LOAN DEFAULT PREDICTION

Juan León[1], Miguel Hernandez[2], Juan Diaz[3], Esteban Avila[4]
Faculty of Engineering - Systems Engineering
Universidad Distrital Francisco José de Caldas, Bogotá, Colombia

INTRODUCTION

The competition requires determining whether a loan will default, as well as the loss that will be incurred in the event of default. Unlike traditional financial approaches that are limited to binary answers, this approach seeks to anticipate and incorporate both the possibility of default and the severity of the resulting losses.

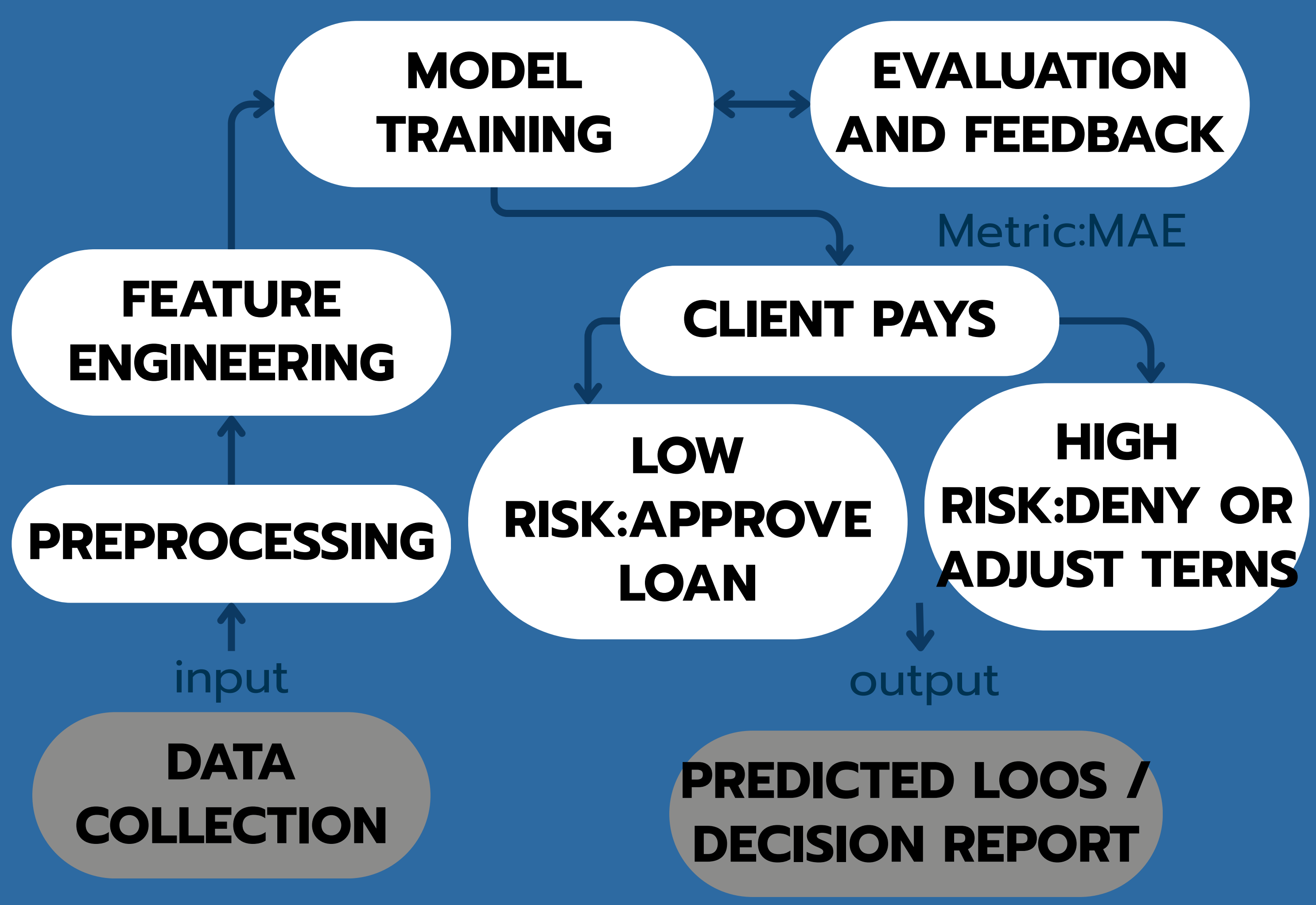
GOAL

Can a dynamic, data-driven learning model improve risk assessment and capital decisions compared to traditional models?

