

Title: Predicting Stock Prices after Earnings Calls

Team name: Yearning for Earnings

Team Members: Shabarish Chenakkod, Zhengjun Liang, Shravan Patankar, Tejaswi Tripathi

Github: <https://github.com/shabarishch/EarningsNLP>

Motivation:

Stock prices are hard to predict. One significant factor that influences the stock market is corporate earnings announcements. They provide crucial insights into a company's financial health, often leading to movement in stock prices. We believe that a more precise understanding of earnings calls gives us better insight into the stock market.

Objective:

1. Create a machine learning model to predict percentage change in stock prices surrounding earnings calls based on sentiment analysis of earnings call transcripts, earnings and revenue data, and stock prices and volume before the earnings call.
2. Devise a simple trading strategy based on our predictions.
3. Find out factors that affect the stock price movement the day after the most.

Data Collection and Processing:

1. Earnings call transcripts, together with earnings and revenue data, are web-scraped using APIs obtained from rapidapi.com.
2. Stock prices and volume data are obtained using Yahoo Finance.
3. Sentiment analysis is done on each transcript, and sentiment scores are assigned.
4. In total, 10 features are considered.

Modeling:

1. A baseline model that predicts no change from the previous day.
2. Linear Regression, XGBoost, and Neural Networks.
 - a. Followed 0.8 - 0.2 training - test split, stratified by company symbol.
 - b. Split based on time for linear regression.
 - c. Used GridSearchCV to tune parameters of XGBoost.
3. We also applied Logistic Regression and Neural Networks for classification.

Results:

1. Models were evaluated based on metrics such as Mean Squared Error (MSE) and Correlation Coefficients between predicted and actual values.
 - a. **MSE:** XGB is better than Linear Regression. NN performance is affected by outliers.
 - b. **Correlation:** NN better than XGB better than Linear Regression.
2. Feature importance plots show that sentiment scores are important in the predictions.

Simple Trading Strategy:*

1. Positive prediction: buy and hold stock for 1 day. Negative prediction: short and then buy for 1 day.
2. Obtained returns of (XGBoost 4.8%, Linear Regression 3%) by trading around 4 earnings call days. Naively buying the stock on earnings day and selling it the next gives a return of 2%.

Future Directions:

1. TF-IDF vectorization
2. Explore more advanced machine learning models such as LSTM
3. Apply our predictive modeling techniques to other financial events and instruments

* Strategy applies only to institutional investors/ earnings during active hours.