Financial report sentiment analysis using Loughran-McDonald dictionary and BERT

SHEETAL R[1], PRAKASH K. AITHAL[2]
[1]Department of Computer Science and Engineering Computer Science and Information Security Manipal Institute of Technology Manipal Academy of Higher Education Manipal, INDIA
[2]IEEE Member Department of Computer Science and Engineering Manipal Institute of Technology Manipal Academy of Higher Education Manipal, INDIA

Abstract: - In the ever-changing world of financial markets, understanding investor behavior and making informed decisions relies heavily on sentiment analysis. This study delves into the integration of traditional techniques, such as the Loughran- McDonald dictionary, with advanced natural language processing (NLP) methods utilizing BERT (Bidirectional Encoder Representations from Transformers). The goal is to enhance the accuracy and depth of sentiment analysis in financial reports.To begin, we employ the specialized Loughran-McDonald dictionary designed for financial sentiment analysis. This lexicon includes domainspecific word lists for positive and negative sentiments, forming a solid foundation for sentiment scoring. Expanding on this foundation, we incorporate BERT, an advanced transformerbased NLP model. BERT's contextual understanding of language and ability to capture intricate semantic relationships within financial texts aim to overcome the limitations of rule-based sentiment analysis. The methodology involves preprocessing financial reports, integrating Loughran-McDonald sentiment scores, and fine-tuning BERT for financial sentiment classification. This hybrid approach leverages both the domain expertise encoded in the dictionary and BERT's contextual comprehension of financial jargon and nuances. We validate and evaluate our implementation using a diverse dataset comprising quarterly earnings releases, annual reports, and other relevant disclosures. Performance metrics such as precision, recall, and F1 score are analyzed to assess the effectiveness of our hybrid approach compared to individual methods. The findings have significant implications for financial analysts, investors, and policymakers by providing a more nuanced understanding of sentiment in financial reports. Our hybrid approach aims to offer improved accuracy in capturing sentiment polarity while facilitating more informed decision-making in today's complex and dynamic realm of financial markets.

Key-words: - Lexicon-based Analysis, Sentiment Analysis, Financial Reports Loughran-McDonald, BERT, Natural Language Processing (NLP).

Received: April 22, 2023. Revised: April 27, 2024. Accepted: May 23, 2024. Published: June 24, 2024.

1. Introduction

In the realm of financial markets, volatility is a well-known characteristic influenced by various factors such as economic indicators, company news, and investor sentiment. Successfully navigating these markets requires the ability to accurately assess the tone and sentiment in financial writings, as well as predict future market trends. Financial analysts typically rely on quantitative information like stock prices, trade volumes, and economic indicators to evaluate market conditions. In today's fast-paced and ever-changing finance world, information is crucial and the lifeblood of decision-making.

Textual data, including financial reports, news articles, and social media posts, has become increasingly popular as a valuable source of sentiment analysis and trend indicators in an era dominated by information overload. To make informed investment decisions and forecast future trends effectively, it is essential to understand the mood and tone of the financial markets.

Investors, analysts, and finance professionals striving for a competitive edge must decipher the underlying sentiment and tone embedded within financial papers, news stories, and market commentary. Sentiment analysis has emerged as a critical tool in meeting this need by providing insights into market sentiment that aid stakeholders in making better decisions.

However, there is still room for improvement in text analysis specifically tailored to the financial industry. The absence of models containing finance-specific vocabulary often leads to a lack of context in financial reports that can cloud investors' perception of a company's current state. Financial reports themselves can be misleading since businesses tend to project a positive image to the public.

Therefore, this study aims to employ machine learning techniques for analyzing sentiment in financial text data. Utilizing sentiment analysis methods effectively applied to determine whether companies performed well or poorly over previous fiscal years.[8]

This research has significant implications for academia, the financial industry, and individual investors in stocks. By leveraging BERT (Bidirectional Encoder Representations from Transformers), a cutting-edge natural language processing (NLP) model known for its contextual understanding of language, this study aims to enhance financial tone analysis and develop prediction algorithms for future market movements. It achieves this by analyzing historical financial texts and market data using the domain-specific Loughran-McDonald dictionary, which assigns sentiment scores to financial terms and phrases commonly used in traditional financial sentiment research. However, recent advancements in NLP, particularly BERT, have highlighted the importance of comprehending language within its context. [9]

This study holds great significance as it aims to bridge the gap between historical sentiment analysis and financial prediction modeling. Our goal is to equip investors, analysts, and financial institutions with a comprehensive and precise toolkit for decision-making. We achieve this by combining the strengths of established financial sentiment research methodologies with the advanced language comprehension abilities of BERT.

In this study, we delve into the realm of financial sentiment analysis, a field that intersects finance and natural language processing (NLP). [2]

The research presented in this study has significant implications for both the academic and business communities. It introduces a comprehensive framework for examining the financial sentiment expressed in various types of financial texts, including news articles and earnings reports. Furthermore, it lays the foundation for developing automated tools and systems that can offer real-time sentiment analysis in the dynamic financial landscape. These advancements will ultimately assist stakeholders in making more informed and effective decisions.

2. Literature Review

In the realm of analyzing financial texts, a comprehensive evaluation was conducted on the tone, utilizing a specialized lexicon known as the Loughran-McDonald dictionary. This dictionary consists of an extensive collection of financial terminology and sentiment scores, allowing for a deep understanding of the emotional connotations associated with specific financial terms and expressions. By leveraging this lexicon, precise sentiment classification becomes achievable. Each financial term or expression within the database is assigned sentiment scores indicating whether it carries positive, negative, or neutral meanings. Consequently, financial documents can be accurately categorized based on their underlying attitudes. Among various models evaluated for performance and accuracy, it was found that the Bidirectional Encoder Representations from Transformers (BERT) model exhibited exceptional results with 90 percent accuracy in predicting sentiments. However, it should be noted that this model does face challenges in terms of loading time and prediction speed due to its intricate nature.[8]

In the fascinating study on sentiment analysis of news articles, they employed a methodology based on the Lexicon-based approach. When it comes to sentiment analysis, there are generally two main approaches: supervised and unsupervised. The supervised approach involves training a classification model using labeled data to classify new data without labels. On the other hand, unsupervised or Lexicon-based approaches do not require any training data. Instead, they rely on inferring the sentiments of words based on their polarity.[9]

In the case of a sentence or document, the collective polarities of individual words determine the overall sentiment conveyed. This is achieved by summing up the polarities of each word or phrase within the sentence. To facilitate this approach, predefined lists of words are used, with each word associated with a specific sentiment. Additionally, there are various methods that can be utilized within this approach.

Overall, research sheds light on an intriguing methodology for sentiment analysis in news articles using the Lexicon-based approach. Their findings provide valuable insights into understanding and interpreting sentiments in textual content without relying on labeled training data.

