sciendo

Business Risk Prediction and Management Based on Artificial Intelligence by Applying Machine Learning Algorithms to Increase Business Flexibility and Social Stability: Banking Credit Risk Approach

Mohammad Sadegh SALEM*
Bucharest University of Economic Studies, Romania
**Corresponding author, sadeghsalem@gmail.com*

Abstract. *Credit risk assessment is vital for financial sustainability in the banking industry. This paper explores the application of artificial intelligence techniques in improving business risk prediction, particularly focusing on the risk of credit default. We constructed and tested five machine learning models – Logistic Regression, Decision Tree, Random Forest, Support Vector Machine, and Neural Network – against the dataset of historical loans. Each model was subjected to performance evaluation through standard classification metrics (accuracy, precision, recall, F1 score) and visual tools (confusion matrices, error distribution plots, and predicted vs. actual charts). The findings indicate that the Random Forest method was superior to the other methods as it produced the highest prediction accuracy and the model with the best balanced performance. The ensemble method did perform better but it was not at the same level as Random Forest. These results showcase that the application of machine learning in lending is great. With these models, banks can assess risk factors of given loans much more accurately. Better predictive accuracy will definitely greater assistance in riskier lending and consequently, aid in social and economic stabilization by building up financial security.*

Keywords: Credit Risk Prediction, Machine Learning, Banking Sector, Artificial Intelligence, Risk Management, Business Flexibility, Social Stability.

Introduction

The modern banking system has become central to economic growth because they provide capital resources and other forms of credit to businesses and individuals. Nevertheless, all loans come with adverse selection which has some rather grave consequences for both the lending institution and the economy. This brings forth the need to accurately evaluate credit risk for the sake of retaining health within the economy and supporting growth. Indeed, poor financial management and credit risk evaluation are the underlying reasons for financial crises and bank failures, making the development of risk assessment models more critical than ever. Common forms of evaluating credit risk rely on statistical techniques (e.g. logistic regression) combined with expert appraisal. To be fair, they have worked to a degree, but they have consistently underperformed when it comes to dealing with complicated non-linear relationships in large datasets. In addition, human evaluations can be very subjective, while simple models often ignore the complex relationships between characteristics of the borrowers and the events of default. While each of these is a single problem, the combination can result in severely flawed credit provisioning policies, be it giving loans to borrowers with a high chance of defaulting or waiting too long for worthy applicants. The new technology in AI and Machine Learning has opened many options to enhance predicting the risk in credit lending. The most advanced ML algorithms can uncover complex interactions among data that are concealed from human eyes as well as analysis based upon linear models. In the past few years, many researchers have tried different approaches using ML techniques on problems related

DOI: 10.2478/picbe-2025-0102

to credit risk, and they were able to achieve results that are much more favorable than the previous techniques AiUe used. The use of ML models has implications for the banking industry, because it can enable better detection of risky loans, thereby enhancing the chances of lowering the default rates. This saves the financial institutions from crises and enables them to extend more credit without worrying too much about default, which would promote economic growth and social stability. The above stated claims were verified using a banking credit dataset in which clinical information from a patient was analyzed to make prediction(s) about Loan Default. We analyze five classification techniques, namely Logistic Regression, Decision Tree, Random Forest, Support Vector Machine (SVM) and Neural Network. We use both quantitative and heuristic metrics to determine the most effective technique. The purpose of doing this is to find out which path taken in AI/ML would lead to the most accurate results in forecasting credit risk. By analyzing the strengths and weaknesses of each model, we provide insights into how AI-based risk assessment can enhance traditional risk management practices. The findings have practical implications for the banking sector, suggesting how machine learning models can be integrated into risk management frameworks to increase business flexibility (through better-informed lending decisions) and promote social stability by mitigating systemic financial risk.

Literature review

Machine Learning Advances in Credit Risk Assessment

Many researchers have proved that machine learning technologies perform better than traditional statistical methods in assessing credit risk. For example, a more recent study on predicting credit default tended to conclude that more advanced machine learning techniques, particularly deep learning, outperform classical models in predicting precision. In particular, it has been shown that certain neural network based methods outperform other linear techniques such as logistic regression in capturing many complex patterns of defaults (Kou et al., 2019; Shoumo et al., 2019). Moreover, well-covered reviews of applications of AI in finance published machine learning as the focus of modern risk management (Aziz & Dowling, 2018; Mashrur et al., 2020). Similarly, algorithms of decision trees, SVMs, and neural networks have also been successfully applied for predictive corporate credit rating and default forecasting (Golbayani et al., 2020; Moscatelli et al., 2020). These studies prove the capabilities of AI models in risk management for banking that are beyond traditional methods. The application of AI technology into management of risks is indeed seen to be the next revolution within the banking sector (Aziz & Dowling, 2018).

Ensemble Methods and Hybrid Approaches

Ensemble methods stands out as machine learning models that are well suited for credit risk classification. Techniques such as Random Forest and gradient boosting have proved their worth by employing the use of multiple learners to heighten accuracy and robustness. As an example, there are studies that show that prediction of loan defaults by ensemble models is more accurate than prediction done by single classifiers (Mashrur et al., 2020). Specifically, the Random Forest algorithm has been given much attention since it is capable of processing complex features while minimizing overfitting which makes it ideal for financial risk problems. Also, the inclusion of domain knowledge is one additional way of improving the results of machine learning algorithms. Munkhdalai et al. (2019) showed how financial knowledge, such as derived credit scores or even economic indicators can aid in machine learning models to improve the predictions of credit risk. These approaches that augment data with industry knowledge tend to improve the accuracy of

models and make them easier to understand than pure statistical models and enable realistic and practical decisions to be made.

