The Future of Lending Starts Here

Revolutionize your lending process with LoanWise – the AI-powered solution for smarter, faster, and more reliable credit approvals. Reduce costs and minimize risk.

Get Started

LoanWise Project

Predicting Credit Eligibility Using Machine Learning

Rafael García

February 2025

LOANWISE

What is LoanWise?

A Smarter Way to Approve Credit

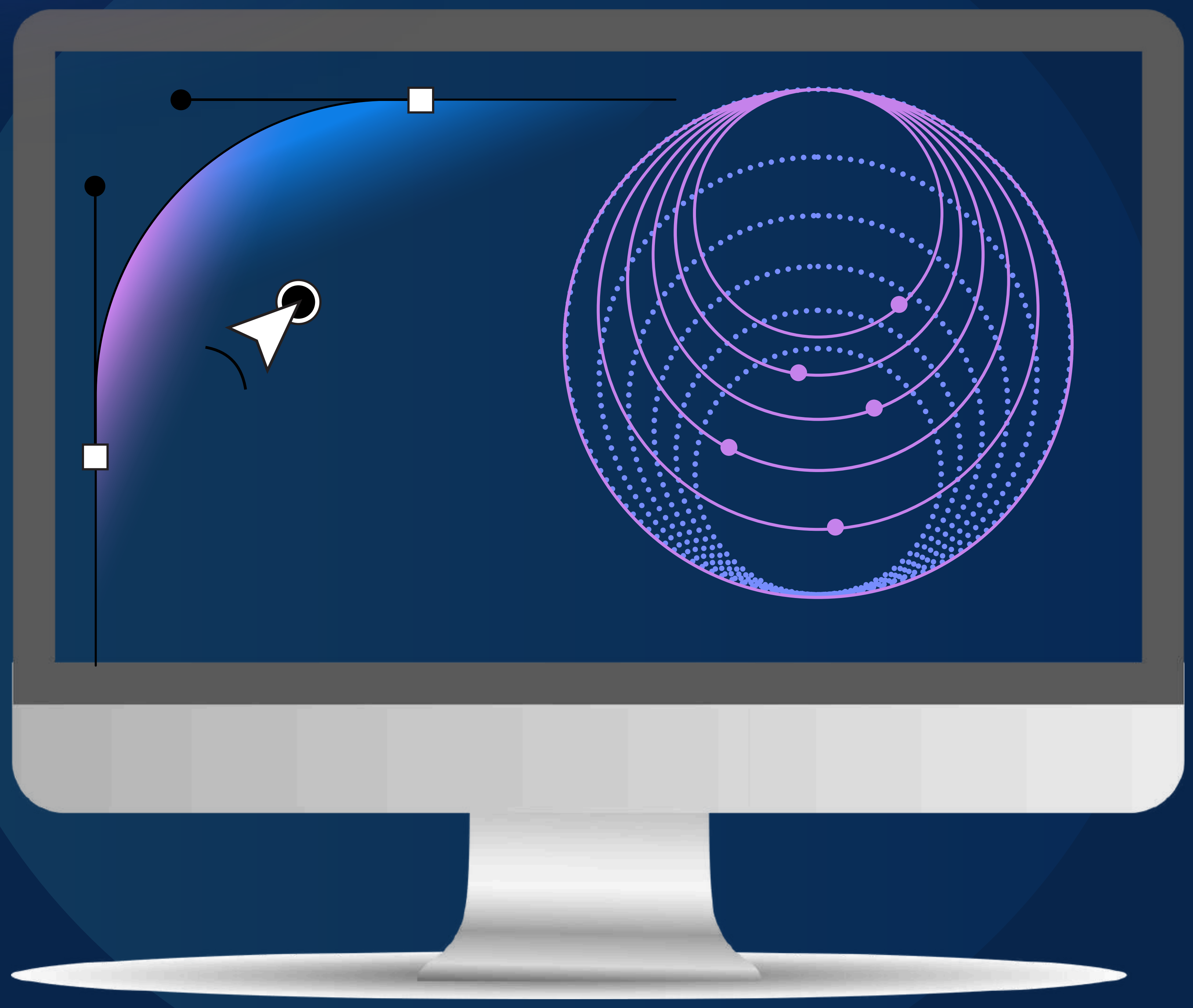
LoanWise is an AI-powered lending assistant that transforms the way financial institutions evaluate credit applications. By leveraging machine learning and predictive analytics, LoanWise helps lenders make faster, smarter, and fairer decisions—reducing costs and improving efficiency.

Why LoanWise?

Traditional credit approval processes are often slow, expensive, and rigid, leading to:

✅ High origination costs due to manual risk assessments.
✅ Long approval times that frustrate potential borrowers.
✅ Inflexible credit profiling that overlooks alternative data points.
✅ Risk of over-indebtedness due to outdated evaluation models.

