HYBRID DATA-DRIVEN FORECASTING SYSTEM FOR REGIONAL CHOCOLATE SALES: AN ENGINEERING APPROACH FOR OPERATIONAL DECISION SUPPORT

Samuel Aljure Bernal – 20202020111; Carlos Alberto Barriga Gámez – 20222020179; David Santiago Aldana Gonzalez – 20222020158; Juan Diego Alvarez Cristancho – 20221020076

INTRODUCTION

The core problem is achieving accurate regional chocolate sales forecasting amidst complex, non-linear market dynamics. Previous solutions rely on established quality frameworks like ISO 9001 and CMMI to ensure process consistency. The main challenge is mitigating instability and chaos, particularly the hypersensitivity of the model to fluctuations in the critical "Time in Region" feature

GOAL

The goal is to design and validate a hybrid Python/R system for reliable regional chocolate sales forecasting,. The expected product is an integrated pipeline that converts complex data into actionable and reproducible predictions

PROPOSED SOLUTION

The forecasting solution employs a hybrid three-layer architecture (Data Processing, Training, Presentation). Python handles data preparation, while R houses the modeling logic,. Key technical considerations include fixing random seeds for reproducibility and utilizing ensemble learning for high predictive accuracy and robustness
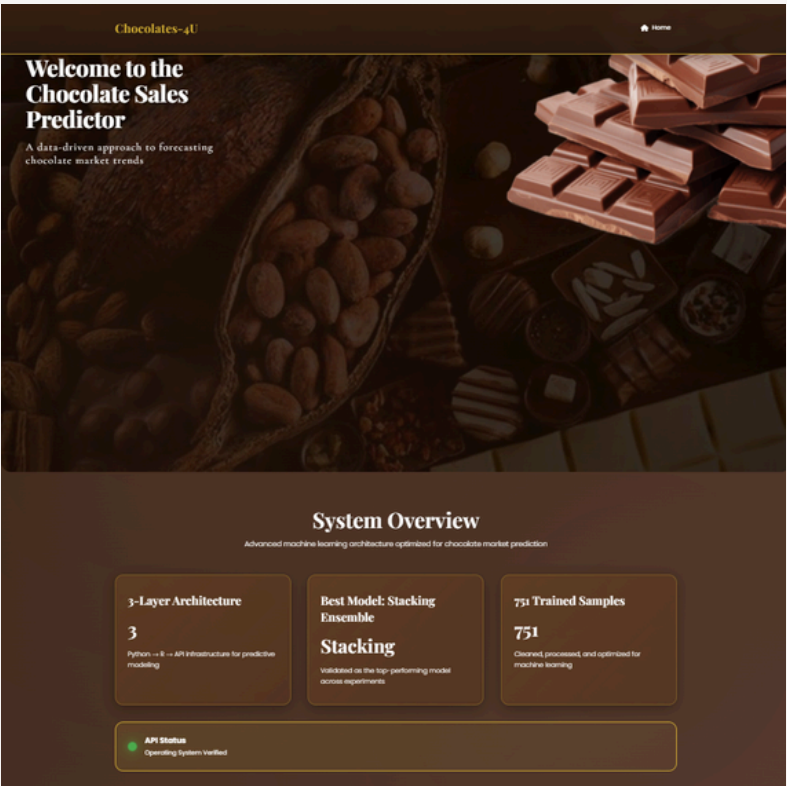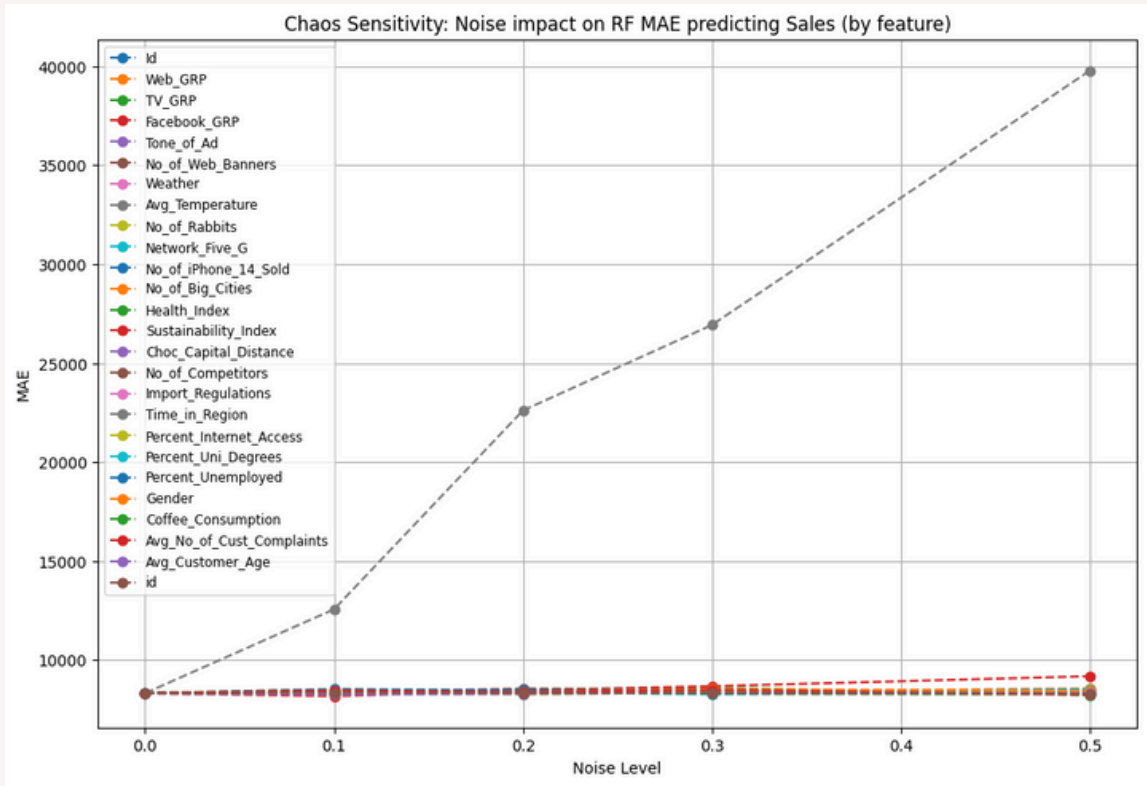
EXPERIMENTS

