Retail Price Optimization - Project Report

Project Objective:

To optimize product pricing in retail by analyzing historical sales, cost, and demand using machine learning models. The goal is to maximize sales and maintain profitability.

Dataset Overview:

- Source: retail_price.csv

- Records: Product sales, cost, category, competition, lag prices

- Features: 30 total

- Target: unit_price

Data Preprocessing:

- Dropped null values

- Removed non-numeric identifiers: product_id, category, month_year

- Correlation matrix plotted

- Feature selection completed

Modeling Approach:

- Models Used:

  - Linear Regression

  - Random Forest Regressor

- Training/Test Split: 80/20

- Target Variable: unit_price

Evaluation Metrics:

- RMSE (Linear): ~0.65

- R2 Score (Linear): ~0.72

- RMSE (Random Forest): ~0.31

- R2 Score (Random Forest): ~0.93

Feature Importance (Random Forest):

- Most influential features:

  - lag_price

  - fp3

  - qty

  - volume

  - total_price

Visual Insights:

- Correlation Heatmap

- Actual vs Predicted Plot

- Feature Importance Plot

Conclusion:

- Random Forest performed best with high accuracy

- Pricing is influenced heavily by lag_price and volume

- Model saved for deployment as rf_model.pkl

Future Scope:

- Deploy using Streamlit web app

- Integrate real-time pricing APIs

- Extend to other product categories

- Add dynamic competitor price tracking