MACHINE LEARNING MODEL EVALUATION REPORT PREDICTING CUSTOMER CHURN

Overview

This report summarizes the performance evaluation of the machine learning model developed for Predicting Customer Churn. The goal of this project was to identify customers likely to churn based on their session behaviors and financial data.

Dataset

The dataset already in used was further to 30 columns, including customer demographics, session details, financial metrics, and product information. Data preprocessing involved several key steps**:**

- Datetime Conversion: SessionStart and SessionEnd columns were converted to datetime format.
- Categorical Encoding: Columns like FullName, City, Product, CampaignSchema, OrderReturn, and ReturnReason were encoded using LabelEncoder.
- **Feature Engineering:**
    - ❖ ProfitMargin was calculated as Price - Cost.
    - ❖ TotalRevenue was derived from Price * Quantity.
    - ❖ Churn was defined as customers with no session in the last 6 months from the latest date (SessionEnd < latest_date - 180 days).
    - ❖ DaysSinceLastSession was created to track the number of days since the last customer interaction.
- Feature Selection: Mutual Information scores were used to select the top 15 features for modeling, including DaysSinceLastSession, ProfitMargin, PaymentMethod, and CreditScore.
- Data Balancing: The dataset was balanced to 1810 samples per class using RandomOverSampler, with added Gaussian noise to reduce overfitting.

Model Description

The following machine learning models were trained and evaluated:

- Random Forest (RF)
- Logistic Regression (LR)
- Decision Tree (DT)
- Extra Tree (ET)

- XGBoost

Performance Metrics

Each model was evaluated using these performance metrics:

- Accuracy: Measures overall correctness.
- Precision: Measures how many positive predictions were correct.
- Recall: Measures how many actual positives were correctly identified.
- F1 Score: Harmonic mean of precision and recall.
- ROC AUC: Measures the model's ability to distinguish between classes.
- Confusion Matrix: Visualizes true positives, true negatives, false positives, and false negatives.

Model	Accuracy	Precision	Recall	F1 Score	ROC AUC
Random Forest (RF)	100%	100%	100%	100%	1.000
Logistic Regression (LR)	98.2%	100%	96.6%	98.3%	0.9998
Decision Tree (DT)	100%	100%	100%	100%	1.000
Extra Tree (ET)	98.6%	100%	97.4%	98.7%	0.987
XGBoost	100%	100%	100%	100%	1.000

Findings and Insights

The best-performing models were Random Forest, Decision Tree, and XGBoost, all achieving perfect scores across every metric. Logistic Regression and Extra Tree also performed exceptionally well but had minor drops in recall and F1 scores.

Key insights:

- Random Forest, Decision Tree, and XGBoost achieved 100% performance, likely due to balanced data and added noise.
- Logistic Regression showed strong precision but slightly lower recall, suggesting it may miss some churn cases.
- DaysSinceLastSession, ProfitMargin, and PaymentMethod emerged as top features based on Mutual Information scores.

Future Improvements

- Hyperparameter Tuning: Further tuning may prevent overfitting in high-performing models.
- Feature Engineering: Explore interaction terms or temporal trends.
- Ensemble Stacking: Combine strengths of multiple models for enhanced performance.
- Explainability: Implement SHAP or LIME for feature importance insights.

Conclusion

The Random Forest, Decision Tree, and XGBoost models provided perfect classification performance, making them strong candidates for deployment. Future efforts should focus on refining the data pipeline and investigating generalizability on unseen data.