LOANWISE



Who can benefit?

🏛️ Banks & Financial Institutions – Improve efficiency and lower credit risks.

📈 Fintech Startups – Scale lending operations with smart automation.

💳 Microfinance & Credit Unions – Expand access to credit responsibly.

The Credit Landscape in Colombia

KEY STATISTICS HIGHLIGHTING THE NEED FOR SMARTER LENDING

Rising Household Debt:

- **58.3%** of Colombian households took on debt or used savings to cover expenses in 2023.
- Households allocate an average of **25.3%** of their income to debt repayment.

Source: Infobae, El Colombiano

HOW LOANWISE ADDRESSES THESE ISSUES?

- ✅ **Reduces Origination Costs:** Automates data analysis, minimizing manual processes and lowering operational costs.
- ✅ **Accelerates Credit Approval:** Processes real-time data, allowing decisions in minutes instead of days.
- ✅ **Prevents Over-Indebtedness:** Ensuring financial sustainability and lowering default risks.

Introducing LoanWise

LoanWise is an AI-driven solution designed to predict customer eligibility for credit, aiming to enhance decision-making in the lending process.

Key Features

- Utilizes machine learning algorithms to analyze applicant data.
- Based on comprehensive Data Analysis
- Dual-Input Prediction Model, predicts on a combination of qualitative and quantitative data.
- Scalable and Customizable, enabling institutions to easily adapt it to new markets and expand with additional features as needed.

Data Collection – Building a Strong Foundation for LoanWise

LoanWise leverages high-quality, diverse financial data to enhance credit decision-making and predict borrower eligibility with accuracy.

Data Source

📌 Dataset Used: Credit Card Capability Data (Source: Kaggle)
📌 Purpose: Evaluate key factors influencing credit eligibility and loan approval.
📌 Key Data Points: Combination of demographic, financial, and behavioral factors.

EDA – Building a Strong Foundation for LoanWise

Key Features in the Dataset

- **Demographic Information:**
  - **Gender** – Helps analyze credit behavior differences.
  - **Number of Children & Family Size** – Impacts financial obligations and repayment ability.
- **Financial & Employment Data**
  - **Employment Status (Unemployed/Employed)** – Stability and income source verification.
  - **Owns a Car / Owns Property** – Potential indicators of financial standing.
- **Contact & Accessibility Indicators**
  - **Work Phone / Personal Phone / Email Availability** – Validates applicant reliability and accessibility.

