Proceedings of the Third International Conference on Intelligent Data Communication Technologies and Internet of Things (IDCIoT-2025)
IEEE Xplore Part Number: CFP25CV1-ART; ISBN: 979-8-3315-2754-9

2025 3rd International Conference on Intelligent Data Communication Technologies and Internet of Things (IDCIoT) | 979-8-3315-2754-9/25/$31.00 ©2025 IEEE | DOI: 10.1109/IDCIOT64235.2025.10915080

Advanced Financial Sentiment Analysis using FinBERT to Explore Sentiment Dynamics

1st S Baghavathi Priya
Department of Computer Science and Engineering,
Amrita School of Computing,
Amrita Vishwa Vidyapeetham
Chennai, India

2nd Madhav Kumar
Department of Computer Science and Engineering,
Amrita School of Computing,
Amrita Vishwa Vidyapeetham
Chennai, India

3rd Nitheesh Prakash J D
Department of Computer Science and Engineering,
Amrita School of Computing,
Amrita Vishwa Vidyapeetham
Chennai, India

4th Krithika N
Department of Computer Science and Engineering,
Amrita School of Computing,
Amrita Vishwa Vidyapeetham
Chennai, India
kritzz2117@gmail.com

***Abstract*—This research investigates the incorporation of sophisticated Natural Language Processing (NLP) methods in the intricate field of finance. In particular, it utilizes FinBERT, a tailored language model created to understand the complex subtleties of financial text data, in conjunction with assessments of DistilBERT and Bidirectional Encoder Representations from Transformers (BERT). The main goal is to create a strong preprocessing framework that improves the interpretative ability of models for sentiment classification designed for financial uses while providing an in-depth evaluation of the comparative performance of the models. Through extensive empirical validation, our approach demonstrates that FinBERT outperforms the other models in accurately capturing subtle sentiment variations and financial terminology. The comparative study underlines the compromises between computational efficiency and precision, with DistilBERT providing a streamlined option. This research emphasizes FinBERT's transformative capabilities in analyzing financial sentiment and simultaneously delivering essential insights into the comparative strengths and weaknesses of other approaches, presenting a comprehensive viewpoint to inform decision-making within the financial sector. Upon comprehensive analysis of the methods used in this paper, it is shown that FinBERT is best at capturing hidden features within the data and able to work well with unseen data also with an average accuracy of 89.6% over all the 3 classes that are present in the dataset. The suggested framework has the potential to greatly improve predictive analytics and risk management approaches, allowing for more informed, precise, and prompt financial decisions in intricate market conditions.**

Index Terms*—DistilBERT, Sentiment Analysis, Natural Language Processing, financial markets, preprocessing, sentiment dynamics, empirical validation, and decision-making, Parts of Speech, Discriminant Correlation Analysis, Convolutional Neural Networks.

I. INTRODUCTION

Sentiment analysis is a fundamental component of NLP that allows for extracting nuanced viewpoints and emotions from textual data. It is critical in financial decision-making since it assesses investor sentiment, detects market trends, and forecasts consumer behaviour. However, traditional methodologies frequently fail to capture the nuanced intricacies of financial terms, resulting in inferior outcomes and restricted insights. These restrictions have driven the development of powerful sentiment analysis technologies designed specifically for financial applications. Among these, FinBERT has surfaced as a notable advancement. Developed on the BERT framework, FinBERT tackles the issues of conventional sentiment analysis by focusing on financial content. Its specialized pre-training on vast financial datasets enables it to recognize nuanced sentiment shifts and contextual interpretations specific to financial terminology, including acronyms, jargon, and unique vocabulary, thereby ensuring improved accuracy and dependability. In addition to FinBERT, BERT and its lighter counterpart DistilBERT have demonstrated efficacy for sentiment analysis tasks. While BERT provides a comprehensive transformerbased framework for comprehending bidirectional text interactions, DistilBERT is a more efficient alternative that reduces processing needs while keeping most of BERT's accuracy. These models, which were pre-trained on broad corpora and fine-tuned for financial sentiment analysis, perform well in tasks like sentiment polarity classification and sentiment intensity evaluation, however they may lack FinBERT's domain specific optimization. These models utilize a transformer-based structure that enables bidirectional text processing, enhancing their capability to understand the relationships between words and their contexts. This allows for a detailed comprehension of sentiment, rendering them useful for diverse financial sentiment analysis purposes. Moreover, these models are exceptionally flexible, rendering them appropriate for tasks related to both general and specialized text data. Advanced machine learning algorithms, such as Long-Short Term Memory (LSTM), Convolutional Neural Networks (CNN), Support Vector Machines (SVM), and Random Forest (RF), have also shown effectiveness in financial text analysis applications. Furthermore, LLMs, with billions of parameters, have transformed the NLP landscape by excelling at sentiment analysis, named entity identification, and machine translation. However, their use is frequently limited by high computational costs and resource needs, especially for domain-specific activities. The relative strengths of these models emphasize their supportive functions in analyzing financial sentiment. While FinBERT provides expertise and exceptional precision in understanding financial terms, BERT and DistilBERT offer greater versatility and efficiency. In summary, FinBERT, BERT, DistilBERT, and various advanced models have transformed sentiment analysis within the finance industry. These advancements have aided the industry in understanding market dynamics, identifying subtle trends, and making informed decisions based on data. As financial markets expand, incorporating these intricate models will remain crucial in shaping sentiment analysis and decision-making.

