Corporate Credit Rating Prediction Using Multi-class Support Vector Machines, Random Forest, and Neural Network

Liyuan Chen, SCPD, liyuan@stanford.edu

PREDICTING

- Corporate credit rating reflects the risk of lending money to the company. Thus, it has been an important assessment and indicator for investors (Wang & Ku, 2021).
- The input to the algorithm is a group of companies' credit ratings and other features in the past.
- A multi-class Support Vector Machines (SVM) model is used to predict the corporate credit ratings. The model is then enhanced with Random Forest. A feed-forward neural network model is also created to compare with the SVM model on the accuracy of the prediction results.
- The output is the predicted corporate credit ratings.
- The model is evaluated on its accuracy score, which is the ratio of the correct prediction (true positive and true negative) to the total number of samples (Baeldung, 2021).

FEATURES

- The features, "Rating Agency", "Corporation", "Rating", "Rating Date", "CIK", "SIC Code", "Sector", and "Ticker", are properties of the company or rating agency. Therefore, they were not included in the experiment since they are not indicators on companies' performance.
- 16 features were selected which are highlighted in yellow.
- The data was split into training and test sets where 70% data were used for training and 30% were used for testing.

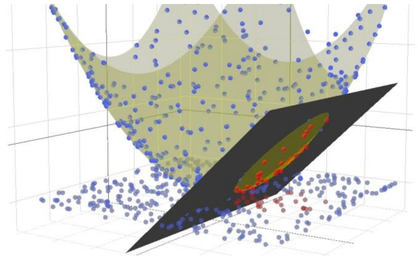
DATA

A set of data on corporate credit ratings is downloaded from Kaggle (Delwadia, 2022). There are 7805 records and 23 features in total.

Col	Feature	Description
1	Rating Agency	The institution who evaluates the corporation
2	Corporation	The entity that is being evaluated
3	Rating	The company's credit score
4	Rating Date	The date when the company is evaluated
5	CIK	Central Index Key, a key representing the entity used by Securities and Exchange Commission's (SEC's) computer systems
6	Binary Rating	The rating given to the company, which is either zero or 1
7	SIC Code	Standard Industrial Classification (SIC) codes by U.S. government
8	Sector	The industry focus of the company
9	Ticker	Company's stock market information / reports
10	Current Ratio	A liquidity ratio that indicates a company's ability to pay back in a short period of time
11	Long-term Debt / Capital	An indicator used to measure how much asset and debt a company has
12	Debt / Equity Ratio	A ratio between company's debt and its value of shareholders' equity
13	Gross Margin	The surplus when the cost is deducted from net sales
14	Operating Margin	The surplus when the cost is deducted from net sales per dollar
15	EBIT Margin	The ratio between company's total cost and total sales
16	EBITDA Margin	The ratio between a company's profit and its revenue
17	Pre-Tax Profit Margin	Company's operating revenue
18	Net Profit Margin	The surplus of company's earnings after all costs, tax expenses, etc. are deducted from the revenue
19	Asset Turnover	The ratio between company's revenue and assets
20	ROE - Return On Equity	The ratio between profit and shareholders' equity
21	Return On Tangible Equity	The earning of investment into the company
22	ROA - Return On Assets	An indicator on the profitability of the company; ratio between EBIT and the company's assets
23	ROI - Return On Investment	Earning from the investment
24	Operating Cash Flow Per Share	Net profit per share where inflation is also counted
25	Free Cash Flow Per Share	The amount of cash generated per company share in the past 12 months

MODELS

(1) Support Vector Machine (SVM)

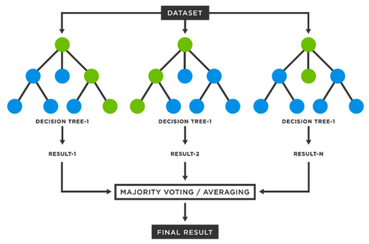


- An SVM model is selected since it is able to map samples to multi-dimensional space to classify them based on different features (Baeldung, 2021).

$$K(X_1, X_2) = \exp\left(-\frac{\| X_1 - X_2 \|^2}{2\sigma^2}\right)$$

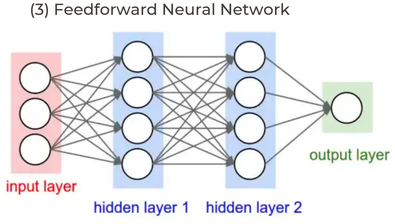
- Radial basis function (RBF) kernel is used since it has proven to be the most effective kernel in most cases (Savas & Dovis, 2019).

(2) SVM + Random Forest



- Random forest model calculates the importance of each feature using the number of samples reaching the end of each branch divided by the total number of samples.
- Thus, greater feature importance indicates that the feature has a stronger influence on the predicted results (Ronaghan, 2019).

(3) Feedforward Neural Network

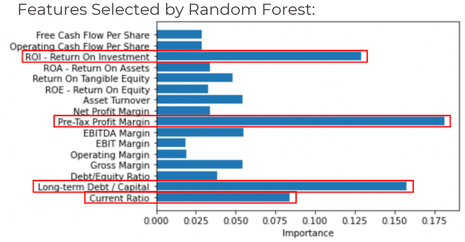


- A neural network refers to a model with multiple layers of processors where each layer of processors taking the output of the previous layer as input (Burns & Burke, 2021).
- A feed-forward neural network which does not form a cycle is selected due to its implicitly (DeepAI, 2019).

