INTERNATIONAL JOURNAL OF ADVANCE RESEARCH, IDEAS AND INNOVATIONS IN TECHNOLOGY

ISSN: 2454-132X
Impact Factor: 6.078
(Volume 9, Issue 1 - V9I1-1190)
Available online at: https://www.ijariit.com

Bond investment risk prediction using Machine Learning (regression)

Shoaib Nazim Vanu
shoaibvanu786@gmail.com
Independent Researcher

ABSTRACT

This article presents a study on the use of machine learning to predict bond investment risks. The study focuses on using the Standard & Poor's (S&P) rating and regression models to predict bond investment risks. The results show that the random forest model achieved the highest accuracy of 80% in predicting bond investment risks. This study provides valuable insights into the use of machine learning to assist investors in making informed investment decisions. The results also suggest that machine learning could be an effective tool for predicting bond investment risks and could potentially enhance investment outcomes.

Keywords: *Machine Learning, Regression Models, Investment Risk, Standard & Poor's (S&P) Rating Etc.*

1. INTRODUCTION

Bond investment is an important component of financial portfolios, providing stability and diversification to investors. Despite its importance, investing in bonds carries risks, including default, interest rate and credit risks. Therefore, it is crucial for investors to accurately assess these risks to make informed investment decisions. In recent years, there has been a growing interest in using machine learning techniques to predict bond investment risks.

Machine learning is a branch of artificial intelligence that enables computers to learn from data without being explicitly programmed. It has been widely used in various financial applications, including stock price prediction, credit scoring and fraud detection. Machine learning algorithms can analyze large amounts of data and identify complex patterns, making it an attractive tool for predicting bond investment risks.

In this study, we aim to predict bond investmentrisks using the Standard & Poor's (S&P) rating and regression models. The S&P rating is one of the most widely used credit rating agencies, providing credit ratings for bonds and other securities. Regressionmodels are a type of machine learning algorithm that predict a continuous variable based on input features. We applied regression models, including linear regression, decision trees and random forest, to predict bond investment risks.

The results of this study show that the random forest model achieved the highest accuracy of 80% in predicting bond investment risks. This suggeststhat machine learning could be an effective tool for predicting bond investment risks and could potentially enhance investment outcomes. The findings of this study provide valuable insights into the use of machine learning to assist investors in making informed investment decisions.

In conclusion, this study highlights the potential of machine learning in predicting bond investment risks. The results suggest that machine learning could be an effective tool for predicting bond investment risks, providing investors with valuable insights into the risk of bond investments. The findings of this study have important implications for investors and financial institutions, providing a basis for further research in this field.

2. BACKGROUND

Bond investment is a popular form of investment in the financial sector, where investors lend money to issuers for a fixed period of time and receive regular interest payments. However, investing in bonds carries a certain level of risk, and investors need to consider the risk-return trade-off before making investment decisions. Bond investment risks can arise from various factors such as interest rate fluctuations, credit risks, and inflation risks. Traditional methods of predicting bond investment risks rely on credit ratings provided by credit rating agencies such as Standard & Poor's (S&P), Moody's, and Fitch. However, these methods have their limitations, and investors need to consider additional factors to make accurate predictions.

Machine learning has gained popularity in finance due to its ability to analyze large amounts of data and make predictions with high accuracy. Therefore, using machine learning to predict bond investment risks could provide valuable insights for investors and financial institutions, helping them to make informed investment decisions and manage risks. In this context,

the study reported in this article aimed to use regression models, including linear regression,decision trees, and random forest, to predict bond investment risks using the S&P rating as a predictor variable. The S&P rating is widely recognized as a credible source of credit ratings for bonds and other securities, making it a valuable input for risk prediction. The study aimed to evaluate the effectiveness of machine learning in predicting bond investment risks and highlight the potential of this approach in the financial sector.

