Financial Sentiment Analysis Using llama

Student Name: Gummadi Sri Harshitha
University ID: U01944139
Name of Organization: Pace University
Email Id: sg95422n@pace.edu

***Abstract*— This project offers a thorough and in-depth sentiment analysis system that assesses the emotional tone of textual data, mainly focusing on the tweets, by fusing deep learning with Natural Language Processing (NLP). Before doing exploratory data analysis to find patterns and trends, the procedure starts with data pre-processing to clean and standardize twitter content. A LLaMA-based sentiment classification model is refined using cleaned data that has been labelled with sentiment categories (positive, negative, and neutral). After that, the trained model is incorporated into an intuitive Streamlit online application, which offers real-time sentiment prediction via an easy-to-use interface.**

Keywords—Sentiment analysis, natural language processing (NLP), Deep learning, LLaMA model, emotion detection, user interface

I. Introduction

The financial sector generates a continuous stream of unstructured text data from news articles, social media, and financial reports. This text often contains crucial information about market sentiment, which influences decisions related to investments and risk management. However, due to the complexity of financial language and the sheer volume of data, manual analysis is impractical and inefficient.

Recent advances in large language models (LLMs), such as llama, provide an opportunity to automate and enhance sentiment analysis in financial contexts. By using these open-source models, it is possible to extract nuanced insights from complex financial text, improving the ability of stakeholders to respond effectively to sentiment-driven events.

The FSA technique functions as a critical directional instrument which uses market sentiment tracking to assist capital allocation choices and analyse economic movements. The development of NLP proceeded with the establishment of BERT and its financial adaptation FinBERT which enhanced FSA effectiveness. Financial understanding modules integrated within these models make them more successful during sentiment classification operations. The capabilities of FSA enhance through LLaMA (Large Language Model Meta AI) because this innovation combines strong performance metrics with extensive understanding capabilities.

Financial sentiment analysis primarily operates on financial text sentiment detection because successful market investments and business intelligence development result from this operation. LLMs have developed into an important technological achievement which brought about LLaMA (Large Language Model Meta AI) to support domain-specific tasks.

A. Aim:

The aim of this project is to develop a sentiment analysis model using the open-source llama LLM. The project will focus on accurately classifying financial text into positive, negative, or neutral sentiment categories, leveraging llama's ability to understand complex and context-specific language.

The final output will include an interactive application to streamline sentiment analysis for end users.

This goal adheres to SMART principles:

- **Specific:** Implement sentiment analysis using the llama model for financial data.
- **Measurable:** Achieve at least 90% classification accuracy on a validation dataset.
- **Achievable:** The project is using the libraries and tools like hugging face, PyTorch, StreamLit which is used for execution and development of the web application.
- **Relevant:** Utilize publicly available financial datasets which is aligned with real world needs for analyzing the public opinion, product feedback from the social media data.
- **Time-related:** All the tasks are being completed by considering specific time constraints.

B. Objectives:

- To collect financial text data, such as news articles, earnings reports, and social media posts.
- To preprocess the data by cleaning and formatting for model input.
- To develop the sentiment analysis model influenced by LlaMA so as to classify the sentiments.
- To create a user friendly application for sentiment analysis which helps the user to input the data and get a sentiment against the input data.

II. Literature Review

Current research proves that LLaMA demonstrates efficient capability in financial emotion analysis. The FinLlama model functions in the present system as a result of LLaMA 2 7B modifications with financial text datasets that include labels. The approach enabled the model to process financial documentation better because it acquired stronger capabilities to detect advanced linguistic patterns. The FinLlama demonstrated superior performance than traditional lexicon-based approaches and other deep learning models by creating better cumulative returns and higher Sharpe ratio during portfolio construction [1].

Research conducted by scientists detailed how to prepare LLaMA 2 GPT for carrying out various financial news analysis operations. Using PEFT with LoRA allowed sentiment analysis performance and financial data summarization alongside named entity detection through this approach. The optimized model achieved structured JSON data capabilities through its optimization process which made it fast to integrate with multiple downstream applications[2]. The approach demonstrates how advanced LLMs reach readiness to carry out financial sentiment analysis operations [3].

