otto group

Product Classification Challenge, A Systematic Analysis

Systems Analysis & Design

Juan Diego Lozada 20222020014
Juan Pablo Mosquera 20221020026
María Alejandra Ortiz Sánchez 20242020223
Jeison Cuenca: 20242020043
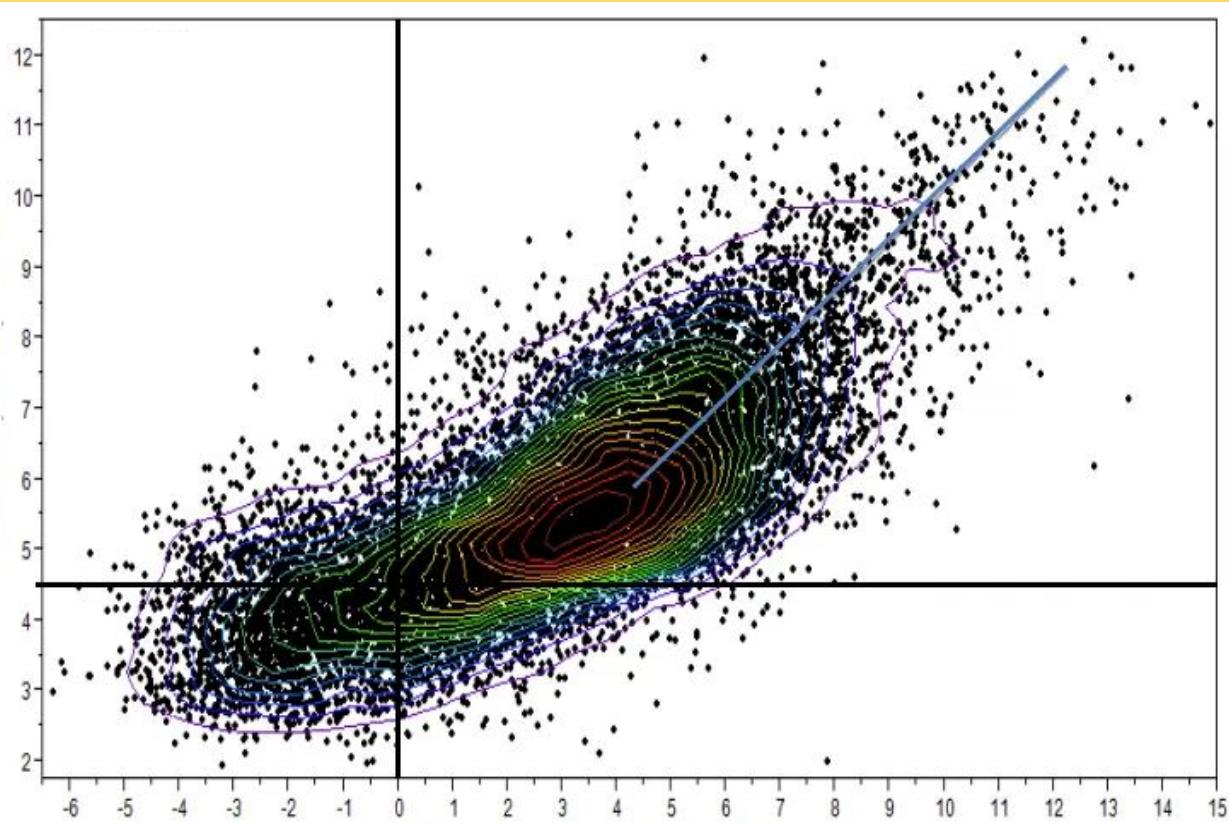
Context and Motivation
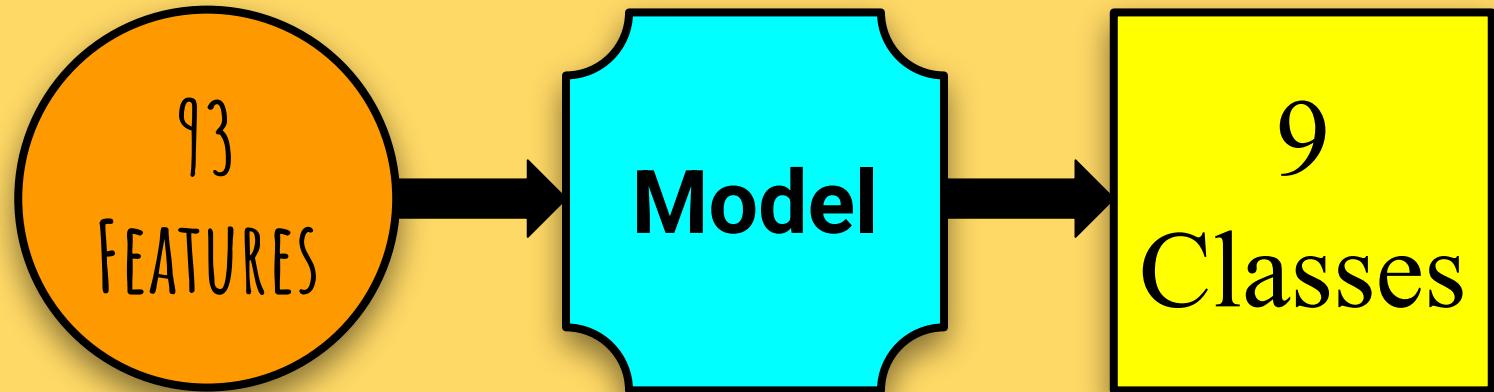
Cuenca, Lozada, Mosquera, & Ortiz

