



Banking Project: Risk Analytics

This project applies risk analytics to banking data, building predictive models to classify customer risk levels and support lending decisions.

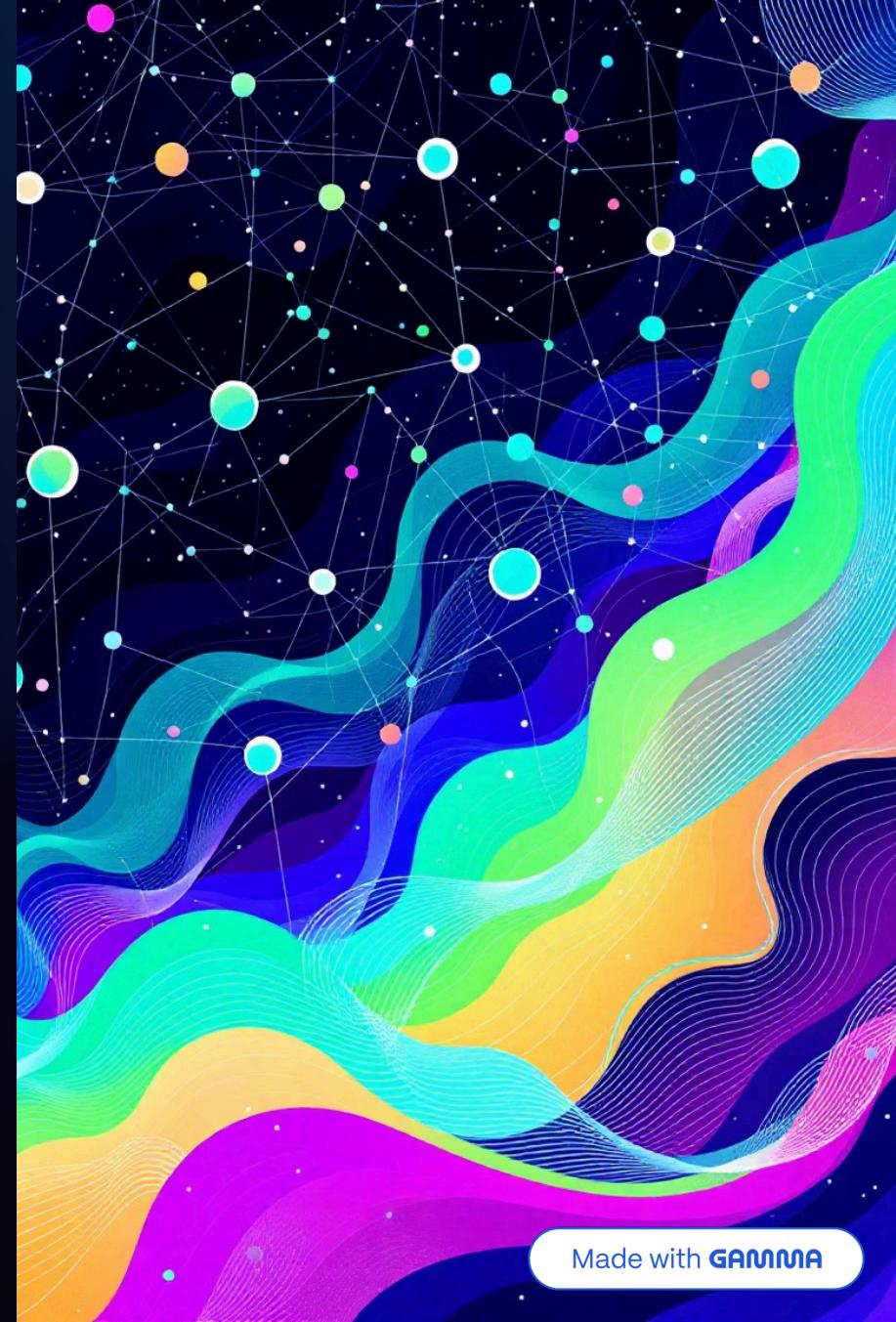
Problem & Solution

Problem Statement

Analyze customer financial data to identify and classify risk levels, supporting effective risk management and lending decisions.