In the realm of financial news analysis, the text found stands out for its authoritative nature and distinct characteristics. To enhance its quality, we have developed a novel workflow framework that incorporates customized text cleanup, fine-tuning of the Bert model, segmentation techniques, and a Chinese enterprise name database. This framework enables us to classify the emotions conveyed in news articles, identify negative financial events, and recognize relevant entities within Chinese financial news texts.

The accuracy of these classification models aligns with their application requirements. Building upon this foundation, we have designed and implemented a comprehensive system for analyzing Chinese financial news texts throughout their entire lifecycle. This system comprises three main modules: the financial news collection module, the financial news analysis module, and the financial news standardization and persistence module.

To ensure scalability and sustainable optimization of our system, we have employed an asynchronous design approach between the financial news collection module and the financial news analysis module. This strategic decision allows for seamless integration while accommodating future growth opportunities.[10]

Kim et al. delved into the fascinating realm of corporate bankruptcy prediction. The aim was to explore whether employing context-specific textual sentiment analysis, specifically BERT, could enhance the accuracy of these predictions. To conduct their study, they meticulously gathered and analyzed data from various sources, including five financial variables derived from stock market data and annual reports, which have been identified as precursors to impending insolvencies.[3]

Additionally, we embarked on a comprehensive examination of a vast collection of MDA narrative disclosures spanning from 1995 to 2020. The objective was to investigate whether incorporating textual sentiment analysis could offer valuable insights into predicting financial distress. The findings were remarkable: textual sentiment analysis demonstrated an augmented predictive capability beyond the well-established financial variables commonly used in such analyses.

Moreover, the study revealed that BERT-based analysis outperformed the dictionary-based approach proposed by Loughran and McDonald (2011), as well as the Word2Vec-based analysis combined with convolution neural network.

This highlights the superior performance of BERT in this domain.

However, there is a challenge associated with domain shifting in current BERT models. To mitigate this limitation, they employed domain-adaptation techniques to fine-tune the existing financial BERT model. This not only reduced computational costs compared to retraining the entire model with a new corpus but also significantly improved prediction accuracy.

In conclusion, the research underscores the potential of contextual textual sentiment analysis for enhancing corporate bankruptcy prediction accuracy. By leveraging advanced techniques like BERT and addressing domain-specific challenges through domain adaptation, we can unlock further insights into predicting financial distress with greater precision and reliability.

3. Methodology

The study delves into a two-pronged approach. Firstly, it examines the historical financial sentiment using the Loughran-McDonald dictionary to gain insights into past market sentiments as seen in Fig 1.[5] Secondly, it leverages BERT to develop predictive models that can anticipate future market movements. By combining these techniques, professionals in finance, analysts, and investors will have a comprehensive toolkit at their disposal to make informed decisions in a dynamic market. This includes risk management, formulating investment strategies, and proactive decision-making support in today's intricate financial landscape. This approach bridges the gap between evaluating historical sentiment and creating predictive models.

The methodology itself commences with preprocessing and extracting features from financial reports. Subsequently, BERT is fine-tuned for sentiment analysis.

3.1 Dictionary-based approach

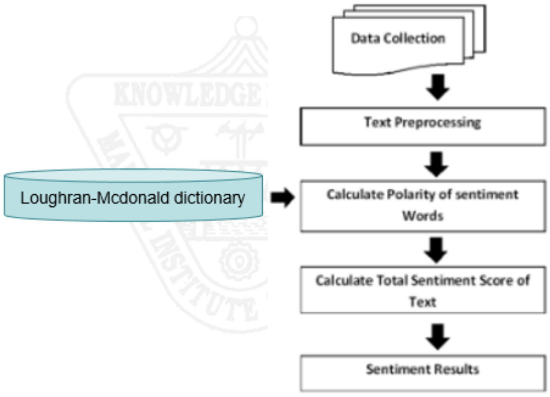
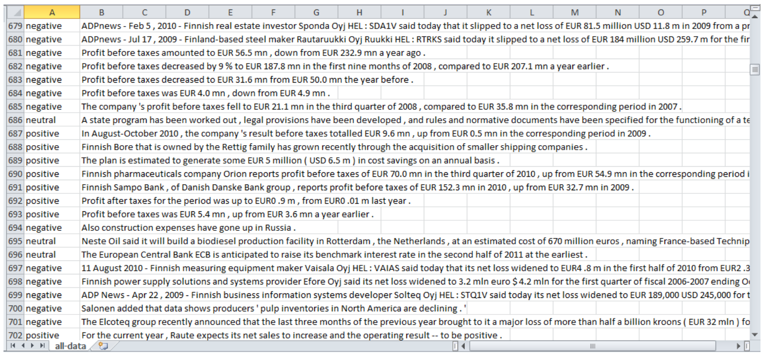

Data Collection
Text Preprocessing
Loughran-Mcdonald dictionary
Calculate Polarity of sentiment Words
Calculate Total Sentiment Score of Text
Sentiment Results

Fig. 1. Loughran-Mcdonald dictionary usage flow

In the realm of finance, Loughran and McDonald (2011) have developed word lists tailored specifically for this domain. These lists consist of both negative and positive words. To gauge the tone of textual disclosures, we follow their methodology by tallying the occurrences of positive and negative words in our dataset. These counts are then adjusted based on the total number of words in each category (DICTPOS and DICTNEG). While this analysis provides valuable information related to market sentiment, it is worth noting that these measures may lack accuracy as they do not take into account the context-specific tone of the texts.[6]

To conduct our study, we utilize a financial phrasebank dataset called "all-data.csv." shown in Fig 2. This dataset has been meticulously labeled by 16 researchers who possess extensive knowledge about financial markets. The sentiment labels assigned to each headline are categorized as either positive, neutral, or negative. In total, there are 4,837 sentiments captured in this dataset, all from the perspective of a retail investor.



