

**Natural Language Processing**  
**-- Project proposal --**

***Financial Sentiment Analysis Using Statistical & Neural NLP Approaches***

*Group 11*

John J Pretz & Manas Ranjan Sahoo

**School of Graduate Professional Studies**

AI 574 – Natural Language Processing

Fall Semester, 2025

Professor: Youakim Badr, Ph.D.

TABLE OF CONTENTS

[1. Introduction 2](#_Toc654715774)

[2. Problem statement 3](#_Toc1500306896)

[3. Related work 3](#_Toc1184169684)

[4. Challenges 4](#_Toc512943047)

[5. Datasets 6](#_Toc1165568556)

[6. Deep Learning Process 7](#_Toc1599775276)

[Methodologies used 7](#_Toc745396764)

[Training and Testing strategy 7](#_Toc73509149)

[Evaluation strategy 7](#_Toc1799240396)

[7. Expected Results 8](#_Toc1902651701)

[8. References and Data sources 8](#_Toc593145036)

[9. Justification for using existing code 10](#_Toc412317026)

# Introduction

Stock markets are a heavy reflection of the public perception about public companies. Public perception - that can change due to a single news article, a social media post, company announcements, earnings reports as well as external macroeconomic and global events. With the world being more digitally connected than ever before and with the amount of data that gets generated every day (402.74 million terabytes[[1]](#endnote-1)), it becomes humanly impossible for investors to manually process and interpret sentiment from these diverse sources. Natural Language Processing tools comes to the rescue by an auto-detection of opinions, attitudes and sentiments expressed across diverse information sources aiding retail and institutional investors to decide on their investment decisions and help manage their risk based on future market trend. While the field of sentiment analysis has been researched for many decades, the continuous evolution of human language keeps it a field that always needs improvements.

# Problem statement

While the availability of financial text data has expanded exponentially, extraction of reliable sentiment analysis is still a challenging task for NLP models. This challenge is compounded by the domain-specific language used, contextual dependency within a sentence or a document as well as the subtle tone used in financial communication to avoid any sudden market movements. Sentiment analysis from such financial data is not a new approach and previously traditional statistical NLP methods, like TF-IDF and n-grams were used to capture surface-level patterns, but these methods fail to capture contextual nuances.

With improvements in machine learning, neural NLP approaches came into prominence and were able to plug the gaps that were evident with statistical NLP methods. Recurrent Neural Networks (RNNs) and their variants like LSTM along with neural embeddings like GloVe provided the much-needed contextual understanding. These were further enhanced through Transformer architecture-based models like FinBERT which were specifically fine-tuned for financial sentiment analysis.

Through this project we aim to chart the progress of sentiment analysis on financial text over the years by comparing statistical and neural NLP methods on multiple sources of financial text like news articles, social media posts, and earnings call transcripts – evaluating their effectiveness in sentiment analysis and predicting future market relevant insights.

# Related work

Over the past decade, financial sentiment analysis has seen a lot of progress in terms of the different sets of tools & methodology used – starting from statistical machine learning approaches modern deep learning architectures. As part of our research, we investigated different research papers to understand how financial sentiment analysis fares when carried out through statistical models and compare that with the research findings where financial sentiment analysis was done using modern deep learning architectures.

One of the studies done by Ahmed et al. (2023) showcased how the usage of simpler classifiers like Logistic Regression along with TF-IDF can provide a comparative result predicted through a complex classifier like SVM and this shows us the benefits of using simple classifiers to handle noisy features. Similar research was also done by Popoola et al. (2024) where social media-based sentiment analysis done through TF-IDF features combined with Random Forest and Naive Bayes and it consistently outperformed KNN across the important metrics. Overall, research papers like these solidify the understanding that classical methods do provide the much-desired reliability and interpretability that they are known for, in real life financial context.

Moving on to research papers on deep neural networks, the advent of transformer-based architecture made them a chosen option for many researchers evaluating them for financial sentiment analysis. One of the noteworthy ones is the paper by Araci (2019) introducing FinBERT which brought in a significant improvement over the sentiment analysis provided by generic language models. There were also multiple research papers that explored the benefits of hybrid architectures in sentiment analysis. A recent paper that stood out of those was by Cheng Yu et al. (2025) wherein a CNN-BiLSTM model was used to predict financial systemic risk analysis with F1 score of 0.88. But hybrid architectures didn’t just stop with non-transformer-based architectures but also used transformer and deep learning architectures as a combination to better determine sentiment analysis. One such standout example was the paper by Cheng et al. (2021) where a combination of FinBERT and BiLSTM was used to attain a F1-score of 0.81.

