Customer Churn Prediction Report

1. Introduction

Customer churn refers to the phenomenon in which clients discontinue their engagement with a company's services. In the telecommunications industry, customer churn directly impacts revenue and increases customer acquisition costs.

This project aimed to develop a predictive system that identifies customers at high risk of churn by analyzing their service usage patterns and account characteristics. Early identification enables the business to implement targeted retention strategies, thereby reducing churn and improving customer lifetime value.

2. Methodology

Data Exploration and Cleaning

• Conducted exploratory data analysis to uncover patterns correlated with churn.

• Addressed missing data, encoded categorical variables, and scaled numerical features for uniformity.

Model Development

Three machine learning models were developed and evaluated:

• Logistic Regression – Simple, interpretable baseline model.

• Decision Tree – Tree-based model capturing nonlinear relationships.

• Random Forest – Ensemble model offering higher accuracy and robustness.

Handling Class Imbalance

The dataset exhibited significant class imbalance, with far fewer churn instances. To mitigate this:

• Class Weighting: Increased the model’s attention to the minority churn class.

• SMOTE (Synthetic Minority Oversampling Technique):

Augmented training data with synthetic churn samples.

Evaluation Metrics

Models were assessed using the following metrics:

• Accuracy: Overall correctness of the model.

• Recall: Effectiveness at identifying actual churners.

• F1-Score: Harmonic mean of precision and recall, balancing false positives and false negatives.

3. Key Findings

Important Features Influencing Churn

The following factors consistently emerged as significant predictors of churn across all models:

• Contract Type: Month-to-month contracts were strongly associated with high churn rates.

• Tenure: Newer customers (shorter tenure) were significantly more likely to churn.

• Internet Service: Users of fiber optic services showed elevated churn, potentially due to service dissatisfaction or cost.

• Monthly Charges: Higher monthly charges increased the likelihood of churn.

• Payment Method: Electronic check users exhibited higher churn rates, suggesting financial constraints or dissatisfaction.

These findings support established business intuition—churn is closely tied to contract flexibility, billing amount, and perceived value of service.

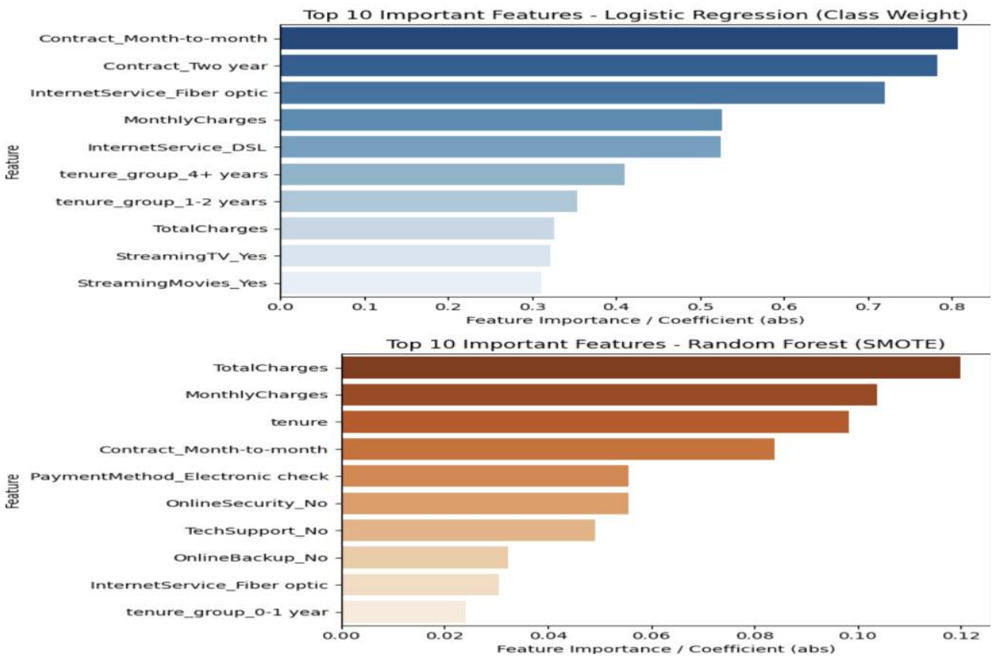
4. Model Performance Summary

• **Logistic Regression** with class weights achieved the highest recall (0.79), making it effective for catching the majority of churners.

