Predicting OAS changes using SEC Filings & News Headlines

Ryan Serrao, Sithal Nimmagadda, Silke Herold, & Tung Ngo
Industry Mentor: Eunice Hameyie
Faculty Mentor: Smaranda Muresan

Data Science Capstone Project with Vanguard

Problem Statement
This work develops a predictive model for the change in companies’ option-adjusted spreads (OAS) for select 5-year corporate bonds. The model mainly uses features extracted from financial news headlines and company filings submitted to the SEC (8-Ks, 10-Ks, and 10-Qs) to predict the daily change in OAS. In addition, a bespoke sentiment lexicon is developed for the SEC filings.

Text Extraction & Features
First, texts from the news headlines and the SEC filings were cleaned, and SEC filings were further parsed into relevant sections of each document. Next, we gathered n-grams and also developed a sentiment score using a bespoke sentiment lexicon based on the Loughran-McDonald dictionary. Our sentiment lexicons are domain-specific and assign polarity scores using frequencies of random walk visits.

Models & Results
The data is balanced in terms of the classes 0/1 for all the bonds. Logistic regression, SVM and Random Forest are used for modeling. The first 80% of the data is used for training while last 20% is used for testing. Metrics considered are accuracy, precision, recall, and F-score. Feature importances are calculated for all five bonds while modeling. The most important features across companies are S&P 500 open, open performance of the company’s stock, days since SEC filing, and sentiment of news headlines. The results listed in the tables below are for the test set.

Conclusion
Overall, the Random Forest is the best performing model with accuracies ranging 58-75%. From the study, we can implement the model with the most important features to predict the change in companies’ OAS spreads. The resources for this experiment, including the model, text methods, and a set of financial lexicons can be used for Vanguard’s future use cases.

Acknowledgments
We would like to thank Eunice Hameyie and Smaranda Muresan for their guidance and feedback throughout this project and Vanguard for their partnership.

References