Challenges in Implementing ML for Credit Risk

The challenges noted in credit risk models are complex yet detectable. One of those challenges is class imbalance. In a given lending portfolio, there often exists a vastly greater number of non-defaulted loans as compared to the number of defaulted ones. This can skew models that aim to achieve high accuracy. For example, an ML model can easily achieve the upper echelon of accuracy if the model only predicts the most populous category, while neglecting the opposite minority class. Another issue is model interpretability. It has already been noted that financial players work under an intensely controlled environment and they have to justify their credit actions to key stakeholders, and regulators for that matter. It is often difficult to understand why high risk and low risk loans are classified as such with more advanced models like neural networks or ensembles, which can tend to operate as a "black box" (Thiel, 2019). This opacity makes it difficult to use these models in much practical settings. Furthermore, there are difficulties of incorporating a broad range of information, including credit reports, transaction files, and even economic data. Large scale customer data brings other complications too like data privacy and the accuracy of the AI models used. Researchers continue investigating how to mitigate these challenges. For example, some are working on explainable AI techniques that assist in interpreting model outcomes and accuracy, fairness, and compliance in model design. (Gupta & Goyal, 2018). Addressing these issues is crucial for developing robust, trustworthy credit risk models that can be deployed in real-world banking operations.

Methodology

Machine Learning Models

This research analyzes five supervised machine learning algorithms that predict the credit risk. Each model offers a different classification and has its advantages and value.

Logistic Regression (LR) is a base logical classifier that predicts the probability of the default using the default logistic function. In addition, it presumes that there exists a linear relationship between the input features and log-odds of default. Logistic regression is efficient, easy to implement and performs calculations whose outputs are in the form of probabilities. Its reimbursement makes it a commonly accepted tool in credit scoring and useful in benchmarking more complex models. (Zhang, 2024)

Decision Tree (DT) is a two-dimensional, recursive model that continuously divides the data into branches which makes the maximum separation between classes. On every node, the algorithm decides the feature to be used as well as the threshold which will give the maximum information gain or impurity reduction (Farrokhi et al., 2020). Decision trees are have logical but non-probabilistic interpretation (each path through the tree defines a set of decision rules), are able to process both quantitative and qualitative data, and capture nonlinear relationships. These models have been applied for many problems in finance like prediction of loan defaults, and customer segmentation. (Zhang 2024, Chaudhary 2020)

The Random Forest (RF) algorithm, as defined (Jiménez in 2023), aggregates decision tree predictions. For RF, each decision tree is trained using random features and data subsets which, in return, reduces overfitting and the variance each decision tree has individually. As most single trees have varying rates of underfitting, RF becomes more accurate for high-dimensional data. Random

Forest models are superb index of choice due to their strong performance and robustness as mentioned by (Zhang, 2023) and (Chaudary, 2020).

Support Vector Machine (SVM), on the other hand is widely popular for being best suited in high-skewed data. It classifies defaulters from non-defaulters and accurately separates them in the feature space using hyperplanes. With the use of kernel functions, SVMs can separate different classes with curve shapes effectively known as non-linear boundaries. Althought SVM does well with overfitting in moderate sized datasets, they are noted to be sensitive with parameter selection and take much longer to train compared to other models (Jiménez 2023).

Neural Network (NN) is a multi-layered model inspired by the human brain's neural architecture. Neural Networks serves as a multi-layered model built on its neural architecture. Input, hidden, and output layers are where interconnected nodes called neurons are organized. Given enough data, they can learn intricate, complex, and multifaceted non-linear mappings from various features to the default probability. Neural networks have the ability to detect elusive patterns that other methods may overlook. This is advantageous for credit risk factors that are interrelated and intricate. On the other hand, they tend to be less interpretable, require careful tuning, and are quite complex and computationally intensive. These neural networks serve as the backbone of deep learning models that have achieved unparalleled success in tasks like fraud detection and credit scoring (Jiménez, 2023; Chen, 2022).

Dataset and Preprocessing

Our recorded history stems from a loan dataset available at a financial institution on borrowers and recent loan activity. Each observation is defined by a loan with multiple attributes related to the borrower's profile and the loan features. The features include the loan amount, interest rate, income of the borrower, employment duration, years in credit history, home ownership status, loan's intent, credit grade of borrower, and prior defaults if applicable. The target variable indicates, in a binary fashion, wither the loan defaulted (an outcome considered most risky) or was fully paid (an outcome that is the least risky). (Lao Tse, 2021)

The dataset was modified during the modeling phase in order to restore its quality and increase its usefulness for the algorithms. Missing entries with employment length were imputed and replaced with the mean value of that specific field. Home ownership, loan intent, loan grade, and the prior default flag as a markup are some values that were categorized to enable the system to easily segment and analyze the data and empower the modeling tools. In addition, features with large magnitudes like income or loan amount were scaled to a specific standard in order to avoid greatly affecting the model within numerical features. After preprocessing, the data was split into a training set and a test set using an 80/20 ratio. The models were trained on the training set and then evaluated on the independent test set to assess how well they generalize to unseen data. It is worth noting that the dataset was moderately imbalanced (relatively fewer defaulted loans than fully-paid loans), so careful attention to evaluation metrics beyond accuracy was necessary (Lao Tse, 2021).