Solution Overview

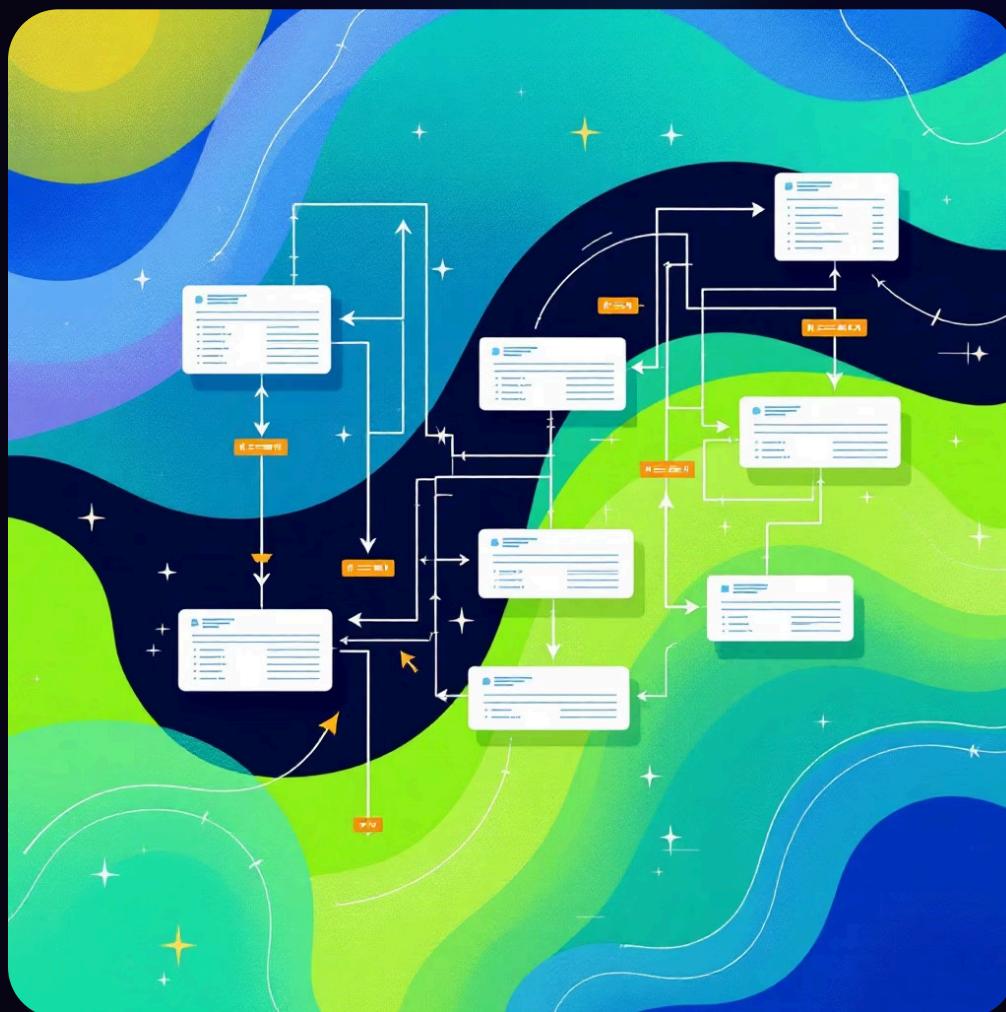
Combines Power BI dashboards for descriptive analysis with ML models for predictive risk assessment, enabling data-driven loan approvals.



Dataset & Connectivity

About the Dataset

Contains interlinked tables (Banking Relationship, Client-Banking, Gender, Investment Advisor) with bank and client details.



Database Connectivity

Banking data stored in MySQL, accessed via MySQL Workbench and Python's `pymysql` library for direct querying into Pandas DataFrames.

```
import pymysql  
conn = pymysql.connect(  
    host="127.0.0.1",  
    user="root",  
    password="",  
    database="banking_Case"  
)  
print("Connected successfully")
```

Exploratory Data Analysis (EDA)

EDA was crucial for understanding data structure, quality, and patterns to guide feature engineering and modeling.

01

Dataset Structure

Examined using `info()`, `describe()`, and `shape` for records, data types, missing values, and distribution.

