# Textual Analysis and Financial Statements

Isaac Liu

March 25, 2024

## Introduction

Corporate credit ratings represent professional estimations of the default risk carried by company debt. These ratings represent critical information for investors - not just institutional investors and financially sophisticated bondholders, but also stockholders, who may be wiped out completely in the event of bankruptcy. Analyzing ways to predict ratings can offer substantial value to a variety of stakeholders. Predictive models may be useful for investors without access to data, companies or potential lenders that seek information about influential factors (there is evidence suggesting financial factors and projections have a causal impact on ratings and are not manipulated by companies in response to forecasted rating changes (He, 2018)), and by any parties seeking interpolated ratings for companies that do not have them.

In this project, we seek to fully leverage the text of earnings calls, along with traditional financial measures and variables, to improve predictions of corporate credit ratings for any given company and quarter and better understand the importance of various influences. Textual features such as pre-trained language model vector embeddings (Araci, 2019) and analyses of sentiment accompany tabular variables as inputs to a variety of supervised machine learning techniques for classification from logistic regression to tree-based methods. If time allows, we will also incorporate advances in the study of graph neural networks to create additional embeddings modelling linkages between firms (Das et al., 2023) implied by calls.

Though much literature has focused on financial statements and reports and credit ratings (as just one example, see Makwana et al. (2022)), our paper takes a relatively underexplored approach, instead incorporating earnings call transcripts. We believe calls offer a richer picture of a firm's financial prospects because they include two-way conversation between company management and financial analysts in form of a Q and A section. This section incorporates the broader beliefs and concerns of the financial community into our predictions. Additionally, in contrast to financial statements, which must be (noisily) parsed to identify sections relevant to management analysis, earnings calls provide more directly valuable and readily available information.

To the best of our knowledge, the closest prior work to ours is Donovan et al. (2021), which leverages the textual content of earnings calls and financial statements to predict credit events such as bankruptcies, interest spread changes, and rating downgrades using unigram and bigram frequencies and the supervised machine learning techniques of Support Vector Regression, Latent Dirichlet Allocation, and Random Forests. The coefficient on a constructed textual measure of credit risk was found to be significant up the 1% level. In contrast to this approach, we focus on predicting the credit ratings themselves, and integrate more techniques from machine learning such as the power of pre-trained language models and a wider variety of algorithms for classification.

# Data and Exploratory Data Analysis

We combine a wide variety of data sources to support our predictions of credit ratings - merging rating data with company earnings calls, financial statement variables, and industry sector. In our final dataset, each observation represents a fixed quarter date (1/1, 4/1, 7/1, 10/1) for a company, with the company's most recent credit rating, earnings call and associated financial statement variables, and sector attached.

Our scope of interest is publicly traded companies from 2010-2016 (a limitation due to the availability of credit rating data). The data is temporally unbalanced, with many companies entering the dataset in later years, after they first receive an observable credit rating (Figure 1).

Figure 1: Observations by Quarter and Year
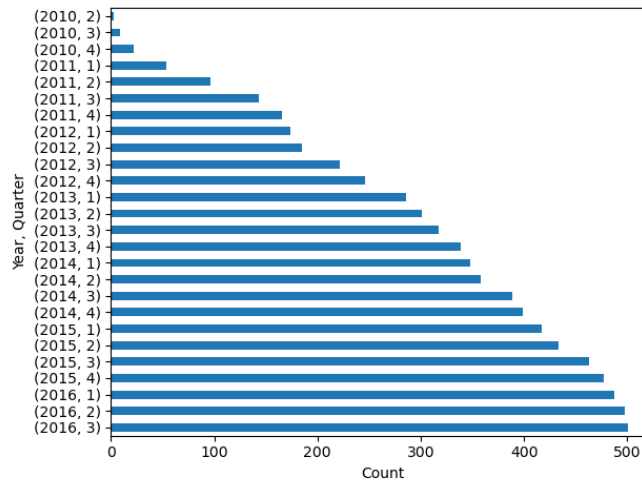


In all, we have 7333 quarters for 536 unique companies.

## Credit Ratings

We make use of long-term credit rating issuances from S and P Rating Services, provided from a combination of two credit rating datasets downloaded in CSV and Excel format from Kaggle (Gewerc, 2020; Makwana, Bhatt and Delwadia, 2022). Each issuance be a change in rating (upgrade, downgrade) or reaffirmation - they occur at ad-hoc intervals. We reshape these rating issuances to a dataset of ratings for each company on each fixed quarter date by creating a rating end date variable that is the date of the next issuance, and joining a list of the fixed quarter dates on the condition that the fixed quarter date is between the issuance date and the end date.