Evaluation Metrics

In order to evaluate model performance, several standard classification metrics were used. Accuracy measures the percentage of all loans which were correctly classified as defaults or non-defaults, which serves as a broad indicator of model performance. Precision measures the percentage of loans predicted as "default" that were indeed defaults; high precision means that a significant number of borrowers regarded as high-risk actually are, which lowers false positive

rates. Recall (or sensitivity) measures the percentage of actual defaulted loans that the model captures with high recall make sure that the truly risky loans are identified, thus lowering false negative rate. The F1 score, which is the ratio of precision and recall, is a clear indication of a balance between two metrics. It is useful in the case of having overwhelming datasets or when both errors - false positive and false negative - would need to be minimal. For instance, relying only on accuracy for an imbalanced dataset like credit risk which has lower default scenarios can be harmful; a model that predicts defaults for all loans would score highly on accuracy but fail to identify high-risk loans. Hence, the union of precision, recall, and F1 score provides a deeper understanding of model performance with special emphasis on defaults and false positives. In addition, each model was analyzed with the confusion matrix, which cross-tabulates the predicted values against actual values and helps in the comprehension of the distribution of true positives, true negatives, false positives, and false negatives (Powers, 2011).

Visualization Techniques

We adopted a few visualization methods to check the models behavior alongside the data under consideration. Moreover, we implemented scatter matrices.

The scatter plot display in pairs was developed alongside other visualization methods for estimating metrics associated with a objective value. Visual pair groupings were plotted according to loan status either default or paid so that groupings and correlation could be seen in the data. This not only allowed for the discovery of multicollinearity or outlier phenomena, but also spotting trends such as the relationship of the amount borrowed in relation to interest rates and the risk of defaulting on repayment.

Another visualization implemented was an error distribution plot that allowed for displaying prediction errors distribution. In this case, the errors for a loan came from allocating a binary classification (1 mark was allocated if the prediction was completely wrong while scoring 0 was allocated if the prediction proved to be correct). The distribution of various errors (misclassification) by model on the test set offers a broad window into the magnitude of challenges faced by the individual model. I where concentration of errors is 0 portrays pairs of predictions at a label with few misclassifications, while concentration of errors around 1 depicts an inverse interpretation. The analysis was undertaken by aggregating all models error distributions into a single chart for easier comparison of the different models.

Predicted vs. Actual Scatter has involves scatter plots in which every model's predicted labels were compared to the actual labels for the test set. Ideally, a model that predicts accurately would only have data points on the diagonal line (representing the criteria that the predicted status must equal the actual status). Off-diagonal data points are errors that were misclassified. Even though this is an oversimplified diagram (with predicted and actual values being 0 or 1 for binary outcome), it allows observers a quick assessment of the model reasoning concerning respect to the accuracy and the question of over-predicting or under-predicting defaults.

All model training and analysis were undertaken in Python, with scikit-learn for modeling and pandas for data handling. Matplotlib and Seaborn were used for the visualizations. We do not provide code snippets in this paper because of space constraints, but the full implementation and analysis code has been posted on an online repository. Interested readers can refer to the project's GitHub repository for the complete code and reproducibility of the results. The entire code for the project or achieve the results can be available in the project's GitHub repository. (https://github.com/SadeghSalem/Business-Risk-Prediction)

Results and discussions

In the Results and Discussion section and in this part of the study, the aim is to analyze the likelihood of a loan default, which translates to whether a borrower goes default on the loan or repaid it. This dataset has a target that is binary where default is treated as the positive class (high-risk) and fully paid is the negative class (low-risk). The institutions are judged on how accurate their predictions are because this helps them manage credit risk more productively and with lesser loss.

Overall Model Performance

Table 1 encapsulates the performance of the five tested models within the defined four techniques of evaluation. It can be observed that the efficiency of the machine learning models is considerably good, but differs in the mix of precision and recall for each model. In this credit risk predication task, the best model was the Random Forest model, which had a better accuracy and F1 score than all other models. He was the most effective model. The Decision Tree was second, as this model also did perform almost as well as the Random Forest in most of the measures. Whereas recall and F1 measures were very low for the Logistic Regression, SVM, and Neural Network models, there were cases where precision was very high. These results show the effectiveness of ensemble learning over single estimators, and at the same time point out that risk prediction models should be measured using many different metrics in order to increase their accuracy.

Table 1. Comparison table of all the metric variables

Model	Accuracy*(%)	Precision*(%)	Recall	F1-Score
Logistic Regression	79.8	71.2	0.1522491349481	0.2508551881414
Decision Tree	88.2	72.3	0.7626297577855	0.7420875420875
Random Forest	92.8	95.6	0.7072664359862	0.8130469371519
SVM	80.3	81.5	0.1467128027682	0.2486803519062
Neural Network	80.6	88.2	0.1446366782007	0.2485136741974

Source: Authors' own research.

- Logistic Regression: Accuracy: 79.8%
- Decision Tree: Accuracy: 88.2%
- Random Forest: Accuracy: 92.8%
- SVM: Accuracy: 80.3%
- Neural Network: Accuracy = 80.6%

As can be seen, the Random Forest model outperforms the others in terms of accuracy, making it the best choice for predicting loan defaults in this dataset.

Figures 1 to 5 present through the graphical representation of confusion matrices and error distribution plots for each classifier. As anticipated, the confusion matrix of Random Forest shows the greatest correct predictions (which have the fewest misclassifications), corresponding with the best accuracy. Moreover, in addition to not flagging as many mistakes, Random Forest was found to have an error histogram that is highly concentrated toward a null error range. In contrast, models such as Logistic Regression, SVM, and Neural Network had more dispersed error distributions, which signify more errors, committed.

