📈 Customer Churn in Telecom: Data-Driven Insights for Customer Retention Strategies

📊 Project Overview

This project analyzes and predicts churn for ABC Telecommunication Company using data from the [SyriaTel Customer Churn` dataset from Kaggle]

By combining data-driven insights with machine learning models, this project provides a framework that enables the company to proactively engage with at-risk customers, enhance loyalty, and minimize revenue loss.

🧠 Business Problem

In 2024, ABC Telecommunication Company reported a churn rate of **26%**(approximately 1,815 customers), resulting in an estimated **$2.4 Million** revenue loss. The challenge is to identify churners early and develop retention strategies. **Project Goals**:

- Minimize financial losses caused by churn.
- Understand churn behavior through Exploratory Data Analysis.
- Build predictive models to identify customers at risk of leaving.
- Identify key churn drivers.

- Provide actionable recommendations to reduce churn and strengthen retention strategies.

🔈 Stakeholders

- **Marketing and Customer Retention Teams**: Design targeted campaigns and loyalty programs.
- **Product and Pricing Teams**: Optimize plans, packages, and pricing.
- **Customer Service Department**: Improve support quality for at-risk customers.
- **Executive Leadership**: Leverage predictive insights for strategic decision-making.

📁 Project Workflow

1. Business Understanding
2. Data Understanding
3. Data Preparation and Preprocessing

- Data Cleaning
- Exploratory Data Analysis(EDA)
- Feature Engineering
- Train-Test Split
- Handling Class Imbalance(SMOTEC)
- Feature Scaling

4. Modeling

- Logistic Regression
- Decision Tree Classifier
- Random Forest Classifier
- XGBoost Classifier
- Model Evaluation
  - Cross Validation
  - Accuracy, Precision, Recall, F1-Score, ROC-AUC
  - Hyperparameter Tuning
  - Feature Importance Analysis

5. Business Recommendation

⚙️ Tools Used

- Language: Python
- Libraries: pandas, numpy, matplotlib, seaborn, scipy, sklearn, imbalanced-learn, XGBoost, collections
- Environment: Jupyter Notebook(learn_env)

✅ Results

Best Model: Tuned XGBoost

- Accuracy: 96%
- Recall(Churners): 82%
- Precision(Churners): 88%
- F1-Score: 85%
- ROC-AUC: 90% XGBoost outperformed all other models by balancing false alarms and missed churners, making it the most reliable for retention strategies.

🔑 Key Influential Features

- `international_plan`
- `cs_calls_intl_plan`
- `high_day_usage`
- `customer_service_calls` . These highlight the main drivers of churn and areas where intervention can be most effective.

📝 Final Recommendation

1. The tuned **XGBoost** model provides strong predictive power(96% accuracy, 82% recall for churners), making it a reliable tool for **proactively identifying at-risk customers**.
2. Predictions are most useful when applied to **customer segments with high service calls, international plans, and heavy daytime usage**. However, since precision is lower than recall, the model may generate false alarms. In contexts where retention resources are very costly, predictions should be combined with business rules before action.
3. **Features-driven actions**

- **Target High-Risk Customers**

- Focus campaigns on customers with international plans, high usage, and frequent service calls.
- Offer personalized incentives such as discounts, loyalty rewards, or tailored plan adjustments.

- **Enhance Customer Service Experience**
  - Resolve issues quickly for customers making multiple service calls.
  - Train support teams to proactively address issues and complaints to improve customer satisfaction.
- **Review Plans and Pricing**
  - Re-evaluate international calling plans and high-usage tariffs for competitiveness.
  - Offer bundled services and offer long-term contract options to increase customer stickiness.
- **Monitor High-Usage and At-Risk Segments**
  - Train usage patterns to flag potential dissatisfaction early.
  - Use predictive modeling to intervene before customers churn.
- **Engage At-Risk Customers with Feedback loops**
  - For customers flagged as high risk by the model, send short, targeted surveys(e.g, about service quality, pricing, or support experience).
  - This provides real-time insights into customer dissatisfaction and shows customers that their feedback matters, which can improve loyalty.
  - Combine survey responses with predictive insights to refine retention offers(discounts, loyalty rewards, or tailored plans

4. **Continuous Model Improvement**

- Regularly retrain the XGBoost model as customer behavior evolves.
- Use predictive insights to guide marketing, product, and service strategies.

💻 Getting Started

To explore or replicate this analysis locally, follow the steps below:

1. 📦 Clone the Repository

```
git clone (https://github.com/Mercykirwa25/End_Phase_3_Project.git)
cd End_Phase_3_Project
```

2. 🐍 Set Up Your Conda Environment

Make sure you have Anaconda installed.

```
conda create -n end_phase3 python=3.11.13
conda activate end_phase3
```

3. 📚 Install Required Packages

```
pip install pandas numpy matplotlib seaborn scikit-learn scipy imbalanced-lear
```

4. 📁 Load the Dataset

Ensure the dataset is in the following structure:

```
Data/
└── bigml_59c28831336c6604c800002a.csv
```

Load it in Python:

```python
import pandas as pd
data = pd.read_csv("bigml_59c28831336c6604c800002a.csv")
```

🤝 Acknowledgements

Special thanks to the dataset contributors: SyriaTel Customer Churn` dataset from Kaggle