

Melchizedek Ackah-Blay

Loan Approval Prediction Insights

2 September 2024

## **Loan Approval Prediction Model Documentation**

### Project Overview

The primary objective of this project is to develop a robust machine-learning model that can accurately predict the approval or rejection of loan applications. By leveraging various application metrics such as credit score, loan amount, loan term, and other relevant features, the goal is to explore the data, engineer meaningful features, build and evaluate predictive models, and provide actionable insights to improve loan approval rates.

### Data Exploration and Preprocessing

1. Imported the necessary libraries for data exploration and preprocessing.
2. Loaded the dataset containing transactional and user-related features.
3. Examined the data shape, data types, and unique values to gain a comprehensive understanding of the available information.
4. Performed exploratory data analysis to uncover insights, such as the distribution of education levels, employment types, and the relationship between credit score and annual income.
5. Encoded the necessary features and prepared the data for model training.

## Feature Selection

To enhance model performance, reduce overfitting, and improve generalization, a feature selection technique (SelectKBest) was employed to extract the most relevant features for predicting loan approval. The top 8 features selected were:

- Number of dependents
- Loan amount
- Loan term
- Credit score
- Luxury assets value
- Annual income

## Model Training and Evaluation

1. Split the data into training and testing sets.
2. Trained and evaluated various classification models, including Logistic Regression, Decision Tree, Support Vector Machine, K-Nearest Neighbor, Random Forest, AdaBoost, XGBoost, and Gradient Boosting.
3. Performed rigorous cross-validation to assess the models' performance.
4. Compared the models based on key evaluation metrics, such as AUC (Area Under the Curve) and classification report (accuracy, precision, recall, F1-score).

## Hyperparameter Tuning

After selecting the top 5 performing models, their respective hyperparameters were tuned to further improve their predictive capabilities.

## Model Selection and Ranking

Based on the comprehensive evaluation, the XGBoost model emerged as the top-performing model, with the highest accuracy (0.99) and a well-balanced precision and recall across both classes.

The final ranking of the models is as follows:

1. XGBoost
2. AdaBoost
3. Random Forest
4. Decision Tree
5. Gradient Boosting

## Feature Importance Analysis

The feature importance analysis revealed that credit score is the single most influential factor in determining loan approval, aligning with real-world lending practices. The analysis also highlighted the relatively lower importance of factors like annual income compared to credit score.

## Lessons Learned

1. Rounding up numbers can significantly impact model performance, so care should be taken when preprocessing the data.
2. Training a diverse set of models and using multiple evaluation metrics is crucial for comprehensive model assessment.
3. Visualizing results and scores helps in better understanding model performance and guiding the model selection process.

## Recommendations

Based on the insights gained from the project, the following recommendations are provided for lending institutions:

1. Focus on credit score improvement programs for potential applicants.
2. Develop more nuanced credit scoring systems that incorporate a wider range of financial and personal data points.
3. Consider alternative data sources for applicants with limited credit history to enhance the inclusivity and fairness of the lending process.

## Conclusion

This project demonstrates the effective development of a loan approval prediction model using various machine learning techniques. The insights gained from the feature importance analysis and the recommendations provided can help lending institutions make more informed, ethical, and equitable financing decisions, ultimately fostering greater financial access and inclusion.