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 9609 entries, 0 to 9608
Data columns (total 19 columns):
 #   Column           Non-Null Count  Dtype
---  ------           --------------  -----
 0   Gender           9609 non-null   int64
 1   Own_car          9609 non-null   int64
 2   Own_property     9609 non-null   int64
 3   Work_phone       9609 non-null   int64
 4   Phone            9609 non-null   int64
 5   Email            9609 non-null   int64
 6   Unemployed       9609 non-null   int64
 7   Num_children     9609 non-null   int64
 8   Num_family       9609 non-null   int64
 9   Account_length   9609 non-null   int64
 10  Total_income     9609 non-null   int64
 11  Age              9599 non-null   float64
 12  Years_employed   9609 non-null   float64
 13  Income_type      9609 non-null   object
 14  Education_type   9609 non-null   object
 15  Family_status    9609 non-null   object
 16  Housing_type     9609 non-null   object
 17  Occupation_type  9609 non-null   object
 18  Target           9609 non-null   int64
dtypes: float64(2), int64(12), object(5)
```

LOANWISE

EDA – Understanding the Dataset

1. Dataset Overview

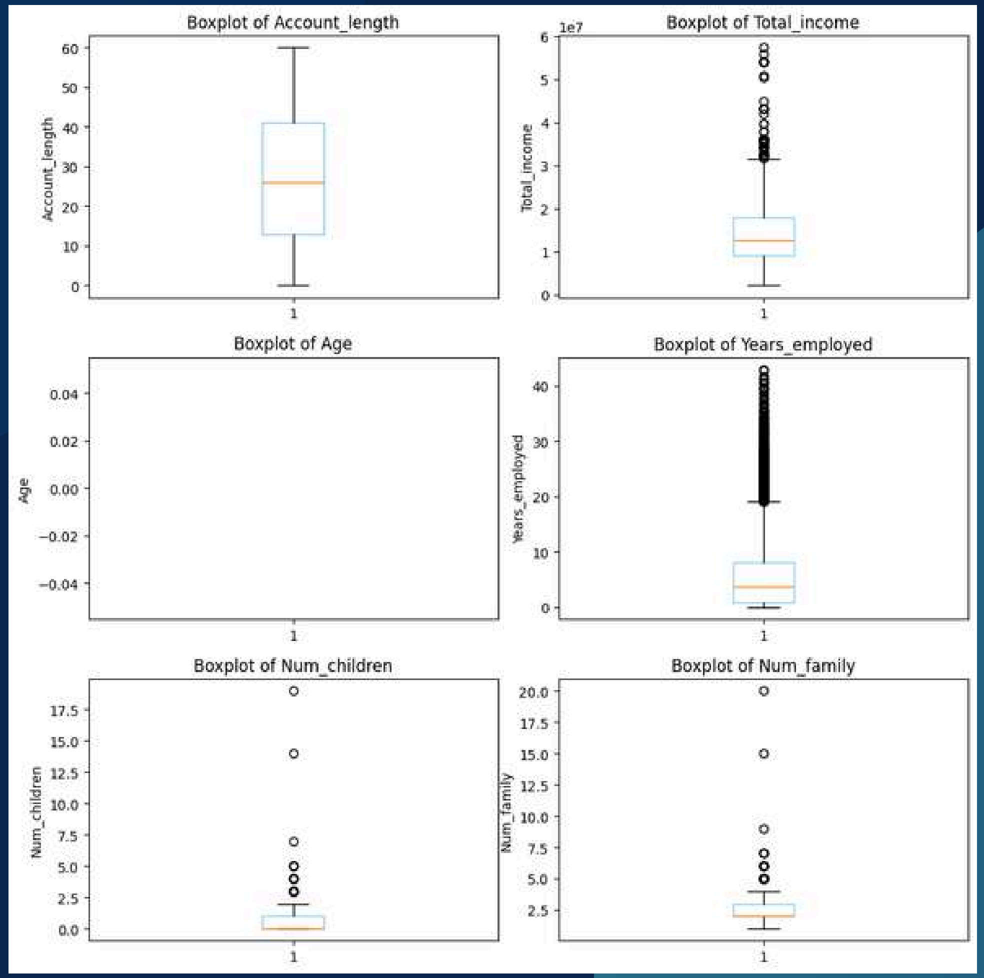
