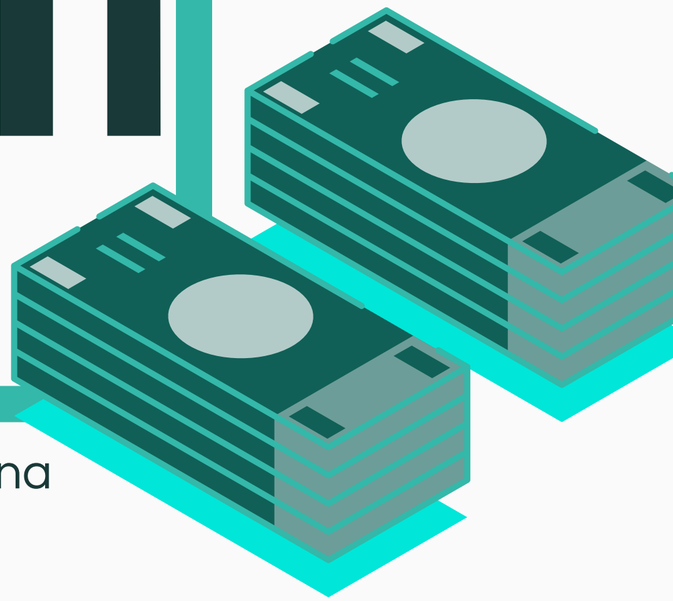
Loan Default Prediction

A Systems Engineering Approach



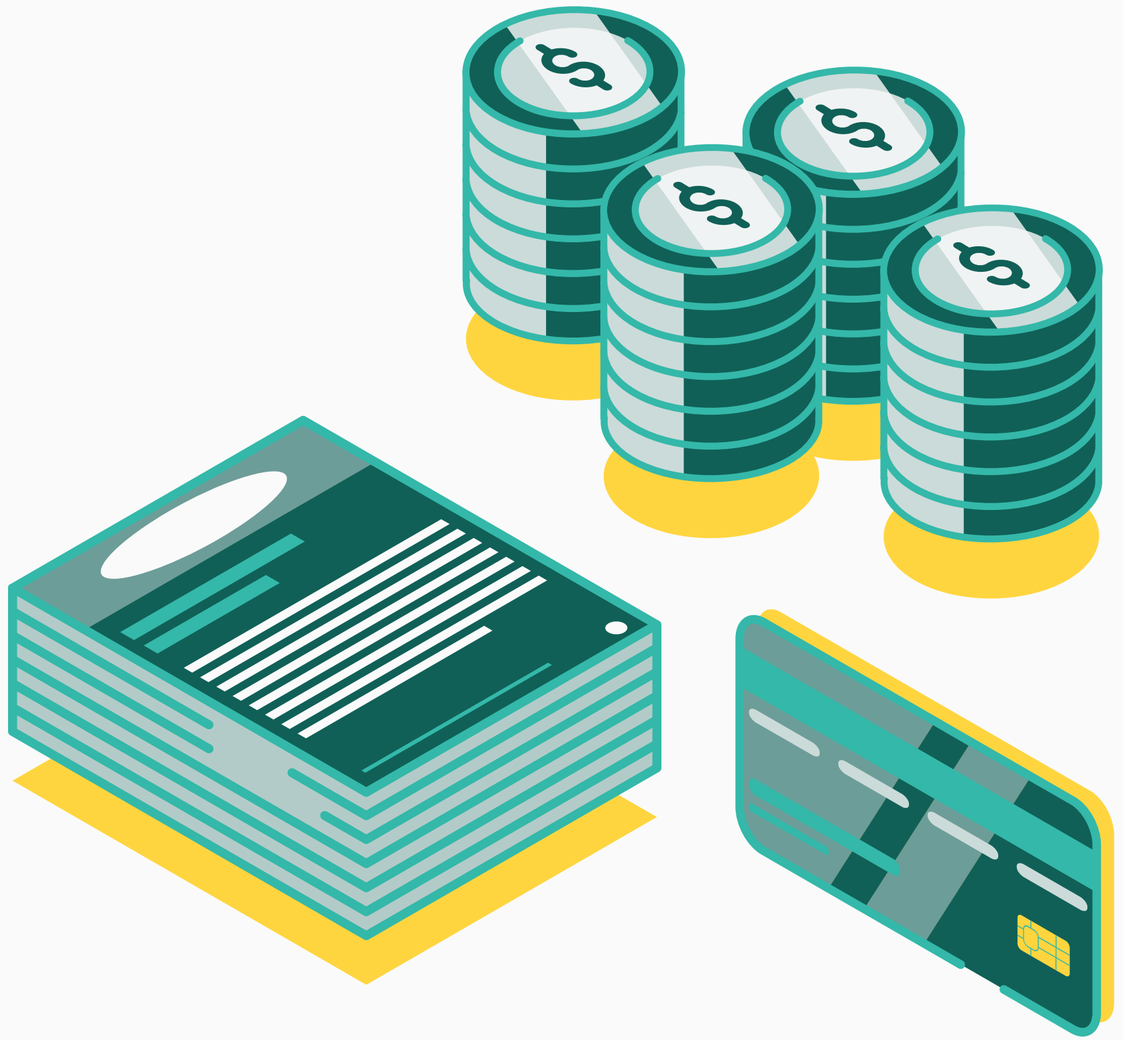
Juan Jose León Gomez

Miguel Angel Hernandez Medina

Juan Pablo Diaz Ricaurte

Juan Esteban Avila Trujillo

INDEX



Introduction

Overview

Traditional risk management is often insufficient

Challenges

Mitigating the inherent system instability

Motivation

Design a system capable of performing a dual prediction

Approach

Systems Engineering principles

SYSTEMATIC ANALYSIS OF THE PROBLEM

1. System Elements

- Financial clients, transactions, and credit histories form a dynamic system where each component influences others.