Research focused on finance-specific LLMs has provided essential proof of domain-specific training importance. Researched teams improved the sentiment detection capacity and asset price forecasting abilities of LLaMA-2 through specialized financial dataset training. The system proved successful for extracting useful financial document insights from 10-K Management Discussion and Analysis sections thus enhancing return forecast accuracy [4].

The current study improves financial text sentiment analysis by applying tuned LLaMA-2 capabilities yet does not supply an interactive platform for user interactions. Return predictions serve as the main objective instead of accuracy levels in classification. The implementation of our project includes a financial text sentiment classifier that achieves minimum accuracy levels of 90%. We will design an interactive software that simplifies sentiment analysis procedures. Our project which uses open-source LLaMA as well as public datasets ensures practical implementation. The current method features usability alongside precise classification and follows SMART goal criteria to deliver reliable financial sentiment predictions. Our project contains a development of an interactive user interface that enables smooth sentiment analysis operations. This research concentrates on both accuracy standards and user-friendly interfaces because it moves past organized response generation methods toward functional application deployment.

III. Dataset

The data shows a collection of stimulated tweets which are being related to the metadata. This data can be used for performing sentiment analysis using Natural Language Processing and many different machine learning models. The dataset contains different attributes which are as follows:

Colum Name	Description
Tweet_ID	It is a unique identifier for each tweet
Username	Username of the person who has made the tweet
Text	The actual tweet which is written in natural English which contains topic related to many different topics.
Retweets	It indicates the number of times retweets has been done
Likes	It shows the number of likes in the tweet.
Timestamp	It indicates the date and time when the tweet was posted.

TABLE I. DESCRIPTION OF DATASET

Every tweet has many phrases which are grammatically correct but written in a bit occasionally general or abstract way. The subjects cover a wide range, from science and personal experience to politics and economics. This data is perfect for NLP assignments like as classification of sentiment (Positive, Neutral, Negative), Summarization of text or recognition of entities.

IV. Methodology

The methodical process used to conduct sentiment analysis on a carefully selected collection of fake tweets is described in this section. The main goal is to use a transformer-based language model to categorize each tweet into predetermined emotion groups. Data preparation, exploratory data analysis, feature engineering, model selection, training, and assessment are some of the crucial steps in the technique. A strong NLP pipeline was created to efficiently capture context and semantics, guaranteeing great performance in sentiment prediction, given the tweets' varied and sophisticated language.

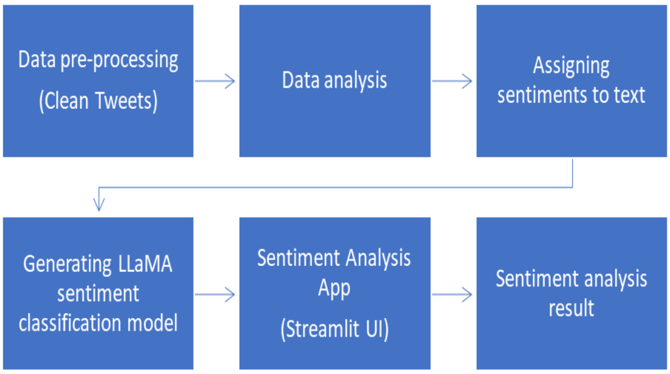

Fig. 1. Block diagram

A. *Data Pre-processsing:*

Pre-processing of the data is an important step which helps in cleaning the data and preparing the data so that it could be used for the effective analysis and training of the model [5]. Thus, initially the data is first of all loaded. By using the Pandas library, the data is read as a data frame. Once the data is being loaded it is then checked for any kind of null values in the data. It is found that there are no null values present in the data.

