Guo, Q., Xie, S., and Zhang, X. 2019. Stock volatility prediction by hybrid neural network. IEEE Access, Vol. 7, 154524-154534.

[23] Alqahtani, S. S., and Rilling, J. Semantic modeling approach for software vulnerabilities data sources. 2019. 17th International Conference on Privacy, Security and Trust (PST), pp. 1-7.

[24] Verma, D., and Muralikrishna, S. N. Semantic similarity between short paragraphs using Deep Learning. 2020. IEEE International Conference on Electronics, Computing and Communication Technologies (CONECCT), pp. 1-5.

[25] Sha, L., Raković, M., Das, A., Gašević, D., and Chen, G. 2022. Leveraging class balancing techniques to alleviate algorithmic bias for predictive tasks in education. IEEE Transactions on Learning Technologies, Vol. 15, No. 4, 481-492.

[26] Zhao, Y., Dong, J., and Peng, T. 2009. Ontology classification for semantic-web-based software engineering. IEEE Transactions on Services Computing, Vol. 2, No. 4, 303-317.

[27] Balakrishnan, V., Shi, Z., Law, C. L., Lim, R., Teh, L. L., and Fan, Y. 2022. A deep learning approach in predicting products' sentiment ratings: a comparative analysis. The Journal of Supercomputing, Vol. 78, No. 5, 7206-7226.

[28] Araci, D. 2019. Finbert: Financial sentiment analysis with pre-trained language models. arXiv preprint arXiv:1908.10063.