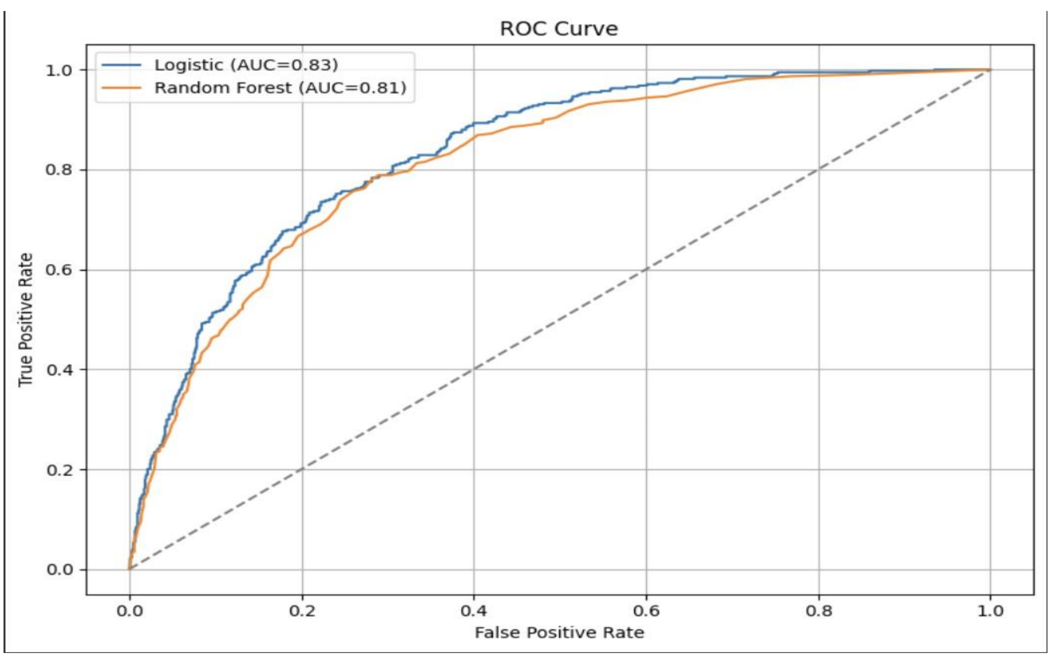
• Random Forest combined with SMOTE achieved the best **balance between recall and precision**, making it the most robust option for real-world deployment.

• **Decision Trees**, while interpretable and intuitive, underperformed slightly in terms of recall and F1-score

Top Features That Predict Churn



Model Comparison Table or ROC Curve



Model Explainability with LIME

To ensure transparency and trust in our machine learning model's predictions, we integrated **LIME (Local Interpretable Model-Agnostic Explanations)**. LIME explains individual predictions by approximating the complex model with a simpler, interpretable one (e.g., linear regression) around the specific instance.

Insights from LIME analysis on a sample prediction:

- **Positive Contributors (reduce churn risk)**:
- **Long Tenure**: Customers with tenure > 56 months are more likely to stay.
- **Contract Type**: Long-term contracts (especially two-year agreements) indicate greater retention.
- **High Total Charges**: Suggests sustained engagement and customer loyalty.

Negative Contributors (increase churn risk):

- **Month-to-Month Contracts**: Strongly correlated with higher churn.
- **Fiber Optic Internet Users**: Slightly more prone to churn, potentially due to pricing or service issues.
- **Low Tenure or Mid-Level Monthly Charges**: Associated with moderate churn likelihood.

These insights enable customer support and marketing teams to proactively design **retention strategies**—for example, offering incentives to newer users or customers on monthly plans.

Web Application Deployment

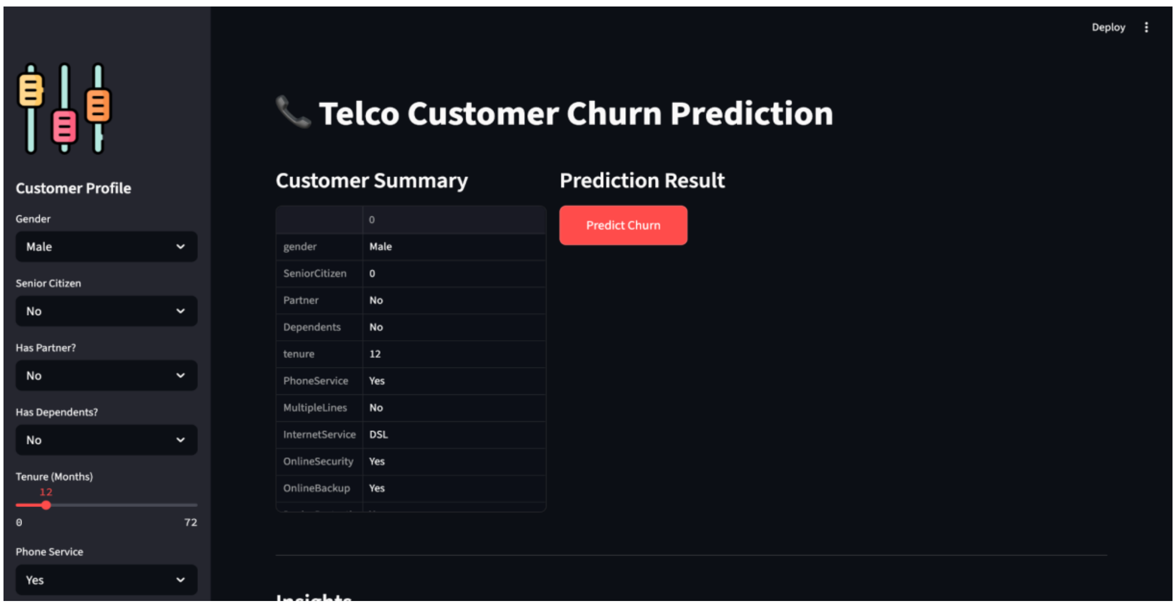
To make the churn prediction system accessible for non-technical users, we developed an interactive web application using **Streamlit**. The app allows non-technical users to input customer attributes and receive real-time churn predictions with visual insights.