II. LITERATURE REVIEW

Sentiment analysis is one of the required methods for extracting emotions through different data sources, including the update of social media, pictures, and videos. Such application includes predicting stock price and market sentiment[2]. Social media sites can use to express feelings as compared to just factual information presented like sarcasm, which may require different difficulties in dealing with sentiment analysis. For example, research on sarcasm recognition in English and Arabic sentiment analysis utilized switch transformers, which frequently surpass traditional models such as BERT, ALBERT, and RoBERTa by integrating switch layers and segmenting input sequences to enhance performance.[1][7] These models have shown promise in more effectively managing intricate linguistic elements, such as sarcasm[1], yet additional improvements are required to boost their efficiency. Decision support from financial news is either of an explanatory or predictive nature.[12]

In financial sentiment analysis, transfer learning solves the problem of scarce training data by leveraging insights from similar fields. For instance, the study ”Stock Prediction through Sentiment Transfer Learning” demonstrated improved accuracy by aligning stock news with a common emotional feature space, making efficient sentiment analysis possible for stocks that have limited news coverage[2]. This reduces the dimensionality of the feature space and improves the predictions for less represented stocks [10]. An increasing trend in financial sentiment analysis is the adoption of hybrid models that integrate various learning algorithms, including ensemble techniques and deep learning methods, to enhance prediction precision and flexibility. The reason for the great attention paid in recent years to the field of sentiment analysis by both the industrial and academic community is to be found in the desire to help decision-making processes as much as possible. [16]

In multilingual environments, researchers have used sentiment dictionaries methods to obtain emotions from texts in a mixture of languages such as Chinese and English.[3][13] Such methods apply large vocabularies and Na¨ıve Bayesian classifiers in order to control polysemous terms that have opposite polarities to improve the precision of sentiment classification in complex linguistic contexts[8]. The combination of multilingual sentiment analysis and transfer learning has demonstrated promise in enhancing accuracy in various languages, particularly in finance where timely news evaluation in several languages is crucial. The integration of text mining into financial analysis represents a significant shift in how researchers approach market predictions.[20]

Visual sentiment analysis provides more emotional depth than text content, mainly because of the rise in visual expressions that are socially media-influenced. Models like ME2M address problems in image-centric sentiment analysis by using Discriminant Correlation Analysis (DCA) to obtain crossmodal emotional semantics. This approach improves data sets and integrates different image characteristics to detect subtle emotional signals, but high-dimensional images can still be challenging[4]. Recent studies have investigated the application of Convolutional Neural Networks (CNNs) alongside DCA to lower dimensionality and enhance the processing speed of visual sentiment analysis systems.

Video sentiment analysis further increases the complexity by extracting emotions from the image, text, and audio elements. Techniques like VAE-based adversarial multimodal domain transfer (VAE-AMDT) create a coherent embedding space for multimodal representations. This makes a huge difference in the intensity of the sentiment being predicted. Future work may be related to making modal encoders more resilient and incorporating multimodal data for better precision in video based sentiment analysis[5]. Furthermore, improvements in video sentiment analysis emphasize real-time processing, with low-latency models being created to assess sentiment during live occurrences, like stock market debates or earnings calls.

Advancements in text-based sentiment analysis, such as POS-based Transformer Learning (POS-TAN), use the incorporation of Part-of-Speech (POS) attention techniques to capture the emotions held in the word order of sentences[6]. It combines POS vectors and self-attention mechanisms to relax traditional constraints of attention-based networks while allowing improved classification accuracy through better attention to emotional input. Recent advancements also investigate the use of BERT-based models for detailed sentiment classification, especially in specialized fields such as financial news analysis, where context can significantly influence sentiment.