	A	B
679	negative	ADPnews - Feb 5 , 2010 - Finnish real estate investor Sponda Oyj HEL : SDA1V said today that it slipped to a net loss of EUR 81.5 million USD 11.8 m in 2009 from a pr
680	negative	ADPnews - Jul 17 , 2009 - Finland-based steel maker Rautaruukki Oyj Ruukki HEL : RTRKS said today it slipped to a net loss of EUR 184 million USD 259.7 m for the fir
681	negative	Profit before taxes amounted to EUR 56.5 mn , down from EUR 232.9 mn a year ago .
682	negative	Profit before taxes decreased by 9 % to EUR 187.8 mn in the first nine months of 2008 , compared to EUR 207.1 mn a year earlier .
683	negative	Profit before taxes decreased to EUR 31.6 mn from EUR 50.0 mn the year before .
684	negative	Profit before taxes was EUR 4.0 mn , down from EUR 4.9 mn .
685	negative	The company 's profit before taxes fell to EUR 21.1 mn in the third quarter of 2008 , compared to EUR 35.8 mn in the corresponding period in 2007 .
686	neutral	A state program has been worked out , legal provisions have been developed , and rules and normative documents have been specified for the functioning of a te
687	positive	In August-October 2010 , the company 's result before taxes totalled EUR 9.6 mn , up from EUR 0.5 mn in the corresponding period in 2009 .
688	positive	Finnish Bore that is owned by the Rettig family has grown recently through the acquisition of smaller shipping companies .
689	positive	The plan is estimated to generate some EUR 5 million (USD 6.5 m) in cost savings on an annual basis .
690	positive	Finnish pharmaceuticals company Orion reports profit before taxes of EUR 70.0 mn in the third quarter of 2010 , up from EUR 54.9 mn in the corresponding period i
691	positive	Finnish Sampo Bank , of Danish Danske Bank group , reports profit before taxes of EUR 152.3 mn in 2010 , up from EUR 32.7 mn in 2009 .
692	positive	Profit after taxes for the period was up to EUR0 .9 m , from EUR0 .01 m last year .
693	positive	Profit before taxes was EUR 5.4 mn , up from EUR 3.6 mn a year earlier .
694	negative	Also construction expenses have gone up in Russia .
695	neutral	Neste Oil said it will build a biodiesel production facility in Rotterdam , the Netherlands , at an estimated cost of 670 million euros , naming France-based Technip
696	neutral	The European Central Bank ECB is anticipated to raise its benchmark interest rate in the second half of 2011 at the earliest .
697	negative	11 August 2010 - Finnish measuring equipment maker Vaisala Oyj HEL : VAIAS said today that its net loss widened to EUR4 .8 m in the first half of 2010 from EUR2 .3
698	negative	Finnish power supply solutions and systems provider Efore Oyj said its net loss widened to 3.2 mln euro $ 4.2 mln for the first quarter of fiscal 2006-2007 ending O
699	negative	ADP News - Apr 22 , 2009 - Finnish business information systems developer Solteq Oyj HEL : STQ1V said today its net loss widened to EUR 189,000 USD 245,000 for t
700	negative	Salonen added that data shows producers ' pulp inventories in North America are declining . '
701	negative	The Elcoteq group recently announced that the last three months of the previous year brought to it a major loss of more than half a billion kroons (EUR 32 mln) fo
702	positive	For the current year , Raute expects its net sales to increase and the operating result -- to be positive .

Fig. 2. Financial phrasebank dataset

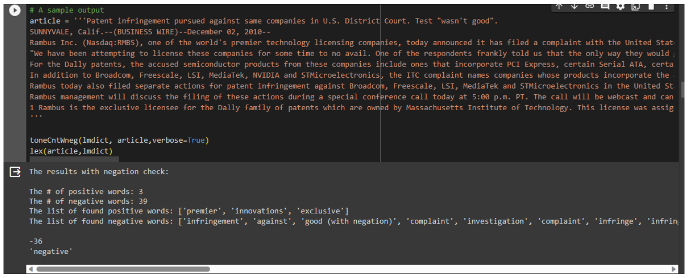
We employ k-fold Cross-Validation, which involves dividing the training set into k smaller subsets. To achieve this, we follow a specific procedure for each of the k "folds": 1. We create a KFold object and utilize it to iterate over the data, acquiring train/test indices for each fold. 2. The subsequent steps are implemented on k-1 of the folds as training data.

1) ***Pre-processing****:* When it comes to preparing text for analysis, preprocessing plays a crucial role in cleaning and refining the data. It involves reducing noise and inconsistencies to ensure that the text can be effectively utilized for tasks such as text mining or sentiment analysis.

The initial step in preprocessing is tokenization, where the text is broken down into individual components called tokens. These tokens can be words, phrases, symbols, or even entire sentences. Punctuation marks and other unnecessary characters are discarded during this process. Additionally, all the text in the documents is converted to lowercase format.

Another important aspect of preprocessing is the negation check. This involves identifying negate words (e.g., isn't, not, never) within three words preceding a positive word. When a negate word is found, it flips the subsequent positive word into a negative one. It's worth noting that negation check only applies to positive words since it's uncommon for double negations (i.e., negate word preceding a negative word) to occur according to Loughran and McDonald.

*2) **Calculate sentiment of tokens and total sentiment of the headings:*** Begin by applying the Loughran-McDonald Dictionary to preprocess the text. For each word in the report, check if it appears in the dictionary and determine its sentiment category, such as positive, negative, or uncertainty. Calculate sentiment scores for each category based on the number of words falling into each. Finally, aggregate these scores to generate an overall sentiment score for the entire document as seen in Fig 3.



```
# A sample output
article = '''Patent infringement pursued against same companies in U.S. District Court. Test "wasn't good".
SUNNYVALE, Calif.--(BUSINESS WIRE)--December 02, 2010--
Rambus Inc. (Nasdaq:RMBS), one of the world's premier technology licensing companies, today announced it has filed a complaint with the United Stat
"We have been attempting to license these companies for some time to no avail. One of the respondents frankly told us that the only way they would
For the Dally patents, the accused semiconductor products from these companies include ones that incorporate PCI Express, certain Serial ATA, certa
In addition to Broadcom, Freescale, LSI, MediaTek, NVIDIA and STMicroelectronics, the ITC complaint names companies whose products incorporate the
Rambus today also filed separate actions for patent infringement against Broadcom, Freescale, LSI, MediaTek and STMicroelectronics in the United St
Rambus management will discuss the filing of these actions during a special conference call today at 5:00 p.m. PT. The call will be webcast and can
1 Rambus is the exclusive licensee for the Dally family of patents which are owned by Massachusetts Institute of Technology. This license was assig
'''


toneCntWneg(lmdict, article,verbose=True)
lex(article,lmdict)
```

```
The results with negation check:

The # of positive words: 3
The # of negative words: 39
The list of found positive words: ['premier', 'innovations', 'exclusive']
The list of found negative words: ['infringement', 'against', 'good (with negation)', 'complaint', 'investigation', 'complaint', 'infringe', 'infrin

-36
'negative'
```

Fig. 3. Sentiment of sample article based on Loughran-McDonald dictionary

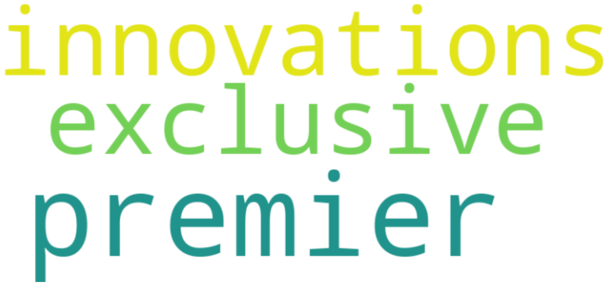


Fig. 4.

