Customer Churn Prediction

Presented By: Junaid Khan

Role: data Scientist Intern at Skilled Score

Project Aim

The goal of this project is to predict which customers are at risk of leaving the telecom service, so the company can take action and reduce customer loss

Data Set Overview

- 7043 customer records
- Columns: Gender, Contract, Tenure, Monthly Charges, Churn (Yes/No)
- Target column: **Churn**

worked with real customer data. The target was to predict whether a customer will churn or not.

Exploratory Data Analysis (EDA)

- **Month-to-month customers churn the most**
- **Fiber optic users churn more**
- **No online security/support → higher churn**
- **High Monthly Charges = more churn**

We looked at the patterns in the data and found common traits of people who leave the service.

Data Preprocessing

- Removed missing values
- Dropped `customerID`
- Encoded categorical variables
- Created new feature: `TotalSpend = MonthlyCharges × Tenure`

We cleaned the data and made it machine-readable. We also created new features to help the model learn better.

Models Trained

We trained 3 models:

- Logistic Regression
- Random Forest
- Gradient Boosting

We compared multiple models to see which one can predict churn more accurately.

Results Before Fixing Imbalance

Model	Recall for Churn	F1-Score for Churn
Logistic Regression	9%	0.14
Random Forest	9%	0.14
Gradient Boosting	15%	0.21

All models struggled to catch churners because there were fewer churners than non-churners.

SMOTE + Gradient Boosting

Metric	Value
Accuracy	61%
Churn Recall	34%
Churn F1	0.35

SMOTE helped the model learn churn better. Recall improved from 9% to 34%.

Business Insights

What causes churn?

- Month-to-month contracts
- Fiber optic service
- No tech support
- High monthly cost

We now know who is at risk. These patterns help the business take action before customers leave.