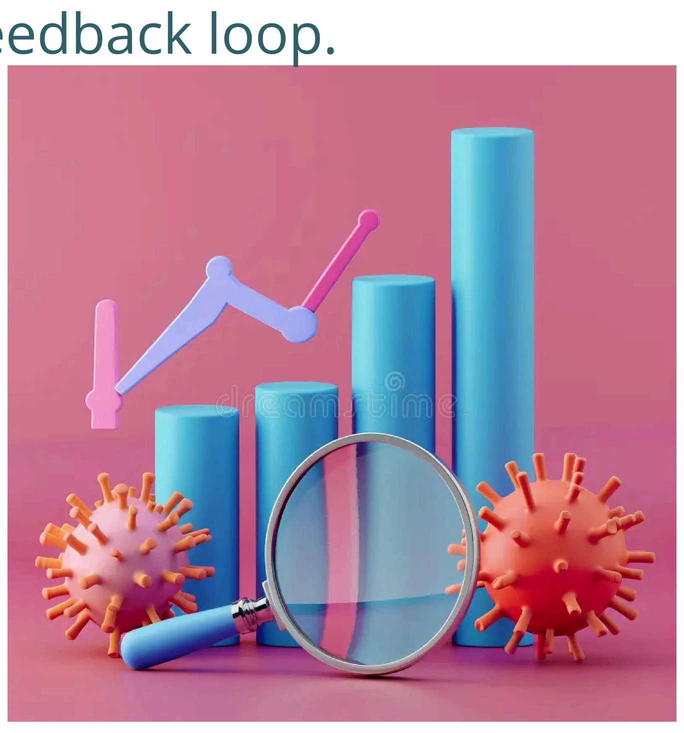
HYBRID ARCHITECTURE FOR EPIDEMIC FORECASTING: INTEGRATING DATA-DRIVEN ML AND SPATIAL CELLULAR AUTOMATA

DANIEL CASTRO, SANTIAGO VARGAS, DAVID SANCHEZ, DILAN TRIANA

UNIVERSIDAD DISTRITAL
FRANCISCO JOSÉ DE CALDAS

INTRODUCTION

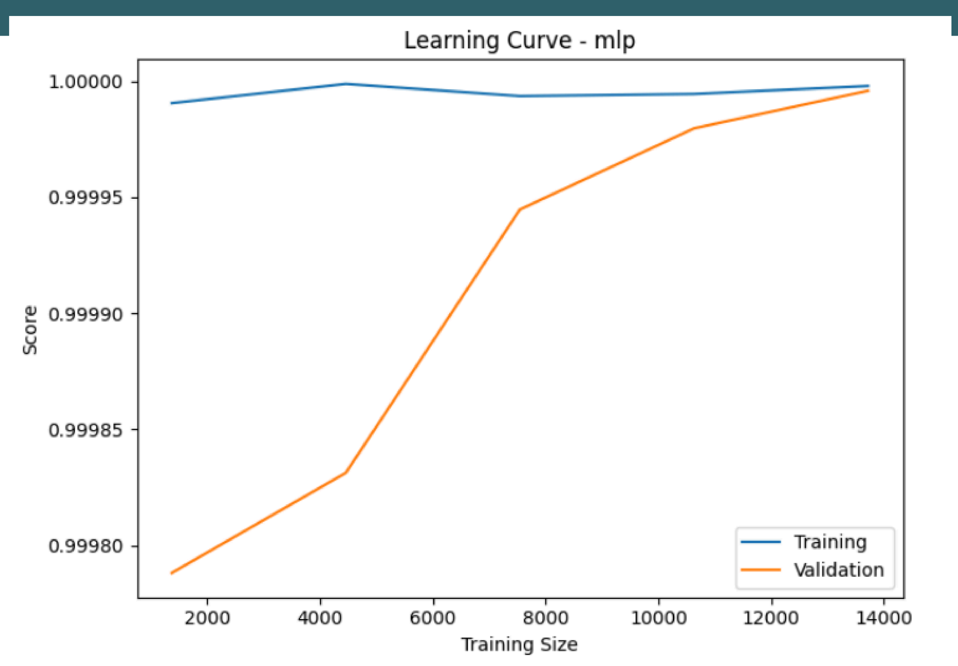
This project addresses the global need for transparent forecasting by implementing a modular system grounded in ***Systems Engineering principles.*** The solution integrates a data-driven pipeline ***(Scenario 1)*** using Machine Learning and an event-based simulation using ***Cellular Automata (Scenario 2).*** The system processes global data to predict cumulative COVID-19 cases, ensuring ***reproducibility, scalability, and stability*** through a validated feedback loop.



GOAL

To develop and validate a ***robust forecasting architecture*** that combines statistical learning ***(MLP, Random Forest)*** with spatial dynamics modeling. The primary goal is to ensure prediction stability under perturbations and demonstrate the technical feasibility of a ***feedback-driven pipeline*** for epidemiological modeling.

