

Predicting Loan Approvals and Analyzing Lender Trends

Prepared by – Farhan A. Abidi

St Lawrence College

School of Business

Objective

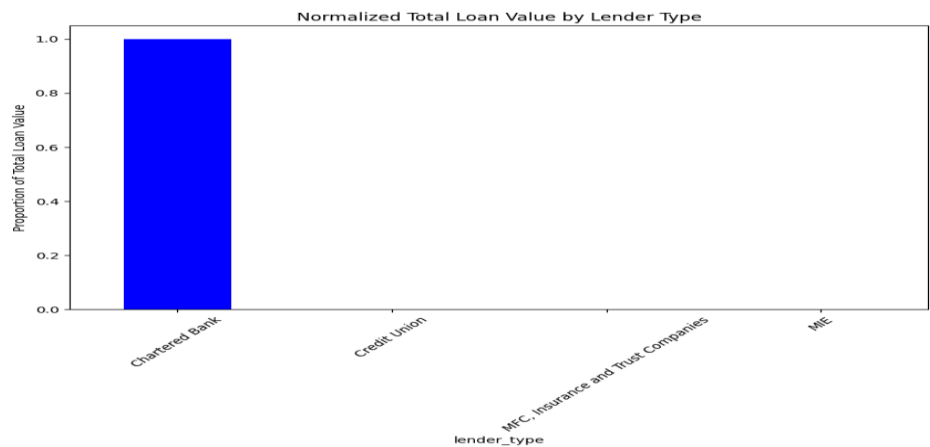
The primary goal of this project is to predict the number of loan approvals (approvals_#) based on key financial metrics, such as the number of applications, lender type, and mortgage value changes. Additionally, the project seeks to analyze lender trends, identify influential factors, and provide actionable insights for stakeholders in the mortgage lending industry.

Dataset Overview

The dataset includes mortgage and housing-related data from multiple worksheets, such as Applications, Approvals, Mortgage Decreases, Total Loan Value Originated, and Total Loan Value Outstanding. Key variables include:

- **reporting_date**: Specifies the reporting period.
- **lender_type**: Categorizes institutions such as Chartered Banks and Credit Unions.
- **applications_#**: Quantifies loan application volume.
- **approvals_#**: Captures the number of approved loans.

The data spans thousands of records, offering a comprehensive view of lending trends. Chartered Banks emerged as the dominant players, consistently contributing the largest share of applications, approvals, and total loan values.



Methodology and Insights

Data Cleaning and Preprocessing

The dataset underwent extensive cleaning to ensure accuracy and consistency:

1. **Standardization**: Column names were standardized for uniformity.
2. **Handling Missing Data**:
 - Numerical columns were filled using mean imputation.
 - Categorical columns were filled with the most frequent value.

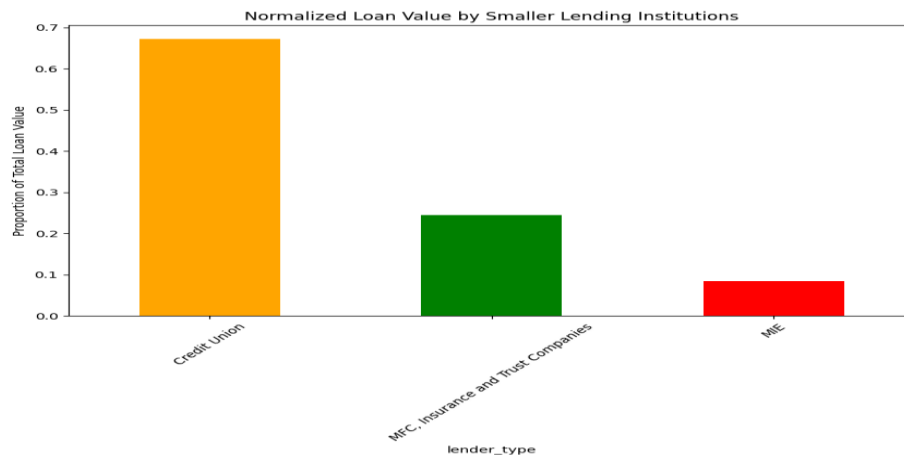
3. **Encoding:** Categorical variables like lender_type were encoded using label encoding and one-hot encoding.

Exploratory Data Analysis (EDA)

EDA revealed significant trends:

- **Lender Dominance:** Chartered Banks dominated loan applications and approvals. A bar chart or pie chart depicting this distribution should be included to emphasize these trends.
- **Smaller Institutions:** Credit Unions and MFCs, while contributing less overall, showed distinct lending patterns that warrant further exploration.
- **Correlation Analysis:** applications_(#) showed a strong correlation with approvals_(#), making it a key predictive feature. A heatmap visualizing the correlation matrix among key variables should be added here.

A secondary analysis was carried out to highlight that smaller institutions contribute significantly less to total loan values but still play an essential role in diversifying lending options.