Problem Definition

Project Goals
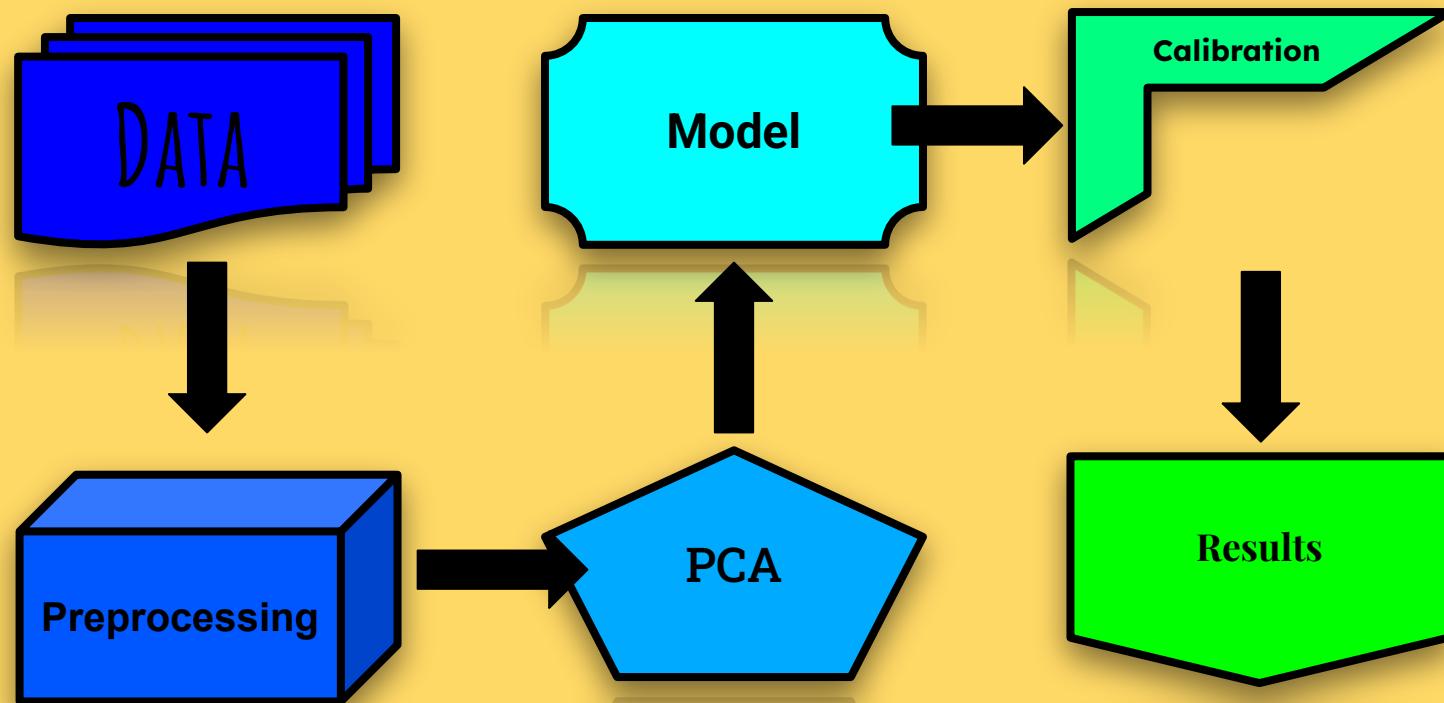
Cuenca, Lozada, Mosquera, & Ortiz