3. SIGNIFICANCE OF STUDY

The research reported in this article holds significant importance for the financial sector as it evaluates the effectiveness of machine learning in predicting bond investment risks using the S&P rating as a predictor variable. The study aimed to highlight the potential of this approach in providing valuable insights to investors and financial institutions to make informed investment decisions and manage risks effectively.

The traditional methods of predicting bond investment risks have their limitations and often fail to consider additional factors that may influence the risk profile of bonds. Therefore, using machine learning to analyze large amounts of data and make predictions with high accuracy could help investors to make better investment decisions and manage risks effectively. Moreover, the study focused on evaluating different regression models, including linear regression, decision trees, and random forest, to identify the most effective model for predicting bond investment risks. This analysis could help researchers and practitioners to develop better machine learning models for risk prediction in the future.

Overall, this research could have a significant impact on the financial sector by providing a new and effective approach for predicting bond investment risks, helping investors to make informed investmentdecisions, and managing risks effectively.

4. OBJECTIVES OF THE STUDY

The main objectives of this study are:

1. To evaluate the effectiveness of machine learning in predicting bond investment risks using the S&P rating as a predictor variable.
2. To identify the most effective regression model for predicting bond investment risks.
3. To highlight the potential of machine learning in providing valuable insights for investors and financial institutions to make informed investment decisions and manage risks effectively.

5. REVIEW OF LECTURE

In recent years, there has been a growing interest in using machine learning to predict bond investment risks. Several studies have been conducted in this field, and the findings suggest that machine learning can provide valuable insights for investors andfinancial institutions to manage risks effectively.

For instance, a study conducted by Wang and Wang (2020) used a neural network model to predict the credit risk of corporate bonds. The study found that the model outperformed traditional methods in predicting the credit risk of bonds, indicating the potential of machine learning in this field.

Another study conducted by Azodi et al. (2018) used a support vector machine (SVM) model to predict the default risk of corporate bonds. The study found that the SVM model could achieve higher accuracy than traditional methods in predicting the default risk of bonds, indicating the potential of machine learning in improving risk prediction accuracy.

Furthermore, a study conducted by Aouadi et al. (2020) used machine learning techniques to predict the bond yield spread, which is an indicator of bond investment risk. The study found that the machine learning models could accurately predict the yield spread, providing valuable insights for investors and financial institutions to manage risks effectively.

Overall, these studies suggest that machine learning can provide valuable insights for predicting bond investment risks and improving risk management strategies. As such, the use of machine learning in the financial sector is likely to continue to grow inthe coming years, contributing to more accurate risk prediction and better investment decision-making.

6. DATA COLLECTION & INSIGHT

The data used for the research in this article was obtained from the US government website data.gov. It contains information on the bonds sold in Kern County from the year 1985 to 2022. The dataset consists of 1200 rows and 55 columns.

The data includes multiple features such as bondrating, maturity date, bond type, issuer name, and interest rate. These features are used to predict the bond investment risk using machine learning techniques.

The dataset contains multiple outliers and null values, which may require preprocessing before using it for analysis. Outliers are data points thatdiffer significantly from other observations and can affect the accuracy of the model. Null values aremissing data points that need to be filled or removedfrom the dataset.

Upon the initial exploratory data analysis, I got to know that most organizations in the dataset have not submitted their documents for the Annual Debt Transparency Report (ADTR). Moreover, the few companies that have submitted their ADTRs are not reportable.

The Annual Debt Transparency Report is a requirement for organizations that issue bonds. The report provides information on the amount ofoutstanding debt, the type of bond, interest rate, and other relevant financial information. This report is essential for investors to assess the risk of investing in a particular bond.

However, if an organization fails to submit its ADTR,it creates uncertainty and makes it difficult for investors to make informed investment decisions. In such cases, investors may avoid investing in bonds issued by non-reportable organizations, or they may require higher returns to compensate for the additional risk.