Logistic Regression

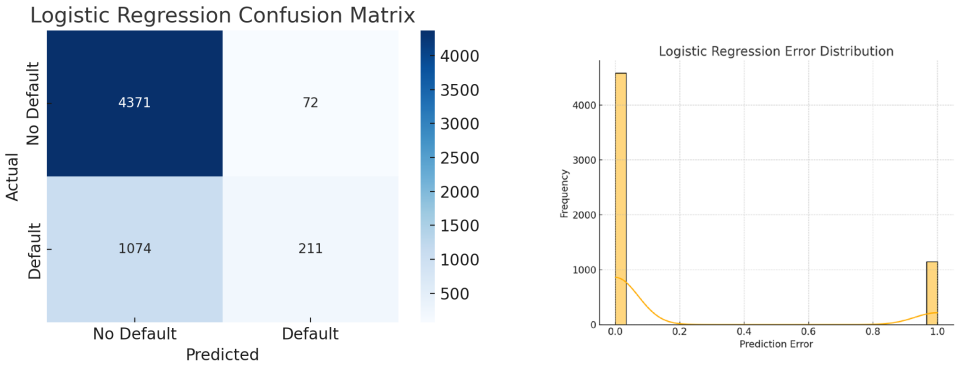

Figure 1. Logistic regression confusion matrix and error_distribution plot

Source: Authors' own research.

Decision Tree

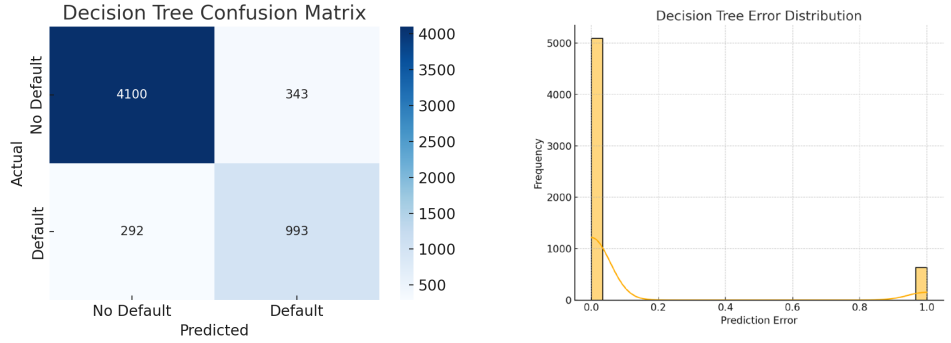

Figure 2. Decision tree confusion matrix and error_distribution plot

Source: Authors' own research.

Random Forest

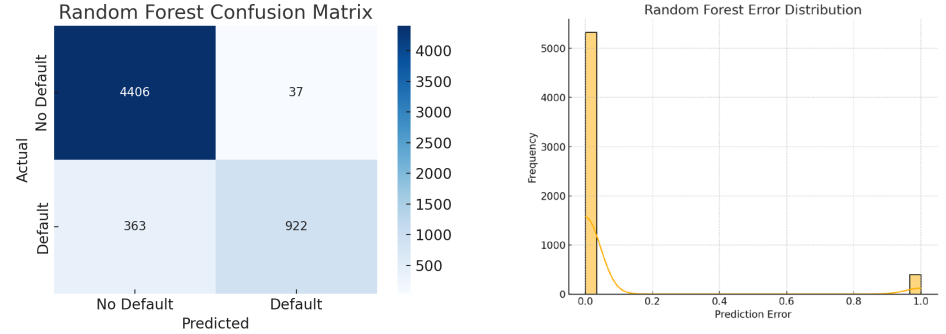

Figure 3. Random forest confusion matrix and error_distribution plot

Source: Authors' own research.

SVM (Support Vector Machine)

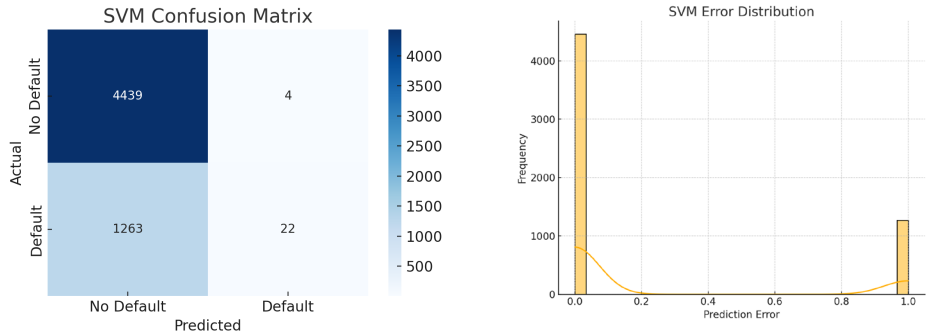

Figure 4. Support vector machine confusion matrix and error_distribution plot

Source: Authors' own research.

Neural Network

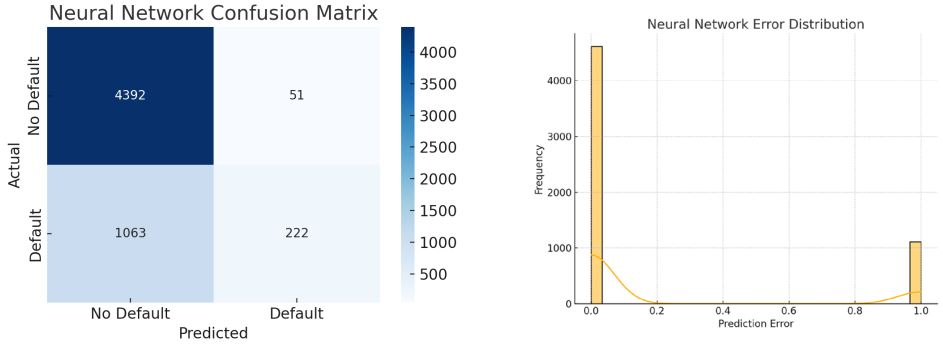

Figure 5. Neural network confusion matrix and error_distribution plot

Source: Authors' own research.