General Methodology
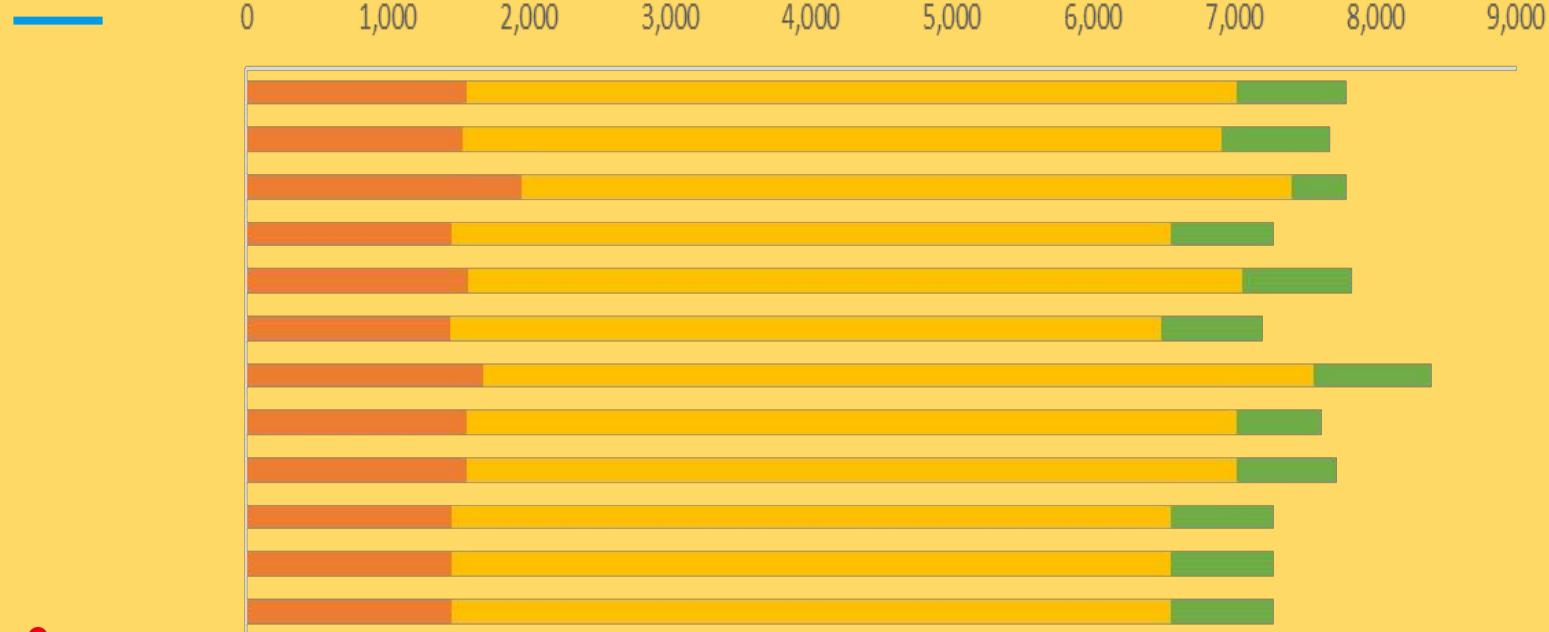
Cuenca, Lozada, Mosquera, & Ortiz