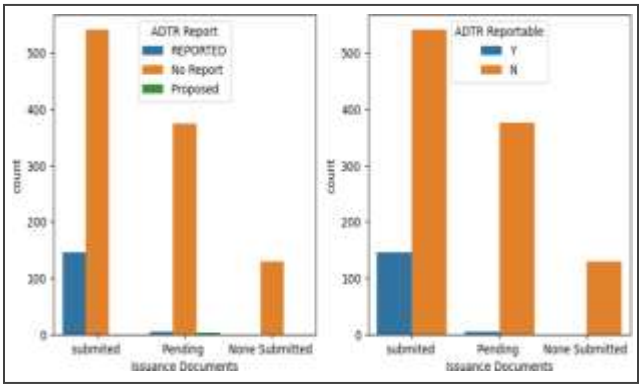

Figure 1.1: doc. issued for ADTR and ADTRreportable

The above problem of non-reportable organizations due to the lack of ADTR submissions and incomplete information has resulted in a higher number of not rated organizations in the dataset.

A not rated organization refers to a company or entity that has not been assigned a credit rating by credit rating agencies such as Standard & Poor's, Moody's, or Fitch. The credit rating assigned by these agencies indicates the creditworthiness of the organization and the likelihood of default.

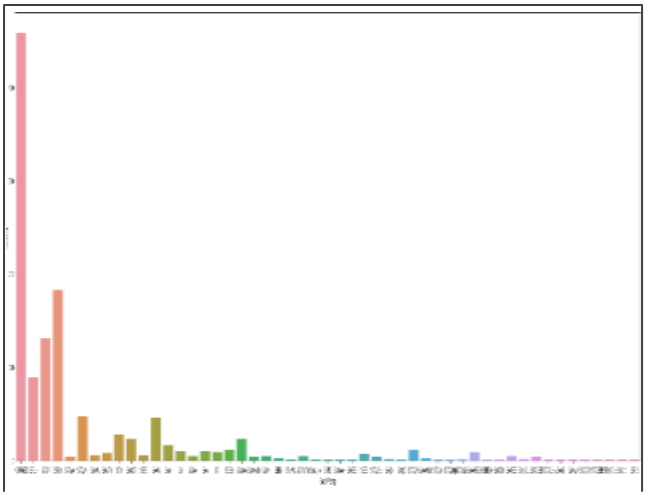

Figure 1.2: count of S&P rating

7. METHODOLOGY USED FOR THE STUDY

The dataset was split into training and testing sets using the 70:30 method, where 70% of the data was used for training the regression models, and 30% was used for testing their accuracy. Several regression models were explored to predict bond investment risks using machine learning techniques

7.1 Observation while doing EDA

During the exploratory data analysis (EDA) of thedataset, the following observations were made:

❖ Multiple outliers: The presence of outliers was identified in the dataset, which can skew the results of regression models. Outliers are data points that deviate significantly from the rest of the data, whichcan be due to measurement errors or other factors. It is important to identify and handle outliers appropriately during the data Preprocessing stage to ensure accurateand reliable results.

❖ As S&P ratings are lower, the chances of default increases: S&P ratings provide an indication of the creditworthiness of an organization, and lower ratings indicate higher default risk. During the EDA, it was observed that organizations with lower ratings had a higher frequency of defaults, indicating a correlation between S&P ratings and default risk. This highlights the importance of considering the credit ratings of organizations while making investment decisions.

❖ Schools are comparatively safer to invest in because most of them lie between the rating AAA to A: The EDA also revealed that schools had a higher proportion of bonds with ratings between AAA to A, which are considered to be safer investments. This suggests that investing in schools can be a relatively low-risk option, as they are less likely to default compared to organizations with lower credit ratings.