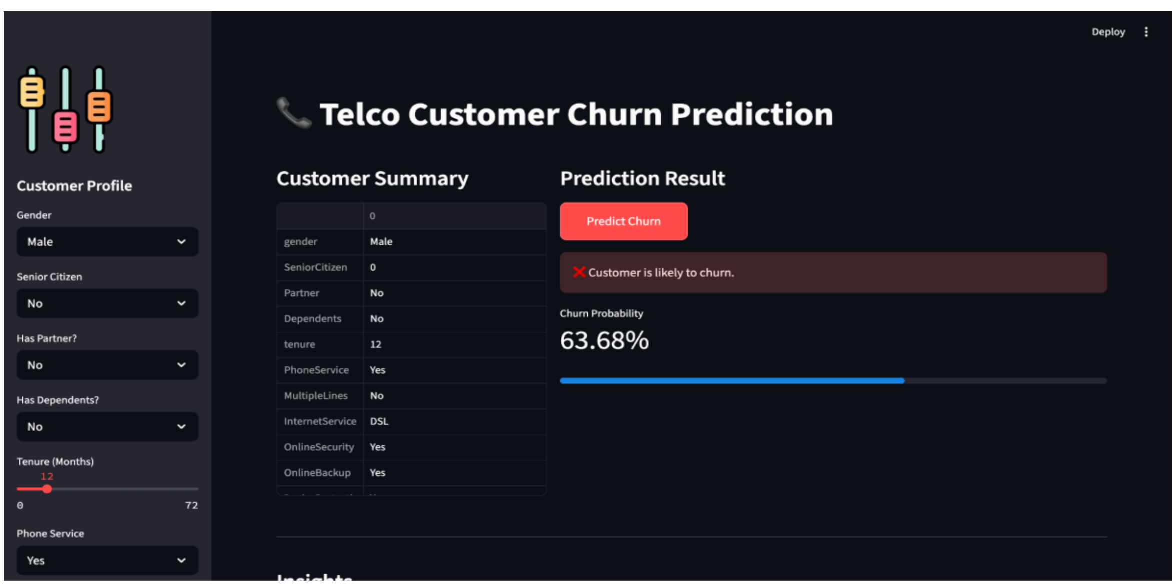
Key Functionalities:

- **Customer Input Interface**: Sidebar form to enter customer demographics and service details.
- **Live Prediction**: Displays churn result (Yes/No) with confidence score instantly.
- **Visual Insights**: Bar chart highlights top influencing factors for each prediction.

- **Interactive Scenario Testing**: Users can tweak inputs and observe how churn probability shifts in real time.
- **User-Friendly UI**: Clean, responsive layout with professional theming, metric highlights, and embedded charts (via Matplotlib/Seaborn).
- **Ready for Deployment**: Easily hostable on **Streamlit Cloud, Render, or Heroku**.

By bridging data science with user experience, this application empowers business users to **simulate churn scenarios, understand the "why" behind predictions**, and take informed action—**without needing direct access to the model code**.





Next Steps – What Could Come Next

1. **Integrate SHAP/LIME in Production**

While LIME was used for exploratory explainability, integrating SHAP or LIME directly into the app can offer **instance-level interpretability** to support decision-making.

2. **Deploy the App Publicly**

Hosting the Streamlit application on a cloud platform (e.g., **Streamlit Community Cloud**, **Render**, or **Heroku**) would allow wider usage by sales or retention teams.

3. **Enhance Data Pipeline**

Build an **automated pipeline** to continuously pull updated customer data from CRM systems and retrain models periodically to adapt to business changes.

4. **Model Monitoring & Drift Detection**

Implement monitoring to detect **concept drift** or a drop in prediction accuracy over time, which can be addressed through retraining or feature updates.

5. **Add Risk-Based Alerts**

Extend the app to notify internal teams when **high churn-risk customers** are predicted, enabling proactive intervention.

6. **A/B Testing of Retention Strategies**

Use the predictions to guide **targeted retention campaigns** and evaluate their effectiveness through **A/B testing** and uplift modeling.

7. **Multi-Model Comparison Dashboard**

Integrate a backend feature to allow **model versioning** and comparison (e.g., Random Forest vs Logistic Regression) to track performance evolution.

8. **Include Customer Feedback Loop**

Introduce a system to capture **actual outcomes** (e.g., whether the customer actually churned), enabling continuous **model evaluation and learning**.