Dataset and Preprocessing
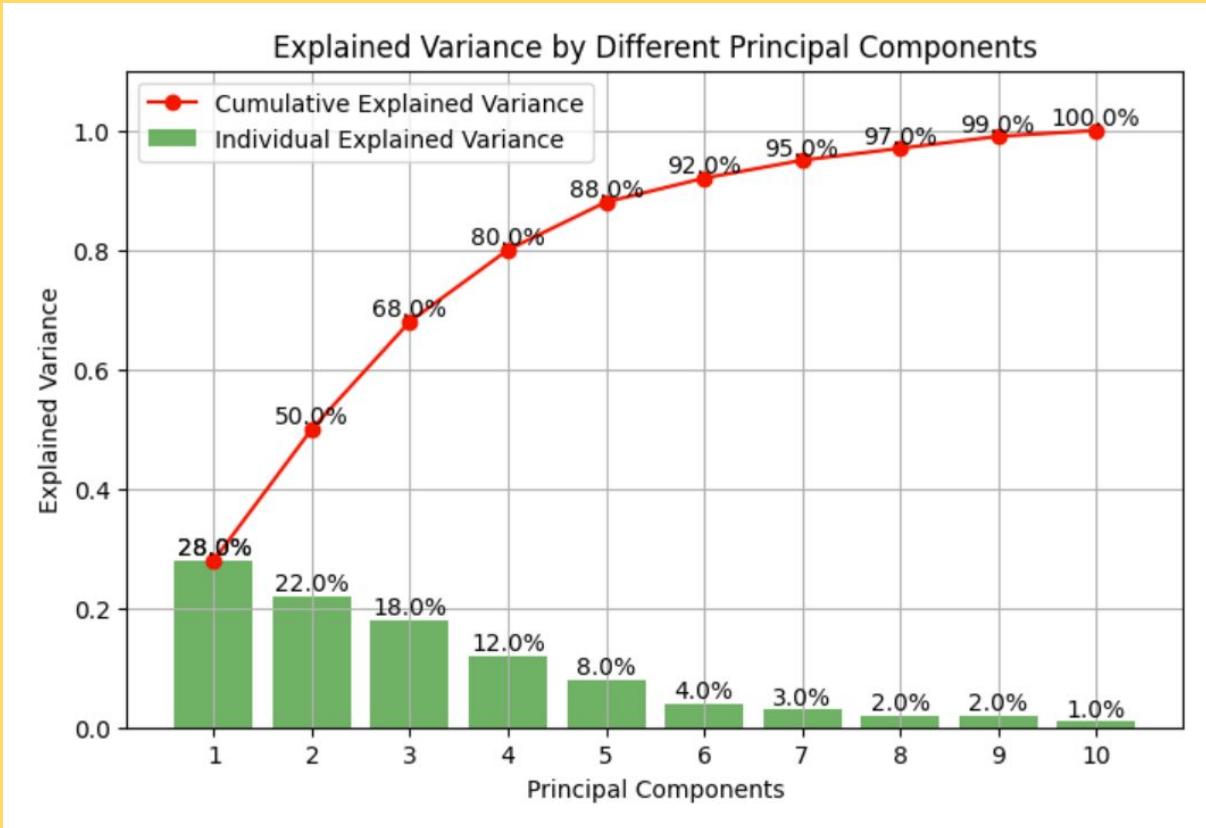
PCA
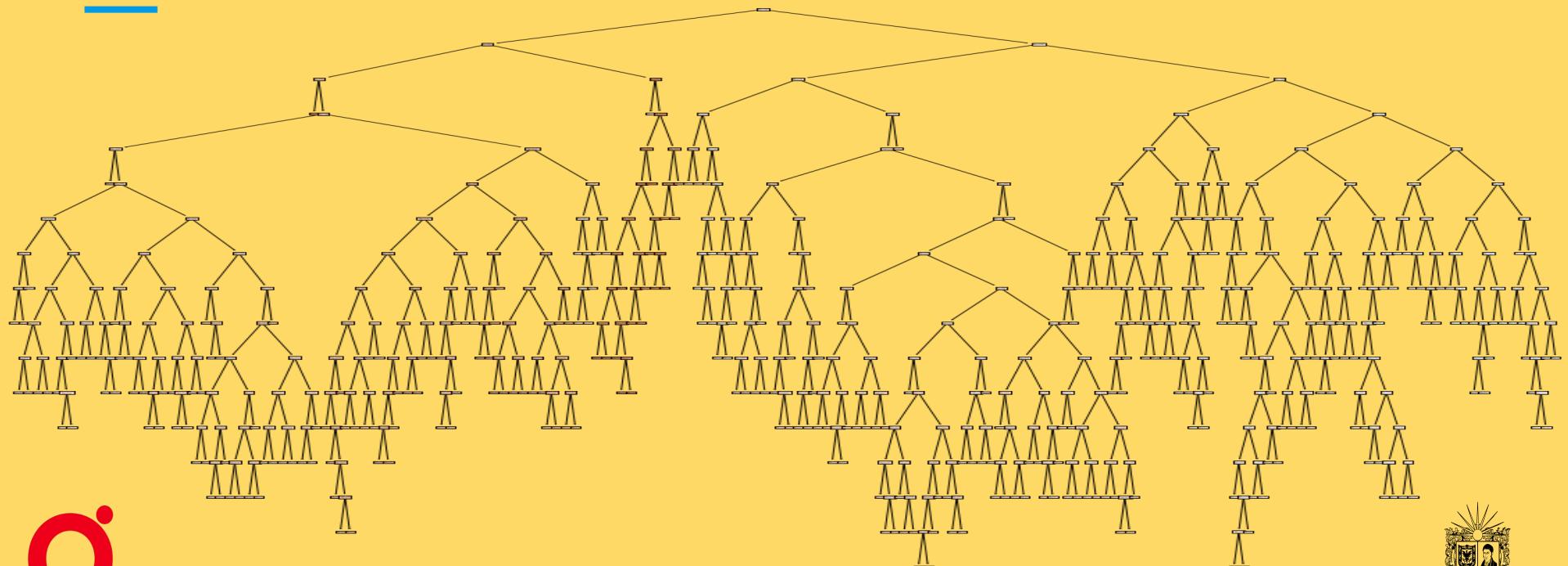
Cuenca, Lozada, Mosquera, & Ortiz

