OW Research Assessment #2

Assessment #9 Annotated Bibliography

Mohit Nair

November 13, 2025

Data Analytics with a concentration in Prescriptive Analytics

MLA Citation:

Petropoulos, Anastasios, et al. ***A Robust Machine Learning Approach for Credit Risk Analysis of Large Loan Level Datasets Using Deep Learning and Extreme Gradient Boosting 1*****. 2018.**

In the article, "A Robust Machine Learning Approach for Credit Risk Analysis of Large Loan Level Datasets Using Deep Learning and Extreme Gradient Boosting," authors Anastasios Petropoulos et. al cover the analysis of corporate financial information using a method involving extreme gradient boosting (XGBoost), an extreme, advanced, version of the gradient boosting model that builds upon the mistakes of the current model by adding on new submodels such as binary decision trees. To conduct this analysis, Petropoulos and researchers first collect the data needed. In this study, only loan information from corporate businesses are used due to the varying nature of financial institutions such as banks. They then define what it means for a customer to default (meaning they can no longer pay off the loan), and this occurs when the loan is either 90 days overdue or is simply rated as defaulted. The researchers cover a span of 10 years (2005-2015), analyzing the default rate for each of the years and comparing it to the number of customers. To further develop the analysis, the researchers determine a set of 354 predictor variables, including Return on Equity (which is found to be the most predictive of financial health for corporations). They also utilize several XGBoost related algorithms, such as Boruta and the XGBoost Package. Intriguingly, Boruta, the XGBoost Package, as well as other packages mentioned are all provided by R, the programming language used for data analysis, which I have to dive into further later on in my OW process.

"A Robust Machine Learning Approach for Credit Risk Analysis of Large Loan Level Datasets Using Deep Learning and Extreme Gradient Boosting" was published in 2018 under the Bank of International Settlements' Irving Fisher Committee on Central Bank Statistics (IFC). The IFC is a research forum that publishes the works of Central Bank Economists and Statisticians. The authors of this article - Anastasios Petropoulos, Vasilis Siakoulis, Evaggelos Stavroulakis, and Aristotelis Klamargias - are experts from the Bank of Greece, connecting their knowledge and personal experience to such a vast matter. In fact, the authors are able to build their research by providing personal examples that involve the work of the Bank of Greece that

helps readers understand and visualize what the researchers aim to present. The purpose of this article is to hone in on the analysis of financial health for corporate businesses using a gradient boosting model that successfully identifies key factors and can help determine how well a corporation is doing economically.

This article is one that guides me in the right direction for my Original Work. For one, it provides data that I can look into, as well as clear and understandable methodologies that I can apply to my OW and further research. However, the biggest takeaway from this article in regards to application for my OW is the fact that many of the packages mentioned in the article, such as Boruta or XGBoost can be found in R. This proves to be a significant advantage - because I have some experience with R and R Studio, I can use these packages to my advantage and do exactly what I want to in regards to analyzing financial data as well as other data. This is something that will be a major part of my OW this year, as this can be one of my main methods of analyzing data.

ANNOTATED ARTICLES:

"A Robust Machine Learning Approach for Credit Risk Analysis…"

https://kami.app/ggw-gTH-2ec-szi