Figure 2 shows the distribution of rating grades used in our final dataset. Finer grades (+, -) are sometimes assigned by agencies, but these grades were removed for this project. Ratings of BBB and above are considered investment grade - these bonds carry empirical one-year default rates of 0 to 1%. Ratings below that are classified as junk, with default rates from 1 to 30, 40, or even 50% for some years (S and P Global Ratings, 2024). Most company-quarters have ratings around the BBB threshold, with very few cases on the extreme ends of the spectrum.

Figure 2: Credit Ratings
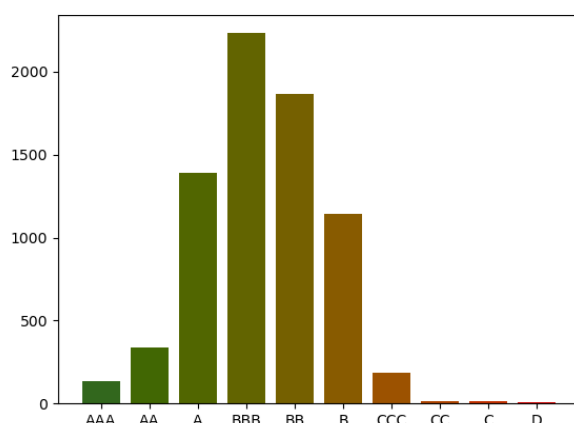


## Earnings Calls

Our earnings call data comes from the Financial Modelling Prep API (**?**), a trusted source widely used in industry. We remove all calls that happened more than 250 days prior and after the year and quarter they are supposed to discuss the results from. Including both prepared remarks and analyst Q and A sessions, the overall average call length in our final data stands at 8688.25 words.

Figure 3: Number of Words in Earnings Calls
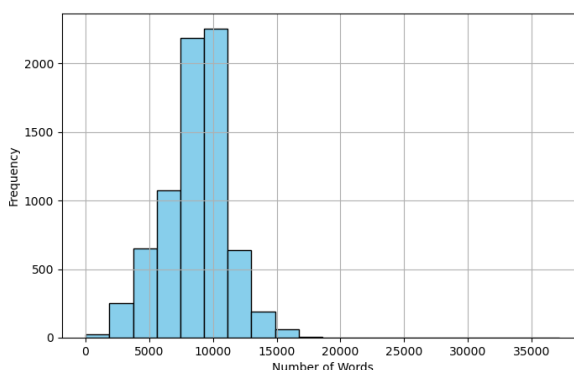


## Financial Statements

Our financial statement variables are also retrieved using the Financial Modelling Prep API. We make use of items from company balance sheets, cash flow statements, and income statements, as well as company market capitalization. We include 124 variables in total, such as revenue, total liabilities, net income, EBITDA.
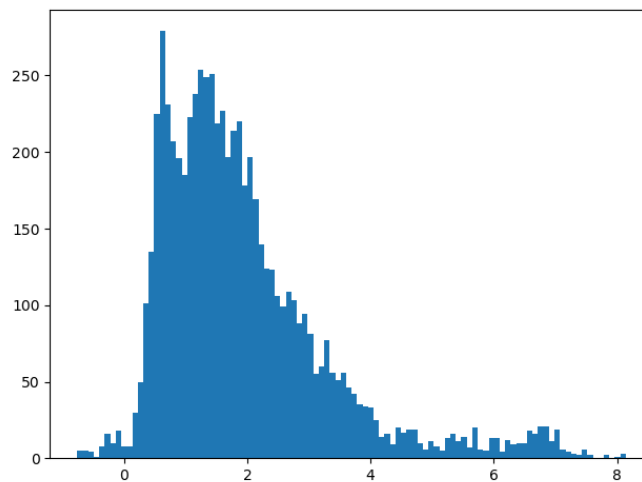
Limit to items reported in USD

Winsorizing: Check for items mis-multiplied by 1,000 in parsing - if last digits are "000.00" and item is above or below 2.5% and 97.5% quantile, divide by 1,000

Tests to ensure the value in income statement and balance sheet are consistent with each other.