02

Univariate Analysis

Analyzed individual variables (age, income, loans) with histograms and bar charts to identify skewness and outliers.

03

Bivariate Analysis

Studied relationships between key variables and target features, uncovering patterns like credit utilization and risk.

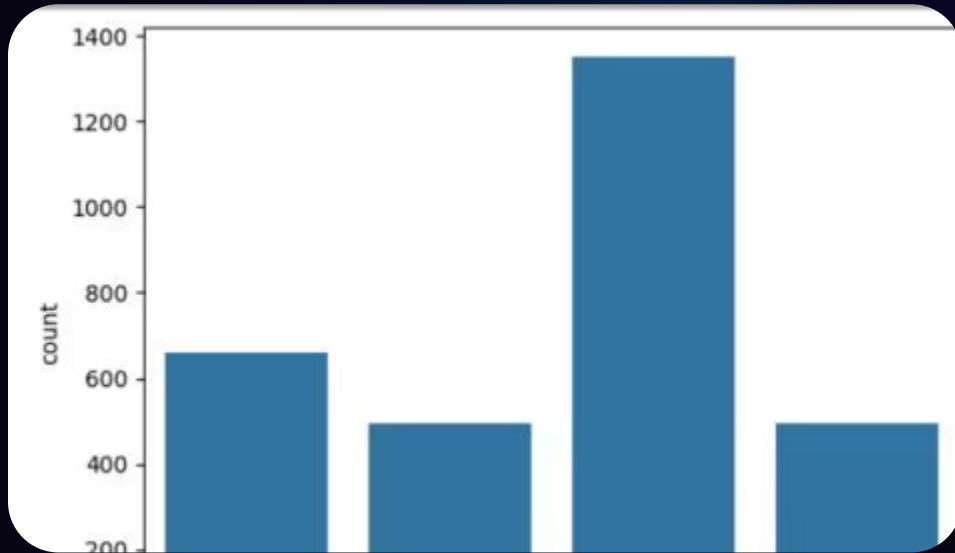
04

Correlation Analysis

Used heatmaps to identify relationships among numerical variables and detect multicollinearity, informing feature selection.

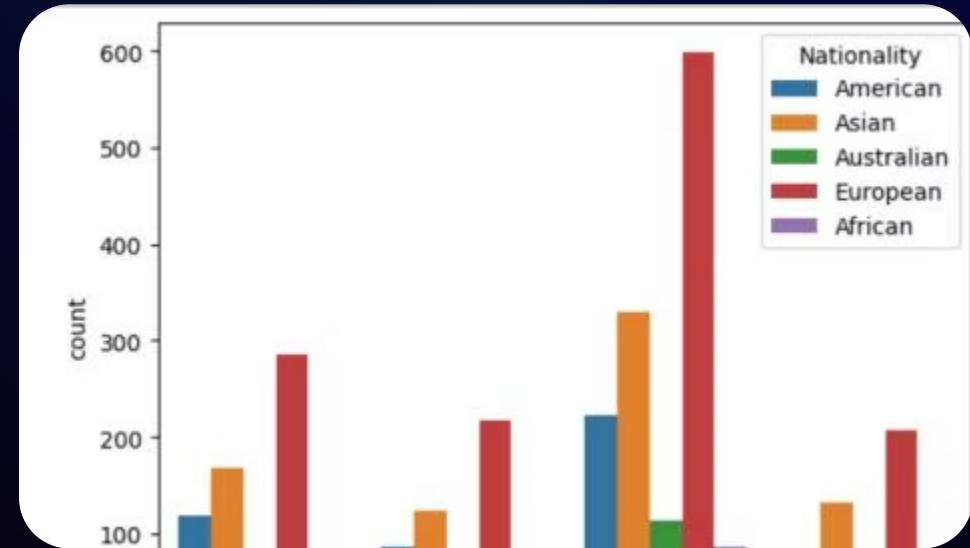
EDA Visualizations

Univariate Distribution



Bar chart showing the distribution of 'Amount of Credit Cards', with the third category having the highest count.

