GitHub

Supply Chain Optimization for Health Commodity Shipment and Pricing

Princy Sutariya

Project Proposal

Objective: Using predictive modeling to optimize costs and efficiency in supply chains for HIV health commodities.

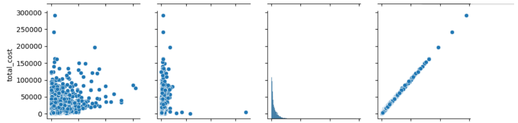
Dataset: Dataset on ARV and HIV lab shipments includes 10,324 rows and 33 columns, detailing costs, logistics, and distribution for developing countries.

Key parameters cover geography, cost details (freight, insurance, unit price), and logistics management (vendor, shipment mode, weight) to enhance supply chain efficiency.

Research Question

1. **How can supply chain costs be minimized for HIV-related health commodities?**
2. **What are the primary causes of delivery delays, and how can they be predicted?**
3. **How can pricing be optimized to improve cost efficiency across different regions?**

Methods



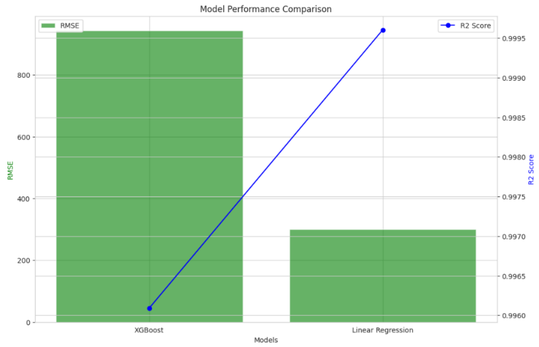
Data Collection & Preprocessing: Sourced data from Data.gov, cleaned and handled missing values, and transformed datetime columns into numerical features for analysis.

Feature Engineering: Applied one-hot encoding for categorical variables and derived new features like delivery_duration to capture delivery delays.Exploratory Data Analysis (EDA): Conducted statistical analysis, visualized correlations, and addressed outliers to understand relationships among key variables.Data Scaling: Used scaling techniques to standardize features for consistent model input.Predictive Modeling: Built and compared XGBoost and Linear Regression models to predict freight_cost_(usd), with XGBoost outperforming due to hyperparameter tuning using Optuna.Model Evaluation: Evaluated models using RMSE, MAE, and R², confirming XGBoost's superior predictive power.

Results

Cost Prediction:
Libraries & Models Used:
XGBoost: XGBoost and Linear Regressor
Evaluation Metrics:
Results:

- XGBoost RMSE =943.64
- XGBoost R^2 =0.996
- LR RMSE = 300.78
- LR R^2 = 0.99960



Conclusions

- While Linear Regression produced a lower RMSE in this scenario, XGBoost's ability to handle complex, non-linear relationships makes it highly suitable for scenarios where supply chain variables exhibit complex interactions.
- The highR 2 scores of both models (close to 1) indicate that the models effectively explain the variance in the freight cost data.
- Both models are viable for predicting freight costs, but for more advanced and scalable applications, XGBoost's robustness makes it a strong candidate.