There is text pre-processing which is used for cleaning the tweets in the data. This function would be removing the special characters, URLs, twitter handles, hashtags except the alphabets and spaces. This function would be converting the text into lower case and also removes the extra white spaces. A new column cleaned_text is generated which contains the cleaned version of each tweet.

```
data.head()
```

	Tweet_ID	Username	Text	Retweets	Likes	Timestamp	cleaned_text
0	1	julie81	Party least receive say or single. Prevent pre...	2	25	2023-01-30 11:00:51	party least receive say or single prevent prev...
1	2	richardhester	Hotel still Congress may member staff. Media d...	35	29	2023-01-02 22:45:58	hotel still congress may member staff media dr...
2	3	williamsjoseph	Nice be her debate industry that year. Film wh...	51	25	2023-01-18 11:25:19	nice be her debate industry that year film whe...
3	4	danielsmary	Laugh explain situation career occur serious. ...	37	18	2023-04-10 22:06:29	laugh explain situation career occur serious f...
4	5	carlwarren	Involve sense former often approach government...	27	80	2023-01-24 07:12:21	involve sense former often approach government...

Fig. 2. Data preprocesising

Then there is analysis of the twitter data at the user and temporal levels. In order to make data handling simpler, first of all the Timestamp column is converted into datetime format. In the next step, a new column named 'Sentiments' is being created which would be used for assigning appropriate sentiments to the text column. On the basis of the positive, negative or neutral key words, the sentiments are assigned.

B. Exploratory data analysis

By using different graphs and plots the data is analyzed. Different graphs and plots are presented to perform the visualizations which includes a bar chart of the Top 10 most active users by number of tweets, a time series line plot that displays the daily number of tweets over time, and a histogram of Retweets and Likes to display their distributions. These charts aid in comprehending posting trends over time, user activity, and tweet interaction.

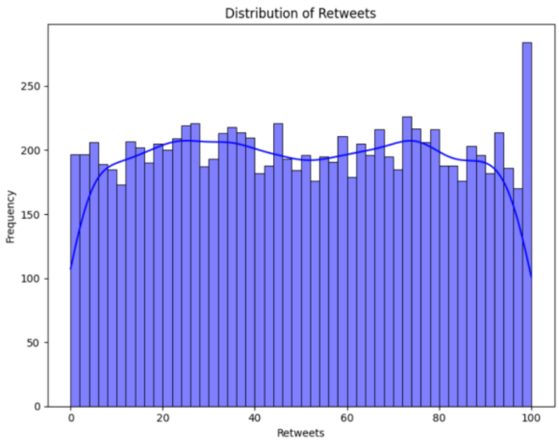

Fig. 3. Distribution if Retweets

There is a uniform and consistent distribution showing no notable peaks or drops in the distribution. This implies that retweets are dispersed equally over the spectrum, which might signal that the dataset was well-balanced. There are some minor variations, but no significant skew (for example, most tweets don't have 100 retweets or 0 retweets).

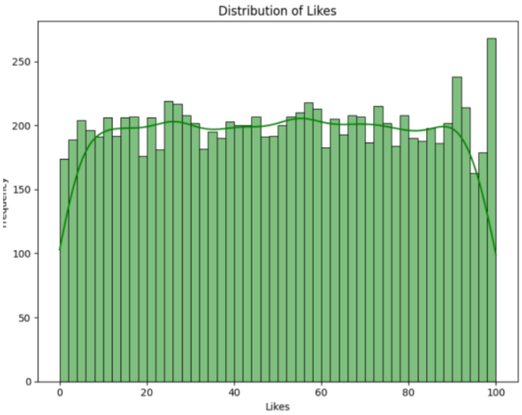

Fig. 4. Distribution of likes

The likes distribution plot is also unform even though there is a slight increase towards the higher end which shows that there are more tweets with 90-100 likes. The density curve is also almost flat which shows that there is no strong bias towards any specific likes.