In addition, models that take into account word polarities, such as love-hate or success-failure, measure sentiment by determining the semantic similarity of target words in their context. This approach captures context-dependent polarity, addressing challenges where terms such as ”small” may be used to convey positive or negative meaning. However, existing models may still struggle with polarity variations that are specific to certain domains. In financial sentiment analysis, greater focus is being placed on domain adaptation methods that aid models in achieving improved generalization when utilized in specialized fields, like forecasting market trends in specific industries. Other BERT models, while effective in general contexts, often struggle to accurately capture the nuanced language and specific terminologies prevalent in financial data, leading to less precise outcomes in this domain.[18]

Finally, creative applications such as steganography in sentiment analysis make use of neural networks to hide messages in text[8]. Such systems rely on pre-trained language models to generate predicted words and securely embed messages, indicating how sentiment analysis merges with secure communication technologies[9]. The rising adoption of Explainable AI (XAI) in sentiment analysis has offered understanding into the decision-making processes of models, thereby improving trust and clarity in applications such as financial sentiment analysis, where transparency in decision support is essential.

These advancements, therefore highlight the move toward broader sentiment analysis across modalities and fields with an improvement in its precision and relevance in striving to understand human emotions and actions.

III. Methodology Used

A. Dataset

TABLE I: Description of columns in the Financial Phrase Bank dataset

Column Name	Description
Sentence	A financial phrase or sentence extracted from documents such as news articles, financial reports, or other sources.
Class Label	The sentiment label associated with the sentence. It can be one of the following: positive, negative, or neutral.

The FinancialPhraseBank dataset, created by Malo et al., is a curated collection of financial text specifically designed for sentiment analysis in the financial domain. This dataset is sourced from a variety of financial documents, including news articles, annual reports, and other financial publications, ensuring a rich representation of language used in the finance industry. Sentences in the dataset are short, sharp, and precise, emphasizing essential financial information that can be of particular use in a sentiment classification task. Notable about this dataset is the quality of annotation performed by financial professionals. Also, to increase reliability, majority voting was performed among multiple annotators to obtain the resultant sentiment labels for each sentence. Such rigorous annotation promotes consistency and relevance to the domain, thereby addressing potential ambiguities in sentiment interpretation. This diversity and consistency make the dataset highly reliable and practical for sentiment analysis research.[17]

The real-world applicability of the dataset is indicated by the distribution of sentiment: 60% of the data falls into the neutral category, which is very much reflective of the objective nature of most financial statements. The critical layers of emotional and contextual nuances are added through the 12% negative and 28% positive sentiments, crucial for applications like risk assessment, investment analysis, and market forecasting.

Besides its primary application in the sentiment analysis realm, the FinancialPhraseBank has also been used extensively as a testing and fine-tuning benchmark in advanced machine learning models. Furthermore, it is considered a relevant platform for tasks as diverse as the summarization of financial text to topic modelling tasks and even the events that move market prices. Very often, financial corpora comprising this dataset find their way in building robust models that adequately address the variegated challenges financial text mining throw up.

Further, the flexibility of the FinancialPhraseBank dataset renders it an unreplaceable asset across academic and industrial circles. While leveraging the use of the data, machine learning practitioners are able to train models not only to extract sentiment but also to uncover the sub-text of financial communication, including the implications of change in policy, economic change, and assessment of company performance. All this essentially emphasizes the value of such data within this rapidly emerging field of financial NLP.

This extended foundation thus allows researchers to rigorously compare models like FinBERT, BERT, or DistilBERT over their ability in capturing the particular language and shifting sentiments that specifically exist in the domain of finance.

Conventional techniques in accuracy and dependability, exhibiting notable gains in comprehending the complex language of financial literature. This research investigates the relative performance of FinBERT in comparison to BERT, DistilBERT, emphasizing the advantages and disadvantages of each approach. The FinancialPhraseBank dataset's detailed and accurately labelled framework acts as a foundation for this study, allowing the models to grasp complex sentiment dynamics vital for enhancing sentiment analysis methods in the financial sector.

B. Data Preprocessing

The subsequent stage in the methodology of this framework involves the preprocessing of the given dataset. Initially, the text and label columns from the dataset (data) are transformed into Python lists for further processing. The dataset is divided into training and testing subsets with the train test split function, ensuring a 75:25 ratio and ensuring reproducibility through a set random state. Tokenization is utilized on both the training and testing texts through a specific preprocess function, which employs the AutoTokenizer from the Hugging Face Transformers library to transform raw text into tokenized sequences. This stage involves truncating to guarantee that the tokenized sequences fit within the model's maximum input length. For BERT-based models, tokenization not only divides the text into word segments but also associates these segments with their respective indices in the tokenizer's vocabulary, incorporating special tokens such as [CLS] and [SEP] when required.

To address the categorical nature of labels, it also specifies a mapping between sentiment labels (e.g., "positive," "neutral," "negative") and numerical IDs (label2id), as well as an inverse mapping (id2label). These mappings convert stringbased labels to the numerical representations required by the model. To facilitate model training, a custom BertDataset class is constructed by extending PyTorch's Dataset. This class receives tokenized encodings and number labels and guarantees that they are indexed correctly.

