Otto Group Product Classification

By :Jeison Cuenca, María Alejandra Ortiz Sánchez, Juan Diego Lozada, Juan Pablo Mosquera Marin

INTRODUCTION

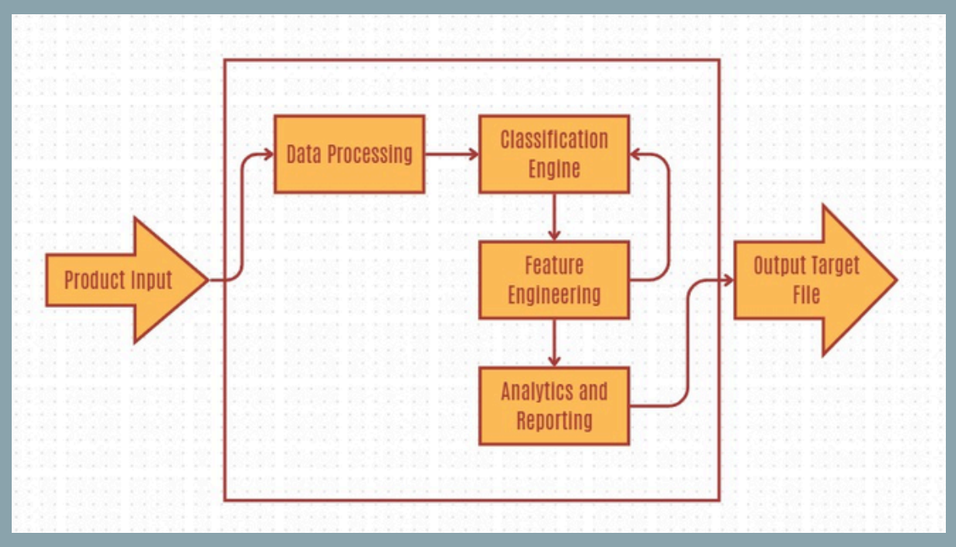
The rapid growth of e-commerce challenges product classification and organization. The Otto Group Challenge focuses on classifying over 200,000 products using machine learning. While advanced models improve accuracy, they often suffer from instability and overfitting. This work highlights the need for adaptive, self-monitoring systems to achieve robust and consistent classification.

GOALS

- Classifying products efficiently and accurately.
- Maintaining stability and interpretability across varying data conditions.
- Incorporating feedback loops for continuous adaptation and improvement.

PROPOSED SOLUTION

A modular and adaptive architecture integrates four subsystems:



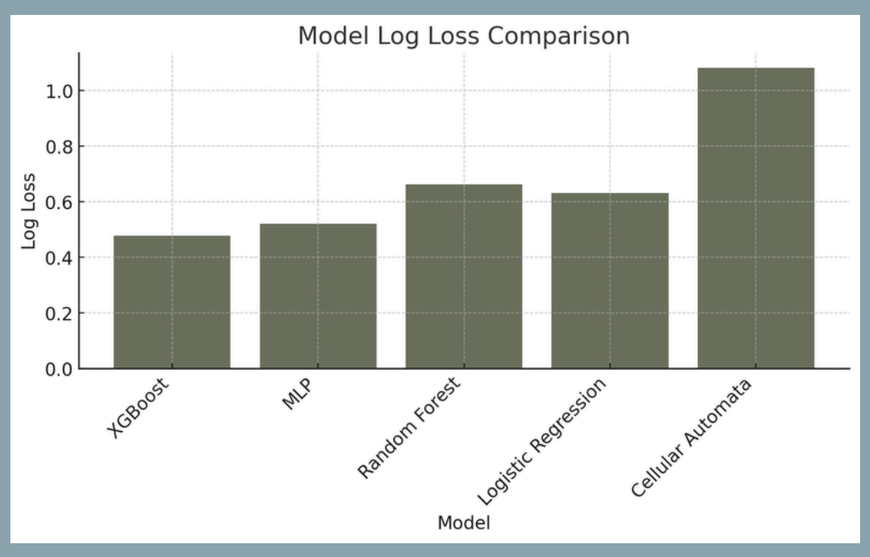
EXPERIMENTS

- Train/validation split
- Models tested (XGBoost, MLP, RF, LR, CA model)
- Number of features, dataset size
- PCA test
- Calibration test
- Ensemble test

Class	Count
Class_2	16122
Class_6	14135
Class_8	8464
Class_3	8004
Class_9	4955
Class_7	2839
Class_5	2739
Class_4	2691
Class_1	1929

RESULTS

The next graphic show the values about log loss using each model and celular Automata, where XGBoost has the minor value



Metric	Value
Internal validation log-loss	≈ 0.50
Kaggle leaderboard log-loss	**1.0877**
Performance gap	$\approx +0.58$

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

CONCLUSIONS

- The experimental pipeline successfully enabled systematic testing of multiple models.
- XGBoost and MLP provided strong baselines, and calibration further boosted performance
- Overall, the project sets a solid foundation for continued optimization and integration of adaptive models.

BIBLIOGRAPHY

- Kaggle Otto Group Product Classification Challenge (2015).
- Pedregosa et al., Scikit-learn: Machine Learning in Python (2011).