CREDIT CARD APPROVAL PREDICTION

BINARY CLASSIFICATION

MACHINE LEARNING PORTFOLIO

DEREK RODRIGO SÁNCHEZ SEGUAME

1. Task Definition
2. EDA
3. Model and Results

01

Task Definition

PROBLEM STATEMENT | CUSTOMER CREDIT CARD APPROVAL

CardSecure Financial is a growing **consumer finance** division of a major Mexican bank that has rapidly expanded its credit card portfolio among salaried professionals and small business owners. Leveraging targeted marketing and flexible credit limits, the division **exceeded its origination goals** in its first two years.

However, CardSecure now confronts a **rising default rate** that is **eroding net interest income** and forcing higher loan-loss provisions, **squeezing profitability** and tying up regulatory capital. To remedy this, CardSecure will deploy a **Machine Learning–driven underwriting engine**—powered by AI.

The project will use the Credit Card Details dataset, which includes demographics, income, asset ownership, employment tenure, and historical approval outcomes. By applying advanced predictive analytics, CardSecure aims to reduce **default-related write-offs**, **streamline approval workflows**, and **enhance customer lifetime value**—driving **sustainable P&L improvement** and reinforcing its competitive edge in Mexico's consumer lending market.

BANKING AND DATA | USING MACHINE LEARNING TO FIND THE OPTIMAL RISK PROFILES

Strategic Approach

Our client

- **CardSecure** is a mexican retail banking division
- Their target market are **young professionals and SMEs**
- They've exceeded their origination goals in the first two years

Pain point

- they are experiencing **excessive defaults** in credit card origination due to **unsound risk model**
- This generates problems in:
  - Income (NII)
  - Higher Loss Provisions (PCLs)

Solution

- Using the provided dataset, a **Machine Learning Binary Classification Model** will be generated
- This model will evaluate each client to decide if they are fit to approval or rejection of a credit card

Note:

02

Exploratory Data Analysis

OVERVIEW | AN UNBALANCED TARGET VARIABLE

of Features | By data type

	#Features
Numerical	11
Categorical	8
Total	**19**

# of Categorical Features | By # of unique observations

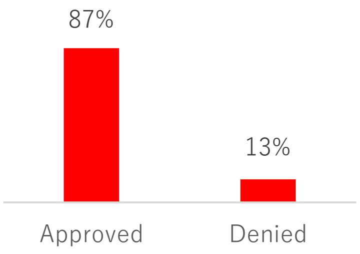
	Two Unique Obs	Three to five Unique Obs	Five+ Unique Obs
#Features	3	3	2