The solution's efficacy was demonstrated through predictive performance and robustness experiments. The predictive test compared four regression models, confirming the Stacking Ensemble achieved the lowest Mean Absolute Error. The chaos analysis injected noise into features, revealing that the Time in Region variable caused exponential instability in the Mean Absolute Error

RESULTS

The Stacking Ensemble demonstrated superior predictive performance, achieving the lowest Mean Absolute Error (MAE) and substantially reducing error compared to the Linear Regression baseline.

Model	MAE
Linear Regression	11.998
Random Forest	8.255
XGBoost	5.107
Stacking Ensemble	**4.017**

Strength: The hybrid system offers high accuracy and operational reliability by aligning with quality standards like ISO 9001 and CMMI. Weakness: Robustness experiments revealed chaotic instability: the Time in Region feature caused exponential MAE degradation when noise was injected, identifying it as a high-risk operational variable

CONCLUSIONS

The core goal was achieved by designing a hybrid Python/R predictive system that provides accurate forecasts. The system answered the research question using a Stacking Ensemble model, substantially reducing prediction error. The process structure maintained operational reliability by aligning with established engineering and quality-management principles

BIBLIOGRAPHY

Asuero, A. G., Sayago, A., & González, A. G. (2006). The Correlation Coefficient: An Overview. Critical Reviews in Analytical Chemistry, 36(1), 41–59. https://doi.org/10.1080/10408340500526766

Fonseca, L. and Domingues, P. (2017). ISO 9001: 2015 edition-management, quality and value. International Journal for Quality Research, Vol. 11, No. 1, pp. 149–158