The ~~get~~item fun~~cti~~on converts each sample into PyTorch tensors with tokenized inputs and labels, enabling smooth data loading during model training. Furthermore, a DataCollatorWithPadding is built to ensure that batches of data are dynamically padded to the maximum sequence length within each batch, hence reducing memory consumption during training. In general, the preprocessing workflow encompasses converting raw text and labels into tokenized sequences, applying truncation, changing sentiment labels into numerical identifiers, and assembling these processed elements into a dataset suitable for PyTorch. This structured preprocessing ensures that the 'data is properly prepared for the subsequent training phase, where it will be fed into a pre-trained BERT model for fine-tuning.

C. BERT

This segment examines the BERT model, which is a transformer and encoder framework created for numerous NLP applications.[14] Unlike conventional models that analyze text sequentially in a single direction, BERT utilizes a bidirectional transformer structure to collect context from both sides of a word at the same time. BERT is made up of several layers of Transformer encoders that collect context from the neighbouring words on either side of each word in a sentence. Masked Language Model (MLM) and Next Sentence Prediction (NSP) are two tasks in unsupervised learning used to pre-train it on an extensive dataset. In pre-training, certain words in a sentence are randomly hidden, and the model predicts these hidden words to understand bidirectional context. The NSP task focuses on assessing whether one sentence logically follows another, aiding the model in understanding inter sentence connections. This model is originally designed for the purpose of financial sentiment analysis especially focus on the need of capital market practitioners and researchers.[19]

The pre-trained BERT model's weights are changed during fine-tuning in order to acquire task-specific representations for sentiment analysis. Tokenizing text input into subwords using before supplying it to the BERT model an appropriate tokenization package. Typically, special tokens '[CLS]' (which stands for "classification") and '[SEP]' (which stands for "seperator") are appended to input sequences. The '[CLS]' token signifies the whole sequence and is utilized for classification purposes. After tokenization, each input is transformed into a numerical representation by use of learnt embeddings for every token. To give information about each token's position in the sequence, positional encodings are added to the token embeddings. These positional encodings are essential for BERT to comprehend the arrangement of tokens in the input sequence. Sentiment analysis and key entities of online financial texts help to understand the public's sentiment state, timely access to public opinions and attitudes, and rapidly get to the subject of information.[15]

Transformer encoders are arranged in several layers to create the architecture of BERT. Each encoder layer makes use of feed forward neural networks and self-attention strategies to gather contextual information. BERT analyzes the input sequence in a bidirectional way, meaning it takes into account the context from both the left and right of each token at the same time. Self-attention, also known as intra-attention, is a mechanism that allows transformers to evaluate each word's relative significance as they process it sequentially. It allows the model to process each token while focusing on the relevant portions of the input sequence. In the selfattention mechanism, BERT allocates varying attention scores to every token, enabling it to concentrate more on the crucial tokens within the sequence. Each token in a sequence is connected with three vectors: query (Q), key (K), and value (V). To construct these vectors, the learning weight matrices are multiplied by the input sequence embedding W_Q, W_K, and W_V. The self-attention mechanism assigns a score to each token in the sequence by computing the dot product of its Query vector and the Key vectors of all other tokens. These scores indicate the significance or likeness of the current token compared to all other tokens in the sequence. To obtain attention weights, the attention scores undergo normalization through the softmax function, ensuring that the weights sum to 1 for each token. The weights dictate how much each token contributes to the representation of the current token.

Ultimately, the output representation for every token is calculated by taking the weighted sum of the Value vectors and adding the attention weights to it. For every token, the contextual data from the complete input sequence is captured by this weighted sum. The following formula can be used to mathematically illustrate the self-attention mechanism:

$$\text{Attention}(Q,K,V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V \tag{1}$$

Here, d_k refers to the dimensionality of the key vectors.

Feedforward neural networks represent a category of artificial neural networks in which the links among nodes do not create cycles. They are made up of an input layer, several hidden layers, and an output layer. Every layer of an FFNN includes a collection of neurons or units, with each neuron linked to every neuron in the preceding layer. In BERT, FFNNs are applied to the output of the self-attention mechanism to further process and refine the representations. Neurons in neighboring layers are completely interconnected, indicating that the output of each neuron is linked to the input of every neuron in the subsequent layer. Neurons use non-linear activation functions such as the rectified linear unit (ReLU), sigmoid, or hyperbolic tangent (tanh) on their inputs. These functions add nonlinearity to the network, enabling it to grasp intricate patterns in the data. FFNNs are trained through methods that modify the network's weights in order to reduce a loss function. Throughout training, the network learns to estimate the relationship between input data and desired outputs by modifying the weights of its connections.