Classification Model
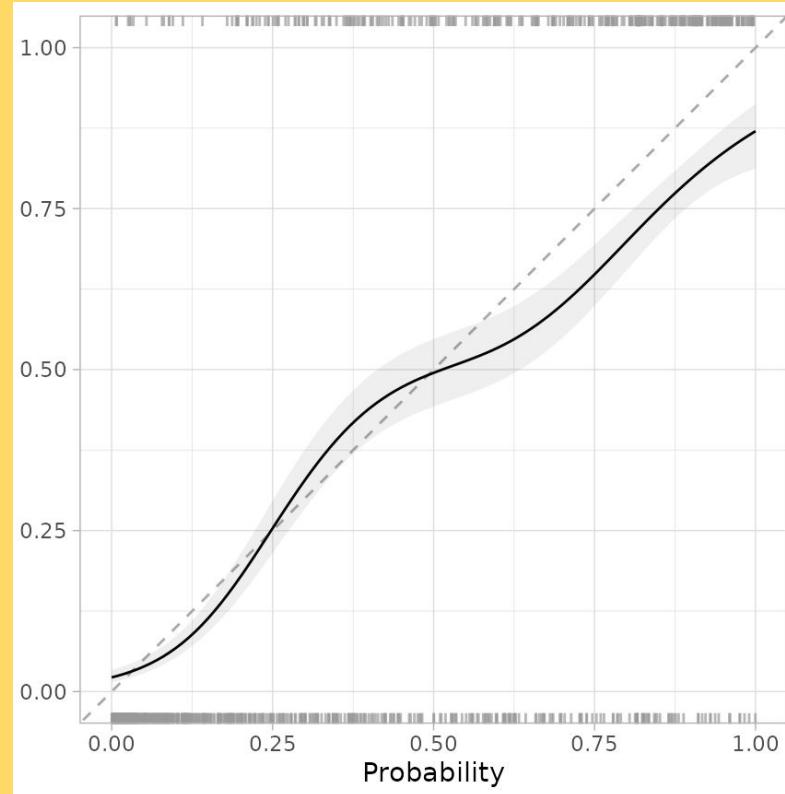
Cuenca, Lozada, Mosquera, & Ortiz