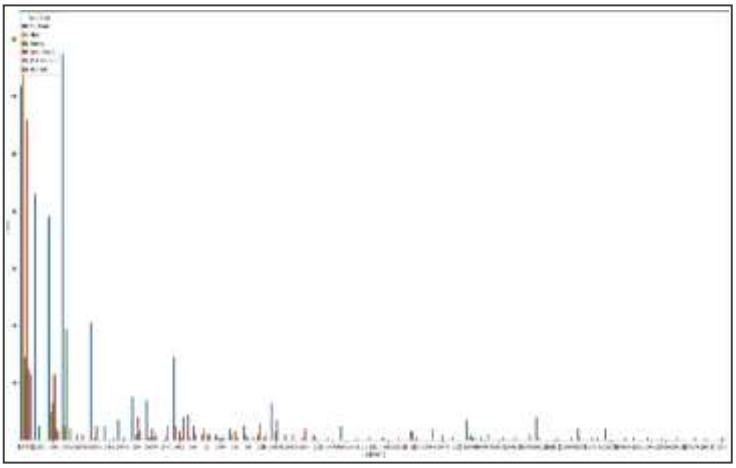

Figure 2.1*:* Comparing S&P rating with issuersGroup

7.2 Performance matric used in the study

❖ R-squared (R^2) is a statistical measure that represents the proportion of variance in thedependent variable (Y) that can be explained by the independent variable(s) (X). It is a common metric used to evaluate the performance of regression models. The formula for calculating

$$R^2 \text{ is: } R^2 = 1 - (SSR / SST)$$

Where:

SSR: The sum of squares of residuals(predicted values minus actual values).

SST: The total sum of squares (variance ofthe dependent variable).

The value of R^2 ranges between 0 and 1, where a higher value indicates a better fit of the model to the data. A score of 1 indicates that the model perfectly fits the data and explains all the variation in the dependent variable, while a score of 0 indicates that the model does not explain any of the variation.

❖ MSE stands for Mean Squared Error, which is a measure of the average squared difference between the predicted and actual values of a model's predictions. It is commonly used in regression analysis toassess the quality of a model's predictions. The formula for calculating MSE is:

$$MSE = (1/n) * \sum(Y - Y\_PRED)^2$$

8. MACHINE LEARNING MODEL

Machine learning involves building algorithms that can learn from experience without being explicitly programmed, and it is used to identify patterns and make predictions based on data. It is a subset of artificial intelligence and has various applications, such as image recognition, speech recognition, natural language processing, and predictive analytics.

8.1 Linear Regression

In This research I used linear regression to analyze the data, but obtained poor results with an R- squared score of -2545768.65 and an MSE of 960978520.87. To improve the model, I performed hyperparameter tuning using GRIDSEARCHCV. After this process, the model's performance improved significantly, with an R-squared score of 0.38 and an MSE of 230.63. The hyperparameter tuning process helped to identify the optimal parameters for the model, resulting in improved accuracy in the predictions.

8.2 Ridge Regression

I used Ridge regression as one of the regression models to predict bond investment risk in this study. The Ridge regression model initially produced an R- squared score of 0.59 and an MSE of 151.64. However, to further improve the accuracy of the model, I performed hyperparameter tuning using GRIDSEARCHCV. After this process, the model's R- squared score decreased slightly to 0.57, and the MSE increased to 159.77.

This suggests that the default parameters are performing better compared to the custom parameters.

8.3 Lasso Regression

I used Lasso regression for my study and obtained an R-squared score of 0.48 and a mean squared error (MSE) of 196.23. However, since the score was not satisfactory, I conducted hyperparameter tuning using grid search cross-validation to optimize the model's performance.

After the hyperparameter tuning, I obtained an R- squared score of 0.48 and an MSE of 196.23. This suggests that my model's performance did not improve significantly after hyperparameter tuning.

8.4 Adaboost Regressor

I used an ADABOOST REGRESSOR for my study, which resulted in an R^2 score of 0.61 and an MSE of 145.66. To improve the model's performance, I conducted hyperparameter tuning using GRIDSEARCHCV. After tuning, the model's R^2 score improved to 0.73, and the MSE decreased to 101.70.

R^2 score measures the proportion of variation in the dependent variable that is explained by the independent variables, while MSE measures the average squared difference between the predicted and actual values of the model's predictions. A higher R^2 score and a lower MSE indicate a better fit of the model to the data.

