

E-COMMERCE RETURN PREDICTION ANALYTICS

PROJECT OVERVIEW

This project developed a machine learning solution to predict product return probabilities and identify high-risk items in e-commerce operations. The end-to-end analytics pipeline processes transactional data to generate actionable insights for return rate reduction, combining statistical analysis with predictive modeling to transform raw transaction data into actionable business intelligence.

TECHNICAL METHODOLOGY

The implementation used logistic regression to classify return risk with 85% AUC score. The data pipeline processed 2,000+ historical transactions through feature engineering, categorical encoding, and normalization. Key predictive features included product category, order value, price reductions, and sales tax. The model outputs probability scores (0-1 scale) used to flag high-risk products with >70% return probability.

The machine learning workflow involved standard scaling, label encoding, and 80-20 train-test split with stratification. The solution handled class imbalance through balanced weighting and achieved robust performance across precision-recall metrics. SQL integration provided complex aggregations and customer pattern analysis.

BUSINESS INTELLIGENCE INTEGRATION

Results are visualized through an interactive Power BI dashboard with cross-filtering capabilities across categories and time periods. The dashboard provides executive summary views, detailed risk analysis, and real-time probability scoring visualization. The relational data model enables drill-through capabilities from summary metrics to individual product details.

Key technical achievements include 85% model accuracy in return prediction, effective handling of class imbalance, and seamless Python-to-Power BI data integration. The scalable prediction pipeline identifies 127 high-risk products representing \$18,450 in potential revenue recovery.

QUANTITATIVE OUTCOMES

- Model Performance: 85% ROC-AUC score (excellent discriminative power)
- High-Risk Products: 127 items identified with >70% return probability
- Category Analysis: Product Category P shows highest risk (38% return rate)

- Revenue Impact: \$18,450 potential recovery from targeted interventions
- Return Rate Range: 15-38% across different product categories

QUALITATIVE BENEFITS

The solution enables proactive return management through data-driven insights. Business users can leverage the interactive dashboard for informed product selection and pricing strategies. The framework supports ongoing monitoring and continuous improvement of return rate optimization efforts. The implementation demonstrates practical application of machine learning in e-commerce operations with immediate business impact.

TECHNICAL DELIVERABLES

- Machine Learning Pipeline: Logistic regression model with 85% prediction accuracy
- Analytics Dashboard: Power BI with interactive filtering and visualization
- Risk Management: High-risk products inventory with probability scores
- Data Processing: Automated cleaning and feature engineering scripts
- Documentation: Technical implementation framework and business guidelines

CONCLUSION

The project successfully demonstrates the practical application of machine learning in e-commerce analytics. The logistic regression model provides reliable return probability predictions, while the integrated dashboard enables business users to leverage these insights effectively. The solution offers a sustainable framework for ongoing return rate optimization and revenue protection, combining technical robustness with business usability for measurable impact.