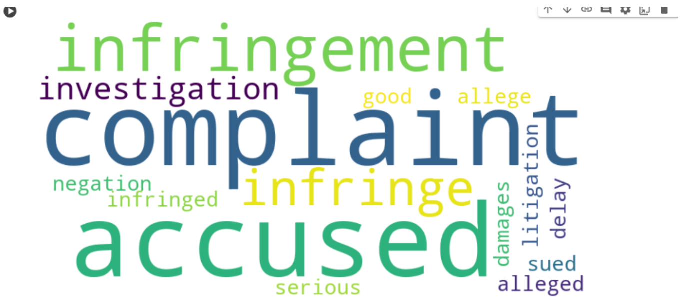
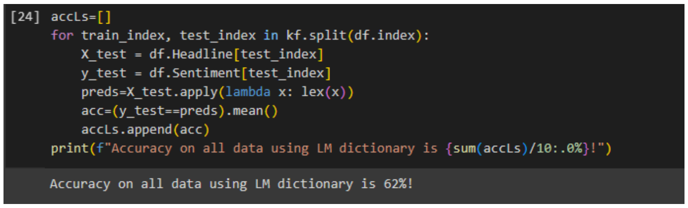


Fig. 5.

The model's validation process involves using the remaining data as a test set to calculate a performance measure like accuracy. This performance measure is then determined by taking the average of the values computed in each iteration of k-fold cross-validation given in Fig 4.

3.2 Bert

*1) **About BERT:*** BERT, which stands for Bidirectional Encoders Representations from Transformers, is a model that allows us to understand text by looking both backward and forward. The Attention Is All You Need paper introduced the Transformer model, which reads entire sequences of tokens simultaneously. Unlike LSTMs, which read sequentially in one direction, the Transformer is non-directional. Through the attention mechanism, the Transformer can learn contextual relationships between words. For example, it can be understood that ”his” in a sentence refers to ”Jim”. The ELMO paper further enhanced this idea by introducing pre-trained contextualized word embeddings. This allows words like ”nails” to have different meanings based on their context, such as referring to fingernails or metal nails.[7]



```
[24] accLs=[]
     for train_index, test_index in kf.split(df.index):
         X_test = df.Headline[test_index]
         y_test = df.Sentiment[test_index]
         preds=X_test.apply(lambda x: lex(x))
         acc=(y_test==preds).mean()
         accLs.append(acc)
     print(f"Accuracy on all data using LM dictionary is {sum(accLs)/10:.0%}!")

     Accuracy on all data using LM dictionary is 62%!
```

Fig. 6. Accuracy of dictionary-based approach

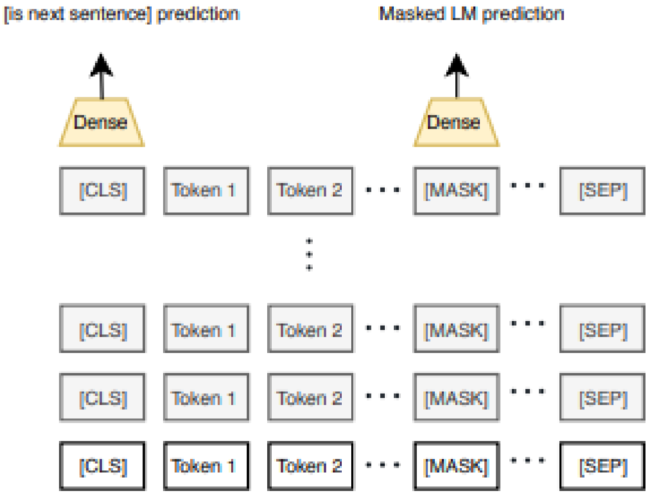


Fig. 7. Bert architecture

There are two types of tasks involved in this process. The first is the classification task, where we determine which category the input sentence belongs to. The second is the Next Sentence Prediction task, where we determine if the second sentence follows naturally from the first sentence. To represent the input sequence, token and position embeddings are used. Two additional tokens, [CLS] and [SEP], as seen in Fig 5 are added at the beginning and end of the sequence respectively. The [CLS] token is used for all classification tasks, including next sentence prediction.

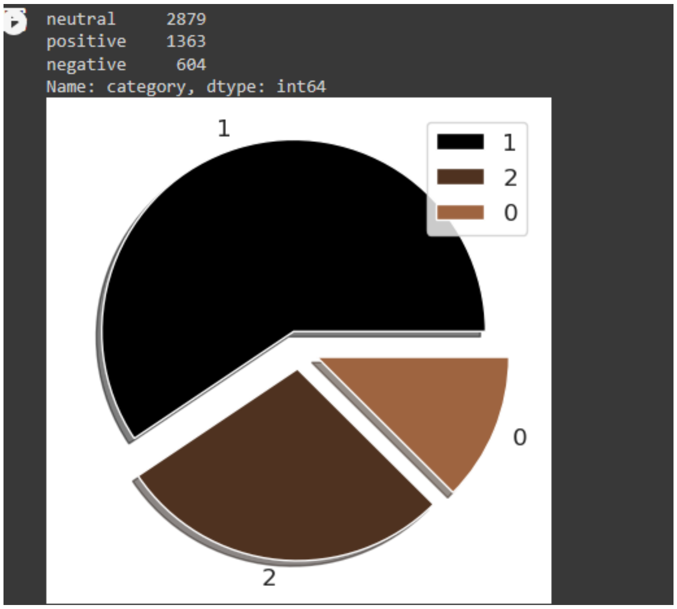
During pre-training, BERT ”masks” a randomly selected 15 percent of all tokens with [MASK] to hide them from the model. BERT is a language model that uses bidirectional transformers and can be applied to downstream tasks after supervised fine-tuning with limited resources.

Similarly, for each sentence in a document, the model assigns probabilities to three classes: positive, negative, and neutral. We then sum up these probabilities for all sentences in a document and normalize them to calculate the sentiment score of each document (referred to as BERTPOS and BERTNEG).[1]

*2) **Steps followed**:* 1. To tackle the Financial Sentiment Problem, we begin by importing all the necessary libraries. In order to implement BERT, we utilize the Bert tokenizer and Huggingface Transformers package with Pytorch. To adhere to BERT's maximum length limit of 512, we set the max sentence length accordingly.

2. Moving on to data handling, we load the dataset using pandas. This allows us to easily access and manipulate the data for analysis.