Probability Calibration
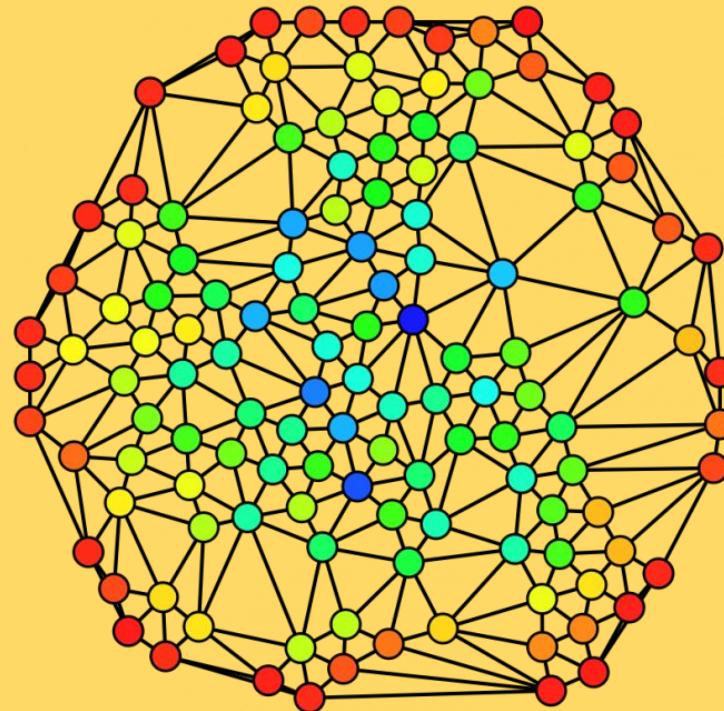
Cuenca, Lozada, Mosquera, & Ortiz

Systemic Simulation

Cuenca, Lozada, Mosquera, & Ortiz

Results
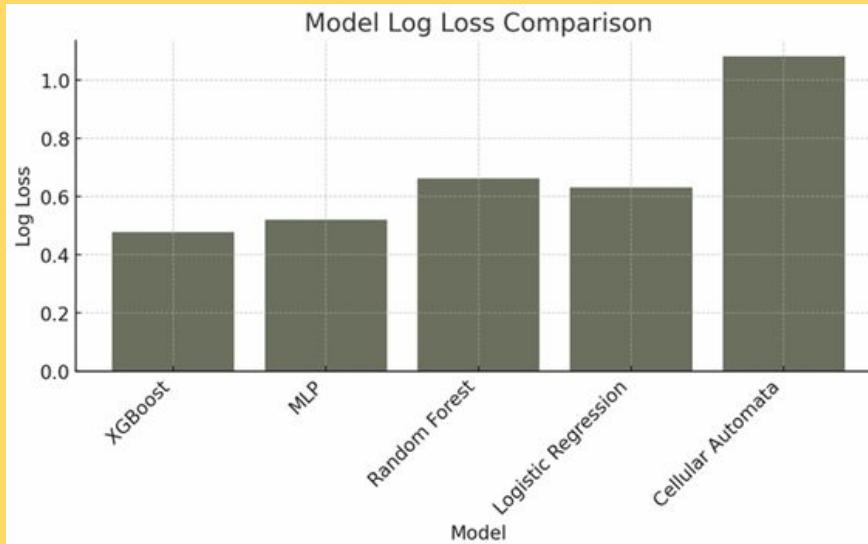
Model	Log Loss	Accuracy
MLP	0.5320	0.7948
XGBoost	0.5327	0.7925
Random Forest	0.6080	0.7853
Logistic Regression	0.6841	0.7511

Model	Before	After	Improvement
Random Forest	0.6632	0.2589	+0.4044
XGBoost	0.4788	0.2378	+0.2409
MLP	0.5208	0.4143	+0.1064
Logistic Regression	0.6311	0.8668	-0.2358

Cuenca, Lozada, Mosquera, & Ortiz

Discussion and Conclusions

Cuenca, Lozada, Mosquera, & Ortiz

END