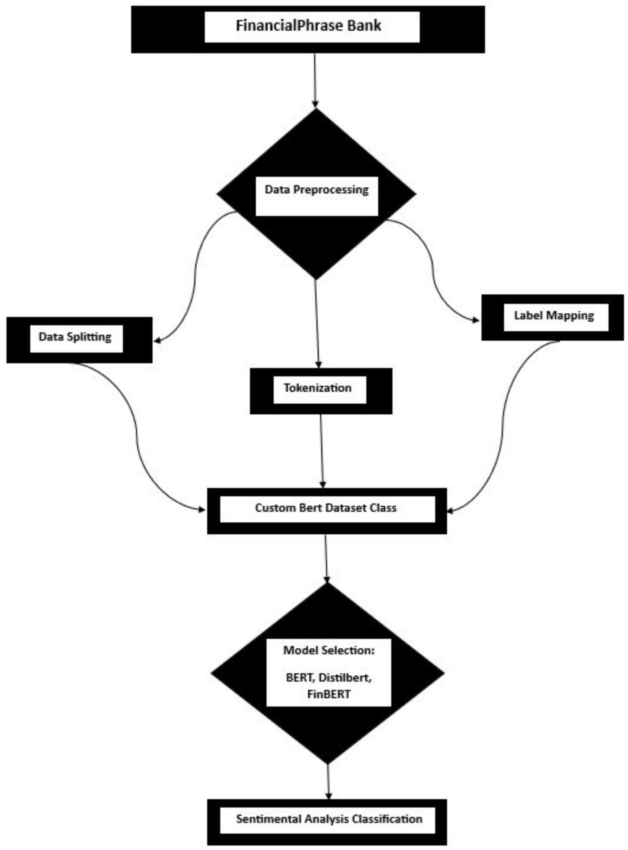

Fig. 1: Architecture Diagram of Sentiment Classification in Financial Data

Mathematically, the computation in an FFNN can be represented as follows:

$$\text{FFNN}(x) = \text{Activation}(xW_1 + b_1)W_2 + b_2 \quad (2)$$

where the input vector is represented by x, the bias vectors are represented by b_1 and b_2 as weight matrices for the first and second layers, and the activation function is applied elementwise to the first layer's output by *Activation*. For sentiment analysis, a task-specific classification head is introduced to the pre-trained BERT model. To derive the probability distribution over sentiment classes, the output representation of the '[CLS]' token is passed through a fully connected layer and then a softmax activation function. Using labeled sentiment data, the BERT model's overall parameters—which comprise the classification head and the pre-trained BERT layers—are adjusted during training. The model is trained to minimize the loss function, such as cross-entropy loss, using methods like stochastic gradient descent (SGD) or Adam optimization. The fine-tuning process ensures that BERT learns the best representations for sentiment analysis tasks. Once training is complete, the model can be applied to new, unseen text data to generate predictions. BERT's capability to produce contextual embeddings enables it to generalize effectively to unfamiliar text, rendering it particularly efficient for tasks such as sentiment assessment.

D. DistilBERT

A condensed form of BERT is called DistilBERT. It seeks to maintain most of BERT's functionality while being lighter and more efficient. This architecture is better suited for deployment in contexts with limited resources, such mobile devices or edge computing devices, because it was created to be a quicker, smaller, and less expensive alternative to BERT. This model, akin to BERT, is built on the Transformer architecture, consisting of an encoder stack. Distilbert, conversely, simplifies its structure by employing fewer layers and attention heads compared to BERT while still utilizing a similar architecture. Unlike BERT, which consists of twelve layers, DistilBERT has just six. To further minimize its size, DistilBERT uses a training technique called knowledge distillation to learn how to replicate the larger BERT model's behavior. DistilBERT is pre-trained on extensive data employing unsupervised learning goals (MLM and NSP), similar to the original model. The model learns to generate contextualized word embeddings during the pre-training phase by considering the surroundings of each word in a text.

Once pre-trained, the model may perform tasks such as classification of text, sentiment analysis, and even question answering. The parameters of the model are then adjusted to better suit the job at hand. The fine-tuning procedure entails adding task-specific layers to the previously trained DistilBERT model. The main purpose of this stage is to improvise the pre-trained model's performance on test data.

The method discussed in this section can successfully achieve a balance between model size and performance. While it might not match the performance of larger models, it still provides a significant reduction in size and computational expense, and it competes effectively across various NLP tasks. Due to its smaller size and quicker interference, it is better suited for real-time applications. In conclusion, DistilBERT presents a robust solution for NLP tasks that require a balance among model size, speed, and effectiveness. Due to its small dimensions and efficient design, it has become a preferred choice for numerous industrial and research uses.