3. To gain insights from the sentiment labels in our dataset, we employ visualization techniques using libraries such as seaborn and matplot as seen in Fig 6. These tools help us effectively present and interpret the sentiment information in a visually appealing manner.



```
neutral     2879
positive    1363
negative     604
Name: category, dtype: int64
```

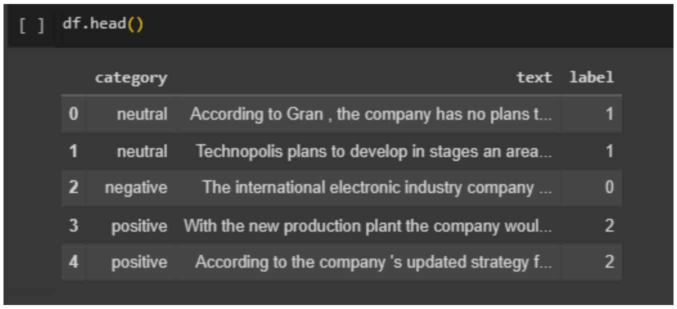
Fig. 8. Sentiment label visualization

4. When it comes to the sentiment label, we have made a change from text to integers shown in Fig 7.

5. Next, we split a given string or text into a list of tokens. BERT utilizes a tokenizer called Word Piece, which divides words into either their complete forms (where one word becomes one token) or into multiple tokens if needed creating a list (token lens) that stores the lengths of tokenized sequences for each text in the 'text' column of the DataFrame. To analyze the distribution of token lengths in a dataset Fig 8.

6. To facilitate text classification, we create a PyTorch dataset.

Fig. 9.



```
[ ] df.head()
```

	category	text	label
0	neutral	According to Gran , the company has no plans t...	1
1	neutral	Technopolis plans to develop in stages an area...	1
2	negative	The international electronic industry company ...	0
3	positive	With the new production plant the company woul...	2
4	positive	According to the company 's updated strategy f...	2

Fig. 10. Converting label to integers

This dataset takes in text and label data, as well as a tokenizer and a maximum length. It provides methods to retrieve the length of the dataset (len) and retrieve a specific item from the dataset (getitem).

Additionally, it includes a function that tokenizes all sentences by defining parameters inside encodeplus. Convert the text and label at the specified index (item) to Python variables. Use the provided tokenizer (self.tokenizer) to tokenize and format the input sequences, including special tokens and handling max length constraints. Flatten the input IDs and attention mask to ensure they are one-dimensional. Return a dictionary with the processed information like inputids which are frequently the only parameters required for model input. They are token indices that represent the numerical values of tokens in sequences used as input by the model. Attentionmask is used to indicate which tokens should be given attention and which should be ignored. This mask indicates which tokens are actual words (1) and which are padding (0).

7.For splitting your DataFrame (df) into training, validation, and test sets, we utilize scikit-learn. It is shown in Fig 9 The test set (dftest) is separate, while the validation set (dfval) is derived from the remaining portion of the original test set.

8.To efficiently load and iterate over batches of data during training or evaluation , we use DataLoader in conjunction with DataSet class to create PyTorch datasets. We create train,test,validation data loaders.

9. Moving on to model-building, we utilize a pre-trained model that serves as the base architecture for BERT. This pre-trained model has been trained on uncased English text.

10. The SentimentClassifier class defines our sentiment

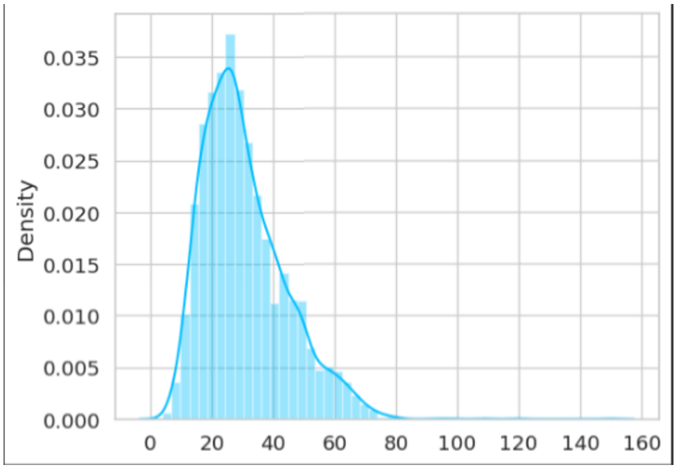
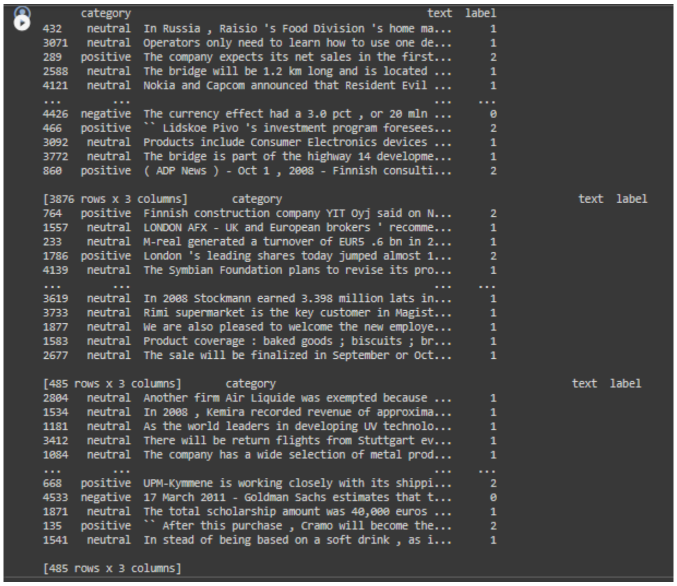


Fig. 11. Distribution of token lengths in a dataset



```
       category                                             text  label
432     neutral  In Russia , Raisio 's Food Division 's home ma...      1
3071    neutral  Operators only need to learn how to use one de...      1
289    positive  The company expects its net sales in the first...      2
2588    neutral  The bridge will be 1.2 km long and is located ...      1
4121    neutral  Nokia and Capcom announced that Resident Evil ...      1
...         ...                                                ...    ...
4426   negative  The currency effect had a 3.0 pct , or 20 mln ...      0
466    positive  `` Lidskoe Pivo 's investment program foresees...      2
3092    neutral  Products include Consumer Electronics devices ...      1
3772    neutral  The bridge is part of the highway 14 developme...      1
860    positive  ( ADP News ) - Oct 1 , 2008 - Finnish consulti...      2

[3876 rows x 3 columns]       category                                             text  label
764    positive  Finnish construction company YIT Oyj said on N...      2
1557    neutral  LONDON AFX - UK and European brokers ' recomme...      1
233     neutral  M-real generated a turnover of EUR5 .6 bn in 2...      1
1786   positive  London 's leading shares today jumped almost 1...      2
4139    neutral  The Symbian Foundation plans to revise its pro...      1
...         ...                                                ...    ...
3619    neutral  In 2008 Stockmann earned 3.398 million lats in...      1
3733    neutral  Rimi supermarket is the key customer in Magist...      1
1877    neutral  We are also pleased to welcome the new employe...      1
1583    neutral  Product coverage : baked goods ; biscuits ; br...      1
2677    neutral  The sale will be finalized in September or Oct...      1

[485 rows x 3 columns]       category                                             text  label
2804    neutral  Another firm Air Liquide was exempted because ...      1
1534    neutral  In 2008 , Kemira recorded revenue of approxima...      1
1181    neutral  As the world leaders in developing UV technolo...      1
3412    neutral  There will be return flights from Stuttgart ev...      1
1084    neutral  The company has a wide selection of metal prod...      1
...         ...                                                ...    ...
668    positive  UPM-Kymmene is working closely with its shippi...      2
4533   negative  17 March 2011 - Goldman Sachs estimates that t...      0
1871    neutral  The total scholarship amount was 40,000 euros ...      1
135    positive  `` After this purchase , Cramo will become the...      2
1541    neutral  In stead of being based on a soft drink , as i...      1

[485 rows x 3 columns]
```

