Predictive Analytics Case Study

David D Berberena

Bellevue University

DSC 630 Predictive Analytics

Andrew Hua

July 7, 2024

<center>**Predictive Analytics Case Study**</center>

Introduction

The case study that has been reviewed revolves around the use of predictive analytics and machine learning algorithms to better aid the banking sector in assessing and managing financial risk. The author wished to understand how certain macroeconomic factors (gross domestic product (GDP), inflation rates, interest rates, etc.) and credit risk factors would affect a bank's ability to identify financial risk factors (Rofi'i, 2023). This identification of financial risk factors is vital to the banking sector due to the potential loss of millions of dollars should a bank fail to be proactive in its approach to combat the mass effects of multiple macroeconomic forces. With data now so prevalent in the modern business world, the banking sector is expected to keep pace with technological advances that could help in more efficient and swift decision-making. The intended goal of the case study was to identify the relationship of major macroeconomic, manufacturing, and credit risk factors to the banking sector's financial risk factors by employing predictive modeling techniques. Should a bank manage to accurately capture, interpret, analyze, and model data related to financial risk management, that bank would then be able to prompt a proactive approach when managing the increased risk in dealing with potential investors in an ever-fluctuating economy.

The author continues by stating that the data acquired for the research within the case study journal article stems from three major avenues: the Data Management and Statistics Department of the Financial Services Authority (provided Indonesian banking statistics data from February 2022 to September 2023), the Central Statistics Agency (provided Indonesian manufacturing industry data from 2022 and 2023 as well as the 2023 Indonesian economic report), and Bank Indonesia (provided Indonesian macroeconomic data from February 2022 to October 2023). The data was procured to view the most accurate and current representation of Indonesian credit and deposit trends, the manufacturing industry, and the macroeconomic conditions of Indonesia as they relate to the

identification of risk factors with the potential to impact financial risk assessment and management for the Indonesian banking sector.

Methods and Results

With the data ready to iterate over using exploratory data analysis (EDA), the author had no trouble moving onto the data mining phase of the task at hand. This is most likely due to the high quality of clean and user-friendly derived from the aforementioned credible business entities. Of the three major data sources, each was addressed with a different approach. The article depicts a Python-generated visualization showcasing the findings of the Financial Services Authority, which is to represent the current standing of the banking sector's clients and their statistical aggregated data. Regarding the risk factors shown on the chart (amount of available credit and number of deposits), both tend to increase together at a steady rate over time, revealing that these two may have a significant impact on the banking sector as the picture the visualization paints is one explaining that the ability and willingness to spend is available. In conjunction with this data, the author also shares that macroeconomic factors that could command a significant effect on financial risk assessment include inflation and GDP. From this first data avenue, we can identify these first few relevant features.

Additional EDA and preliminary model generation using the Central Statistics Agency data takes us through the manufacturing industry in Indonesia during the period mentioned in the case study article. It was stated that a strong correlation existed between increased production output from the manufacturing sector and increased credit demand from the banking sector. Using ordinary least squares regression, the author attempted to confirm the high correlation as indicative of being either significant or irrelevant, with the short-term goal of adding another feature to the list of relevant features that could affect financial risk management. Looking at the results of the model, a coefficient of determination value of 1.0 and a P-value of 0.00 confirms that the ordinary least squares model fit the

data well and that the manufacturing output predictor variable is highly significant to the increase of credit demand. Applying the macroeconomic factors seen above as well as interest rates, it is easy to see how the banking industry is reactionary toward these influencers, implementing less lenient credit policies and assessing the liquidity of bank assets.

While the Bank Indonesia macroeconomic data has been cross-analyzed with the two previous data sources, these factors do receive a visualization of their own to view how they each fluctuate compared to each other over a 12-month period of analysis. The author makes several intuitive assumptions about the results, cementing the opinion that inflation rates, interest rates, and GDP growth may indeed be of significance as they relate to the banking sector's financial risk factors and how they might be skillfully placated.