Providing an F1-score around 81%, the Random Forest achieved the highest accuracy (about 92.8%). Moreover, it is worth noting that the F1 score represents simultaneously balancing the index. The Decision tree also performed strongly with about 88.2% accuracy and an F1 score of 74.2%. In contrast with the other models, the Logistic Regression, SVM, and Neural Network models all had much lower F1 scores of 25%. This indicates that those models were not effective on the dataset as learners were not able to consolidate precision and recall working simultaneous.

The combined error comparison in Figure 6 also presents the case vividly: The Random Forest and Decision Tree errors gravitate towards zeros while these errors for the less accurate models get dispersed. These results clearly support the measures contained in Table 1. Therefore, there is strong evidence that these methods outperform all others in terms of predicting credit risk from the available data as a credit risk dataset.

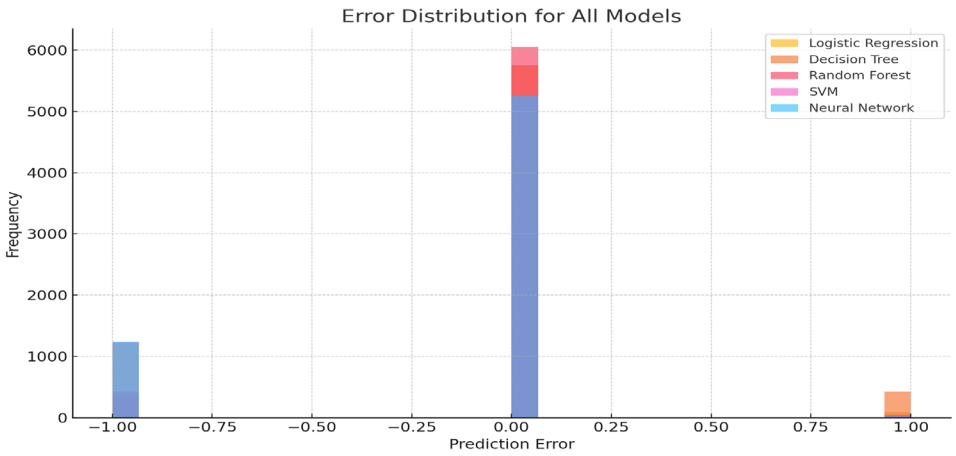

Figure 6. Combined error distribution chart

Source: Authors' own research.

Accuracy Comparison

Using every model, the prediction accuracy achieved did not occur uniformly. Forest Model achieved the highest accuracy with roughly 93% of loans correctly classified. This accuracy is better than the other models and can be explained by the Forest Model being an ensemble model which combines many decision trees and is able to capture the plethora of signals present in the data. The accuracy of Decision Tree model was also good, performing roughly 88% correct classification due to being able to model non-linear decision boundaries which simple linear models struggle with. However, Neural Network, SVM, and even Logistic Regression did comparatively poorly with roughly 80% correct classification. These results, especially the performance of the simpler linear model (Logistic Regression) and the more complex SVM and NN models, demonstrate that these models were able to capture the majority, but not all correct classified cases. As indicated from the side-by-side comparison in figure 7, it is clear that tree-based models outperform models with simpler and deeper structures. The variation is noticeable. The gap between Random Forest and the rest suggests that ensemble learning provided a tangible advantage in capturing the underlying patterns of default risk in the data.

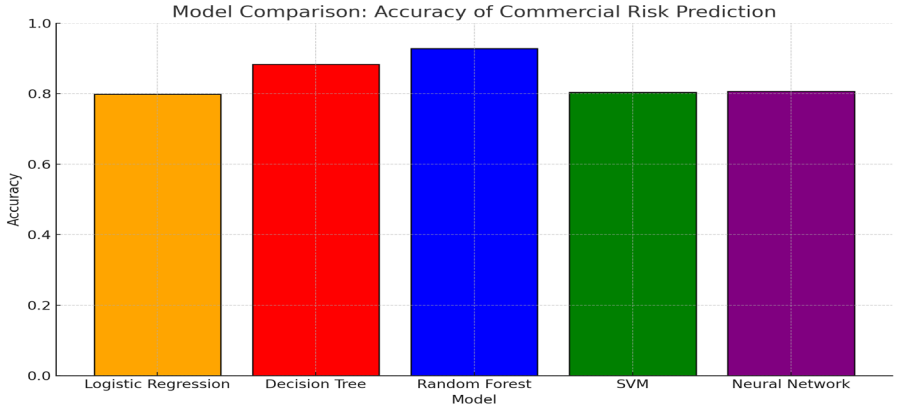

Figure 7. Accuracy of commercial risk prediction chart

Source: Authors' own research.

Precision and Recall Trade-offs

When looking closer at precision and recall, it is clear how each model aligns itself against the balance of false alarms against missed detection. The Random Forest model had an outstanding precision of approximately 95.6% Table 1as shown in figure 7, which means that all loans classified as default (high risk) tended to be actual cases of defaults. This suggests that Random Forest has a very low false positive rate, which is great when one wants to avoid labeling healthy loans as unnecessarily risky. The Neural Network also had a relatively higher precision as well (88%) and SVM's precision was moderately high (81.5%) which shows that these models were highly credible in predicting a default (when they predicted a default, they usually were correct). Logistic Regression and Decision Tree had lower precision (around 71–72%), which shows that they are suffering from high false positive rates – they misclassified non defaulting loans as defaults more often than not.