Analysis of the different approaches followed by these research papers helps us understand how statistical models help create an efficient and interpretable baseline whereas transformer-based and hybrid architectures help capture the nuances of human language and domain specific language better. A study of these architectures helps us devise our own novel architecture for sentiment analysis as well as compare different model variants in a stand-alone manner as well. This would help us contribute to the field of machine learning through our novel architecture as well as help showcase pros and cons of each model variant used.

# Challenges

Financial sentiment analysis does pose its own sets of challenges – some which are common across the field of sentiment analysis and some which are specific to the financial industry. Some of those challenges are listed below:

* Nuances in human languages – Human language is a continuously developing field and available data could be rife with slang, acronyms, sarcasms, irony, idioms which could make sentiment analysis tougher for machines.
* Domain specific language – Financial texts could contain domain-specific terminologies which could pose a challenge for any machine learning model unless it is previously trained on it.
* Imbalanced datasets – Presence of imbalanced datasets for each of the classification categories could pose challenges for the machine learning model to properly ‘learn’ from them.
* Sentiment classification depth – Human sentiments are more nuanced than just a positive, negative and neutral classification. As per American psychologist Robert Plutchik, there are 8 primary emotions[[2]](#endnote-2) which serve as the foundation for all other emotion variants. So, while we try to classify the sentiment from financial text within 3 categories, their effectiveness in predicting future market insights would depend on a more fine-grained understanding of the sentiment behind the texts.
* Overfitted models – Models can get overfit on the training data during the multiple tests they go through. This can be avoided by keeping the training, test, and validation datasets separate without any cross contamination by replacing data among each dataset.

Apart from domain specific and model specific challenges, there are a few physical challenges that we can foresee. They are listed below:

* CPU processing times – Deep learning models are usually process intensive and could be taxing for systems that do not have GPUs in them. As neither of our systems is powered by a GPU, we intend to focus on choosing models that do not consume a lot of processing power and maintain the quality of results for the best possible versions.
* Network connectivity – Loss of network connectivity can lead to incomplete training of models. This would be rectified by creating save points within the training steps.

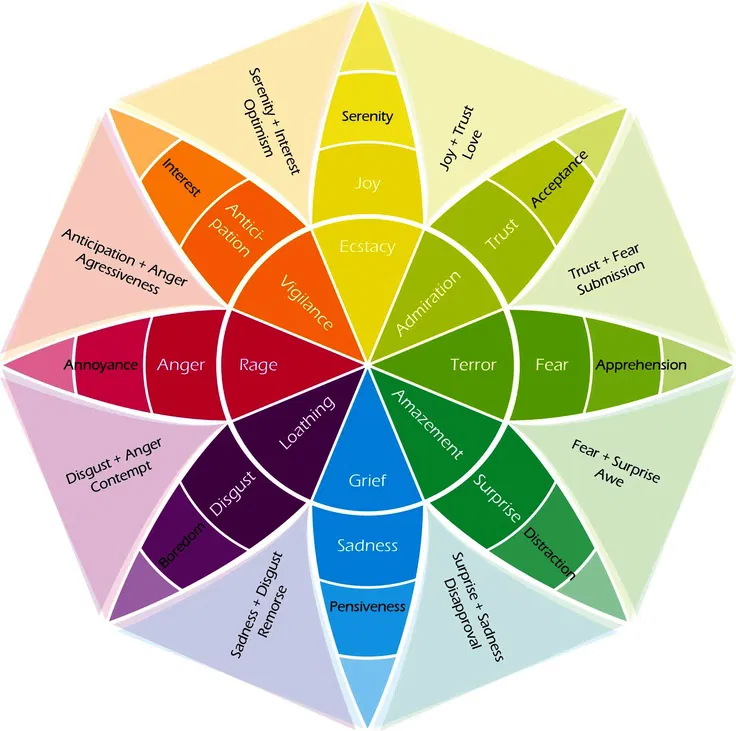


Image Courtesy – Positive Psychology[[3]](#endnote-3)

# Datasets

Throughout this project we intend to leverage a combination of publicly available and domain-specific datasets to train and validate the model created. Some of the datasets that we intend to use are:

1. Financial Phrase Bank - Domain-specific datasets consisting of over 4800 financial news phrases categorized by sentiment.
2. Social media datasets on financial news from Kaggle
3. Sentiment analysis datasets from Hugging Face
4. Earnings filing datasets from Hugging Face

These different datasets would form the bulk of the training and validation datasets. These datasets would vary in size and structure, and we would have to ensure to clean and process them to keep a consistency for the model to learn from it accurately.

# Deep Learning Process

As our objective is to compare sentiment analysis across different approaches, we intend to use different techniques and model architectures throughout this project.

## Methodologies used

* Statistical NLP methods – Logistical regression using TF-IDF, SVM with TF-IDF
* Neural NLP methods – LSTM or BiLSTM for a richer contextual modeling
* Transformer-based methods – FinBERT for sentiment analysis and possibly compare it with the base BERT or fine-tuned RoBERTa model to compare effectiveness.