With several first-round picks of potentially impactful factors to the identification and management of the banking industry's financial risk factors, the problem was solved by crafting a predictive classification model that was instrumental in confirming whether the factors in question made a significant impact in the banking sector's risk management efforts or not. The model used for this task was a logistic regression model, and the data had been split into training and test sets at an 80/20 ratio. The sigmoid function was used in the author's formula to convert the output variable to a probability, strengthening the assumption of a multinomial logistic regression model. After reviewing the article, the model chosen was the most logical choice as the EDA already revealed that all of these features were relevant to the banking industry, and the model simply needed to run the variables and detect whether they were significant in predicting the banking sector's financial risk factors. This way of finding a solution to the task here is indicative of a classification problem, and logistic regression is one of the best ways to handle numerical input variables to realize a categorical probability outcome. The sigmoid function smooths the output to a scale between 0 and 1, which is extremely favorable for the output needed to confirm or deny the feature's significance past a certain threshold.

Once the model was created, it was tuned, cross-validated, and evaluated using the accuracy, precision, and recall statistics attached to the model's results. These metrics were chosen as the standard to evaluate the success of the model in how well it performed with the test data after being trained to classify the features as significant or not and at what probability were the features impactful. The accuracy statistic here is meant to identify the percentage of which the model correctly identifies the feature as significant or not impactful. Precision is measured by the percentage of significant instances identified by the model against those actually proven to be correct, and the recall metric highlights the proportion of the model's ability to identify all truly significant instances. After being adjusted through the tuning and cross-validation stages, each of these metrics is indicative of how well the model can perform.

Conclusion

With the results in hand, the true goal of the article's research was never to put the model into deployment or any similar action but to simply show the effectiveness of predictive analytics and machine learning algorithms in identifying the Indonesia banking sector's financial risk factors using data that presented predictor variables such as Indonesian manufacturing industry factors, banking factors, and macroeconomic factors. However, the implications of the results were discussed at length within the journal article, as the banking sector can learn to harness predictive analytics and machine learning algorithms to shift from a reactive stance to a proactive stance when it comes to the external factors seen in this study. The research was performed and completed in an ethically sound and legal way, with the data being accumulated from reputable sources who are held to higher data standards so as not to disclose any confidential and/or sensitive data, so the belief is that there may be no actionable consequences involved in the formulation of the results of the case study.

The author learned from the results that while the model generated performed well on the data, there is still room to grow. This is easily seen with the metrics of accuracy, precision, and recall coming in at 85%, 80%, and 75%, respectively. Digging deeper into the model's results, it was found that inflation and interest rates were the two most significant predictor variables for the banking sector's financial risk factors. Operational efficiency also was taken into account with the final model being trained and tested, and the results showed that the model generalized well for the operational efficiency in addition to the macroeconomic factors. A recommendation that the author made to banks in Indonesia was that they should incorporate a more involved macroeconomic data analysis strategy into their financial risk assessments, as this would allow for a more proactive approach to allow for better weathering of economic conditions.

Overall, the opinion is that the article's stated research covered many aspects of predictive analytics and machine learning techniques with a great level of detail relating to the business problem at hand to come up with a data-driven recommendation and plan of action. The author handled the research well and was very meticulous in the search and subsequent display of aggregated data. The model chosen was effective and yielded a favorable set of statistics. Looking over each aspect of the study a final time, one thing that could have been done differently was during the OLS model analysis of manufacturing output to credit demand, more features and/or data should have been added. While having an R-squared value of 1.0 is good, a perfect score like that typically means the model is either under or overfitted, and in this case, running the model on only one feature could highlight the model's underfitted result. Adding other manufacturing sector factors would help the model properly analyze each one with increased accuracy as more data would allow the model to generalize better. All in all, the case study was an excellent example of what predictive analytics has the power to tell businesses and in what direction those businesses should go to prosper from the information the data has revealed.

References

1. Rofi'i, Y. U. (2023). Financial Risk Management in Indonesian Banking: The Integrative Role of Data Analytics and Predictive Algorithms. International Journal Software Engineering and Computer Science (IJSECS), 3(3), 300-309. https://doi.org/10.35870/ijsecs.v3i3.1823.