STOCK SENTIMENT ANALYSIS

Dhanush Bitra[2021BCS0052]
Nived Krishna[2021BCS0028]
Sasi Kiran Reddy[2021BCS0150]

Abstract

The Stock Sentiment Analyzer is a project aimed at predicting the sentiment of financial news headlines and providing users with insights into the performance of stocks. The project utilizes sentiment analysis techniques and machine learning algorithms to analyze news articles and generate sentiment graph of the stock. The application allows users to search for specific stocks and obtain sentiment analysis results for the past 7 days.

1 INTRODUCTION

Are you aware of **pump and dump scam**.Let me explain ,Social media has a huge impact on our lives in this digital era. Businesses are using influencers to advertise stocks. This leads to people- investing without knowing the company's finances. People buy stocks, causing prices to go up. This helps stake-holders and partners, who then cash out. After they leave, stock prices drop suddenly. Many investors lose money. Choosing stocks is important. What's our mission? To provide clear, honest info for you. We utilize innovative technology to analyze market opinions on specific stocks. Understand the chatter. This aids beginners in making informed decisions. Acquiring stocks blindly is hazardous. Our initiative offers insights into stock sentiments. Novices can make decisions more intelligently. Trustworthy sentiment data is vital. Studying stocks becomes simple once you understand the present reputation of the stock. We also help people who need fast stock study. Our system gives opinions on companies, so you can see trends quickly instead of taking lots of time. In this report, we will present you the methodology, implementation, and results of our stock sentiment analysis project. We believe our project will serve as a valuable tool for both new and experienced investors, offering them a deep understanding of the sentiment driving stock prices and aiding them in making more well judged investment decisions

2 LITERATURE SURVEY

Researchers have closely looked at the link between emotions on social media and predicting stock markets in recent years. Social media sites like Twitter have become key sources to understand public opinion and feelings on various topics, including financial markets. Past studies show a relationship between changes in market indicators like the Dow Jones Industrial Average (DJIA) and overall public mood gleaned from Twitter. This link suggests social media mood can indicate upcoming shifts in the stock market. As a result, academics have delved deeper into sentiment analysis techniques to gain valuable insights from social media data. Figuring out people's thoughts from text is called opinion mining or sentiment analysis. Word2Vec and N-grams are two ways to do this. Word2Vec is a simple computer system that looks at how words relate to each other. It does this by making word pictures using lots of numbers. N-grams check word patterns in text to find feelings. People use these methods to understand emotions in tweets and social media. Sentiment analysis' ability to find emotions in text data improved by using machine learning ideas. Labeled datasets train models through supervised techniques. This lets them predict emotions correctly. The models categorize feelings like good, bad, or neutral, by using features from the text. Examples: word embeddings, sentence structures. Sentiment analysis gets better at understanding feelings in text when using machine learning. Models learn from labeled datasets with supervised machine learning. They can then predict if sentiment is positive, negative, or neutral - quite accurately. The models use word embeddings and sentence structures from the text for categorizing the feelings. Sentiment analysis has great potential to enhance trading strategies and investment decisions. It helps investors evaluate market sentiment and forecast stock movements. The Stock Sentiment Analyzer project explores using sentiment insights for stock predictions. This is beneficial for financial institutions and individual investors. However, more research is needed to examine sentiment analysis robustness and scalability across asset classes and market conditions.

3 PROPOSED MODEL

3.1 Base Model description

The BERT (Bidirectional Encoder Representations from Transformers) model is a transformer-based model trained on a vast amount of English text using self supervised learning techniques. It serves as a foundational model pre-trained completely on raw text data, without any human labeling. During pre-training, it employs two key objectives: masked language modeling (MLM) and next sentence prediction (NSP). In MLM, 15% of words in input sentences are randomly masked, and the model predicts these masked words. In NSP, pairs of sentences are concatenated, and the model predicts whether they naturally follow each other. BERT can be fine-tuned for various downstream tasks such as sequence classification, token classification, and question answering.

3.2 Fine Tuned Model

In This project we are scraping the recent articles for the keyword and return it as dictionary with keyword as key and list of text extracted from scraping various sites related to keyword as values (basically scrapping title and text from <p> paragraph tag),and also the dataset which we are using has sequence of text.So to train our model we need to finetune the bert model based on Sequence Classification(machine learning task to predict a category for sequence of data) We have stored the dataset we created in hopsworks ,so that we can add new data by scraping articles of some famous stocks for which articles will be published frequently,to the existing data which will allow us to retrain the model every week for better accuracy and to stay up-to-date (fetching new features every week and adding them to existing dataset is done by github actions and model retraining is done by modal). We have performed hyper parameter tuning