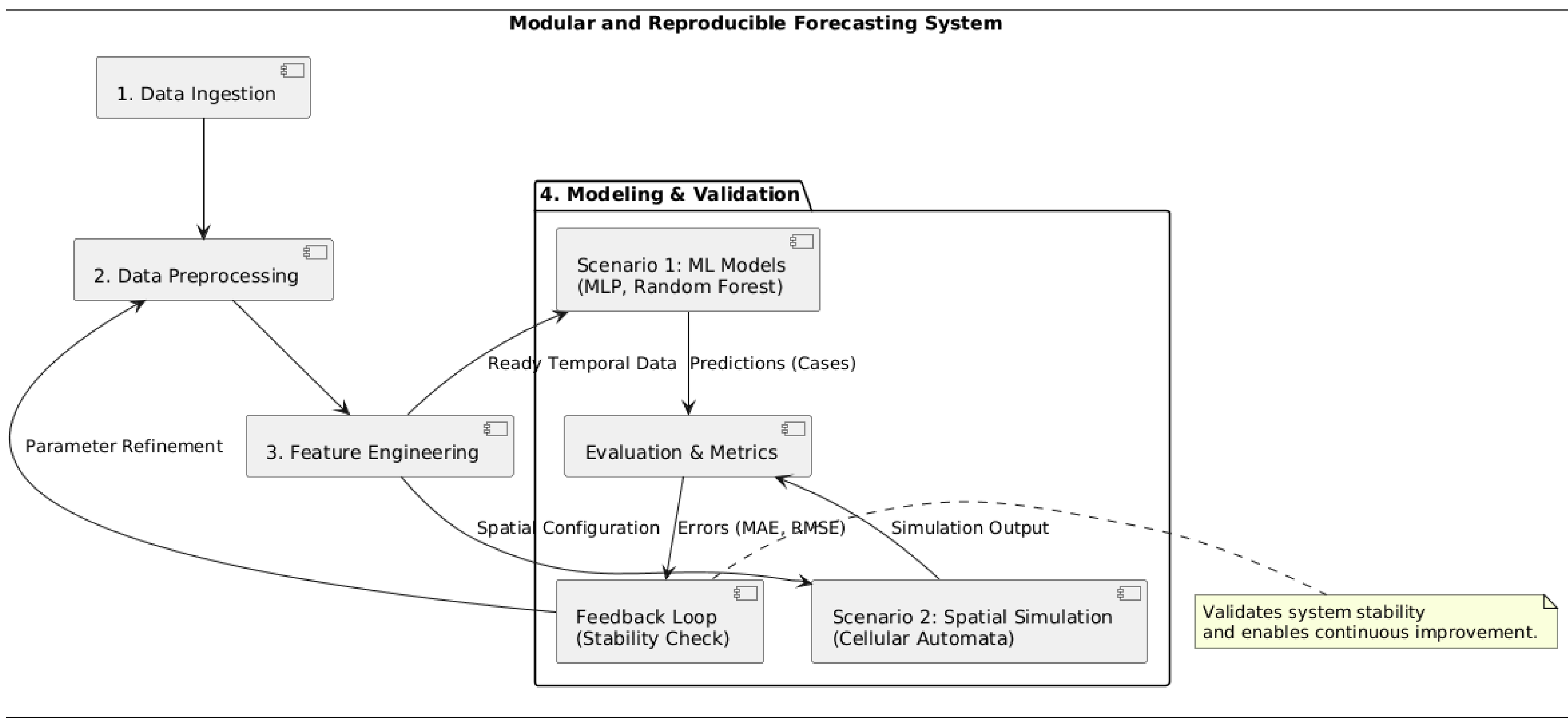
RESULTS



Quantitative Metrics:
MLP Model: MAE = 1.16, RMSE = 10.56 (Most Accurate).
Random Forest: MAE = 3.04, RMSE = 75.91 (High Variance).

Stability Analysis:
Robustness: The ML models showed high resilience to ***±5% input perturbations.***
Feedback Loop: Demonstrated ***convergence*** without oscillatory instability.
Cellular Automata: Exhibited ***high sensitivity*** and emergent spatial patterns, contrasting with the numerical stability of the ML models.

PROPOSED SOLUTION



EXPERIMENTS

Scenario 1: Machine Learning Performance Three models were evaluated using RMSE and MAE metrics. The ***Multi-Layer Perceptron (MLP)*** achieved the best balance of accuracy and stability, significantly outperforming ***Random Forest*** and Linear Regression.

Scenario 2: Spatial Simulation A 25x25 grid Cellular Automata was simulated. The experiment tested the system's sensitivity to initial conditions and infection probabilities, observing ***rapid convergence*** to homogeneous states, contrasting with the gradual trends of the ML models.

CONCLUSIONS

The project successfully validated a ***hybrid forecasting system.*** Results demonstrate that ***Machine Learning models (MLP)*** provide numerical stability and precision for trend forecasting, while ***Cellular Automata*** offer insights into non-linear spatial dynamics and emergence the ***modular architecture*** proved effective for handling real-world data, ensuring reproducibility and enabling continuous refinement through ***feedback loops.***

Kaggle. (2020). COVID-19 Global Forecasting (Week 1–5) [Dataset]. Kaggle.
Scikit-learn Developers. (2023). Scikit-learn: Machine learning in Python [Software documentation]. Scikit-learn.
Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (pp. 785–794). ACM.