Categorical Features Distributions

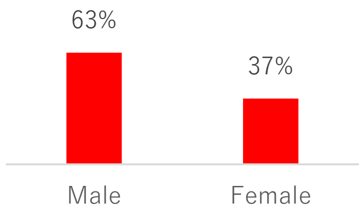
Target

Features

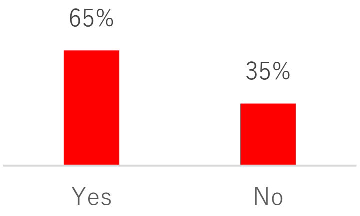
Churn



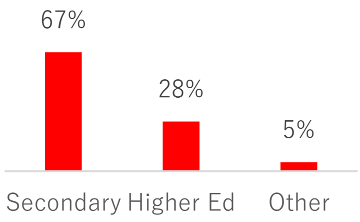
Gender



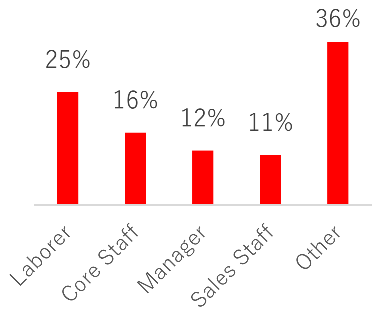
House Owner



Education



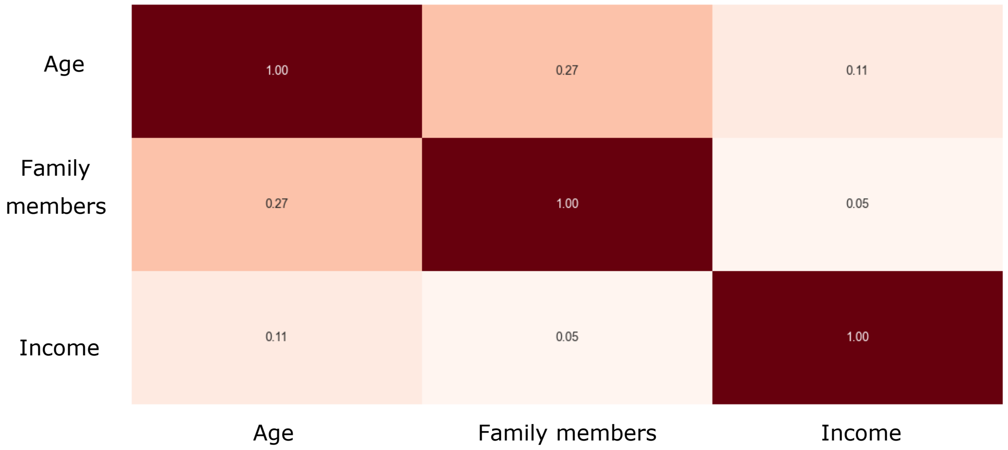
Occupation



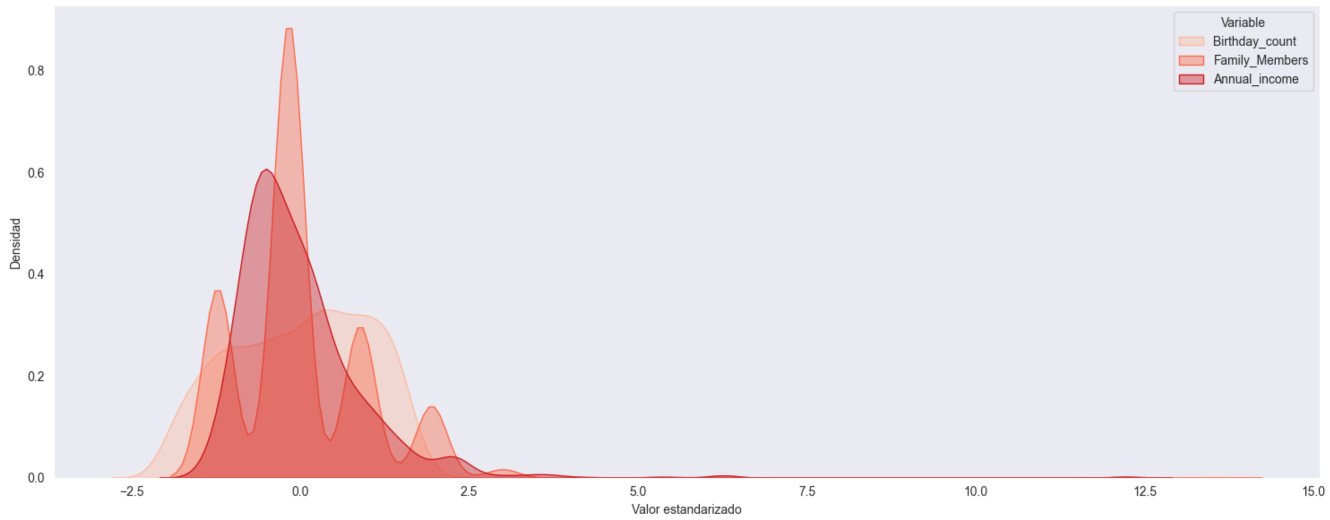
Note:

NUMERICAL VARIABLES | HEATMAP, DISTRIBUTIONS AND BOXPLOTS

Correlation Matrix Heatmap

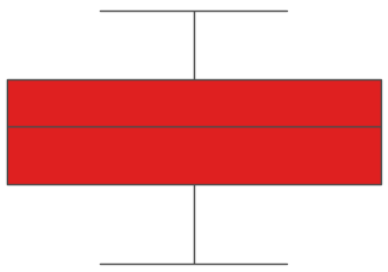


Distribution Graph

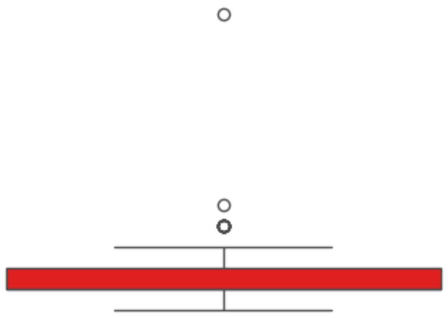


Box plot | by feature

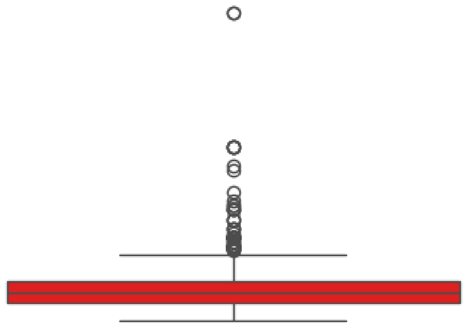
Age



Family Members



Annual Income



03

Model Selection and Results

MODEL SELECTION | GRADIENT BOOSTING PERFORMED THE BEST IN F1 SCORE; OPTIMAL FOR UNBALANCED DATASETS

Model	Tuning	Parameters	Metrics
Gradient Boosting	Randomized Search	learning rate = 0.01 n_estimators = 100 Max_Depth = 4	**F1: 98%** Precision: 97% AUC ROC: 99%
Random Forest	Randomized Search	n_estimators = 427 max_depth = 17	**F1: 98%** Precision: 97% AUC ROC: 99%
Logistic Regression	Grid Search	L1 Regularization C = 5.2	**F1: 97%** Precision: 97% AUC ROC: 99%
SVM	Grid Search	C = 1.0 kernel = rbf	**F1: 49%** Precision: 92% AUC ROC: 98%

Note:

FROM DATA TO ACTION ITEMS | BUSINESS VALUE PROVIDED BY MACHINE LEARNING

Findings

- **Gradient Boosting** has the best scoring overall
- In **unbalanced** datasets **traditional scoring** metrics (accuracy) **won't reflect real performance** for the model

Conclusions

- In general, **the drivers** of approval are
  - **Occupation**
  - **Age**
  - **Income**
- A solid classification model will improve **approval criteria** and **allow the operation to escalate**

Next Steps

- **Deploy and monitor** the model for unaccounted scenarios
- Keep track of KPIs such as
  - **Past due credit card volume**
  - **PCL ratio**

Note: