1. Data Extraction

1.1 Source of Data

The data for this sentiment analysis project was extracted from earnings calls using the Bloomberg Terminal. Earnings call transcripts were accessed and saved in Word document format.

1.2 Data Processing

The Word documents containing the earnings call transcripts were then processed to extract relevant text data for sentiment analysis. The extraction involved converting the Word documents into CSV format to conduct further analysis.

2. Modeling

2.1 Sentiment Models

Several sentiment analysis models were tested to evaluate their performance on the extracted earnings call data. The following models were included in the analysis.

- **DistilRoBERTa -** The DistilRoBERTa model is a distilled version of RoBERTa-base, with 6 layers, 768 dimensions, and 12 heads, totaling 82M parameters (compared to RoBERTa-base's 125M parameters). It follows the same training procedure as DistilBERT. Notably, DistilRoBERTa is case-sensitive, distinguishing between lowercase and uppercase English. It offers a speed advantage, being, on average, twice as fast as RoBERTa-base.
- **FinBERT -** FinBERT is a pre-trained NLP model designed for analyzing sentiment in financial text. It achieves this by fine-tuning the BERT language model using a substantial financial corpus, specifically utilizing the Financial PhraseBank by Malo et al. (2014) for this purpose. The model is tailored to excel in financial sentiment classification.
- **FinancialBERT -** FinancialBERT is a BERT (Bidirectional Encoder Representations from Transformers) model that has undergone pre-training using an extensive collection of financial texts. The primary goal is to improve research and practical applications in financial natural language processing (NLP). The intention is to provide financial practitioners and researchers with a readily available, pre-trained model, eliminating the need for substantial computational resources typically required for training such models.

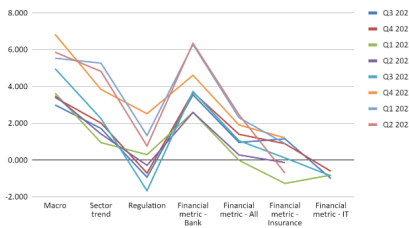The models that we are using are located at
https://github.com/HarisMahmood8/BERT_Keyword_Extractions

2.3 Model Evaluation

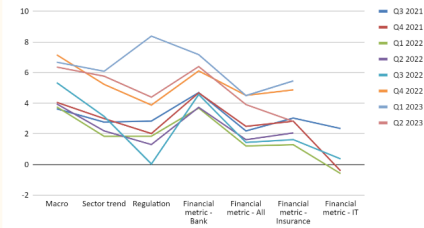# 3. OutPut and Regression Analysis

3.1 Model Outputs

distilroberta-finetuned-financial-news-sentiment-analysis

Models

Smaller batch size
for training &
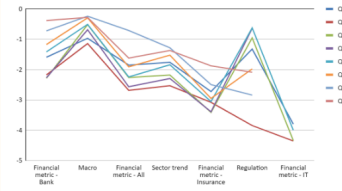evaluating

ahmedrachid/FinancialBERT-Sentiment-Analysis

- fine-tuned version of distilroberta-base on
the financial_phrasebank dataset.

- BERT model fine-tuned on
financial_phrasebank

Training hyperparameters

The following hyperparameters were used during training:

- learning_rate: 2e-05
- train_batch_size: 8
- eval_batch_size: 8
- seed: 42
- optimizer: Adam with betas=(0.9,0.999) and epsilon=1e-08
- lr_scheduler_type: linear
- num_epochs: 5

ProsusAI/finbert

Fine-tuning hyper-parameters

- learning_rate = 2e-5
- batch_size = 32
- max_seq_length = 512
- num_train_epochs = 5

pre-trained NLP model to analyze sentiment of financial
text. It is built by further training the BERT language
model in the finance domain

3.2 Regression Analysis

Sentiment analysis MLOps

Get data from Seeking Alpha or Bloomberg → Google Cloud → Extract data from Google Cloud (BigQuery)

Enable necessary APIs → Google Cloud

authentication required → Google Cloud

Extract data from Google Cloud (BigQuery) → Create Git repository

Push the data to BERT model ← Cloning the Git to python environment ← Create Git repository

Push the data to BERT model → Transform the BERT model → Result → Store to CSV file → Analysis