E. FinBERT

FINBERT is another BERT variation designed specifically for finance-related tasks and datasets. It makes use of BERT's pre-trained language representation to give contextual embeddings for financial content. FINBERT is trained on large volumes of general text data during the pre-training phase. It improves the pre-trained BERT model by using a corpus of financial text data that is specific to the financial domain. FINBERT gains the capability to generate contextualized embeddings for financial terms and expressions during pretraining, thereby capturing their meanings and connections within the financial sector. This model could be enhanced for different financial tasks, such as sentiment analysis, Named Entity Recognition (NER), and classification, following pretraining.

The algorithm utilized in FINBERT builds upon the Transformer-based architecture underlying BERT, incorporating specialized pre-training tailored to financial data. The model utilizes self-attention mechanisms, similar to BERT, while also integrating training specific to the financial domain. The mathematical formulation of FINBERT's algorithm is as follows:

Let $X = (x_1, x_2, \ldots, x_n)$ represent the input sequence (e.g., financial sentences), where each x_i is a token. During pretraining, FINBERT tokenizes the input sequence and builds contextualized embeddings for each token x_i depending on its connection with surrounding tokens.

Similar to BERT, the self-attention mechanism computes a weighted sum of the embeddings for each token using three vectors: Query (Q), Key (K), and Value (V).

For the reasons listed above, task-specific layers are added to the FINBERT's pre-trained layers. The tokenization, input representation, and classification process is performed in the same way as it is done in BERT.

F. Optimization Process

A FINBERT model's efficiency can be optimized through a number of strategies implemented systematically. Tokenization and data handling can be enhanced through efficient tokenizers and preprocessing the dataset to reduce noise, and batching and data loading can be optimized using parameters like batch size, 'num-workers', and caching preprocessed data to avoid redundant computations. Hyperparameter tuning, either by hand or through automated tools such as Optuna, helps determine optimal values for learning rates, batch sizes, and dropout rates. Mixed-precision training using PyTorch's 'torch.cuda.amp' can accelerate computations and decrease memory usage significantly. Fine-tuning the model by freezing early layers and using learning rate schedulers further optimizes training. Gradient accumulation enables larger effective batch sizes without exceeding memory limits, and training on multiple GPUs accelerates the process. Collectively, these optimizations improve the balance between computational efficiency and model performance, making it possible to train and evaluate faster with the same level of accuracy.

G. Testing

The testing and validation processes within the FINBERT code are devised to ensure strong performance evaluation and fine-tuning of the model. During the validation process, a heldout portion of the dataset is utilized to monitor the performance of the model after every training epoch. This process will help prevent overfitting because the metrics to be evaluated, including loss, accuracy, precision, recall, and F1 score, will be calculated on unseen data. Hyperparameter tuning requires a validation step to ensure that there is a good measure of how the model generalizes to data it was not trained on. Testing uses an entirely separate dataset, which was not involved in either training or validation, to provide an unbiased assessment of the final performance of the model. The results on the test set have been discussed in the next section.

Metrics are calculated to assess the effectiveness of the model in real-world applications. Testing and validation processes use tokenized inputs to keep aligned with the pretrained FINBERT architecture, ensuring that embeddings are consistent. The workflow makes sure that model performances have a well-balanced balance between training accuracy and generalization performance through comparison of result patterns across those datasets.

IV. Results

This section presents the results obtained from the evaluation of the data set using the BERT, DistilBERT, and FinBERT algorithms. The following metrics are used in this paper: Accuracy, Precision, Recall, and F1 Score. These metrics are frequently used to assess how well models perform on classification tasks. The following provides the formulae for the measures in question:

$$\text{Accuracy} = \frac{\text{Number of Correct Predictions}}{\text{Total Number of Predictions}} \times 100\% \tag{3}$$

$$\text{Precision} = \frac{\text{True Positives}}{\text{True Positives + False Positives}} \tag{4}$$

$$\text{Recall} = \frac{\text{True Positives}}{\text{True Positives + False Negatives}} \tag{5}$$

$$\text{F1 Score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision + Recall}} \tag{6}$$

Where:

- The number of accurately anticipated positive cases is known as True Positives (TP).
- The quantity of positive cases that were mispredicted is known as False Positives (FP).
- The quantity of negative cases that were erroneously forecasted is known as False Negatives (FN).
- The number of accurately anticipated negative cases is known as True Negatives (TN).