- The dataset consists of 9,609 instances with 18 features covering demographic, financial, and employment information.
- The target variable indicates whether an applicant is eligible for credit (1 = Eligible, 0 = Not Eligible).

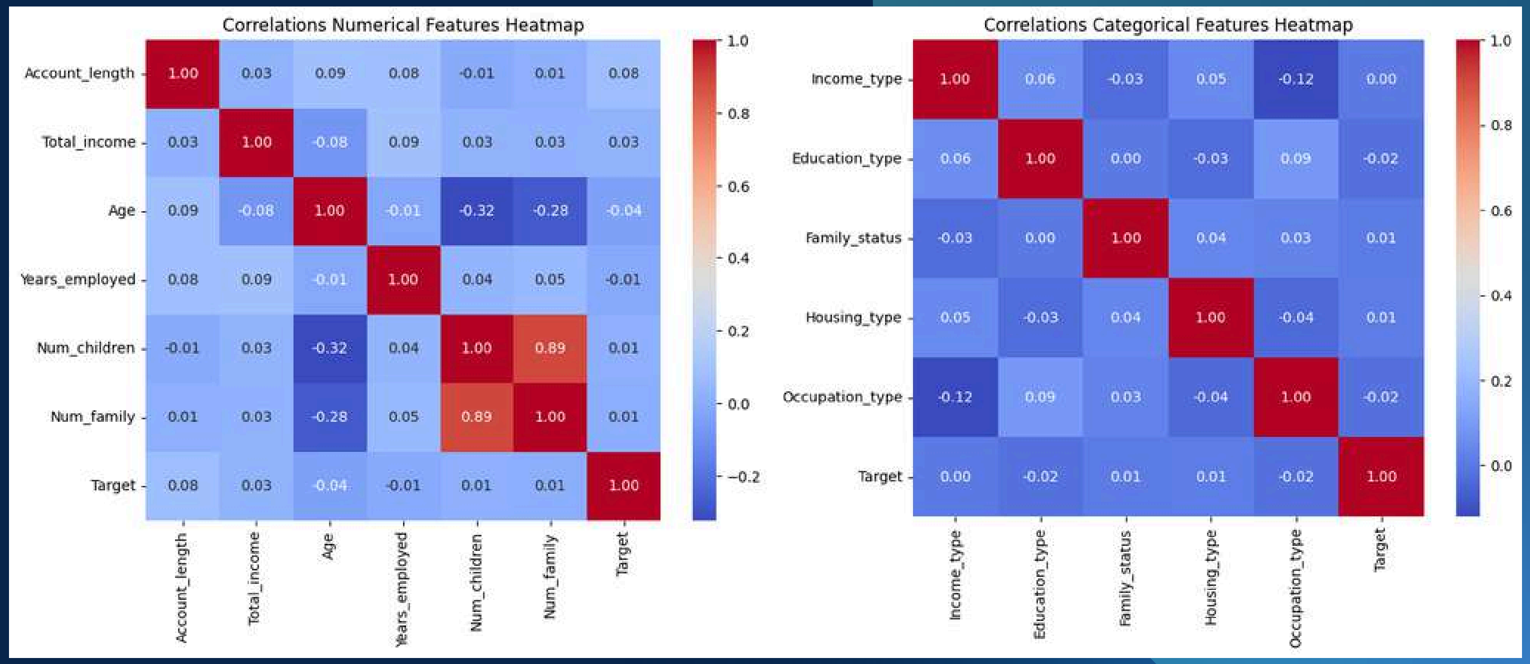
2. Outliers and Distributions

- Boxplots reveal high variance in total income and years employed, indicating potential outliers that could impact model performance.
- Certain demographic features (e.g., number of children) show skewed distributions, requiring standarization for better model accuracy.
- Age are not available due to null values.

3. **Correlation Insights**

- Account Length and Total Income show the strongest positive correlation with credit eligibility, highlighting their importance in decision-making.
- Low correlation between some categorical variables and credit approval suggests potential feature engineering opportunities.





LOANWISE

Data Processing – Preparing for Accurate Predictions

1. Encoding Categorical Variables

- Applying One-Hot Encoding to transform categorical features (e.g., income type, occupation type, education level) into numerical values.
- This ensures the model can effectively interpret non-numeric data without introducing bias.

