Textual Analysis and Financial Statements

Isaac Liu

March 25, 2024

Introduction

high-level subject area info

(**?**)

problem statement and question

Company ratings and creditworthiness are important 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.

Are ratings based on hard numbers, or do company outlooks and sentiment also matter? Are they predictable?

note credit rating data access is limited and our model can be used to interpolate

In this project, we seek to explore whether incorporating the text of earnings calls improves predictions of corporate credit ratings.

high level data description

roadmap we then

Data and Exploratory Data Analysis

We combine a wide variety of data sources to support our predictions of credit ratings. Our scope of interest publicly traded companies from 2010-2016 (a limitation due to the availability of credit rating data). All data not from APIs was received as CSV or Excel files, which we converted to parquet files for efficient intermediate data storage throughout the project.

Data is at the level of

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

temporal imbalance

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 from Kaggle (**??**).

Each issuance be a change in rating (upgrade, downgrade) or reaffirmation - they occur at ad-hoc intervals.

Figure 1: Observations by Quarter and Year

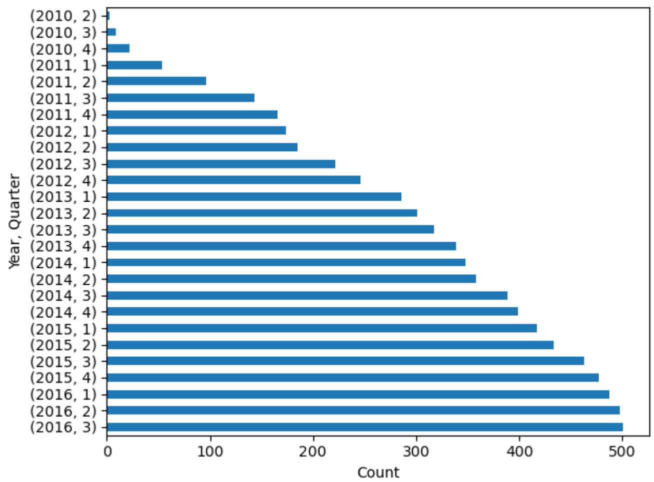
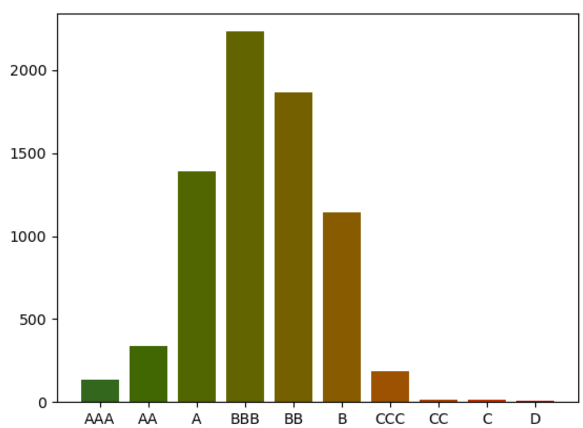


Figure ?? shows the distribution of rating grades used in our final dataset. Finer grades (+, -) are sometimes assigned by agencies, but these grades were removed for this project. Ratings of BBB and above are considered investment grade - these bonds carry empirical one-year default rates of 0 to 1%. Ratings below that are classified as junk, with default rates from 1 to 30, 40, or even 50% for some years (?). 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

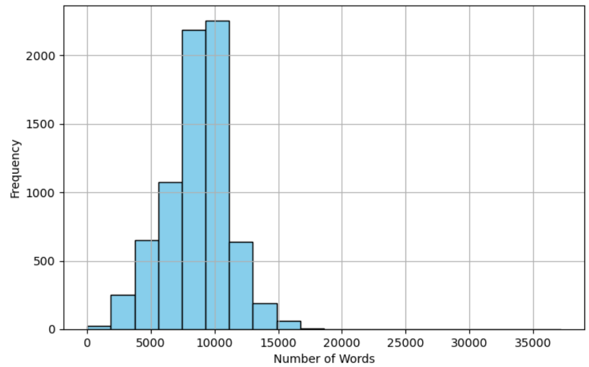
API source

Quarterly conference call transcripts that contains speaker remarks and Q and A session from 2010 - 2016.

Remove all calls that happened more than 250 days prior and after the fixed quarter date

The overall average call length stands at 8688.25words.

Figure 3: Number of Words in Earnings Calls



Financial Statements

API source

Items from balance sheet, cash flow statement, income statement, and company market capitalization

124 variables in total. Examples: revenue, total liabilities, net income, EBITDA

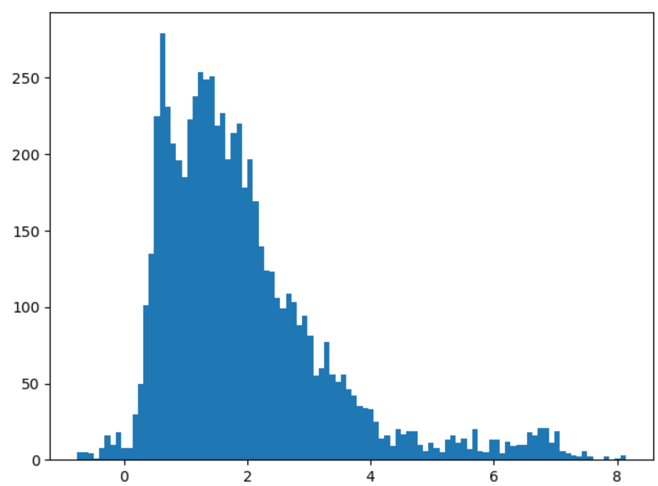
Limit to items reported in USD

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

Figure 4: Altman Z-Score

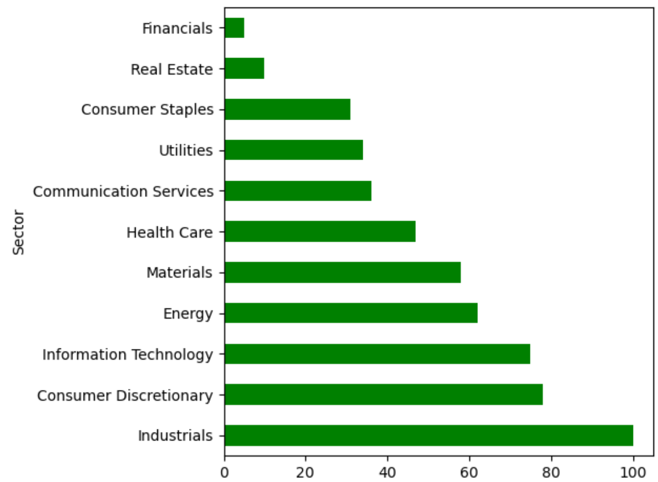


Sector

GCIS developed by S and P

Obtained from Kaggle with supplementary manual lookup

Figure 5: Firms by Sector



sectoral imbalance

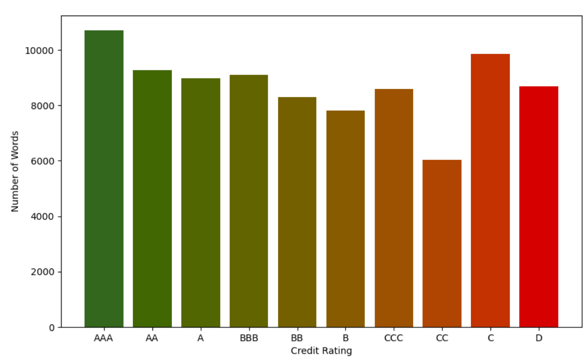
Quality Control

quality control code review of all data cleaning code numerous investigations

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

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

Alan Gewerc. Corporate Credit Rating with Financial Ratios, July 2020.

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

S and P Global Ratings. S and P Global Ratings, 2024.