The table II depicts the value of the metrics for the experiment performed in this paper:

TABLE II: Performance comparison of different models

Method	Accuracy	Precision	Recall	F1 Score
BERT	0.843	0.818	0.828	0.822
DistilBERT	0.843	0.818	0.828	0.822
FinBERT	0.896	0.884	0.896	0.89

A. Interpretation of Results

A confusion matrix is shown in figure 2 that represents the model's performance for three sentiment categories, which are positive, neutral, and negative. Along the diagonal are correct classifications for 131 positive instances, 648 neutral instances, and 307 negative instances. Off-diagonal elements indicate misclassifications. For example, 13 neutral instances were misclassified as positive, while 44 negative instances were predicted as neutral. Neutral exhibits the highest classification accuracy of this category, as represented by the maximum diagonal value, while there are certain confusions between neutral and negative classes. Overall, the robustness of the model has been proved from the overall matrix, with higher strength towards the dominant class-neutral. We find that performance can be substantially improved by training the model longer, with bigger batches over more data.[11]

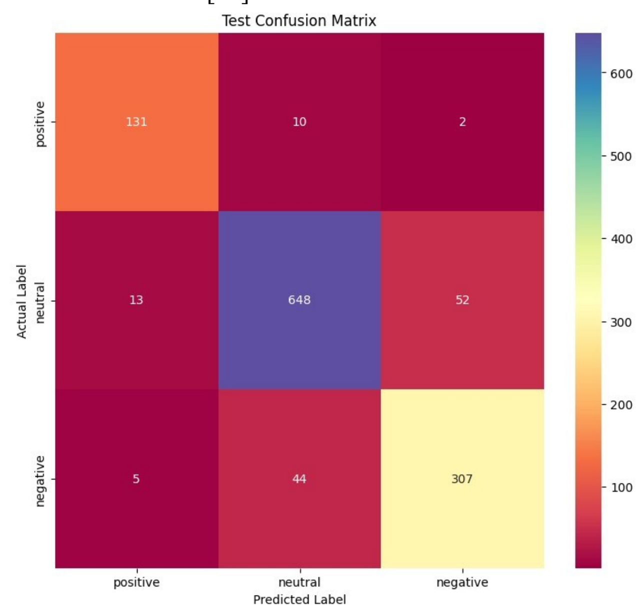

Fig. 2: Confusion Matrix for FinBERT

To enhance accuracy even more, the subsequent methods might be explored:

- Longer training periods with larger datasets could assist the model in acquiring more extensive features.
- Adjusting hyperparameters (such as learning rate and batch size) can enhance the model's performance.
- Techniques for data augmentation, like paraphrasing or back-translation, can be employed to improve the training dataset.
- Ensemble techniques such as merging several models can enhance overall accuracy by utilizing the advantages of each individual model.

The FinBERT model makes its predictions with an average accuracy of 0.896 across all the 3 class labels and with a precision of 0.87 for the positive class, 0.92 for the neutral class, and 0.85

for the neutral class which indicates that the model is capable of performing well on unseen data.

B. Future Work

There are several promising avenues for further research into this, with prospects for improving the precision and efficiency in using FINBERT for financial sentiment analysis. Domain pretraining on larger and more diverse financial corpora can enhance the model's contextual understanding of nuanced financial terminology, while the integration of multimodal data builds a more comprehensive framework for sentiment analysis: textual financial reports combined with numerical stock metrics or social media trends. Knowledge distillation can be a gateway to obtaining leaner, more efficient models that do not affect performance and can be deployed in resource-constrained environments. Adversarial training also may strengthen the robustness against noisy or ambiguous financial data, and even low-rank adaptation or adapter layers could significantly mitigate the computational overhead of fine-tuning. The use of active learning methods can further optimize model training by focusing on uncertain or highvalue samples for labeling, while federated learning allows for fine-tuning on decentralized datasets in a secure manner while preserving privacy. Collectively, these techniques bridge existing gaps in precision and efficiency to enable FINBERT for a broader scope of financial tasks such as market forecasting, portfolio management, and risk assessment. Handling missing data is also critical for transformer-based models. While we did not specifically address missing values in our current framework, incorporating methods like placeholder text for missing entries or excluding null instances could enhance robustness and consistency in future iterations.

V. CONCLUSION

Extracting emotions from a financial news data set can be useful in the real world for many good reasons, such as tracking market sentiments, managing risks in the financial sector, or identifying trends. Gaining insight into how traders, investors, and everyone else feel about specific stocks, companies, or the market as a whole may be quite beneficial. While negative sentiment could point to worries or possible selling pressure, positive sentiment could be related to anticipation and possible purchasing interest. The assessment and management of investment hazards can be aided by sentiment research. Unexpected changes in attitude might be signs of impending volatility or market downturns, enabling investors to reduce losses by modifying their portfolios appropriately. Sentiment analysis can be used to spot new patterns or modifications in consumer tastes that may have an effect on particular sectors or companies. Such data may be used to make strategic investment decisions or change corporate policy. Machine learning models may be trained to anticipate market changes or identify potential investment opportunities by analyzing historical sentiment data in conjunction with market performance. Predictive analytics can help investors remain ahead of the curve and capitalize on market events. Actual market and shareholder sentiment may be derived by tracking sentiment in news stories, social networking sites, and a variety of other internet-based resources. This comprehension can significantly aid computational trading strategies and high-frequency traders. Techniques in statistical trading that involve sentiment analysis can generate algorithms capable of automatically executing trades based on set emotional indicators. These strategies can help traders capitalize on temporary fluctuations in the market triggered by emotional changes. Examining feelings towards specific brands or companies is another application of sentiment analysis. By tracking the sentiment related to corporate earnings announcements, product launches, and changes in leadership, investors can gain valuable insights into the potential impacts on stock prices and the overall performance of a company. Businesses can assess shareholder emotion regarding their shares and the general impression of their company's image by using sentiment analysis. Companies can use such data to better target stakeholder relationships messaging and tactics to allay client fears or build on favourable emotions.

