📘 Telecom Customer Churn Analysis — SkyWave Telecom

📌 Project Overview

This project presents a comprehensive **Exploratory Data Analysis (EDA)** and **machine learning-based predictive modeling** approach to analyze and forecast customer churn for **SkyWave Telecom**.

The goals of this project are:

- Identify key factors driving customer churn.
- Understand customer behavior through data exploration.
- Build predictive models to classify customers based on their likelihood to churn.
- Provide actionable insights for improving retention strategies.

📂 Repository Structure

```
Telecom_Customer_Churn_Analysis/
│
├── data/
│   └── CustomerChurn.csv        # Original dataset
├── notebooks/
│   └── Telecom_Customer_Churn_Analysis_NOV.ipynb
├── reports/
│   └── README.md                # Project summary and results
```

🗃️ Dataset

- **Source:** Kaggle
- **Records:** 7,043 customers
- **Features:** 20 columns (demographics, service usage, account info)
- **Target variable:** Churn (Yes/No)

Key Features:

- Demographics: Gender, Senior Citizen, Partner, Dependents
- Service Info: Phone Service, Internet Service, Online Security, Device Protection, Tech Support, Streaming TV/Movies
- Billing: Contract Type, Paperless Billing, Payment Method, Monthly Charges, Total Charges
- Tenure: Customer tenure in months

🔎 Exploratory Data Analysis (EDA)

Data Cleaning & Preprocessing:

- Dropped duplicates and missing values in Total_Charges.
- Converted categorical variables into professional format.
- Created Tenure_Group bins for easier analysis.

Insights:

- **Demographics:** Gender has minimal impact; senior citizens are more likely to churn (~41%). Customers without partners/dependents show higher churn.
- **Services:** Fiber optic internet users churn more (42%) than DSL (19%). Absence of add-on services (Online Security, Backup, Tech Support) correlates with higher churn.
- **Contracts & Billing:** Month-to-month contracts see the highest churn (43%). Electronic check users churn the most (45%).
- **Financial Factors:** Higher monthly charges increase churn likelihood, while higher total charges (long-term customers) reduce churn.

Visualizations:

- Bar charts for categorical variables vs. churn.
- Histograms and KDE plots for numerical variables (Monthly_Charges, Total_Charges).

- Correlation heatmap shows strongest churn drivers: Tenure, Payment_Method_Electronic check, Contract_Two year, Internet_Service_Fiber optic.

⚙ Data Preprocessing

1. **Binary Encoding**: Yes/No → 1/0
2. **One-Hot Encoding**: Multi-category features (Gender, Contract, Payment_Method, Tenure_Group, etc.)
3. **Standardization**: Numerical features scaled using StandardScaler
4. **Final Feature Set**: 36 numeric columns (int & float)

🤖 Machine Learning Models

Three tree-based classifiers were implemented and evaluated:

Model	Accuracy	Precision	Recall	F1 Score
Decision Tree	0.734	0.497	0.496	0.497
Random Forest	0.792	0.648	0.466	0.542
XGBoost	0.802	0.662	0.512	0.578

Analysis:

- **XGBoost** achieves the best overall performance, balancing high accuracy, precision, and recall.
- **Random Forest** performs well, but slightly lower recall indicates some churned customers are missed.
- **Decision Tree** performs weakest, less reliable for deployment.

Conclusion:

XGBoost is the most suitable model for predicting churn and supporting retention strategies.

📌 Key Takeaways & Recommendations

1. **Enhance Retention for Short-Term Customers:**
   Onboarding programs, loyalty points, or proactive support for 1–12 month tenure customers.
2. **Reassess Fiber Optic Pricing & Quality:**
   Address higher churn in this segment through pricing review and service improvements.
3. **Promote Add-on Services:**
   Bundled offers for Online Security, Backup, Device Protection, and Tech Support increase retention.
4. **Encourage Long-Term Contracts:**
   Benefits or discounts for 1- or 2-year contracts reduce month-to-month churn.
5. **Improve Payment Convenience:**
   Incentivize automated payments to reduce churn associated with electronic check users.
6. **Monitor High-Charge Customers:**
   Target high monthly charge customers with satisfaction surveys or personalized offers.

📊 Visualizations

- **Churn Distribution**: Bar chart of churned vs retained customers
- **Tenure Groups**: Churn by tenure bins
- **Service Types**: Churn across internet and add-on services
- **Financial Factors**: KDE plots of monthly and total charges vs churn
- **Correlation Heatmap**: Visualizing key features affecting churn
- **Confusion Matrices & ROC Curves** for all models

(All plots are included in the notebook Telecom_Customer_Churn_Analysis.ipynb.)

🛠 Tech Stack

- Python, Pandas, NumPy, Matplotlib, Seaborn
- Scikit-learn: DecisionTreeClassifier, RandomForestClassifier, StandardScaler
- XGBoost: XGBClassifier
- Metrics: Accuracy, Precision, Recall, F1-Score, ROC-AUC
- Jupyter/Colab for notebooks

👤 Author

D. Saif