Telco Customer Churn Analysis Report

Date: August 13, 2025

This report provides an analysis of customer churn prediction models, offering actionable insights for retention strategies.

For inquiries, contact via project communication channels.

Executive Summary

This report evaluates three machine learning modelsLogistic Regression, Random Forest, and XGBoostfor predicting customer churn in the Telco dataset. Key findings include XGBoost's superior balance (AUC: 0.842) and Logistic Regression's high recall (82.6%) for risk screening. Recommendations focus on a hybrid deployment strategy to optimize retention efforts, potentially reducing churn by 20-30%.

- **Dataset**: Telco Customer Churn, focusing on retention vs. churn.
- **Models**: Logistic Regression (high recall), Random Forest (high precision), XGBoost (balanced).
- **Business Impact**: Optimize retention budgets, protect revenue, and enhance operational efficiency.

Model Performance Comparison

The models were evaluated on validation data using Accuracy, Precision, Recall, F1-Score, and AUC-ROC.

Table 1: Model Performance Metrics

Model	Accuracy	Precision	Recall	F1-Score	AUC-ROC
Logistic Regression	75.2%	51.8%	82.6%	63.5%	0.812
Random Forest	79.7%	66.9%	52.3%	58.9%	0.830
XGBoost	80.1%	62.4%	70.2%	66.1%	0.842

Insights:

· Logistic Regression excels in capturing churners, ideal for broad screening.

· Random Forest minimizes false positives, suitable for high-value customers.

· XGBoost offers the best overall predictive power.

Visual Analysis: Confusion Matrices

Confusion matrices illustrate model predictions for "Retained" vs. "Churned" customers. Due to LaTeX limitations, detailed matrix visuals are described:

- **Logistic Regression**: High true positives, but more false positives.
- **Random Forest**: Balanced predictions, fewer false positives.
- **XGBoost**: Strongest balance, optimal for general use.

Note: In the final PDF, high-resolution confusion matrix images would be embedded using the `graphicx` package. For now, assume blue-themed heatmaps (Blues cmap) with bold labels and no grid lines for a modern aesthetic.

Strategic Recommendations

Model-Specific Applications

Logistic Regression (High Recall Strategy)

- **Use Case**: Early warning system for customer retention.
- **Metrics**: 82.6% Recall, 51.8% Precision.
- **Actions**:
  - Implement broad retention campaigns.
  - Use for initial customer screening.
  - Prioritize capturing all potential churners.
  - Budget for higher false positive rate.

Random Forest (High Precision Strategy)

- **Use Case**: Premium customer retention programs.
- **Metrics**: 79.7% Accuracy, 66.9% Precision.
- **Actions**:
  - Target high-value customers.
  - Deploy expensive retention offers.
  - Focus on quality over quantity.
  - Optimize resource allocation.

XGBoost (Balanced Strategy)

- **Use Case**: General retention strategy.
- **Metrics**: 0.842 AUC, 70.2% Recall.
- **Actions**:
  - Implement as default model.
  - Balance resource allocation.
  - Use for automated decision making.
  - Combine with business rules.

Implementation Strategy

1. **Phase 1**: Deploy Logistic Regression for broad screening.
2. **Phase 2**: Add Random Forest for premium segments.
3. **Phase 3**: Implement XGBoost for automated decisions.

Expected Business Impact

- **Cost Reduction**: Optimize retention budgets, reduce unnecessary interventions.
- **Revenue Protection**: Prevent high-value churn, increase retention rates.
- **Operational Efficiency**: Automate risk assessment, streamline processes.

Next Steps & Considerations

- **Deployment**: Integrate models into CRM via APIs.
- **Monitoring**: Track model drift quarterly; retrain as needed.
- **Ethics**: Ensure predictions are fair across demographics.
- **Enhancements**: Explore deep learning or ensemble methods.

For a full visual PDF, this document can be enhanced with embedded confusion matrix images and additional charts using the `graphicx` package.