## Training and Testing strategy

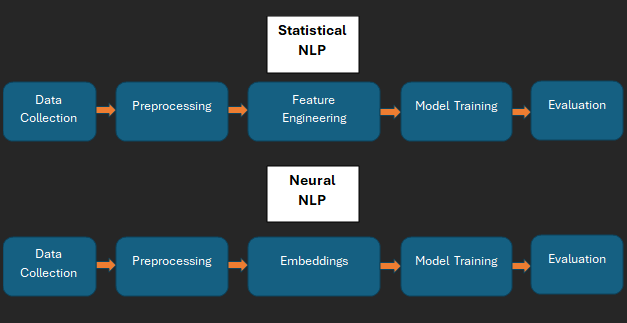
The training and testing strategy would first involve cleaning up the available dataset and carefully catering to any imbalance present within it. The steps involved would be:

* **Data preprocessing** – We will be collecting data from multiple sources, remove noise and normalize the text
* **Feature engineering** – We will be extracting features using methods like CBOW, TF-IDF
* **Dataset distribution** – An 70/15/15 split with stratification to train/validate/test the model for effectiveness
* **Reliability & robustness** – We will use k-fold cross validation for reliability and robustness
* **Handling imbalance** – We will be using oversampling techniques like SMOTE or by class weighting to handle imbalances
* **Regularization** - To prevent overfitting, we would be employing early stopping and dropout regularization
* **Hyperparameter tuning** – Grid search to determine the optimized values

## Evaluation strategy

After training the datasets against different approaches, we intend to compare their effectiveness using different metrics. Some of the metrics that we intend to use are:

* Accuracy
* Precision
* Recall
* F1-score
* ROC AUC
* PR AUC



Workflow Diagram

# Expected Results

What we expect to see is that statistical NLP models will create strong baselines when trained on simple datasets like Financial Phrase Bank but will lack performance on more complex and context-rich data like social media posts or earnings calls. Neural models like BiLSTM would improve upon the baseline performance but they do lack the ability to retain contextual awareness on really long texts. Transformer-based models are expected to outperform the other variants due to their ‘self-attention’ heads allowing contextual understanding over much longer texts and showcasing better understanding of domain-specific meanings and sentiment shifts. Through this project we aim to showcase how statistical models provide interpretability but how neural models provide superior performance over large, diverse financial text corpora.

# References and Data sources

Our data sources for this project will consist of different sources like:

* Kaggle datasets for financial sentiment analysis:
  + Financial Sentiment Analysis[[4]](#endnote-4)
  + Sentiment Analysis for Financial News[[5]](#endnote-5)
  + Twitter Financial News Sentiment Dataset[[6]](#endnote-6)
  + Sentiment Analysis on Financial Tweets[[7]](#endnote-7)
* Hugging Face datasets
  + Financial Phrase Bank[[8]](#endnote-8)
  + Synthetic Financial Tweets for Sentiment Analysis[[9]](#endnote-9)
* Stock Values and Earnings Call Transcripts for Sentiment Analysis[[10]](#endnote-10)

Apart from different data sources, we also studied and will continue to explore additional resources based on prior work done on the same topic and the references for those are listed below:

* Yu, Cheng, Xu Zhen, Chen Yuan, Wang Yuhan, Lin Zhengao, and Liu Jinsong. ‘A Deep Learning Framework Integrating CNN and BiLSTM for Financial Systemic Risk Analysis and Prediction Keywords-Financial Systemic Risk, Convolutional Neural Network, Bidirectional Long Short-Term Memory Network, Deep Learning’, 2 2025. <https://doi.org/10.48550/arXiv.2502.06847>.
* Cheng, Wenjuan, and Siyi Chen. ‘Sentiment Analysis of Financial Texts Based on Attention Mechanism of FinBERT and BiLSTM’. In *2021 International Conference on Computer Engineering and Application (ICCEA)*, 73–78. IEEE, 6 2021. <https://doi.org/10.1109/ICCEA53728.2021.00022>.
* Cai, Ren, Bin Qin, Yangken Chen, Liang Zhang, Ruijiang Yang, Shiwei Chen, and Wei Wang. ‘Sentiment Analysis About Investors and Consumers in Energy Market Based on BERT-BiLSTM’. *IEEE Access* 8 (2020): 171408–15. <https://doi.org/10.1109/ACCESS.2020.3024750>.
* Ahmed, Wesam, Noura Semary, Khalid Amin, and Mohamed Adel Hammad. ‘Sentiment Analysis on Twitter Using Machine Learning Techniques and TF-IDF Feature Extraction: A Comparative Study’. *IJCI. International Journal of Computers and Information* 10 (11 2023): 52–57. <https://doi.org/10.21608/ijci.2023.236052.1128>.
* Wahyuningsih, Tri. ‘Analyzing Sentiment Trends and Patterns in Bitcoin-Related Tweets Using TF-IDF Vectorization and K-Means Clustering’. *Journal of Current Research in Blockchain* 1 (6 2024): 48–69. <https://doi.org/10.47738/jcrb.v1i1.11>.
* Popoola, Gideon, Khadijat-Kuburat Abdullah, Gerard Shu Fuhnwi, and Janet Agbaje. ‘Sentiment Analysis of Financial News Data Using TF-IDF and Machine Learning Algorithms’. In *2024 IEEE 3rd International Conference on AI in Cybersecurity (ICAIC)*, 1–6. IEEE, 2 2024. <https://doi.org/10.1109/ICAIC60265.2024.10433843>.
* Fatouros, Georgios, John Soldatos, Kalliopi Kouroumali, Georgios Makridis, and Dimosthenis Kyriazis. ‘Transforming Sentiment Analysis in the Financial Domain with ChatGPT’. *Machine Learning with Applications* 14 (12 2023): 100508. <https://doi.org/10.1016/j.mlwa.2023.100508>.
* Ardekani, Aref Mahdavi, Julie Bertz, Cormac Bryce, Michael Dowling, and Suwan(cheng) Long. ‘FinSentGPT: A Universal Financial Sentiment Engine?’ *International Review of Financial Analysis* 94 (7 2024). <https://doi.org/10.1016/j.irfa.2024.103291>.
* Moradi-Kamali, Hamid, Mohammad-Hossein Rajabi-Ghozlou, Mahdi Ghazavi, Ali Soltani, Amirreza Sattarzadeh, and Reza Entezari-Maleki. ‘Market-Derived Financial Sentiment Analysis: Context-Aware Language Models for Crypto Forecasting’, 3 2025. <http://arxiv.org/abs/2502.14897>.
* Shen, Yanxin, and Pulin Kirin Zhang. ‘Financial Sentiment Analysis on News and Reports Using Large Language Models and FinBERT’, n.d.
* Nasiopoulos, Dimitrios K., Konstantinos I. Roumeliotis, Damianos P. Sakas, Kanellos Toudas, and Panagiotis Reklitis. ‘Financial Sentiment Analysis and Classification: A Comparative Study of Fine-Tuned Deep Learning Models’. *International Journal of Financial Studies* 13 (5 2025): 75. <https://doi.org/10.3390/ijfs13020075>.
* Du, Kelvin, Frank Xing, Rui Mao, and Erik Cambria. ‘Financial Sentiment Analysis: Techniques and Applications’. *ACM Computing Surveys* 56 (10 2024): 1–42. <https://doi.org/10.1145/3649451>.
* Araci, Dogu. “Finbert: Financial Sentiment Analysis with Pre-Trained Language Models.” arXiv.org, August 27, 2019. <https://doi.org/10.48550/arXiv.1908.10063>.

# Justification for using existing code

Using any available open-source code would help us in improving our efficiency across the different steps of the model creation. Usage of established libraries like scikit-learn for preprocessing, spaCy and NLTK for tokenization, TF-IDF for vectorization, helps speed up our implementation of statistical NLP methods, which would form our baseline. Existing transformer-based frameworks from Hugging Face would provide us fine-tuned models for financial text sentiment analysis which would reduce our overall development time. These would allow us to focus on data preprocessing, experimentation on model hyperparameters and analyze the results with a fine-tooth comb.

1. https://rivery.io/blog/big-data-statistics-how-much-data-is-there-in-the-world/ [↑](#endnote-ref-1)
2. https://www.6seconds.org/2025/02/06/plutchik-wheel-emotions/ [↑](#endnote-ref-2)
3. https://positivepsychology.com/emotion-wheel/ [↑](#endnote-ref-3)
4. [↑](#endnote-ref-4)
5. https://www.kaggle.com/datasets/ankurzing/sentiment-analysis-for-financial-news [↑](#endnote-ref-5)
6. https://www.kaggle.com/datasets/borhanitrash/twitter-financial-news-sentiment-dataset [↑](#endnote-ref-6)
7. https://www.kaggle.com/datasets/vivekrathi055/sentiment-analysis-on-financial-tweets/data [↑](#endnote-ref-7)
8. https://huggingface.co/datasets/takala/financial\_phrasebank [↑](#endnote-ref-8)
9. https://huggingface.co/datasets/TimKoornstra/synthetic-financial-tweets-sentiment [↑](#endnote-ref-9)
10. https://dataverse.nl/dataset.xhtml?persistentId=doi:10.34894/TJE0D0 [↑](#endnote-ref-10)