# CAPSTONE PROJECT: APPLICANT DETAILS FOR LOAN APPROVAL

**Problem Statement**

Financial institutions, such as banks, face the challenge of efficiently and accurately assessing the creditworthiness of loan applicants. Without robust loan approval systems in place, banks encounter several issues that hinder their lending operations and risk management practices. A cumbersome and lengthy loan approval process can negatively impact the customer experience, leading to frustration and potential loss of business for banks.

Organizations with slow approval systems risk losing customers to competitors offering faster services, leading to reduced market share and revenue.

**Executive Summary**

Goal

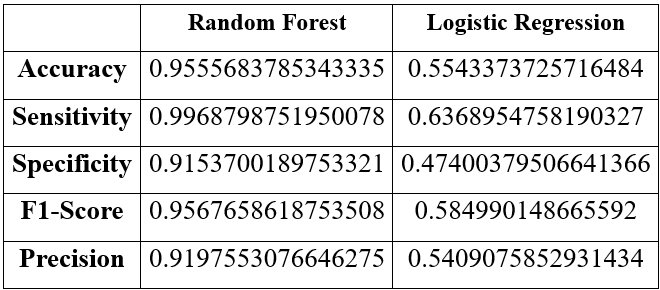
Implementing efficient loan approval systems allows banks to provide faster decisions to customers, enhancing satisfaction and loyalty and next, contribute in reducing bankruptcy rate in Malaysia.

Data Source

The data was obtained from Kaggle.com.

Metrics

* Accuracy: Proportion of correctly predicted loan approvals and denials among all loan applications.
* Precision: Proportion of correctly predicted approved loans among all predicted approved loans.
* Recall: Proportion of correctly predicted approved loans among all actual approved loans.
* F1 Score: Harmonic mean of precision and recall for predicting loan approvals
* Specificity: Proportion of correctly predicted denied loans among all actual denied loans.



Findings

* Random Forest model achieved high accuracy (95.5%) and performed well in terms of precision, recall, and specificity.
* Logistic Regression model performed poorly, indicating that it is not suitable for this classification task.
* Under sampling was used to balance the data, resulting in equal representation of both classes.

Risks/Limitations/Assumptions

* The performance of the models may vary depending on the quality and representativeness of the data.
* Under sampling may lead to loss of information and potential bias in the dataset.
* Assumptions underlying the logistic regression model, such as linearity and independence of predictors, may not hold true in the real-world scenario.

**Statistical Analysis**

Data Acquisition and Preprocessing

* Data was collected from the financial institution's database, including various applicant details.
* It consists of applicant details for loan approval, including features such as age, annual income, employment history and loan default risk.
* Under sampling technique was applied to balance the classes of loan default and non-default cases.
* Several unnecessary columns such as Applicant ID, Occupation, Residence City and Residence State removed from the dataset.
* Categorical variables were encoded using one-hot encoding.

Model Implementation

* Two models were implemented: Random Forest and Logistic Regression.
* Features were standardized in Logistic Regression to ensure uniformity in model training.
* Hyperparameters were tuned using techniques such as cross-validation and grid search to optimize model performance.

Model Evaluation

* Models were evaluated using cross-validation.
* Metrics such as accuracy, precision, recall, F1-score, and specificity were calculated to assess model performance.
* Confusion matrices were generated to visualize the distribution of predicted outcomes.

Inference

* The Random Forest model outperformed the Logistic Regression model in predicting loan default risks.
* The Random Forest model demonstrated higher accuracy, precision, recall, and specificity, indicating its superiority in identifying loan default cases.
* Feature importance analysis revealed significant predictors of loan default, including income, employment history, and age.

**Conclusion**

Overall, the Random Forest model showed promising results in predicting loan default risks based on applicant details. Further refinement of features and model parameters could potentially improve predictive performance. It's essential to monitor model performance over time and update the model periodically to adapt to changing trends and patterns in loan applicant data. Additionally, considering domain expertise and incorporating feedback from financial experts can enhance the robustness and applicability of the predictive model in real-world scenarios.

**References**

1. Pino, I. (2024, January 27). How long does it take to be approved for a personal loan? *Yahoo! Finance*. Retrieved 2024, from https://finance.yahoo.com/personal-finance/how-long-does-it-take-to-get-a-personal-loan-213010800.html.
2. Viswanatha, V., Ramachandra, C. A., Vishwas, N. K., & Adithya, G. (2023). Prediction of Loan Approval in Banks using Machine Learning Approach. *International Journal of Engineering and Management Research*, *13*(4). <https://doi.org/10.31033/ijemr.13.4.2>

**Link to Github**