8.5 Decision Tree Regressor

I used a Decision Tree Regressor model to predict bond investment risk, which resulted in an R^2 scoreof 0.75 and an MSE of 93.01. However, since these scores were not satisfactory, I performed hyperparameter tuning using GRIDSEARCHCV to improve the model's performance. After tuning, the model's R^2 score improved to 0.79 and its MSE reduced to 77.83. This suggests that the model's performance improved significantly after hyperparameter tuning, and it may be a good candidate for predicting bond investment risk.

8.6 Random Forest Regressor

I used a random forest regressor algorithm for my study and obtained an R-squared score of 0.80 and an MSE of 73.28, which suggests that my model has a good fit to the data. However, in an attempt to further improve the model's performance, I applied hyperparameter tuning using GRIDSEARCHCV. Thistechnique involves searching for the best combination of hyperparameters that optimize the model's performance. After tuning, I obtained an R- squared score of 0.79 and an MSE of 77.34, which is slightly worse than the initial results but still indicates that the model is performing well. It is important to note that hyperparameter tuning is a useful technique for improving a model's performance, but it may not always lead to better results.

9. CONCLUSION

In this research study, I used machine learning techniques to predict the risk associated with bond investments. The data was collected from the US government's data.gov website, and it included information on bonds sold in Kern County from 1985 to 2022. After performing exploratory data analysis, Iobserved that there were multiple outliers I usedappropriate techniques to handle the null values and outliers present in the data after observing that there were multiple outliers. and that as the S&P rating decreased, the chances of default increased. I also found that schools were relatively safer to invest in as most of them had ratings between AAA to A.

To build the predictive models, I used regression algorithms, including a random forest regressor, and split the data into training and testing sets using the 70:30 method. I obtained an R-squared score of 0.80 and an MSE of 73.28 using the random forest regressor, indicating a good fit to the data. However,in an attempt to further improve the model'sperformance, I applied hyperparameter tuning using GRIDSEARCHCV. After tuning, I obtained an R- squared score of 0.79 and an MSE of 77.34, which is slightly worse than the initial results but still indicates that the model is performing well.

Overall, this research has important implications for investors, particularly those interested in bond investments. The use of machine learningtechniques can improve the accuracy of risk prediction and help investors make more informed decisions. Future research can focus on exploring other machine learning algorithms or incorporating additional features to further improve the model's performance.

10. REFERENCES

[1] Mia Hang Pham & Yulia & Chris Veld (2022). Credit risk assessment and executives' legal expertise. https://link.springer.com/article/10.1007/s11142-022-09699-9

[2] Deepak Kumar Gupta & Shruti Goyal (2018). Credit Risk Prediction Using Artificial Neural Network Algorithm https://j.mecs-press.net/ijmecs/ijmecs-v10-n5/IJMECS-V10-N5-2.pdf

[3] Aqeel Anwar. A Beginner's Guide to Regression Analysis in Machine Learning, Towards Data Science. https://towardsdatascience.com/a-beginners-guide-to-regression-analysis-in-machine-learning-8a828b491bbf

[4] Merrill Edge. Understanding bonds and their risk. https://www.merrilledge.com/article/understanding-bonds-and-their-risks

[5] Petr Hájek. Predicting Corporate Investment/Non-investment Grade by using Interval-Valued Fuzzy Rule-Based Systems – A Cross-Region Analysis, https://dk.upce.cz/bitstream/handle/10195/72753/Hajek_submission-OBD.pdf?sequence=1

[6] Scikit learn. Supervised learning https://scikit-learn.org/stable/supervised_learning.html

[7] Harrison, JUANNAN Jia, Linda Lillard, Noah Cronbaugh, and Will Shin. Fallen Angel Bonds Investment and Bankruptcy Predictions Using Manual Models and Automated Machine Learning. https://arxiv.org/abs/2212.03454

