Problem Statement:

By analyzing the earning call transcripts, would classification model provide value for investors next day for trading?

Data Source:
1. Text sources: Earning call transcripts in the major automobile maker such as Ford, General Motors, Tesla, Fiat Chrysler are retrieved from seekingalpha.com with essential subscription.
2. Numerical sources: Stock price and volumes are retrieved from Yahoo Finance

Modeling Objective:
1. While the call transcripts are released, this model will predict either the stock price will be up or down next day.
2. Since most Earing per share and quarter calls will be held after the market closes or before the market. Investors have a short period of time to analyze how the fluctuation could before trade. The value of this model will provide buy or sell signal to the speculative investors.

EDA and Modeling Process
1. After merging the text and numerical dataset, transcripts will be splitted into Management Discussion (DA) and Analyst Question and Answer (QA). Features engineering will be derived from these two sessions. New features such as syntax sentiment, polarity, subjectivity, number of difficult words, readability indices, etc. will be added. After all, the numerical columns will be part of the data analysis
2. Second, text alone will be vectorized with ngrams for a new modeling process. Features of importance could provide investors what phrase or words matters
3. When the generalized model is completed, confusion matrix will be adapted to examine the performance of the classification. Additionally, a specialized model that only included dataset for one company will be modeled, followed by above engineering process. This aims to audit whether the classification model may provide a better prediction in one company rather than the entire industry.
4. Lastly, new stop words will be collected and introduced to the model for another round of evaluation.

Recommendation
1. Investors could receive up or down signal from multiple classification. They could compare whether these signals will be helpful to validate their intuition on next day trading after the earning call releases.

Limitation
1. More text data need to be added. Only four automobile makers' quarterly call transcripts from 2014-2019 were collected.
2. Neural network and tuning could be implemented

3. Simplify the modeling and pipeline process that provide a better manipulation on processing.
4. Third-party research equity reports, dividends amount, and LDA topic could be included as one of the features