REFERENCES

[1] S. Muhammad Ahmed Hassan Shah and S. Faizan Hussain Shah, ”Arabic Sentiment Analysis and Sarcasm Detection Using Probabilistic Projections-Based Variational Switch Transformer,” June 2023.

[2] X. Li, H. Xie, R. Y. K. Lau, T.-L. Wong, and F.-L. Wang, ”Stock Prediction via Sentimental Transfer Learning,” November 2018.

[3] G. Xu, Z. Yu, H. Yao, F. Li, Y. Meng, and X. Wu, ”Chinese Text Sentiment Analysis Based on Extended Sentiment Dictionary,” April 2019.

[4] H. Zhang, J. Wu, H. Shi, Z. Jiang, D. Ji, T. Yuan, and G. Li, ”Multidimensional Extra Evidence Mining for Image Sentiment Analysis.”

[5] Y. Wang, J. Wu, K. Furumai, S. Wada, and S. Kurihara, ”VAE-Based Adversarial Multimodal Domain Transfer for Video-Level Sentiment Analysis,” May 2022.

[6] K. Cheng, Y. Yue, and Z. Song, ”Sentiment Classification Based on Part-of-Speech and Self-Attention Mechanism,” January 2020.

[7] K. Abdalgader and A. Al Shibli, ”Experimental Results on Customer Reviews,” October 2020.

[8] C.-C. Chang, ”Reversible Linguistic Steganography With Bayesian Masked Language Modeling,” April 2023.

[9] B. Liu, ”Sentiment Analysis and Opinion Mining,” Synthesis Lectures on Human Language Technologies, vol. 5, no. 1, pp. 1-167, 2012.

[10] X. Li, H. Xie, L. Chen, J. Wang, and X. Deng, ”News Impact on Stock Price Return via Sentiment Analysis,” Knowledge-Based Systems, vol. 69, pp. 14-23, 2014.

[11] Y. Liu, M. Ott, N. Goyal, J. Du, M. Joshi, D. Chen, O. Levy, M. Lewis, L. Zettlemoyer, and V. Stoyanov, ”RoBERTa: A Robustly Optimized BERT Pretraining Approach,” arXiv preprint arXiv:1907.11692, 2019.

[12] M. Kraus and S. Feuerriegel, ”Decision Support from Financial Disclosures with Deep Neural Networks and Transfer Learning,” Decision Support Systems, vol. 104, pp. 38-48, 2017.

[13] Y. Cui, W. Che, T. Liu, B. Qin, and Z. Yang, ”Pre-Training with Whole Word Masking for Chinese BERT,” arXiv preprint arXiv:1906.08101, 2019.

[14] Z. Lan, M. Chen, S. Goodman, K. Gimpel, P. Sharma, and R. Soricut, ”ALBERT: A Lite BERT for Self-Supervised Learning of Language Representations,” arXiv preprint arXiv:1909.11942, 2019.

[15] L. Zhao, L. Li, X. Zheng, and J. Zhan, ”A BERT-Based Sentiment Analysis and Key Entity Detection Approach for Online Financial Texts.”

[16] M. Pota, M. Ventura, R. Catelli, and M. Esposito, “An effective BERTBased pipeline for Twitter Sentiment Analysis: A case study in Italian,” Sensors, vol. 21, no. 1, p. 133, Dec. 2020, doi: 10.3390/s21010133.

[17] Y. Shen and P. K. Zhang, “Financial sentiment analysis on news and reports using large language models and FinBERT,” arXiv (Cornell University), Oct. 2024, doi: 10.48550/arxiv.2410.01987.

[18] A. Saxena, A. Santhanavijayan, H. Shakya, G. Kumar, B. Balusamy, and F. Benedetto, “Nested Sentiment Analysis for ESG Impact: Leveraging FinBERT to Predict Market Dynamics Based on Eco-Friendly and NonEco-Friendly Product Perceptions with Explainable AI,” 2024.

[19] T. Jiang and A. Zeng, “Financial sentiment analysis using FinBERT with application in predicting stock movement,” 2023.

[20] K. Kirtac and G. Germano, “Enhanced financial sentiment analysis and trading strategy development using large language models,” 2024.