Fig. 12. Training,Validation and test sets

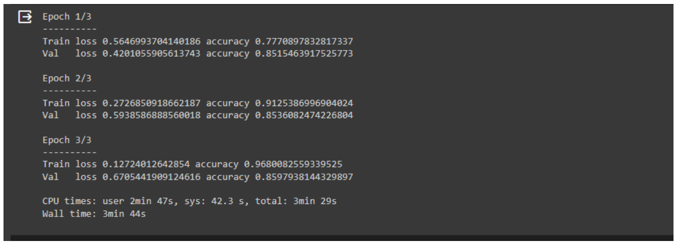
analysis model using BERT. It incorporates a dropout layer (self.drop) with a dropout probability of 0.3 . Adds a linear layer (self.out) with mapping from the BERT hidden size to the number of classes (nclasses). The forward method handles the forward pass of our model. By applying BERT to the input data, both hidden states and pooled output are obtained. In this case, only pooled output is utilized. The pooled output goes through the dropout layer which is a regularization technique that randomly sets a fraction of input units to zero during training in order to prevent overfitting and subsequently passed through linear layer, which performs a linear transformation based on learned weights. This generates the final output.

11.Then we define a loss function (such as cross-entropy loss) and an optimizer (such as Adam) for model optimization.[4]

12. We define training and evaluation loop for one epoch.

4. Result Analysis

When it comes to training a sentiment analysis model, it is important to consider a few key factors. One of these factors is the number of epochs that the model should be trained for. It is also essential to monitor the performance of the model on both the training and validation sets shown in Fig 10.



```
Epoch 1/3
----------
Train loss 0.5646993704140186 accuracy 0.7770897832817337
Val   loss 0.4201055905613743 accuracy 0.8515463917525773

Epoch 2/3
----------
Train loss 0.2726850918662187 accuracy 0.9125386996904024
Val   loss 0.5938586888560018 accuracy 0.8536082474226804

Epoch 3/3
----------
Train loss 0.12724012642854 accuracy 0.9680082559339525
Val   loss 0.6705441909124616 accuracy 0.8597938144329897

CPU times: user 2min 47s, sys: 42.3 s, total: 3min 29s
Wall time: 3min 44s
```

Fig. 13. Result of training and evaluating the model

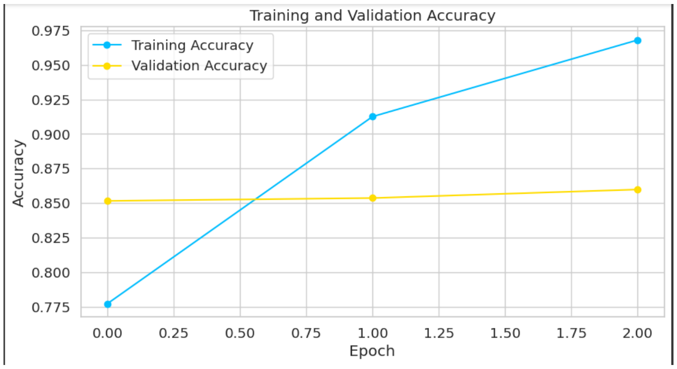


Fig. 14. Training and validation accuracy

As seen in Fig 11 and Fig 12 We use Matplotlib to create two plots: one for accuracy and another for loss. The x-axis represents the epochs, and the y-axis represents either accuracy or loss. It will provide visual insights into how the model is learning over time. Ideally, you want to see the training accuracy increasing and the training loss decreasing. The validation metrics will help you understand how well your model generalizes to unseen data.

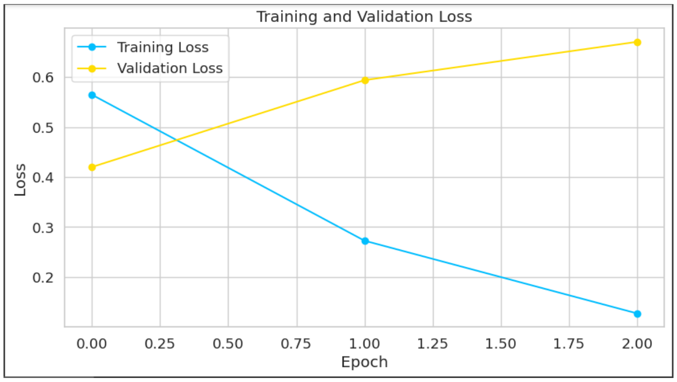
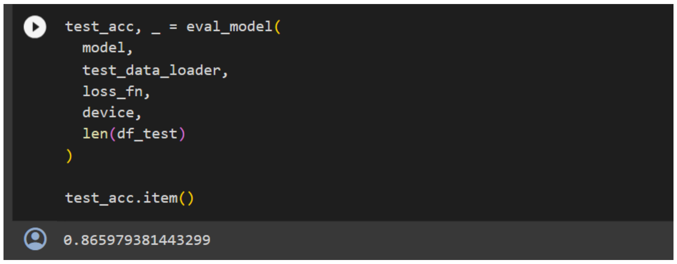


Fig. 15. Training and validation loss

In order to evaluate the effectiveness of the model, it is necessary to test it on a separate test set and extract accuracy metrics from this evaluation given in Fig 13.



```
test_acc, _ = eval_model(
  model,
  test_data_loader,
  loss_fn,
  device,
  len(df_test)
)

test_acc.item()
```

```
0.865979381443299
```

Fig. 16. Evaluating the model on the test set and extracting the accuracy.