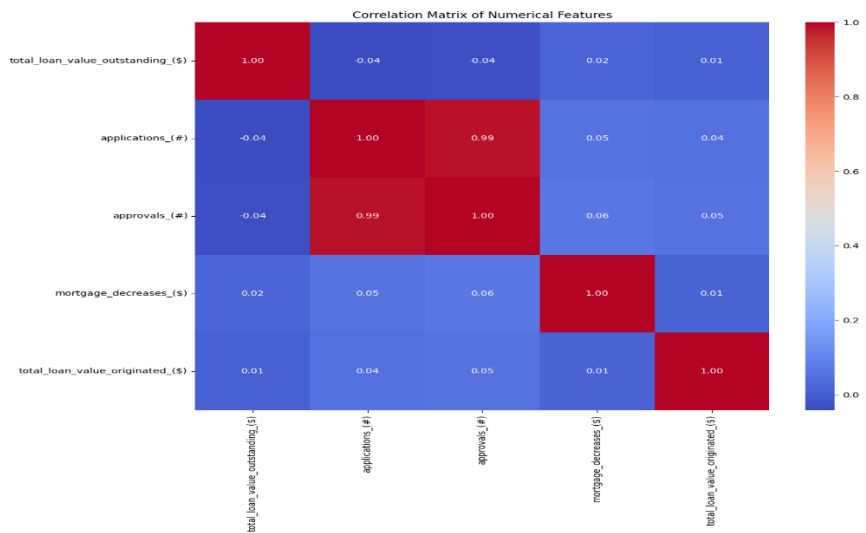
Recommendations for Smaller Institutions Based on Lending Patterns

Credit Unions: Leverage their strong market share among smaller institutions by focusing on niche customer segments and offering competitive loan packages tailored to regional needs.

MFCs (Mortgage Finance Companies): Strengthen their position by exploring partnerships with real estate developers and targeting underserved markets with flexible loan options.

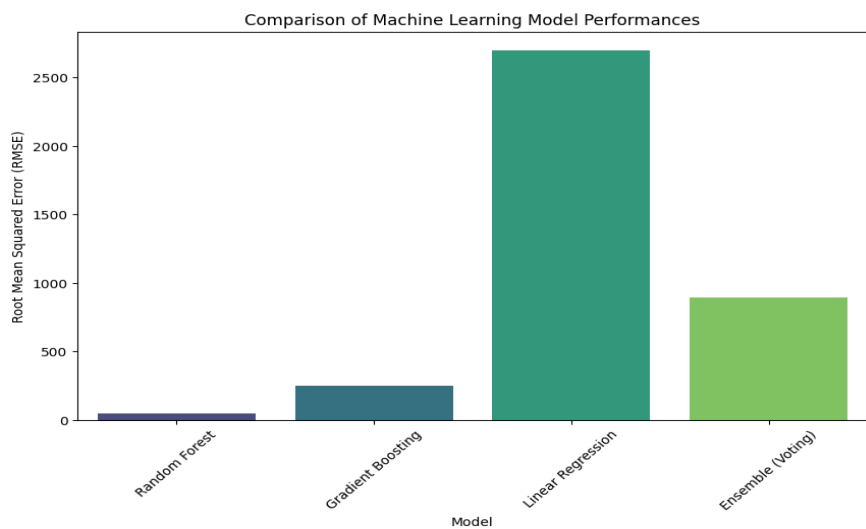
Strategic Growth: Both institutions should enhance digital platforms and customer service to attract younger, tech-savvy borrowers and compete effectively with larger players like Chartered Banks.

A Correlation analysis confirmed that applications_(#) is highly correlated with approvals_(#), indicating its importance as a predictive feature.



Machine Learning Models

To predict loan approvals, multiple machine learning models were deployed. The primary objective was to leverage historical data to identify patterns and provide actionable insights for financial institutions. Three models were evaluated:



- **Random Forest Regressor:**
 - RMSE: **49.23**
 - R²: **0.999999**
 - MAE: **22.41**

- MAPE: **0.0096%**
- The standout performer, excelling at capturing non-linear relationships and feature interactions.
- **Gradient Boosting Regressor:**
 - RMSE: **253.31**
 - R^2 : **0.999985**
 - MAE: **177.35**
 - MAPE: **0.0574%**
 - Competent but less precise than Random Forest.
- **Linear Regression:**
 - RMSE: **2697.07**
 - R^2 : **0.9983**
 - MAE: **1884.17**
 - MAPE: **0.767%**
 - Struggled with non-linear patterns and performed poorly overall.

The models provide a robust framework for predicting approval counts, allowing lenders to optimize resource allocation and operational strategies. For example, lenders can anticipate high-volume periods and adjust staffing levels or marketing efforts accordingly.

Discussion

The analysis underscored the dominance of Chartered Banks in the lending market, highlighting their pivotal role in shaping industry trends. While smaller institutions like Credit Unions and MFCs contribute less overall, their presence is essential for offering diverse lending options and addressing niche market needs. The separate analysis of these smaller institutions revealed unique lending patterns that could inform targeted strategies.

Feature importance analysis confirmed that `applications_#` was the most influential variable, reinforcing its predictive value. However, the Random Forest model's near-perfect R^2 score suggests potential overfitting. Future work should prioritize addressing this limitation through:

1. **Validation:** Testing the model on unseen datasets or employing cross-validation.
2. **Feature Expansion:** Incorporating external economic indicators, such as interest rates and regional housing trends, to enhance predictive power.
3. **Regularization:** Applying techniques like L1 or L2 regularization to reduce model complexity and improve generalizability.

Conclusion

This project effectively demonstrates the application of machine learning in predicting loan approvals and analyzing lender trends. The Random Forest model emerged as the most reliable, achieving exceptional performance metrics. However, overfitting concerns must be addressed to ensure robustness in real-world applications.

The analysis also provides actionable insights into the dominance of Chartered Banks and the niche roles of smaller institutions. By refining the model further and validating it with unseen data, this framework can be adapted to support strategic decision-making in the mortgage lending industry. Visuals such as a heatmap for correlations and bar charts for model performance and lender contributions will further enhance the report's impact.