Construct Altman Z-score
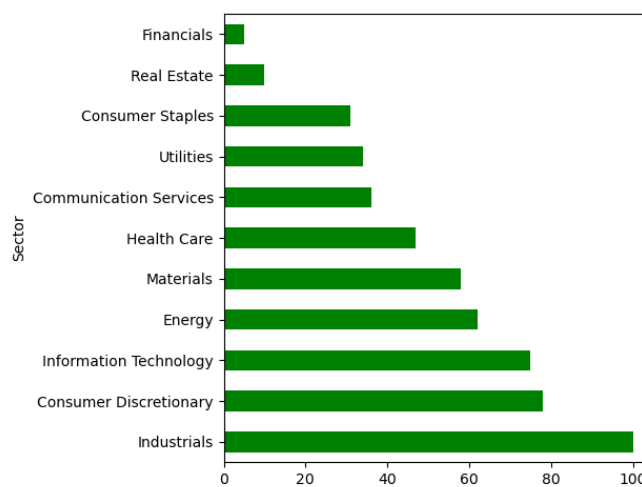
Altman (1968)

Figure 4: Altman Z-Score



## Sector

GCIS developed by S and P

Obtained from Kaggle with supplementary manual lookup

CSV and Excel format

Figure 5: Firms by Sector



sectoral imbalance

## Quality Control

quality control code review of all data cleaning code numerous investigations
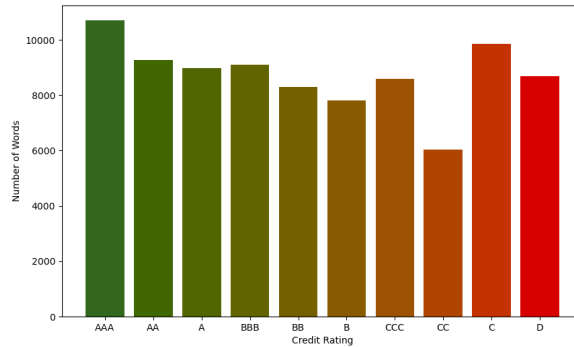
date gaps investigations

company dropout and the company that leaves

## NLP Features

Average call length

Figure 6: Average Call Length by Credit Rating



outliers and errors

correlations and patterns

identification of good machine learning methods

## Modelling

Our overall model architecture is of the form

$$\text{Predicted Credit Rating} = f(\text{Financial Statement Variables}, \text{Sector}, \text{NLP Features})$$

functions began with logistic regression

XXX logistic regression predictors

multinomial, balanced class weights, l1 penalty

table of predictions

fitting and output

assumptions

interpretation

## Next Steps

Ensembling and Auto-ML

more classifiers

first steps using AutoML

a good starting point for diving deep on more algorithms

algorithms and accuracy from them

outputted feature importance

Graph Neural Network incorporating the relationships between companies, trained end-to-end with both tabular financial data and NLP features

Fine tune the pre-trained LLMs for NLP feature construction

# References

**Altman, Edward I.** 1968. "Financial Ratios, Discriminant Analysis and the Prediction of Corporate Bankruptcy." *The Journal of Finance*, 23(4): 589–609. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1540-6261.1968.tb00843.x.

**Araci, Dogu.** 2019. "FinBERT: Financial Sentiment Analysis with Pre-trained Language Models." arXiv:1908.10063 [cs].

**Das, Sanjiv, Xin Huang, Soji Adeshina, Patrick Yang, and Leonardo Bachega.** 2023. "Credit Risk Modeling with Graph Machine Learning." *INFORMS Journal on Data Science*, 2(2): 197–217. Publisher: INFORMS.

**Donovan, John, Jared Jennings, Kevin Koharki, and Joshua Lee.** 2021. "Measuring credit risk using qualitative disclosure." *Review of Accounting Studies*, 26(2): 815–863.

**Gewerc, Alan.** 2020. "Corporate Credit Rating with Financial Ratios."

**He, Guanming.** 2018. "The Impact of Impending Credit Rating Changes on Management Earnings Forecasts." *Global Journal of Management and Business Research*, 18: 1–18.

**Makwana, Ravi, Dhruvil Bhatt, and Kirtan Delwadia.** 2022. "Corporate Credit Rating."

**Makwana, Ravi, Dhruvil Bhatt, Kirtan Delwadia, Agam Shah, and Bhaskar Chaudhury.** 2022. "Understanding and Attaining an Investment Grade Rating in the Age of Explainable AI."

**S and P Global Ratings.** 2024. "S and P Global Ratings."