# Financial Sentiment Analysis Using the Recurrence Neural Network

Sayan Das
sayand2909@gmail.com || linkedin.com/in/sayandass/

# Contents

# 1 Introduction

Sentiment analysis in finance refers to the application of natural language processing (NLP) techniques to analyze textual data, such as news articles, social media posts, earnings call transcripts, and other sources of financial information, to extract sentiment or opinions related to financial assets, markets, or events. This analysis aims to understand the collective sentiment of market participants and its potential impact on financial markets, investment decisions, and risk management strategies.

The usefulness of sentiment analysis in finance stems from its ability to provide valuable insights and actionable information to various stakeholders in the financial industry:

1. **Market Participants** - Sentiment analysis helps traders, investors, and financial analysts gauge market sentiment and investor emotions, which can influence market movements and asset prices. By analyzing sentiment data, market participants can identify market trends, sentiment shifts, and potential trading opportunities.

2. **Risk Management** - Sentiment analysis is valuable for assessing market sentiment and identifying potential market risks and vulnerabilities. By monitoring sentiment indicators, financial institutions can better manage risks associated with market volatility, sentiment-driven price fluctuations, and investor behavior.

3. **Investment Decisions** - Sentiment analysis can inform investment decisions by providing insights into investor sentiment towards specific stocks, sectors, or asset classes. Investors can use sentiment data as one of the factors in their investment decision-making process, alongside fundamental and technical analysis.

4. **Financial News Analysis** - Sentiment analysis can help analyze financial news and media coverage to extract sentiment-related information and identify sentiment-driven market reactions. By monitoring news sentiment, financial institutions can stay informed about market sentiment trends and news sentiment impact on financial markets.

Previous work in sentiment analysis in finance has focused on various aspects, including:

1. **Sentiment Lexicons and Dictionaries** - Developing sentiment lexicons and dictionaries tailored for financial text analysis, which contain words and phrases annotated with sentiment scores or polarity labels.

2. **Machine Learning Models** - Building machine learning models, such as support vector machines (SVM), random forests, and deep learning models (e.g., recurrent neural networks, convolutional neural networks), to classify financial text and extract sentiment features.

3. **Event Studies** Conducting event studies to analyze the impact of sentiment on financial markets around specific events, such as earnings announcements, mergers and acquisitions, economic indicators, and geopolitical events.

4. **Sentiment Visualization** Creating visualizations and dashboards to visualize sentiment trends, sentiment scores, and sentiment-related metrics for better understanding and decision-making.

Overall, sentiment analysis in finance plays a crucial role in understanding market sentiment, sentiment-driven market dynamics, and their implications for financial markets and investment decisions. Continued research and innovation in this field are essential for advancing sentiment analysis techniques and their applications in finance.

# 2 Objective

The objective of this project is to develop a sentiment analysis model using Recurrent Neural Networks (RNNs) to analyze financial sentiment data extracted kaggle. Specifically, the aim is to build a predictive model that can accurately classify the sentiment of financial texts (such as news articles, earnings call transcripts, and social media posts) into positive, negative, or neutral categories.

Financial sentiment analysis plays a crucial role in understanding investor sentiment, market dynamics, and potential investment opportunities in the financial markets. By leveraging machine learning techniques, particularly RNNs, we aim to extract valuable insights from textual data to inform investment decisions, risk management strategies, and market sentiment analysis.

# 3 Dataset Description

The Financial Sentiment Analysis dataset is a consolidated collection of financial sentences with sentiment labels, aimed at advancing research in financial sentiment analysis. It combines two datasets, FiQA (Financial QA) and Financial PhraseBank, into a single CSV file for easy accessibility and usability. The dataset is sourced from the research paper titled "Good debt or bad debt: Detecting semantic orientations in economic texts" by Malo et al. (2014).

## 3.1 FiQA Dataset:

The FiQA dataset consists of financial questions and answers extracted from online forums, social media platforms, and other sources. Each data instance in the FiQA dataset contains a financial sentence along with its associated sentiment label. The sentiment labels in the FiQA dataset indicate the sentiment polarity of the financial sentences, categorized as positive, negative, or neutral.

## 3.2 Financial PhraseBank:

The Financial PhraseBank dataset contains a collection of financial phrases and sentences from various financial documents, reports, and news articles. Similar to the FiQA dataset, each data instance in the Financial PhraseBank is annotated with a sentiment label indicating the sentiment polarity of the financial text.

## 3.3 Combined Dataset:

The combined dataset merges the FiQA and Financial PhraseBank datasets into a single CSV file, providing a comprehensive collection of financial sentences with sentiment labels. Each data instance in the combined dataset consists of the financial sentence text and its corresponding sentiment label. The sentiment labels are encoded as categorical variables representing the sentiment polarity of the financial text (positive, negative, neutral).

## 3.4 Data Columns:

**Sentence:** Textual content of the financial sentence.
**Sentiment Label:** Categorical variable representing the sentiment polarity of the sentence (positive, negative, neutral).

The Financial Sentiment Analysis dataset contains a total of 5842 data instances, with each instance comprising a financial sentence and its associated sentiment label.

# 4 Methodology

We followed the given method to carry out out task of detecting the sentiment of any text -

## 4.1 Data Preprocessing:

We had undergone the following pre-process steps to change the data from a textual format to a numerical format, which is more relevant to a machine for the analysis purpose.

1. **Converting to small cases:** Converted all textual data into lowercase to ensure consistency and simplify text processing.

2. **Remove Stop Words:** Removed common stop words (e.g., ”the”, ”is”, ”and”) from the text as they do not contribute to sentiment analysis and may introduce noise.

3. **Lemmatization:** Applied lemmatization to reduce words to their base or root form, which helps in standardizing the text and reducing dimensionality.

4. **Vectorization using Word Embedding:** Utilized word embedding technique to represent words as dense vectors in a high-dimensional space. Converted preprocessed textual data into word embeddings to capture semantic relationships between words and phrases.

## 4.2 Imbalance Handling:

It is observed for the dataset caontains clear indication of the presence of imbalance in the distribution of the response value (figure - 1).

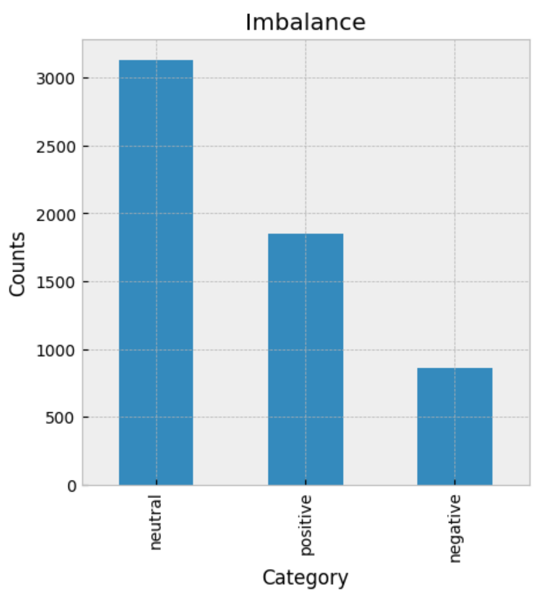


Figure 1: baplot showing the mbalance in the dataset

Therefore, we address class imbalance by using random oversampling to increase the number of minority class samples (negative and positive sentiments) to match the majority class (neutral sentiment). Random oversampling technique randomly duplicate instances from the minority classes until the dataset is balanced.

## 4.3 Encoding of the Response Variable:

Label encoding is required for the conversion of the response variable (since in our case the response is given in the textual format) into the numerical value, since it is usefule for the machine to understand.

We encoded the sentiment labels into numerical format using label encoding, that mapped the sentiment categories to numerical values as following:

- Negative: 0
- Neutral: 1
- Positive: 2

This encoding scheme facilitates model training as it converts categorical data into a format suitable for machine learning algorithms.

## 4.4 Modelling using Neural Network:

Implemented an RNN-based sentiment analysis model with the first layer being an RNN (SimpleRNN) to capture sequential dependencies in the textual data. Next we have added multiple dense layers after the RNN layer to extract and learn higher-level features from the sequential data. Utilized activation functions such as ReLU (Rectified Linear Unit) to introduce non-linearity and improve model expressiveness and Softmax in the last layers to get the how likely each input is to be classified in each response categories.

# 5 Conclusion

In this project, we developed a sentiment analysis model for financial texts using a neural network architecture with dropout layers in each of the dense layers to stabilize the model performance. We also employed EarlyStopping as a callback mechanism to prevent overfitting and achieve optimal model performance.

After training the model for a total of 37 epochs, we obtained the following results:

- Training Accuracy: 89.71%
- Validation Accuracy: 97.23%

The following figure 5 indicates how the training and the validation accuracy has changed through the training process -


Figure 2: training and testing accuracy through the epochs

These results indicate that the model achieved high accuracy in both training and validation phases, demonstrating its effectiveness in predicting sentiment categories for financial texts. The use of dropout layers helped in regularizing the model and reducing overfitting, while EarlyStopping ensured that the training process stopped when the validation accuracy stopped improving, thus preventing unnecessary training iterations.

Overall, the developed sentiment analysis model shows promising performance and can be deployed for real-time sentiment analysis of financial texts, aiding investors, financial analysts, and other stakeholders in making informed decisions and assessing market sentiment trends.