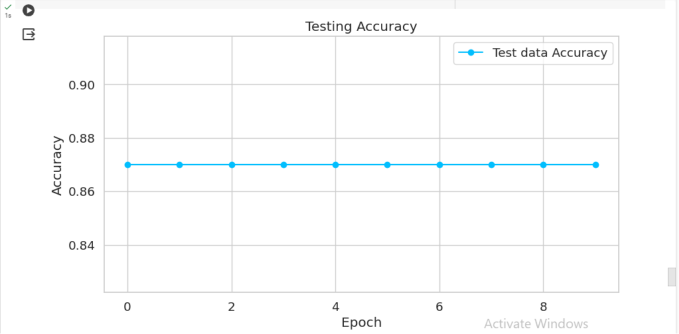


Fig. 17. Test data Accuracy

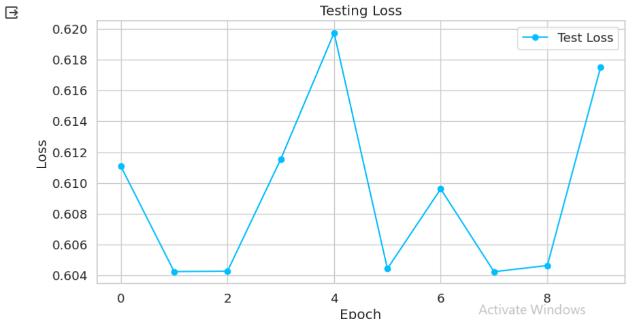
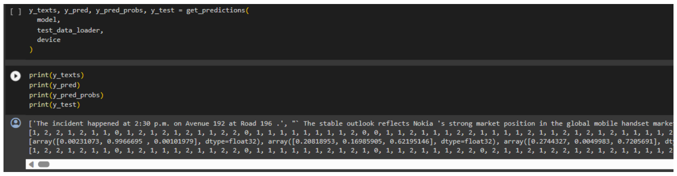


Fig. 18. Test data Loss

This involves obtaining predictions, prediction probabilities, and real values from the model using a test data loader shown in fig 14.



```
y_texts, y_pred, y_pred_probs, y_test = get_predictions(
  model,
  test_data_loader,
  device
)
```

```
print(y_texts)
print(y_pred)
print(y_pred_probs)
print(y_test)
```

```
['The incident happened at 2:30 p.m. on Avenue 192 at Road 196 .', "` The stable outlook reflects Nokia 's strong market position in the global mobile handset marke
[1, 2, 2, 1, 2, 1, 1, 0, 1, 2, 1, 2, 1, 2, 1, 1, 2, 2, 0, 1, 1, 1, 1, 1, 1, 1, 1, 2, 0, 0, 1, 1, 2, 1, 1, 1, 2, 2, 1, 1, 1, 1, 2, 1, 1, 2, 1, 2, 1, 2, 1, 1, 1, 1, 2
[array([0.00231073, 0.9966695 , 0.00101979], dtype=float32), array([0.20818953, 0.16985905, 0.62195146], dtype=float32), array([0.2744327, 0.0049983, 0.7205691], dt
[1, 2, 2, 1, 2, 1, 1, 0, 1, 2, 1, 1, 1, 2, 1, 1, 2, 2, 0, 1, 1, 1, 1, 1, 1, 2, 1, 2, 1, 0, 1, 1, 2, 1, 1, 1, 2, 2, 0, 2, 1, 1, 2, 1, 2, 2, 1, 2, 1, 2, 1, 1, 1, 1, 2
```

Fig. 19. Predictions for test data set

Here are the key findings from the classification report as seen in Fig 15:

1. Precision: With a precision score of 0.89, we can conclude that 89 percent of the predicted negative instances were correct. This measures the accuracy of the positive predictions. Formula: Precision = TP / (TP + FP)

2. Recall: The recall score of 0.88 indicates that 88 percent of the actual negative instances were captured by our model. This measures the ability to identify all relevant instances. Formula: Recall = TP / (TP + FN)

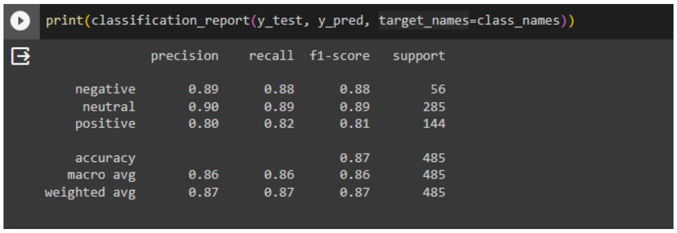
3. F1-Score: The F1-Score, which is a harmonic mean of precision and recall, is determined to be 0.88. It provides a balanced measure between precision and recall, especially when there is an imbalance between classes. Formula: F1-Score = 2 * (Precision * Recall) / (Precision + Recall)

4. Support: In our dataset, there are a total of 56 instances in the negative class. This represents the number of actual occurrences for each class.

5. Accuracy: The overall accuracy score is calculated to be 0.87, meaning that 87 percent of all predictions made by our model were correct. Formula: Accuracy = (TP + TN) / (TP + TN + FP + FN)

6. Macro Avg F1-Score: With an average F1-score across classes at 0.86, this metric calculates individual metrics for each class and then takes their average without considering class size differences.

7. Weighted Avg F1-Score: The weighted average F1-score is determined to be 0.87, taking into account the class imbalance in our data by assigning weights based on each class's presence in true data samples.



```
print(classification_report(y_test, y_pred, target_names=class_names))
```

```
              precision    recall  f1-score   support

    negative       0.89      0.88      0.88        56
     neutral       0.90      0.89      0.89       285
    positive       0.80      0.82      0.81       144

    accuracy                           0.87       485
   macro avg       0.86      0.86      0.86       485
weighted avg       0.87      0.87      0.87       485
```

Fig. 20. Classification report

When it comes to evaluating the model's performance across different classes, two important metrics to consider are F1-Scores and accuracy. These metrics provide a comprehensive assessment of how well the model is performing on this specific dataset. The high F1-Scores and accuracy indicate that the model is doing well in terms of its performance on this particular dataset.

A confusion matrix in Fig 16 is a table that is often used to describe the performance of a classification model on a set of test data for which the true values are known. In a confusion matrix, each row represents the instances in an actual class, while each column represents the instances in a predicted class. The diagonal elements (49, 253, 118) represent the correct predictions for each class (True Positives). The off-diagonal elements represent misclassifications.