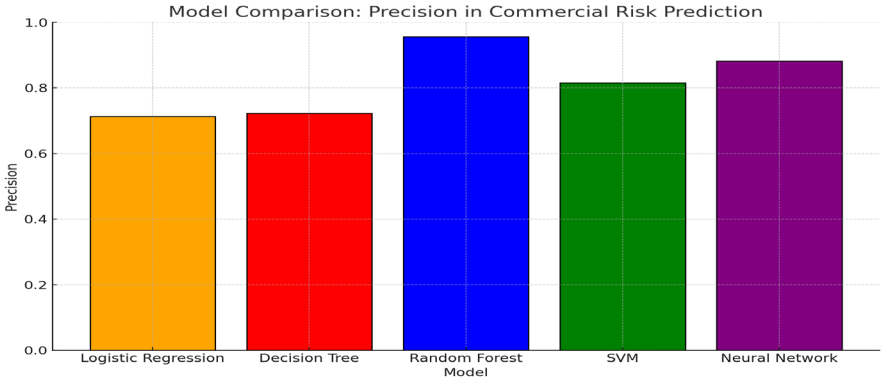

Figure 8. Precision in commercial risk prediction chart

Source: Authors' own research.

Nonetheless, once we analyze the recall metric, the situation is flipped for certain models. According to the Decision Tree, 76% of actual defaults were captured. This was the highest recall claim among the models while also capturing as much risky loans as possible. Unlike Random Forest, which recorded a recall of approximately ~71%, Decision Tree beat in recalling defaults. Random did not capture many defaults as with such high precision. On the other hand, Logistic Regression, SVM, and Neural Network all had low values of recall (somewhere between 14 - 15%), signifying that these models had missed most of the defaults. In reality, such low recall would mean that, out of all the truly risky loans, too many are being ignored by these models which should not be the case for an effective risk management system.

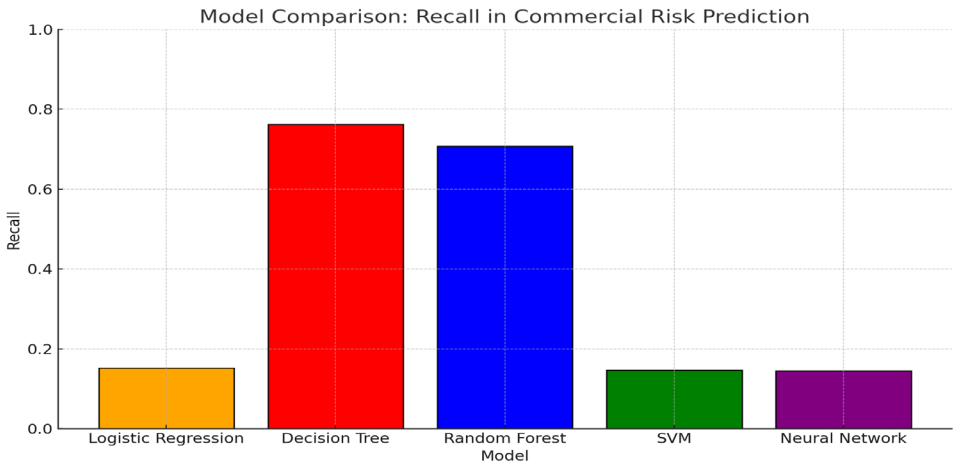

Figure 9. Recall in commercial risk prediction chart

Source: Authors' own research.

These differences in precision and recall scores reveal an important trade-off: the Neural Network and SVM were tuned (implicitly or explicitly) to be more take fewer risks in the decision windows made. While they were exceedingly cautious and managed to have low rates of false positives, that very caution led them to missing out on a significant amount of real defaults. On the other hand, the Decision Tree achieved high recall at the expense of accuracy or precision of the net because such nets used to catch defaults had high levels of false positives. Random Forest improved on this ability by achieving high precision while maintaining fairly good rates of recall. These figures demonstrate this trade off and show recall and precision for all models on the two figures. The ideal credit risk model seeks maximum precision to mitigate false positive cases and maximum recall to cover all problem loans. Out of the models analyzed, Random Forest achieved the most optimal results.

F1-Score as a Balanced Metric

The F1-score provides a single metric that balances precision and recall, and it clearly differentiates the models in our study. As shown in Table 1 and Figure 10, Random Forest achieved the highest F1-score (~81.3%), indicating that it offers the best compromise between avoiding false positives and false negatives. The Decision Tree's F1-score (~74.2%) was the next best, reflecting its strong recall and reasonable precision. In contrast, the F1-scores for Logistic Regression, SVM, and Neural Network were all approximately 25%, which is very low. This was not so strong here, but definitely slashing what could be achieved. Neural Networks achieved good precision, but poor recall lowered the F1 Score, which indicates weak precision; Neural Networks model did not work as expected as well. The F1 score paints a clear picture of how dire the situation is for with trust many people claim that there is very little risk does it take to get one in bottom credit score high neck without worrying that both errors cut giving predictors straight edge.

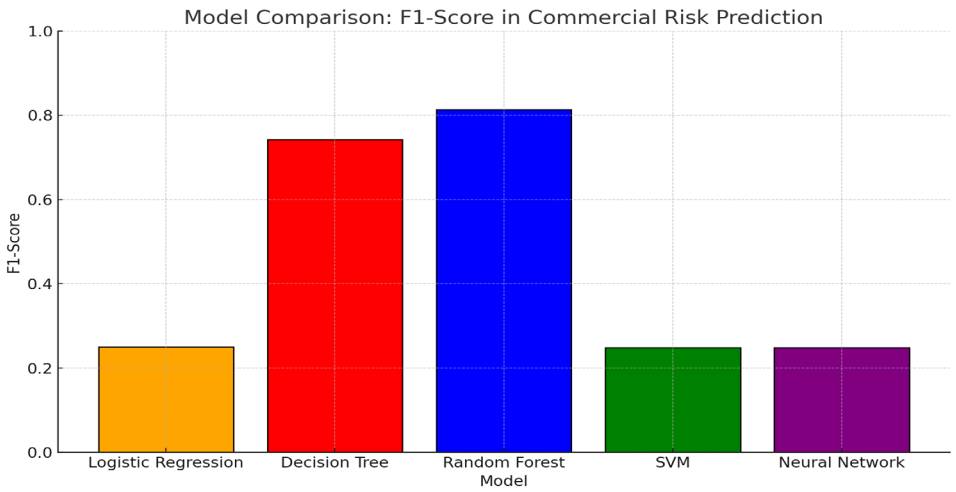

Figure 10. F1-Score in commercial risk prediction chart

Source: Authors' own research.