```
#Applying One-Hot encoding to transform categories to numbers
data_encoded = pd.get_dummies(data_imputed, columns=categorical_columns, drop_first=False)
encoded_columns = data_encoded.columns.difference(data_imputed.columns)
data_encoded[encoded_columns] = data_encoded[encoded_columns].astype(int)
print(data_encoded.info())
print("Number of columns generated by One-Hot Encoding:", len(data_encoded.columns))
```

2. Handling Missing Data – Imputation with Regression

- Some records contained missing age values.
- Instead of discarding these records, we applied regression-based imputation, using available features to predict missing values.
- Comparison:
  - Original Data (Before Imputation): Contained missing or inconsistent values.
  - Processed Data (After Imputation): Predicted missing values based on patterns in the dataset.

	Age
0	38.737382
1	46.343359
2	39.589841
3	60.686169
4	36.848857
5	40.505002
6	35.063542
7	43.694067
8	38.138898
9	33.091896

Age
32.86857362
58.79381507
52.3214029
61.504343
46.19396702
48.67451077
29.21072986
27.46394519
30.02936405
34.74130201

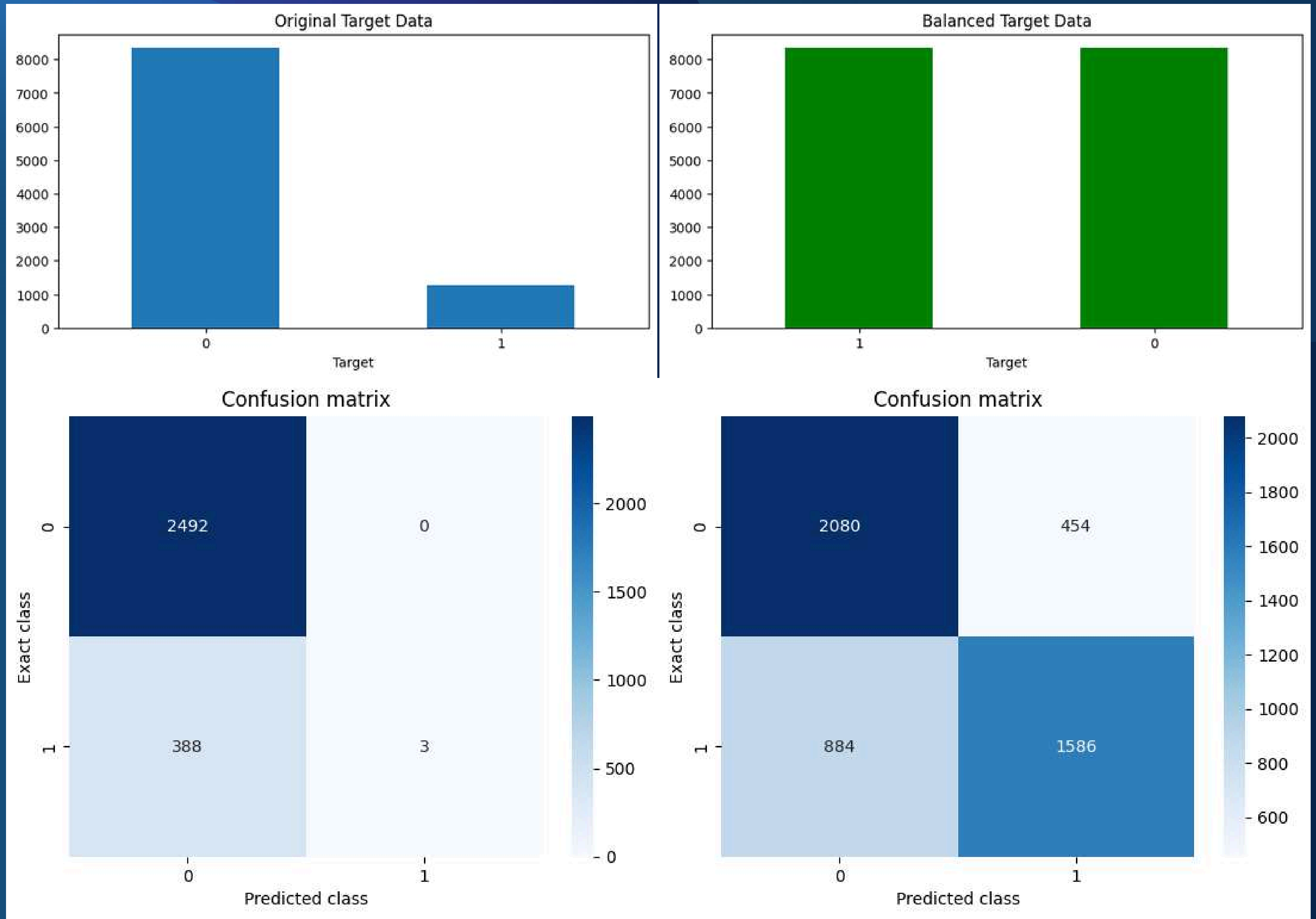
	Account_length	Total_income	Age	Years_employed
0	0.250000	0.581004	0.377033	0.289060
1	0.483333	0.124035	0.534313	0.072170
2	0.066667	0.352520	0.394661	0.194170
3	0.333333	0.372104	0.830899	0.000000
4	0.083333	0.352520	0.337982	0.048940
...	...	...	...	...
9604	0.800000	0.117507	0.296326	0.072042
9605	0.316667	0.124035	0.472626	0.171450
9606	0.350000	0.091394	0.657419	0.109527
9607	0.533333	0.189316	0.277303	0.084325
9608	0.216667	0.124035	0.096190	0.075924

3. Standardization and Normalization