RESULTS

(a) SVM vs. SVM + Random Forest on All Data

Features Selected by Random Forest:

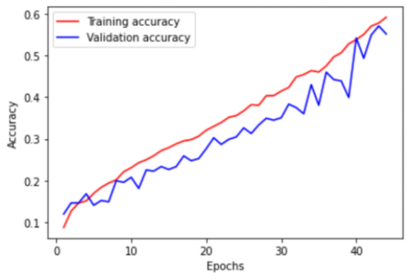


- By using the selected features, the prediction accuracy of SVM model increased from 12.1% to 16.2%

(b) SVM vs. SVM + Random Forest on Different Rating Agencies

Agency	No. of Samples	Accuracy (%)			Selected Features
		SVM	SVM + RF	↑ in Accuracy	
Standard & Poor's Ratings Services	2813	11.5	12.9	+1.4	'Current Ratio', 'Long-term Debt / Capital', 'Pre-Tax Profit Margin'
DBRS	26	25.0	0.0	-25.0	'Current Ratio', 'Debt/Equity Ratio', 'Gross Margin', 'Asset Turnover', 'Return On Tangible Equity', 'ROI - Return On Investment'
Moody's Investors Service	1636	15.7	12.6	-3.1	'Current Ratio', 'Long-term Debt / Capital', 'EBITDA Margin', 'Pre-Tax Profit Margin', 'Net Profit Margin', 'ROI - Return On Investment'
Fitch Ratings	477	16.0	19.4	+3.4	'Current Ratio', 'Long-term Debt / Capital', 'Gross Margin', 'EBITDA Margin', 'Pre-Tax Profit Margin'
Japan Credit Rating Agency Ltd.	22	28.6	57.1	+28.5	'Current Ratio', 'EBIT Margin', 'Asset Turnover', 'Free Cash Flow Per Share'
HR Ratings de Mexico S.A. de C.V.	5	0.0	50.0	+50.0	'Long-term Debt / Capital', 'Debt/Equity Ratio', 'Gross Margin', 'Operating Margin', 'EBIT Margin', 'EBITDA Margin', 'Pre-Tax Profit Margin', 'Net Profit Margin', 'Return On Tangible Equity'
Egan-Jones Ratings Company	2826	12.1	17.0	+4.9	'Current Ratio', 'Long-term Debt / Capital', 'ROE - Return On Equity', 'ROI - Return On Investment'

(c) Feedforward Neural Network for All Data



- Although 8000 epochs were set to be used to determine the best parameters, the model would stop iterations when there is no improvement in the validation loss for 10 consecutive epochs.
- The model stopped after the 44th epoch with a training accuracy of 59.2% and a validation accuracy of 55.2%

FUTURE

- In the future, since corporate credit rating varies with time, a Long Short-Term Memory model could be used and compared with SVM, random forest and neural network models.
- A white reality check could also be performed to evaluate the accuracy of the algorithm to make sure the predictions do not match the actual values by chance.

DISCUSSION

- The experiment shows that random forest is able to improve SVM model on corporate credit rating forecast in most cases. However, a large sample size is required to build the model.
- Feed-forward neural network is proved to be more accurate in predicting corporate credit rating than SVM and random forest.
- The model results showed volatility between different rating agencies, which could be caused by correlation between features.
- It is recommended to repeat the experiment using a larger sample size and independent features.

REFERENCE

- Baesse de Araujo, I., & Rezende Mustifaga, Y. (2021). Structure and experiments to recognition and classification of fake news about citizens in Brazilian government positions. Proceedings of The 11th International Conference on Research in Engineering, Science & Technology. https://doi.org/10.33422/11th.restconf.2021.03.15
- Baeldung. (2021, August 25). Multiclass classification using support Vector Machines. Baeldung on Computer Science. Retrieved October 28, 2022, from https://www.baeldung.com/cs/svm-multiclass-classification
- Burns, E., & Burke, J. (2021, March 26). What is a neural network? explanation and examples. Enterprise AI. Retrieved December 4, 2022, from https://www.techtarget.com/searchenterpriseai/definition/neural-network
- DeepAI. (2019, May 17). Feed Forward Neural Network. Retrieved December 4, 2022, from https://deepai.org/machine-learning-glossary-and-terms/feed-forward-neural-network
- Rathi, M. (2018, August 29). Feedforward Neural Networks. Mukuls Blogs RSS. Retrieved December 5, 2022, from https://mukulrathi.com/demystifying-deep-learning/feed-forward-neural-network/
- Ronaghan, S. (2019, November 1). The mathematics of decision trees, random forest and feature importance in Scikit-learn and Spark. Medium. Retrieved December 4, 2022, from https://towardsdatascience.com/the-mathematics-of-decision-trees-random-forest-and-feature-importance-in-scikit-learn-and-spark-f2861df67e3
- Savas, C., & Dovis, F. (2019). The impact of different kernel functions on the performance of scintillation detection based on support Vector Machines. Sensors, 19(23), 5219. https://doi.org/10.3390/s19235219
- TIBCO Software. (n.d.). What is a random forest? Retrieved December 5, 2022, from https://www.tibco.com/reference-center/what-is-a-random-forest
- Wang, M., & Ku, H. (2021). Utilizing historical data for Corporate Credit Rating Assessment. Expert Systems with Applications, 165, 113925. https://doi.org/10.1016/j.eswa.2020.113925