

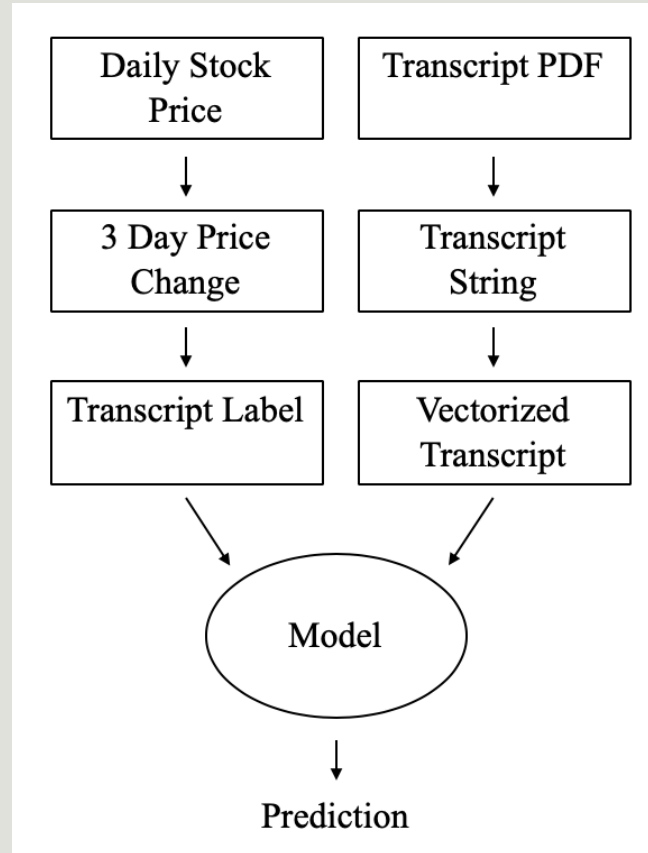
Understanding Market Overreactions using NLP

Overview: We wanted to see how text analysis could help predict overreactions/underreactions of stock prices based on corresponding Earnings Calls

Data: 378 Transcripts off Factset for 11 companies + corresponding share prices for 10 years. Utilized 5 fold cross validation to increase training size.

Final goal: Develop the basis of a trading strategy that may be rule based or discretionary

Process



Classification timeframe: Delta of price at market open and price at market close

- 1 if positive, 0 if negative

Classifiers Used:

- SVM with multiple kernels
- Logistic Regression
- Gaussian Naive Bayes
- **Multinomial Naive Bayes* (Best)**
- Adagrad
- Adaboost
- Random Forest

MNB Parameters

- Ngram range (1-2)
- Max features - 25,000
- Stop words - English
- Alpha – 0.03

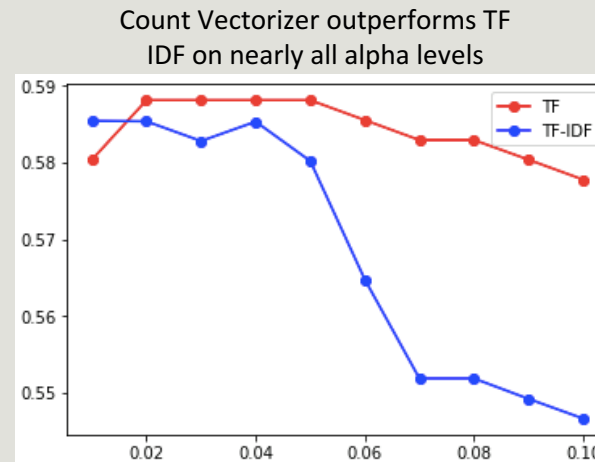
Results

Count Vectorizer

Classifier	Accuracy %
MNB	58.81%
SVM	49.50%
LOG REG	51.04%
GNB	53.64%

TF IDF Vectorizer

Classifier	Accuracy %
MNB	58.28%
SVM	52.34%
LOG REG	51.06%
GNB	52.61%



Analysis of Results

- At first, anything below 60% accuracy may not seem fascinating. However, in the financial markets, every basis point (0.01%) can mean millions.
- A 58% accuracy is game changing. A simple systematic trading bot that uses leverage can consistently beat benchmarks.
- A trading strategy that plays off these results could be implemented via Quantopain.

Limitations and Future Research

- Understand the impact of trading costs on this trading strategy to make it realistic.
- Use only a singular companies to predict that companies reactions. This allows for standardization of speech.
- Train on a greater set of transcripts for a better accuracy number
- Test out different types of time periods, rather than the strict 1 day, 3 day, etc.
- Understand the effect of leverage on returns
- Sentiment as a feature
- Neural nets, especially converting words to vectors