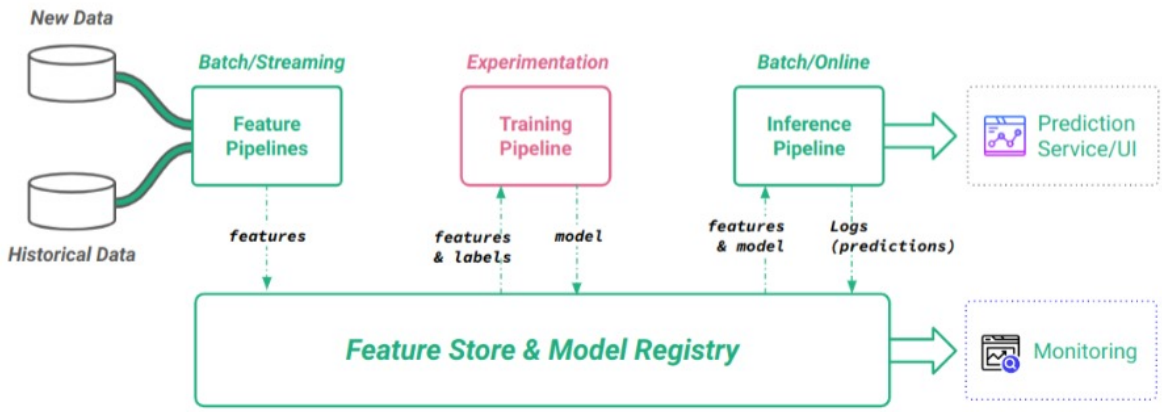


Figure 1: Architecture Diagram

using optuna(a machine learning framework that uses hyperparameter optimization to automatically find optimal values) we have performed the hyperparameter tuning for ten iterations, each iteration consists of 100 epochs which took us around 6 hours to get the final parameters when trained on Google Colab T4 Gpu. With the parameters we obtained from the hyper parameter search we have trained our model to get an accuracy of 89%.

3.3 WEEKLY PIPELINE ARCHITECTURE

Every week using github actions new features will be extracted and will be added to existing dataset in hopsworks and model retraining will be performed in Modal and the new model will be pushed to huggingface from where the api is called to the frontend.

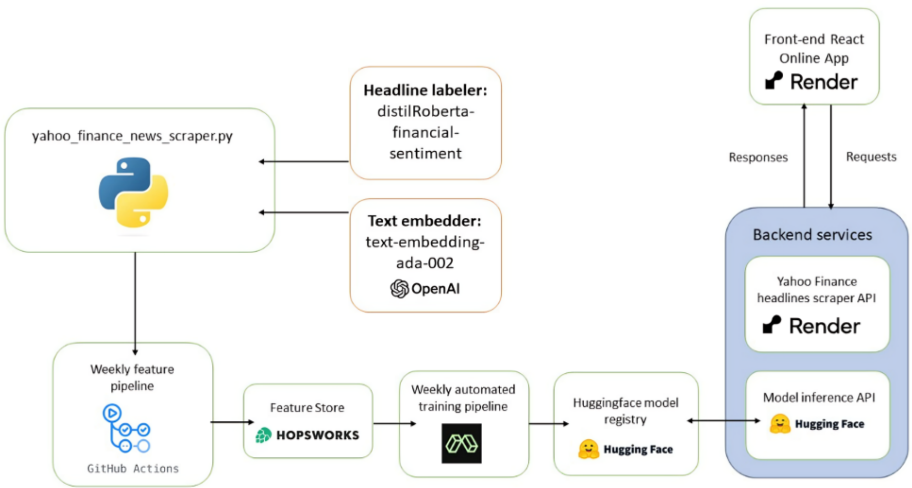


Figure 2: Weekly Pipeline

4 EXPERIMENTAL RESULTS AND ANALYSIS

4.1 DATASET INFORMATION

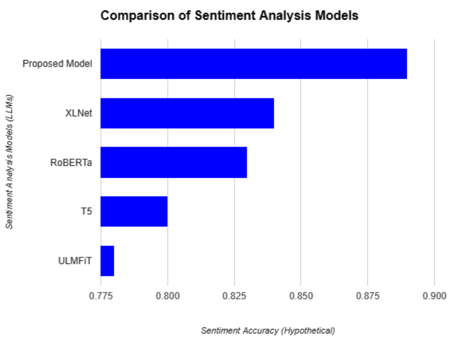
Two datasets formed the basis of our base dataset. Zeroshot Twitter and Financial Phrasebank are the first and second, respectively. In essence, both datasets contain a text and a sentiment label. It is negative, neutral, and positive for FinancialPhraseBank and bearish, bullish, and neutral for Zeroshot. Four subsets of the financial phrase bank dataset—50%, 66%, 75%, and 100%—were made depending on the annotators' agreement regarding labeling. To increase the size of our final dataset, we chose the one with 75% agreement. Preprocessing was necessary to make the datasets compatible with each other's various labeling schemes so they could be merged. Finally, we mapped the labels to 0, 1, and 2 in the Hopsworks feature store: negative, positive, and neutral.