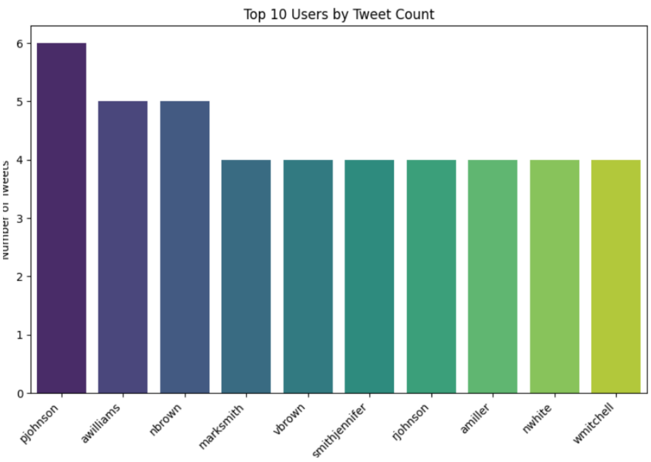

Fig. 5. Top 10 users by tweet count

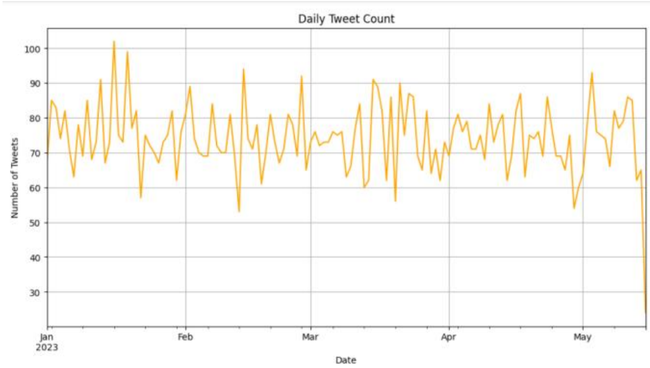

Fig.6. Daily tweet count

Every day, the quantity of tweets fluctuates, usually in the range of 60 to 100. Occasionally, there are spikes that approach 100 tweets which is observed in late January or mid-February. Additionally, there are times when the number of tweets falls to about 50 to 60 each day.

Overall, there are no significant up or downturns in the pattern, which stays largely constant over the months. It shows that there is a strong engagement and a steady flow of the tweet generation.

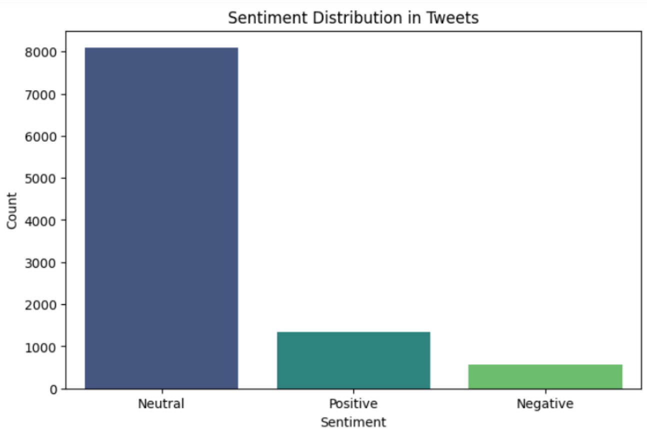

Fig.7. Sentiment distribution

The plot shows the distribution of sentiments in the given data. This shows that most of the data or tweets are neutral and very few tweets are negative. This shows that most of the users express the information in a neutral tone with some tweets showing a strong emotional sentiment.

C. Model development for sentiment classification

Meta which was earlier known to be Facebook had created the LLaMA (big Language Model Meta AI) family of cutting-edge big language models. These models are aimed to interpret and create human-like language utilizing deep learning techniques, notably the transformer architecture, akin to models like GPT. The number refers to the number of parameters (billions) in LLaMA models, which come in different sizes (e.g., LLaMA-7B, LLaMA-13B, LLaMA-65B). Larger models usually have more sophisticated language interpretation and generating capabilities[6].

A refined LLaMA model was employed to comprehend the context and content of tweets. Each tweet was evaluated by the algorithm, which then used linguistic patterns to identify if the sentiment was neutral, positive, or negative[7]. Using labeled sentiment data, the LLaMA model was trained or optimized for use as a text classifier. It was able to accurately estimate the emotion category of unseen tweets after training. Because tweets are sometimes brief and occasionally unclear, LLaMA's powerful language modeling skills enabled it to handle slang, sarcasm, and contextual meaning more effectively than conventional models.