- Variables like total income, years employed, and age showed large variations in scale.
- Standardization (MinMaxScaler) was applied to ensure fair comparisons between features, improving model performance.

LOANWISE

First Experiment – Initial Model and Data Challenges



- Accuracy: 0.8654
- Precision (PPV): 1.0000
- Recall (Sensibilidad, TPR): 0.0077
- Specificity (TNR): 1.0000
- F1-Score: 0.0152

- Accuracy: 0.7326
- Precision (PPV): 0.7775
- Recall (Sensibilidad, TPR): 0.6421
- Specificity (TNR): 0.8208
- F1-Score: 0.7033

1. Initial Model – Logistic Regression

- The first model used Logistic Regression as a baseline approach.
- However, results were not satisfactory due to dataset imbalance, which led to unrealistic performance metrics.
- The model showed high accuracy but failed to correctly classify credit-eligible applicants, as the majority class dominated predictions.

2. Addressing Class Imbalance – Applying SMOTE (Synthetic Minority Over-sampling Technique)

- To mitigate dataset imbalance, SMOTE was applied to generate synthetic samples for the minority class.
- This technique helped the model learn more representative patterns from both credit-eligible and non-eligible applicants.

3. Model Performance After Rebalancing

- Once the dataset was balanced, the model was re-trained, leading to:
  - Lower precision for minority class predictions
  - More realistic F1-score and ROC-AUC values
  - Improved overall decision-making for credit eligibility

LOANWISE

Model Optimization – Nine Experiments for Performance Improvement

Overview of the Experimentation Process