From a bank's point of view, a low F1 score model would be problematic due to its lack of comprehensive classification performance for the minority class (defaults). This is why models such as Random Forest (and, although to a lesser degree, Decision Tree) are more useful, as they seem to strike a balance between precision and recall. This significantly helps in managing credit risk.

Implications for Credit Risk Management

The analysis provides numerous theft effective resources for maintaining credit risk. First, the efficient use of the Random Forest method shows that these learning approaches will be very useful as modernization moves forth in the risk assessment tools for banks. In addition, a random forest model will greatly reduce the number of false positives - where credit worthy clients are marked as high risk - enabling the bank to feel second in increasing the credit base for borrowers, (adding business agility). Furthermore, a strong recall means the model will capture the maximum number of risky loans leading to lower losses due to defaults and increasing the robustness of the portfolio.

Second, the results suggest a strategy where, if capturing every single defaults is the strategy, obtrusively trying to avoid the defaulting clients will steer clients towards the Decision tree model with high risk tolerance. Accepting more good clients and declining some highly risky loans.

These models show extremely low precision in regard to working with defaults, which is indicative of extreme financial loss SVM and Neural Network need to be treated with caution. These complex workings necessitate further data and could result in financial risk if not addressed. Modification of credit risk calculation models is key regarding the bounds of predictive risk SVM neural models is calculated on.

A single measurement of percentage achieved could have led us into perceiving that all models were efficient which sets a range of 80-93% accuracy. However, utilizing WPS earned by F1 score helped in computing the accuracy and recall levels as well as measuring precision. In banking this means needless rejection of credit applications and accepting bad loans should never happen. Effectiveness in the context of quality means catching bad loans and good clients while

not allowing valuable markers of clients be misjudged. These results are aligned with other findings stressing the fact that AI can greatly improve traditional risk management practices.

Banks can use machine learning models such as Random Forest to enhance the predictive accuracy of their credit evaluations. This could result in more stable financial performance over time and a lower rate of loan defaults, which is good for individual institutions and the economy as a whole. The economy benefits from higher social stability as defaults are reduced and lending is done in a more responsible manner to mitigate the risk of banking crises and to make sure that credit is granted to those who deserve it. Don't forget though, that model choice may be influenced by the organizational context — some institutions may prefer simpler and easier to understand models like logistic regression or decision trees, even with less predictive accuracy, if meeting regulatory requirements and being transparent is more important. Model interpretability vs performance is a more subtle issue than pure metric scores and is a very important aspect of machine learning.

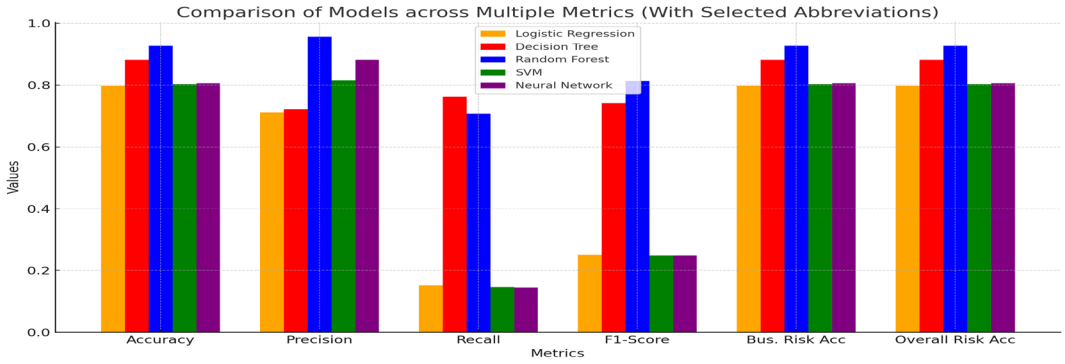

Figure 12. Multi-metric comparison chart

Source: Authors' own research.

Conclusion

This research illustrates how machine learning algorithms can enhance credit risk evaluation processes within banking institutions. After assessing five different models with a real world loan dataset, it was clear that Ensemble Methods are far more effective than traditional methods for the detection of riskier loans. For example, the Random Forest model dominated the competition by having the best score for both accuracy and F1, meaning that it had the best combination of detecting defaults while minimizing false positives. The Decision Tree model also did well (in particular, in recall) but the Logistic Regression, SVM, and a simple Neural Network did poorly because these models do not take into account all of the relevant features of default risk.

The results are relevant for many banks and other financial institutions. Using Machine Learning methodologies in credit risk assessment facilitates decision making and allows lenders to identify high risk and low risk borrowers more precisely rather than making guesses. By improving risk prediction, banks can remain proactive while confidently issuing credit to customers without compromising on business quality, helping to stimulate economic growth. The social benefits include improved stability as there will be fewer defaults and financial suffering that can disrupt communities and markets. Despite the optimistic results, our research has a number of shortcomings.