In Fig 18 We are encoding a text in the dataset using the encodeplus method from the Hugging Face Transformers

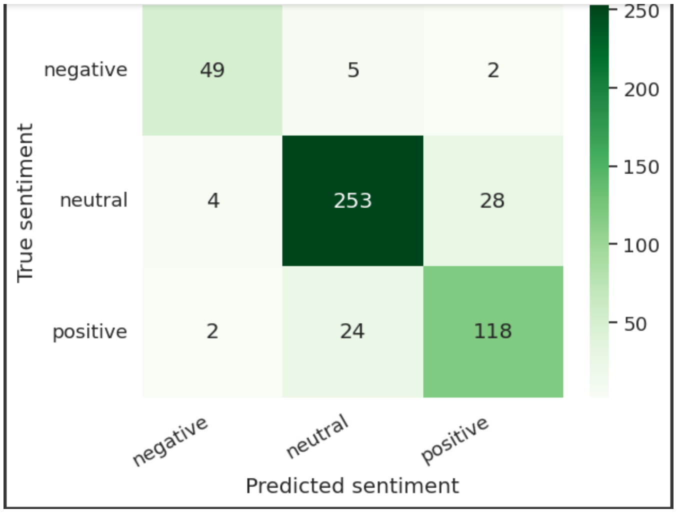
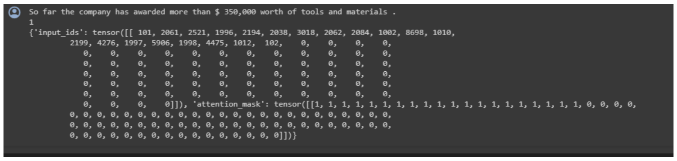


Fig. 21. Confusion Matrix

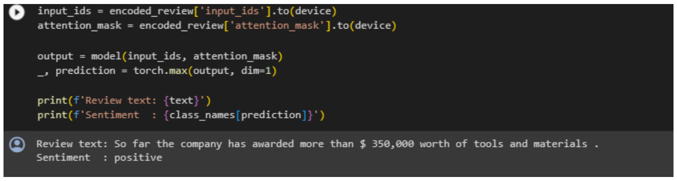
library. This method is commonly used to convert raw text into tokenized and formatted input that can be fed into a pre-trained transformer model.



```
So far the company has awarded more than $ 350,000 worth of tools and materials .
1
{'input_ids': tensor([[ 101, 2061, 2521, 1996, 2194, 2038, 3018, 2062, 2084, 1002, 8698, 1010,
         2199, 4276, 1997, 5906, 1998, 4475, 1012,  102,    0,    0,    0,    0,
            0,    0,    0,    0,    0,    0,    0,    0,    0,    0,    0,    0,
            0,    0,    0,    0,    0,    0,    0,    0,    0,    0,    0,    0,
            0,    0,    0,    0,    0,    0,    0,    0,    0,    0,    0,    0,
            0,    0,    0,    0,    0,    0,    0,    0,    0,    0,    0,    0,
            0,    0,    0,    0,    0,    0,    0,    0,    0,    0,    0,    0,
            0,    0,    0,    0]]), 'attention_mask': tensor([[1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 0, 0, 0, 0,
         0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
         0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
         0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0]])}
```

Fig. 22. Tokenizing a particular random text in the dataset giving the output dictionary of tensors, including the input IDs, attention mask, etc., that can be used as input for prediction.

Using the encoded output with a pre-trained model to make sentiment predictions in Fig 19.



```
input_ids = encoded_review['input_ids'].to(device)
attention_mask = encoded_review['attention_mask'].to(device)

output = model(input_ids, attention_mask)
_, prediction = torch.max(output, dim=1)

print(f'Review text: {text}')
print(f'Sentiment  : {class_names[prediction]}')
```

```
Review text: So far the company has awarded more than $ 350,000 worth of tools and materials .
Sentiment  : positive
```

Fig. 23. Prediction for the raw data

Comparing Fig 18 and 19 We can see that the true sentiment is 1 - neutral and predicted is 2 - positive hence it is one of the erroneous prediction.

5. Future Scope

To enhance the research, it is possible to extend it by incorporating a sentiment analysis model that combines the Loughran McDonald dictionary with bert. This combined model would possess the ability to comprehend the intricacies of financial data and effectively extract sentiment. Moreover, it would be resilient against market manipulation and noise, leading to more accurate predictions regarding the future impact on financial markets based on identified sentiments.

6. Conclusion

Sentiment analysis in the field of finance is a promising avenue of study, offering practical applications and the potential to enhance decision-making and risk management within the financial industry.

There exist several research challenges when it comes to sentiment analysis in finance. These include effectively handling noisy data, addressing issues related to market manipulation, and ensuring the accuracy, reliability, and contextual understanding of sentiment data sources. By overcoming these challenges, sentiment analysis in finance can play a significant role in empowering investors, traders, and financial institutions to make more informed choices.

The impact of sentiment analysis in finance extends beyond individual stakeholders. It has the potential to contribute to more efficient and stable financial markets as a whole. Additionally, it can aid in mitigating the risk of financial crises. In this regard, sentiment analysis holds great promise for positively shaping our world by fostering better decision-making practices within the realm of finance.

References

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

[2] Fatehjeet Kaur Chopra and Rekha Bhatia. Sentiment analyzing by dictionary based approach. *International Journal of Computer Applications*, 152(5):32–34, 2016.

[3] Alex G Kim and Sangwon Yoon. Corporate bankruptcy prediction with domain-adapted bert. In *EMNLP 2021, 3rd Workshop on ECONLP*, 2021.

[4] Menggang Li, Wenrui Li, Fang Wang, Xiaojun Jia, and Guangwei Rui. Applying bert to analyze investor sentiment in stock market. *Neural Computing and Applications*, 33:4663–4676, 2021.

[5] Tim Loughran and Bill McDonald. The use of word lists in textual analysis. *Journal of Behavioral Finance*, 16(1):1–11, 2015.

[6] Tim Loughran and Bill McDonald. Textual analysis in finance. *Annual Review of Financial Economics*, 12:357–375, 2020.

[7] Muhammad Talha Riaz, Muhammad Shah Jahan, Sajid Gul Khawaja, Arslan Shaukat, and Jahan Zeb. Tm-bert: A twitter modified bert for sentiment analysis on covid-19 vaccination tweets. In *2022 2nd International Conference on Digital Futures and Transformative Technologies (ICoDT2)*, pages 1–6, 2022.

[8] Gim Hoy Soong and Chye Cheah Tan. Sentiment analysis on 10-k financial reports using machine learning approaches. In *2021 IEEE 11th International Conference on System Engineering and Technology (ICSET)*, pages 124–129. IEEE, 2021.

[9] Soonh Taj, Baby Bakhtawer Shaikh, and Areej Fatemah Meghji. Sentiment analysis of news articles: A lexicon based approach. In *2019 2nd international conference*

on computing, mathematics and engineering technologies (iCoMET), pages 1–5. IEEE, 2019.

[10] Zhixiong Tan, Bihuan Chen, and Wei Fang. Analysis and application of financial news text in chinese based on bert model. In *Proceedings of the 2020 Asia Service Sciences and Software Engineering Conference*, pages 35–39, 2020.

Contribution of Individual Authors to the Creation of a Scientific Article (Ghostwriting Policy)

The authors equally contributed in the present research, at all stages from the formulation of the problem to the final findings and solution.

Sources of Funding for Research Presented in a Scientific Article or Scientific Article Itself

No funding was received for conducting this study.

Conflict of Interest

The authors have no conflicts of interest to declare that are relevant to the content of this article.