The goal is to create a model capable of providing accurate and consistent answers, reducing the mean absolute error (MAE), even in the face of changing data, and offering a more useful tool than the traditional model, thereby facilitating decision-making and capital management.

PROPOSED SOLUTION

We built a classic machine learning model (Random Forest) and a Cellular Automaton trained with previously standardized data, developed in Python under Systems Engineering principles to guarantee Robustness, Scalability and Homeostasis.



Key technical steps:
-Imputation of missing values
-Scaling and normalization
-Feature engineering and selection
-Training with LightGBM and CatBoost
-Inference and generation of submit.csv

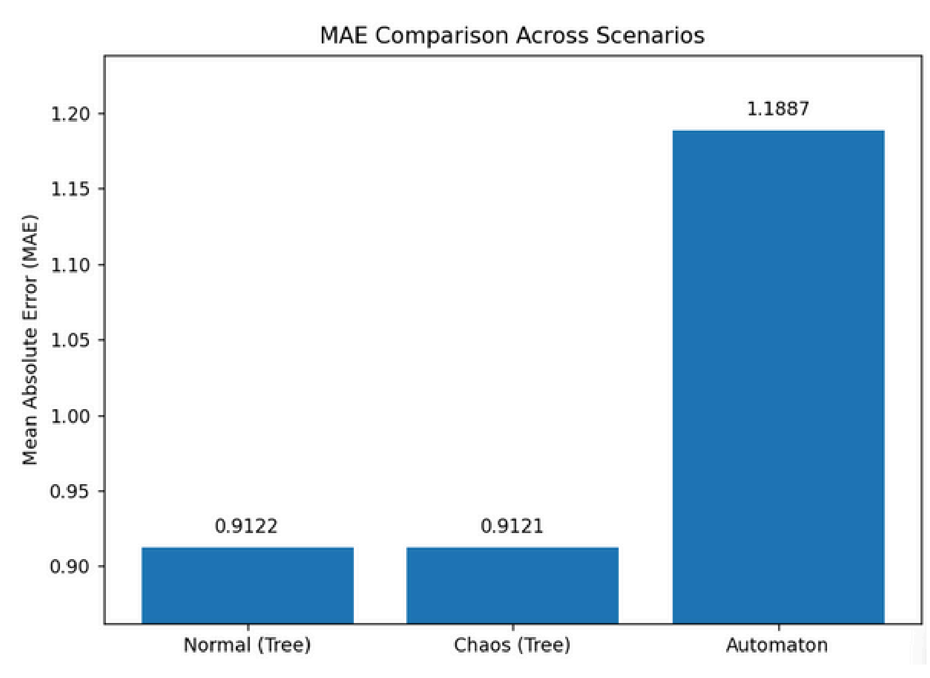
EXPERIMENTS

The system was evaluated in three scenarios to measure accuracy, robustness, and dynamic behavior:
-Normal Case: stable simulation with preprocessed data.
-Chaos Case: controlled disturbances to measure the system's sensitivity and robustness.
-Financial System Case: macro-financial shock and nonlinear dynamics.

Case	Description	Key parameters	Metric evaluated
Normal	Baseline simulation with preprocessed and undisturbed data	Data properly preprocessed	MAE
Chaos	Controlled induced perturbation of the inputs to assess sensitivity	Multiplication of the values by a random factor between 0.5 and 1.5	MAE
financial system	Macro-financial shock to observe non-linear dynamics	eigenvalue persistence=0.8 probability of contagion=0.6	MAE

RESULTS

We used a large dataset (train_v2) with more than 100,000 records, which was later preprocessed down to a clean subset of 5,000 instances. The evaluation metric (MAE) showed that the deviation between simulated and real outcomes remained minimal across all scenarios, staying well below 2% relative difference.



CONCLUSIONS

Our goal was to design a robust machine learning process to predict loan defaults from anonymized data.
The final LightGBM model achieved competitive MAE defaults through advanced preprocessing and consistent transformations.
While we achieved our goal, the model's interpretability remains limited due to feature anonymization.