

Corporate Credit Rating Prediction in the Energy Sector

Rodrigo Palmaka

Models

Goal: input financial data, output credit rating. Currently done manually by analysts of ratings agencies. Prone to bias/human error.

Feature Engineering: Dataset has 600+ financial features, how to pick?

SVM: Max margin classifier. Uses soft margins and penalty to handle non-linearly separable data. Used extensively in literature.

Neural Network: Build upon literature and try to implement custom architecture
MLP (Multi-Layer Perceptron).

Data

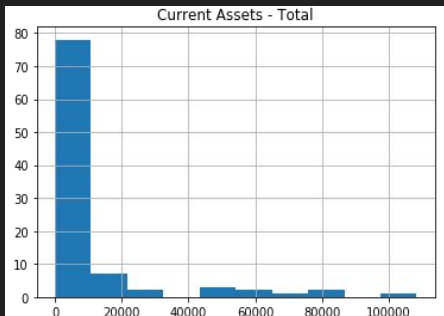
Label Distribution:

BBB	1525
BBB+	996
BBB-	951
A-	721
BB	707
BB+	706
A	463
AA	336
AA-	302
AAA	246
BB-	229
A+	221
B+	147
B	141
AA+	64
CCC+	57
B-	9
SD	1

Features:

- 96 companies
- $N = 3677$
- D (# of features) = 39
- Quarterly figures in range [01/01/2009, 01/01/2017]

Notable Features (grouped by company):



Stockholders Equity - Total

count	97.000000
mean	18514.691371
std	34030.046364
min	-98.244000
25%	1631.334000
50%	4706.000000
75%	17760.000000
max	164628.620900

Cash

count	97.000000
mean	1953.577573
std	4019.292973
min	0.000000
25%	77.400000
50%	348.034091
75%	1627.807775
max	19107.086950

Next Steps:

- Join features and labels on company and dates (labels are monthly, features quarterly)
- Understand time series assumptions