To achieve the best credit eligibility predictions, we conducted ***nine experiments***, progressively refining models through hyperparameter tuning, balancing data, and feature selection.

Key Steps in Model Evolution

➡ Baseline Model – **Logistic Regression** (Unbalanced Data)

- High accuracy but severely biased due to class imbalance.
- Precision = **1.0**, but Recall = **0.007**, indicating poor minority class detection.

➡ Balancing the Data – **Logistic Regression**

- Applied SMOTE to address imbalance.
- Recall improved to **0.642**, making the model more reliable.

➡ Introducing More Advanced Models – **XGBoost**

- Base XGBoost model outperformed Logistic Regression in Recall and F1-score.

➡ Hyperparameter Tuning – **XGBoost Optimization**

- Two stages of tuning increased performance further.
- **F1-score reached 0.871**, with optimized recall of **0.840**.

➡ **Feature Selection** – Reducing Complexity

- Selected most relevant variables to enhance interpretability while maintaining high accuracy.

➡ Exploring Other Models – **Random Forest & LightGBM**

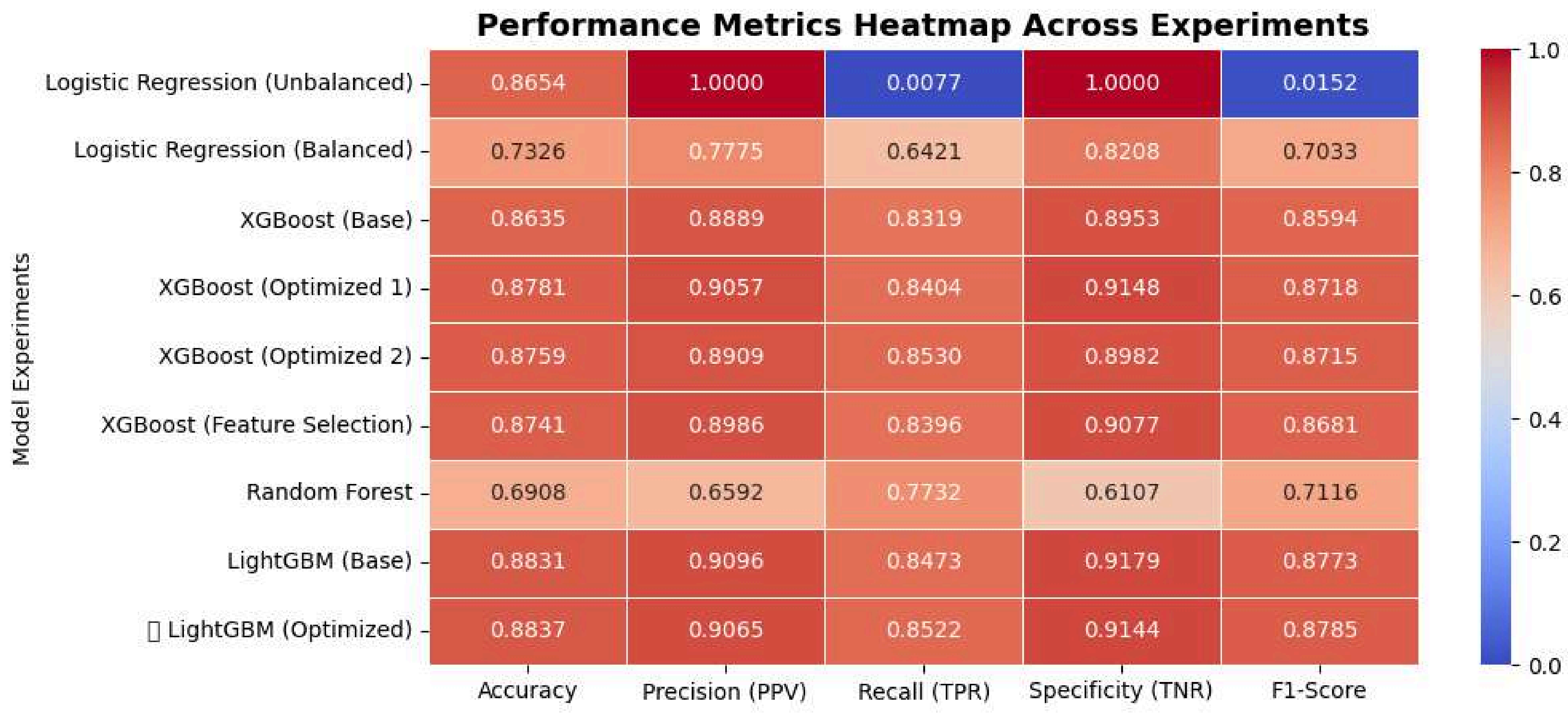
- Random Forest showed lower accuracy and recall, proving less effective.
- LightGBM demonstrated strong performance, competing closely with optimized XGBoost.

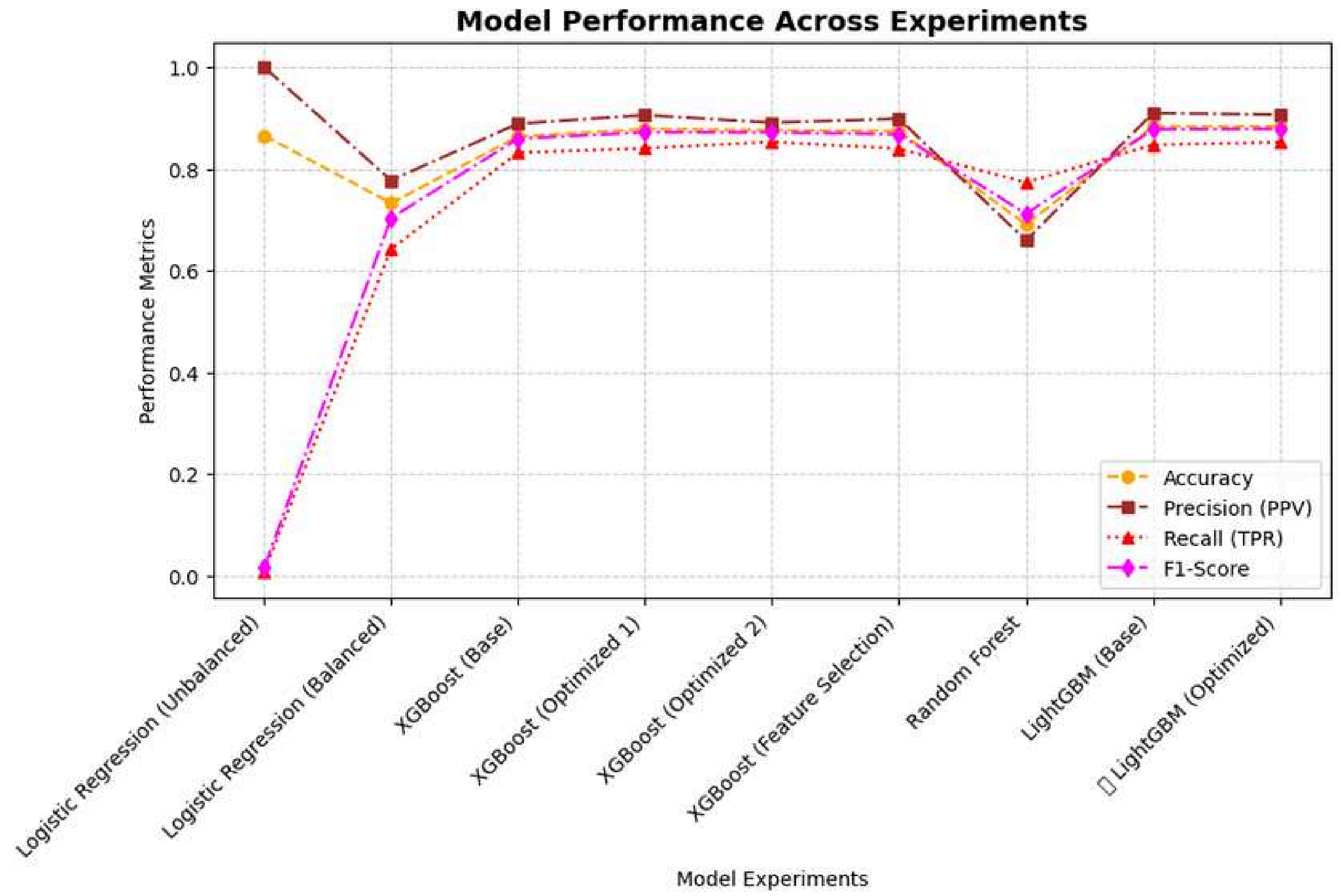
➡ Final Model – **Optimized LightGBM**

- Highest specificity (**0.914**) and balanced recall (**0.852**).
- Best trade-off between accuracy, precision, and recall.

LOANWISE

Model Optimization Results





Through systematic model improvements, we achieved significant enhancements in credit risk prediction accuracy. By addressing data imbalances, optimizing hyperparameters, and selecting relevant features, LoanWise ensures fair, reliable, and data-driven lending decisions.