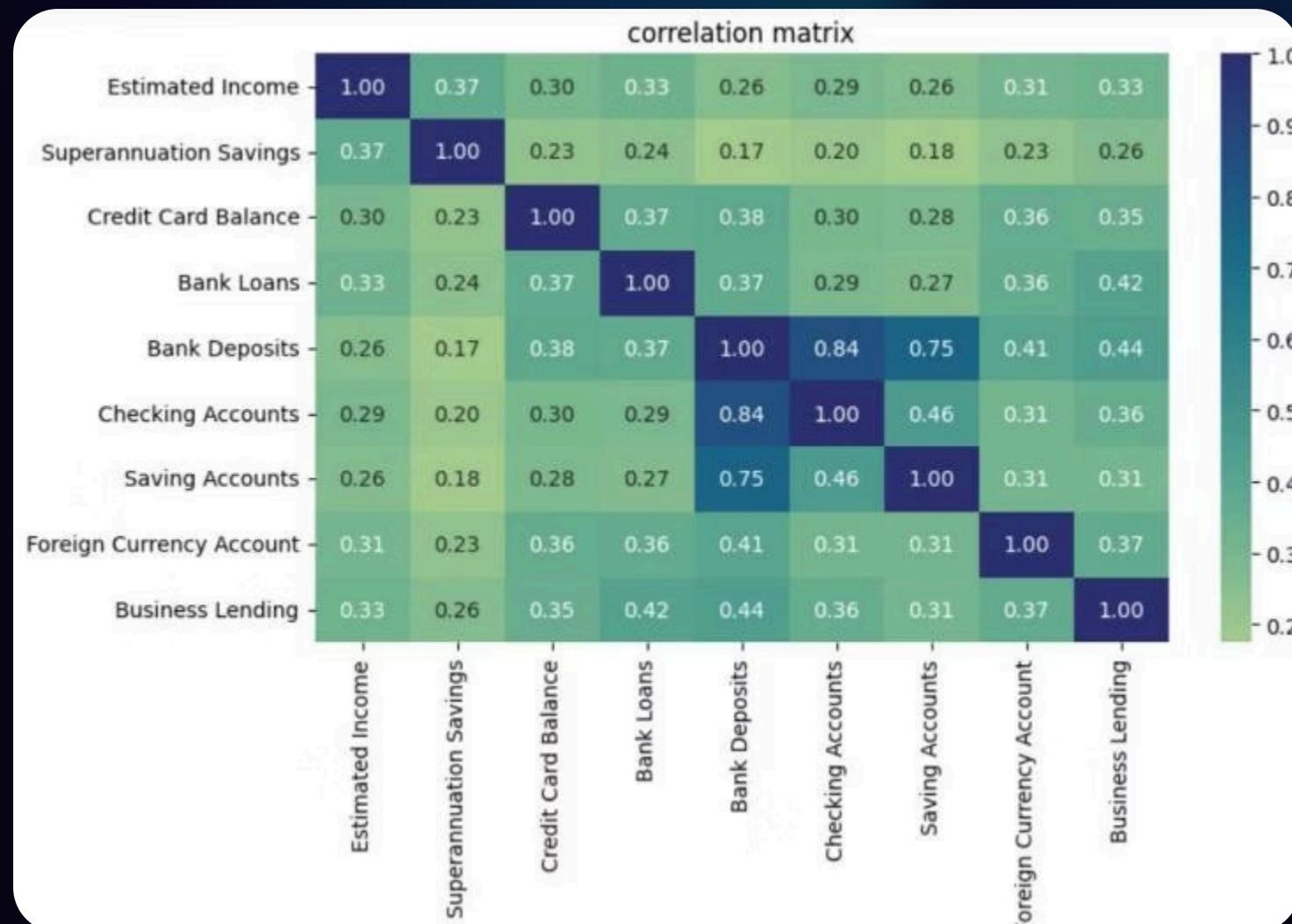
Bivariate Analysis: Nationality



Distribution of a predictor by 'Nationality', with Europeans showing the highest count.

Correlation Matrix

Heatmap revealing relationships between financial variables, crucial for feature selection.