Initially, the dataset, while representative, pertains to a specific institution and time frame; the model's performance may vary for other demographics or sets of principles. Future models

should be validated on different datasets like those from different banks or regions so that the models built can be generalized. On the other hand, the more elaborate SVM and Neural Network models might perform better with some further tuning, or the addition of some features, such as incorporating macroeconomic factors or borrower behavioral aspects. There is no doubt that we did not devote our time to hyperparameter optimization, thus performance can be improved upon. Further, explaining the model's outcomes for a test case is indeed an interpretability aspect that can enhance its predictive performance but requires elaborate scrutiny, which the authors did not attempt. In practice, the banking industry, AI models are highly sensitive to explainability, integration into existing systems, and adherence to regulations such as fairness and transparency to be used. There are multiple ways that future research can expand upon this work. One such option is working with existing models that are more sophisticated like gradient boosting machines such as XGBoost or LightGBM, or expanding to deep learning that may further enhance the predictive performance. It is possible to explore new enhancements by synthesizing AI learning systems with input from domain experts to efficiently utilize both data patterns and the domain's underlying structure. Researchers need to develop better techniques to handle class imbalance; such as cost sensitive learning methodology and anomaly detection technique which enhance the recall for rare events while not excessively sacrificing precision. Furthermore, ensemble intelligence or neural network model predictions could be explained using XAI techniques, which would make the results more acceptable to the stakeholders. Finally, longitudinal research can analyze the effect on the finances of the institutions on the use of AI driven credit risk models, such as reduction of defaults, or gaining in the performance of the loan portfolio over the period of time. Therefore, machine learning is a potent tool to predict and manage business risk. Our comparative analysis certainly proves that, assuming there is a proper algorithm and evaluation system set in place, AI models are capable of adjusting the measurements of credit risk with relative ease, accuracy, and reliability. Putting aside the hurdles faced during the formulation of these models, with proper planning, banks can utilize artificial intelligence to streamline the operations of their institutions, which would increase the competitiveness of the financial industry.

References

Mashrur, A., Luo, W., Zaidi, N., & Robles-Kelly, A. (2020). Machine Learning for Financial Risk Management: A Survey. *IEEE Access, 8*, 203203-203223.

Munkhdalai, L., Munkhdalai, T., Namsrai, O., Lee, J.Y., & Ryu, K. (2019). An Empirical Comparison of Machine-Learning Methods on Bank Client Credit Assessments. *Sustainability*.

Thiel, D.V. (2019). Artificial Intelligent Credit Risk Prediction: An Empirical Study of Analytical Artificial Intelligence Tools for Credit Risk Prediction in a Digital Era. *Journal of Accounting and Finance*.

Moscatelli, M., Parlapiano, F., Narizzano, S., & Viggiano, G. (2020). Corporate default forecasting with machine learning. *Expert Syst. Appl., 161*, 113567.

Golbayani, P., Florescu, I., & Chatterjee, R. (2020). A comparative study of forecasting Corporate Credit Ratings using Neural Networks, Support Vector Machines, and Decision Trees. *ArXiv, abs/2007.06617*.

Gupta, D.K., & Goyal, S. (2018). Credit Risk Prediction Using Artificial Neural Network Algorithm. *International Journal of Modern Education and Computer Science, 10*, 9-16.

Shoumo, S.Z., Dhruba, M.I., Hossain, S., Ghani, N.H., Arif, H., & Islam, S. (2019). Application of Machine Learning in Credit Risk Assessment: A Prelude to Smart Banking. *TENCON 2019 - 2019 IEEE Region 10 Conference (TENCON)*, 2023-2028.

Farrokhi, A., Shirazi, F., Hajli, N., & Tajvidi, M. (2020). Using artificial intelligence to detect crisis related to events: Decision making in B2B by artificial intelligence. *Industrial Marketing Management, 91*, 257 - 273.

Kou, G., Chao, X., Peng, Y., Alsaadi, F.E., & Herrera-Viedma, E.E. (2019). MACHINE LEARNING METHODS FOR SYSTEMIC RISK ANALYSIS IN FINANCIAL SECTORS. *Technological and Economic Development of Economy*.

Aziz, S., & Dowling, M.M. (2018). Machine Learning and AI for Risk Management. *Disrupting Finance*.

Zhang, Z. (2024). An Application of Logistic Regression in Bank Lending Prediction: A Machine Learning Perspective. *Highlights in Business, Economics and Management*.

Li, C., & Zhang, J. (2024). Research on Credit Risk Prediction Models Based on Machine Learning. *2024 6th International Conference on Machine Learning, Big Data and Business Intelligence* (MLBDBI), 1-4.

Jiménez, O., Jesús, A., & Wong, L. (2023). Model for the Prediction of Dropout in Higher Education in Peru applying Machine Learning Algorithms: Random Forest, Decision Tree, Neural Network and Support Vector Machine. *2023 33rd Conference of Open Innovations Association* (FRUCT), 116-124.

Zhang, Q. (2023). Loan Risk Prediction Model based on Random Forest. *Advances in Economics Management and Political Sciences*, 5(1), 216–222. https://doi.org/10.54254/2754-1169/5/20220082

Chaudhary, S., Baliyan, V., & Katheria, Y. (2020). Loan Prediction System Using Decision Tree and Random Forest Algorithms.

Lao Tse. (2021). Credit risk dataset [Data set]. Kaggle. https://www.kaggle.com/datasets/laotse/credit-risk-dataset

Powers, D. M. W. (2011). Evaluation: From precision, recall and F-measure to ROC, informedness, markedness & correlation. *Journal of Machine Learning Technologies*, 2(1), 37–63. https://bioinfopublication.org/files/articles/2_1_1_JMLT.pdf