Next Steps – Enhancing LoanWise's Capabilities

After successfully **onboarding the first early adopters**, LoanWise will focus on enhancing its predictive capabilities and expanding its features to deliver even greater value to financial institutions and borrowers.

1 Further Model Optimization

- Continue improving model accuracy by **exploring advanced deep learning techniques** and refining feature engineering.
- Enhance **explainability** of predictions to increase trust and transparency in lending decisions.

2 Predicting Loan Amounts Based on Debt Capacity

- Develop a model that **estimates optimal loan amounts** considering a user's **financial stability** and **debt-to-income ratio**.
- Prevent over-indebtedness by offering **realistic and sustainable credit limits**.

3 Personalized Credit Utilization Recommendations

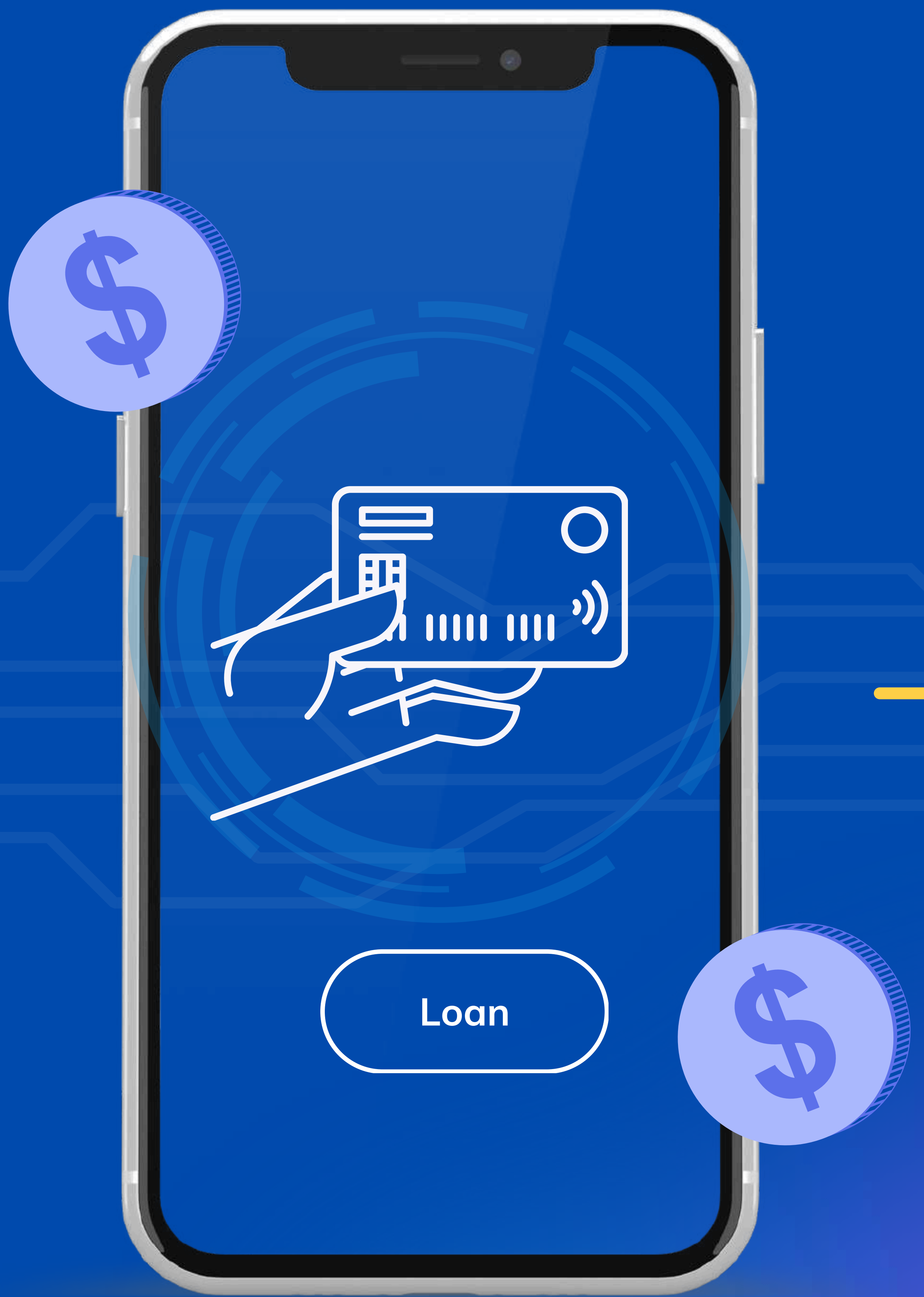
- Leverage qualitative and quantitative data to **provide tailored credit usage suggestions**.
- Categorize borrowers based on spending behavior and financial goals, offering customized credit options that align with their needs.

4 Real-Time Adaptive Scoring System

- Implement a dynamic credit scoring approach that **adjusts over time** based on an applicant's financial behavior.
- Enable continuous updates and **early risk detection** for lenders.

LOANWISE

LoanWise is not just a credit eligibility tool—it is shaping the future of **smarter**, **fairer**, and more **responsible** lending



Thank You!

Try LoanWise

github.com/ragarciaga/loanwiseapp

linkedin.com/in/ragarciaga/

Colombia, South America