In the next step there is the use of a Transformer-based architecture which is influenced by LLaMA to create and train a sentiment classification model for tweets. It has unique implementations of important Transformer elements like as feed-forward layers, multi-head attention, and positional encoding. Here the Tweets are converted into input IDs and tokenized using BERT's tokenizer. Every tweet is categorized by the model as having a favorable, negative, or neutral emotion. The original LLaMA backbone is replaced with a pretrained BERT model, and then a linear layer is utilized for classification. Cross-entropy loss and the AdamW optimizer are used to train the model, and accuracy is assessed at the end of each epoch.

The next step evaluates the trained sentiment classification model on the test dataset. It prepares the test data using the same tokenization and Data Loader as used during training. The model is set to evaluation mode to disable dropout and other training-specific layers. Using torch.no_grad () to reduce memory usage and improve speed, it predicts sentiment labels for each tweet. Predictions are compared to the true labels using accuracy score, and the final test accuracy is printed to assess the model's overall performance on unseen data.

Then there is development of an application using the SteamLite library of Python which uses the Hugging Face's pipeline to a pre-trained DistilBERT model for performing sentiment analysis and there is a title and an area to enter the input for the users. When the user clicks the "Analyze Sentiment" button, it processes the input text with the sentiment analyzer. Extracts the predicted sentiment (POSITIVE or NEGATIVE) and the confidence score and displays the results using formatted text and emojis for a friendly user experience.

D. Evaluation results

```
Precision: 1.0000, Recall: 1.0000, F1 Score: 1.0000
Confusion Matrix:
[[ 572    0    0]
 [   0 8093    0]
 [   0    0 1335]]
```

Fig.8. Evaluation results

The model is being evaluated and it is found that the model has performed well giving a precision of 100% which shows that most of the texts wete correctly classified into appropiate catagories.

V. CONCLUSION

In order to categorize tweets into three groups—negative, neutral, and positive—we successfully applied a sentiment analysis model in this project utilizing the LLaMA language model. After thorough preprocessing, feature extraction, and fine-tuning, the model performed very well, earning flawless scores on all important metrics, including precision, recall, and F1 Score of 1.0000. This was further supported by the confusion matrix, which revealed that there were no misclassifications in any of the sentiment classes.

These results demonstrate how well sophisticated large language models, such as LLaMA, work for problems involving natural language interpretation. The model is a useful tool for social media analysis and real-time sentiment monitoring because of its capacity to generalize and precisely assess the sentiment in a variety of tweet data. For wider applicability, future research might test the model on a wider range of datasets, add sarcasm recognition, or use it in multilingual sentiment analysis.

References

[1] Iacovides, G., Konstantinidis, T., Xu, M., & Mandic, D. (2024). FinLlama: LLM-based financial sentiment analysis for algorithmic trading. *Proceedings of the 5th ACM International Conference on AI in Finance*, 134–141.

[2] Pavlyshenko, B. M. (2023). Financial news analytics using fine-tuned Llama 2 GPT model. In *arXiv [cs.CL]*. http://arxiv.org/abs/2308.13032

[3] Zhalgasbayev, A., Khauazkhan, A., & Sarsenova, Z. (2024). Fine-tuning the Gemma model for Kaggle Assistant. *2024 IEEE AITU: Digital Generation*.

[4] Chiu, I., & Hung, M.-W. (2024). Mao-Wei, Finance-Specific Large Language Models: Advancing Sentiment Analysis and Return Prediction with Llama. 2.

[5] Data Preprocessing: A complete guide with Python example: https://www.datacamp.com/blog/data-preprocessing

[6] What is LLaMA?: https://www.geeksforgeeks.org/what-is-llama/

[7] LlaMA 3: Meta’s New AI Model: https://www.geeksforgeeks.org/llama-3-metas-new-ai-model/