4.2 RESULTS AND COMPARISONS

In this section we will be dealing with the performance of our model and the comparison of our model with the existing solutions.

Metric	**Value**
Loss	0.5985
Accuracy	0.8947

Table 1: Model Metrics



Models	Sentiment Accuracy (Hypothetical)	Notes
Proposed Model (LLM)	0.89	**Strongest Performer**
XLNet	0.84	Powerful LLM, might require intensive fine-tuning for sentiment analysis compared to BERT
RoBERTa	0.83	Effective LLM, computationally expensive than BERT for similar accuracy
T5	0.80	Versatile LLM, requires specific task adaptation for sentiment analysis
Universal Language Model (ULMFiT)	0.78	Earlier LLM, might require more data or customization compared to newer models

Figure 3: Comparison between Proposed Model and Existing Models

5 Final App

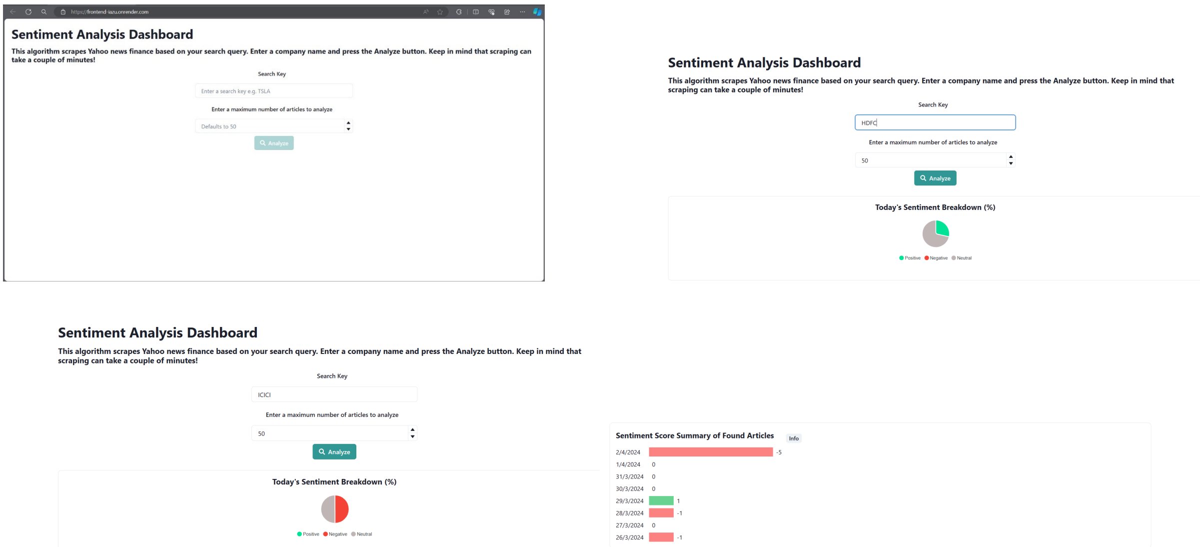


Figure 4: Some screenshots of the results obtained from our deployed model

6 CONCLUSION

The project regarding Stock Sentiment Analyzer marks an important step forward. It utilizes sentiment analysis and machine learning to comprehend social media sentiment, offering actionable insights for investors. Analyzing financial news headlines, our project serves as a valuable tool. It helps navigate the complex modern financial markets where social media discourse shapes investor sentiment and market behavior significantly. Through meticulous dataset preparation, model training, and cloud deployment, we've established a solid foundation. Users can benefit from real-time sentiment analysis insights, empowering them effectively.

7 FUTURE ASPECT

Our project's next step involves predicting stock prices using sentiment analysis and historical data. We'll analyze social media discussions to find sentiment trends. Then, we'll use historical stock prices to create predictive models that forecast price changes based on those trends. This added approach will help investors make better decisions. They'll be able to identify promising opportunities accurately and confidently. Building these capabilities advances our project's usefulness in an impactful way. We plan to research complex AI algorithms. We'll use deep learning methods to enhance our predictive models' accuracy. Our goal is capturing intricate market shifts precisely. Through ceaseless innovation and adaptation, we stay dedicated to progressing the Stock Sentiment Analyzer's abilities. We aim to provide substantial value to investors amidst an ever-changing financial environment.

8 References

1.Financial Phrasebank

2.Zeroshot Twitter

3.Sentiment Analysis on 10-K Financial Reports using Machine Learning Approaches

4.Research on sentiment analysis of financial texts

5.Public Sentiment Analysis in Twitter Data for Prediction of a Company's Stock Price Movements

6.Sentiment Analysis of Twitter Data for Predicting Stock Market Movements

7.Sentiment Analysis in Financial Texts

8.Sentiment analysis of financial news using unsupervised approach

9.Pre-trained Large Language Models for Financial Sentiment Analysis

10.A Framework to utilize the Human Ability of Word Association for analyzing Stock Market News Reports