2. Interdependencies

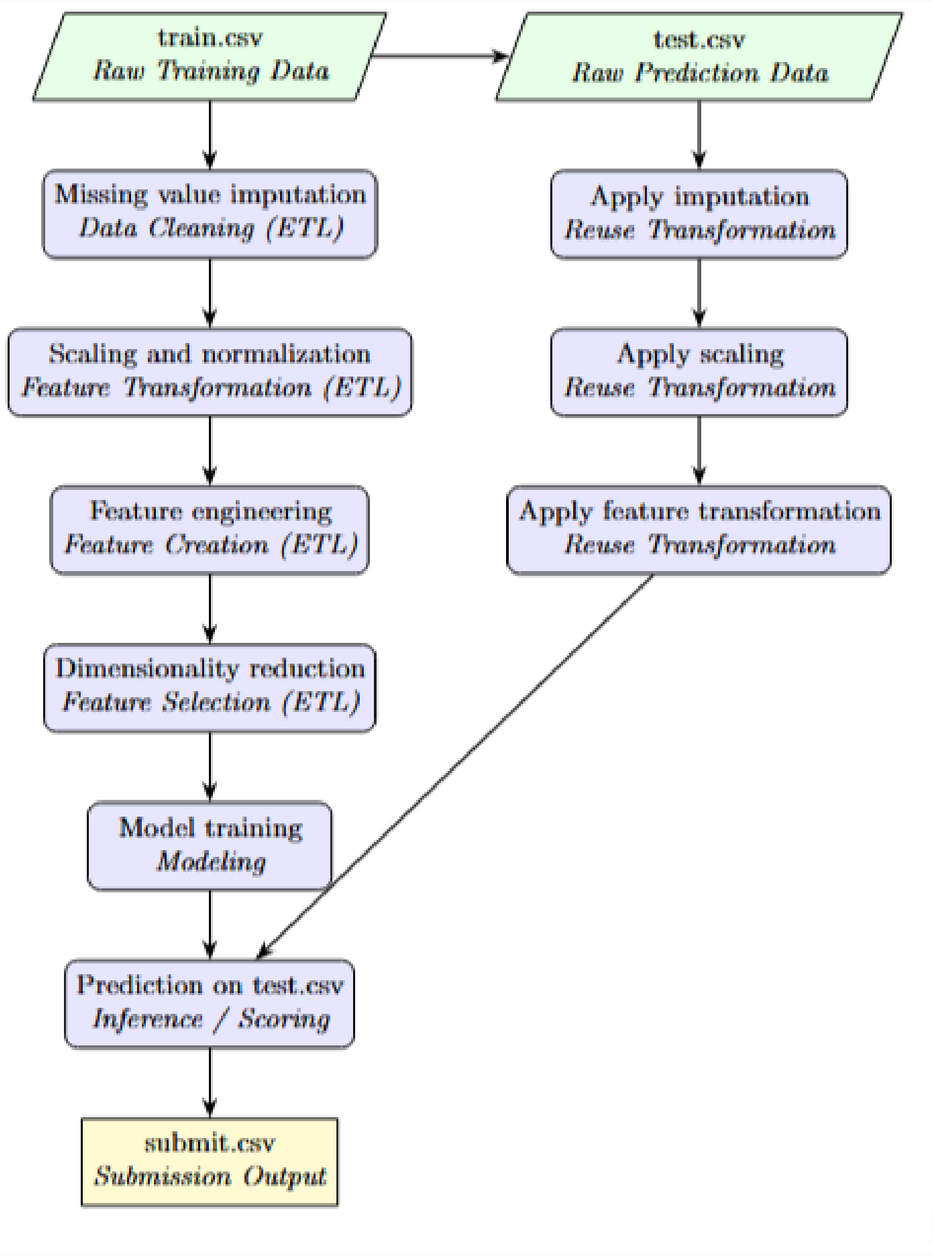
Variables like income, debt, and payment behavior are deeply connected — a small change in one alters the whole structure.

3. Sensitivity and Chaos

- The system shows nonlinear behavior: minor data variations can lead to large prediction shifts, reflecting chaotic dynamics.

4. Systemic Risk

- When unstable patterns emerge in data, errors amplify through the model, producing unreliable or biased outcomes.



System Design

Data Preprocessing

- Imputation of missing values
- Feature scaling and normalization

Feature Engineering

- Creation and selection of relevant variables

Dimensionality Reduction

- Removal of noise and irrelevant features

Model Training

- Gradient Boosting (LightGBM, CatBoost)

Prediction & Submission

- Apply trained model to test.csv
- Generate submit.csv for Kaggle

conclusions

- A robust data processing pipeline is essential when working with anonymized and high-dimensional datasets. Ensuring consistency between training and test sets significantly improves model reliability.
- Gradient boosting models like LightGBM and CatBoost demonstrated strong performance and adaptability, even in the absence of domain-specific feature names or labels.
- Despite the challenges, the final solution achieved competitive accuracy by focusing on systematic preprocessing, model tuning, and error minimization under the MAE metric.

THANK You!