Estimated Income	1.00	0.37	0.30	0.33	0.26	0.29	0.26	0.31	0.33
Bank Loans	0.33	0.24	0.37	1.00	0.37	0.29	0.27	0.36	0.42
Bank Deposits	0.26	0.17	0.38	0.37	1.00	0.84	0.75	0.41	0.44



Power BI Dashboard

Interactive dashboard for visualizing customer profiles and financial patterns.

1

Home

High-level overview of key metrics and navigation.

2

Loan Analysis

Focuses on customer loan behavior and credit exposure.

3

Deposit Analysis

Analyzes deposit patterns and savings behavior.

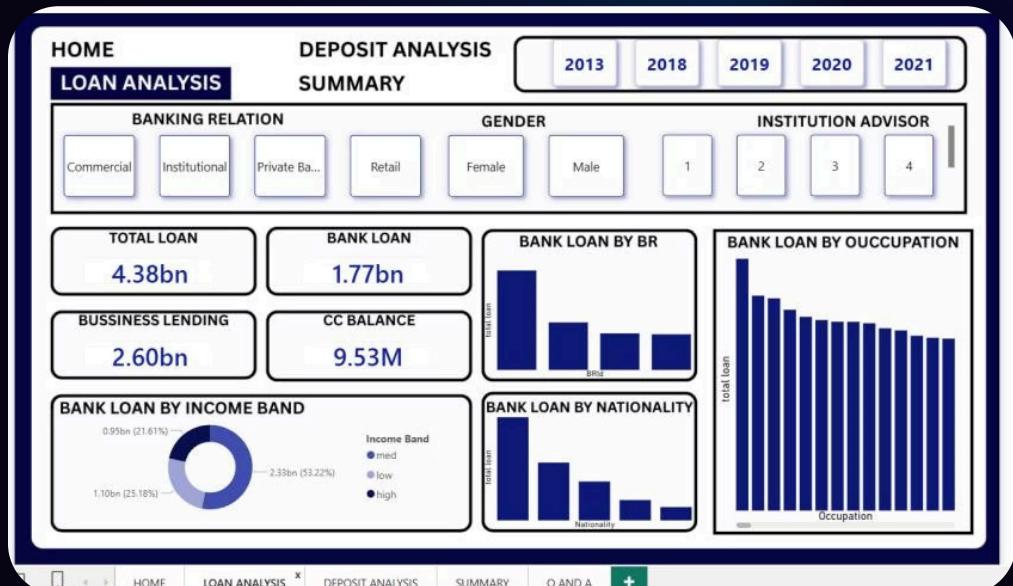
4

Summary, Q&A

Consolidated insights and interactive data querying.

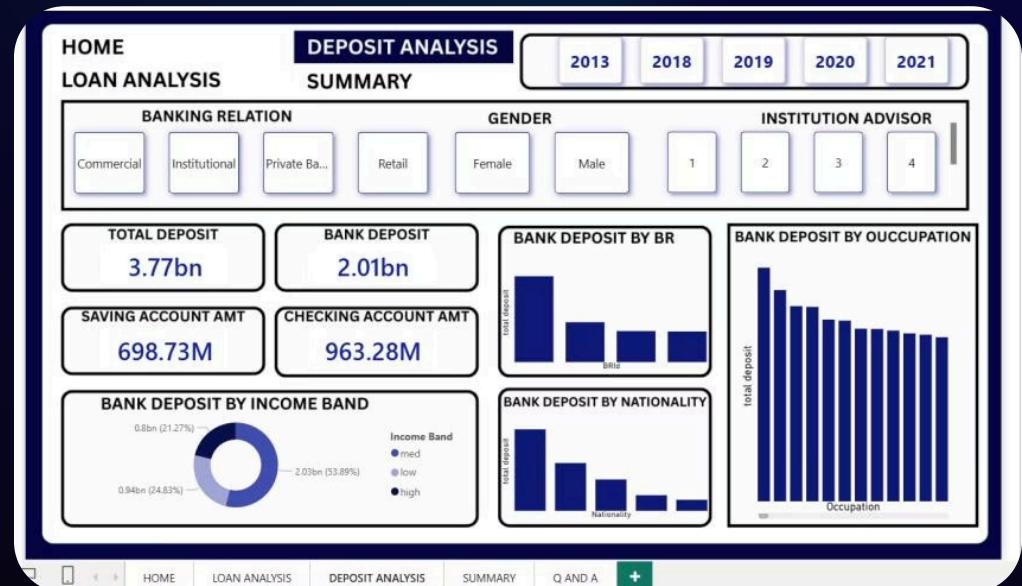
Loan & Deposit Insights

Loan Analysis Dashboard



Visualizes loan distribution by banking relation, occupation, income band, and nationality.

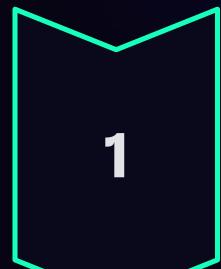
Deposit Analysis Dashboard



Illustrates deposit patterns by banking relation, occupation, income band, and nationality.

Machine Learning Methodology

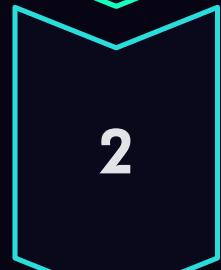
A pipeline to predict customer risk levels based on financial and demographic attributes.



Feature Engineering

1

Created Total Balance, Credit Utilization, and Loan-to-Income Ratio.



Data Preparation

2

Encoded categorical variables and standardized numerical features.



Model Development

3

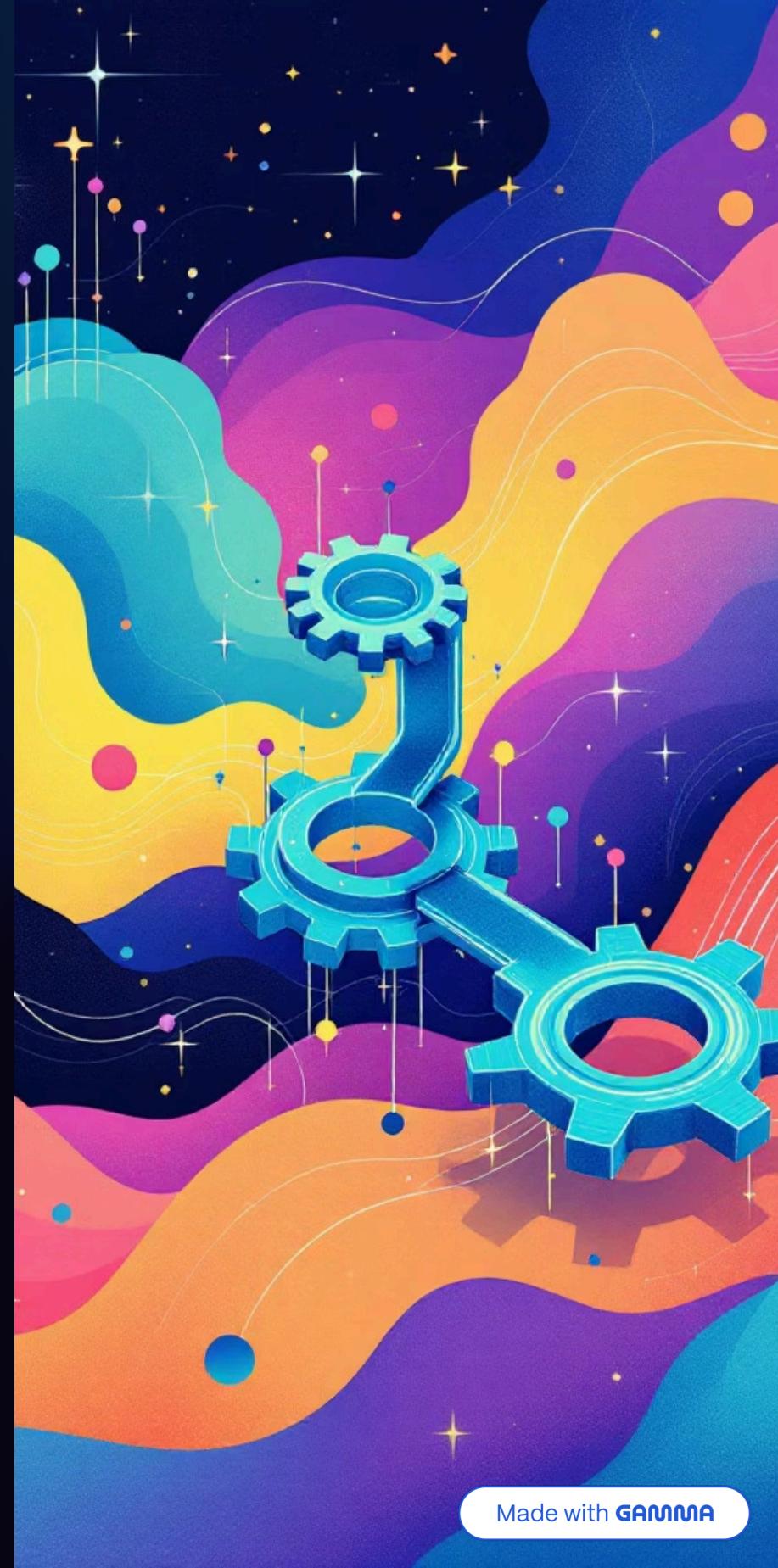
Implemented Logistic Regression, Random Forest, and Gradient Boosting.



Model Evaluation

4

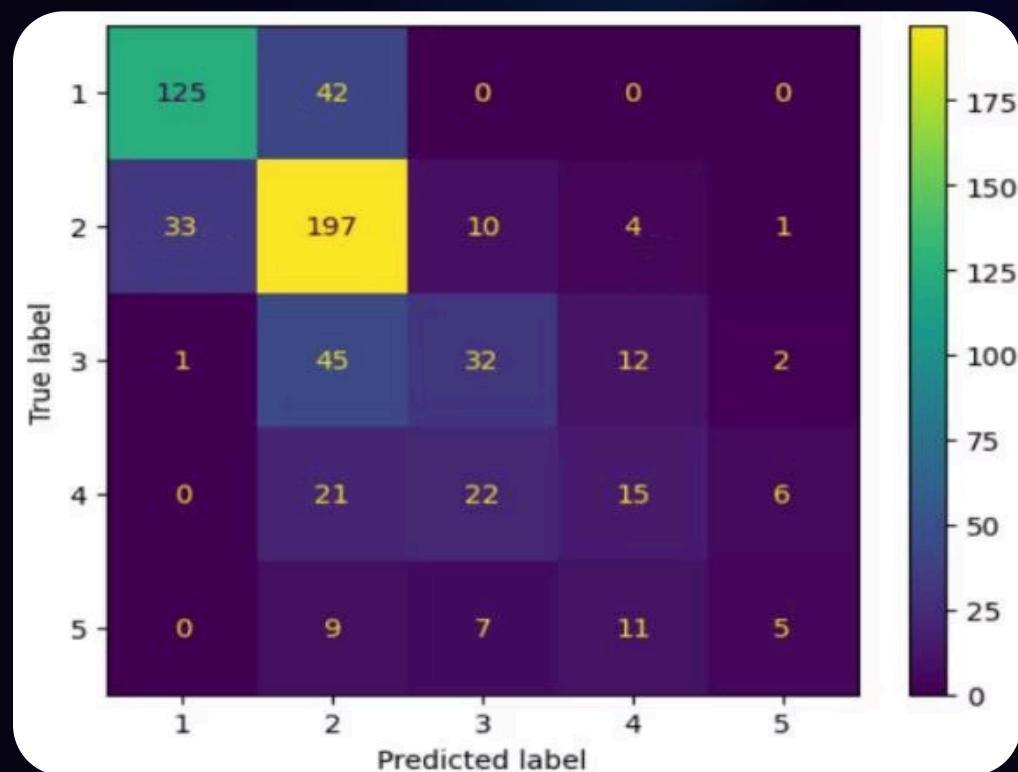
Evaluated using accuracy, precision, recall, F1-score, and ROC-AUC.



ML Outcome & Future

Gradient Boosting Performance

Demonstrated superior performance in classifying risk categories.



Future Scope

- Deploy model as a web application for real-time assessment.
- Integrate explainable AI (SHAP) for transparency.
- Expand dataset with real-time data for better generalization.
- Automate data pipelines for continuous retraining.
- Extend to credit scoring, customer segmentation, fraud detection.

ROC-AUC: 0.8546724217082096