Telecom Customer Churning

ML Project

Team: Charon, Vatsal and April
Group Number: 15

Overview

01 Introduction

02 Approach

03 Evaluation

04 Insight

Introduction

Telecom Churning ML Project	
Target Industry	Telecom
Target Metric	Customer Churn Rate
Goal	Analyze historical customer data Predict whether a customer is likely to leave the service
Dataset	Customer Account Data Subscription Details User Behavior Target Variable (Churn) https://www.kaggle.com/datasets/mnassrib/telecom-churn-datasets

Approach: EDA (1)

Visualization of Feature Distribution

Checking Class Imbalance in the Target Variable

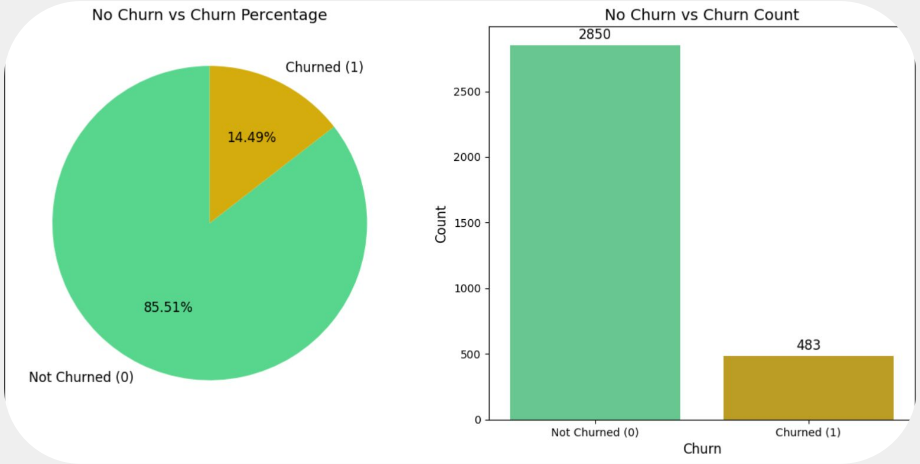
Correlation Between Features

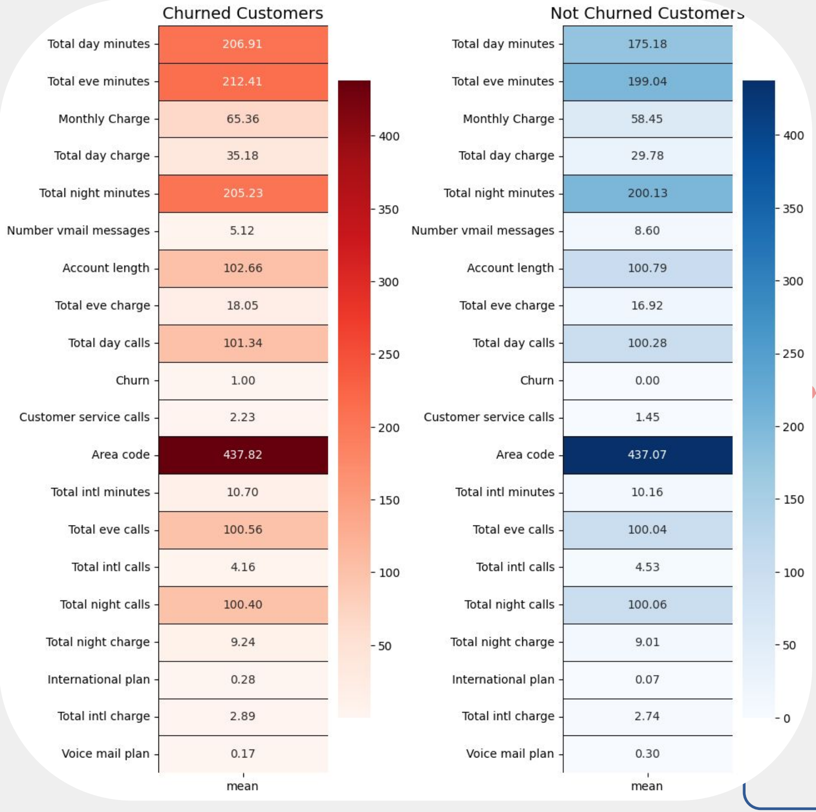
Churn Distribution of certain features

Mean feature values by churn status

Checking for Outliers

Approach: EDA (2)





<u>**Conclusion of EDA:**</u>

- No missing values or duplicate records
- Most variables follow a normal distribution
- Features have different scales
- Class imbalance: majority are non-churned customers
- Multicollinearity exists (e.g., minutes & charges)
- Higher service usage & charges increase churn probability
- Geographic location (State) significantly impacts churn
- Outliers retained as they reflect valid behavior

Evaluation

Best Strategy for handling imbalance-Class Weights

Unbalanced Dataset

Reflects the real-world scenario

Class-Weighted

Pay more attention to ‘churn’

Balanced Dataset (SMOTE)

Changes the data distribution by oversampling

Unbalanced Dataset Results:

	Model	Accuracy	Precision	Recall	F1 Score	ROC AUC
1	DecisionTree	0.946027	0.944564	0.946027	0.943086	0.913583
2	RandomForest	0.934033	0.934978	0.934033	0.927063	0.939085
0	LogisticRegression	0.866567	0.842035	0.866567	0.835073	0.823259
3	XGBoost	0.854573	0.730295	0.854573	0.787561	0.927202

Unbalanced Dataset (Class-Weighted Models) Results:

	Model	Accuracy	Precision	Recall	F1 Score	ROC AUC
3	XGBoost	0.962519	0.961916	0.962519	0.961192	0.939935
2	RandomForest	0.943028	0.940876	0.943028	0.940862	0.944583
1	DecisionTree	0.910045	0.921348	0.910045	0.914029	0.894574
0	LogisticRegression	0.785607	0.866687	0.785607	0.811100	0.831868

Balanced Dataset (SMOTE) Results:

	Model	Accuracy	Precision	Recall	F1 Score	ROC AUC
3	XGBoost	0.935532	0.932988	0.935532	0.933581	0.917544
2	RandomForest	0.916042	0.911989	0.916042	0.913286	0.923214
1	DecisionTree	0.868066	0.883909	0.868066	0.874322	0.824941
0	LogisticRegression	0.812594	0.833751	0.812594	0.821768	0.810020

Evaluation

Best Model-XGBoost

Highest Recall & F1

Critical for identifying churners accurately.

Strong ROC AUC

Excellent at distinguishing churners vs. non-churners.

Consistent Performance

Maintains top performance on both *class-weighted* and *SMOTE* datasets.

```
Unbalanced Dataset Results:
```

	Model	Accuracy	Precision	Recall	F1 Score	ROC AUC
1	DecisionTree	0.946027	0.944564	0.946027	0.943086	0.913583
2	RandomForest	0.934033	0.934978	0.934033	0.927063	0.939085
0	LogisticRegression	0.866567	0.842035	0.866567	0.835073	0.823259
3	XGBoost	0.854573	0.730295	0.854573	0.787561	0.927202

```
--------------------------------------------------
Unbalanced Dataset (Class-Weighted Models) Results:
```

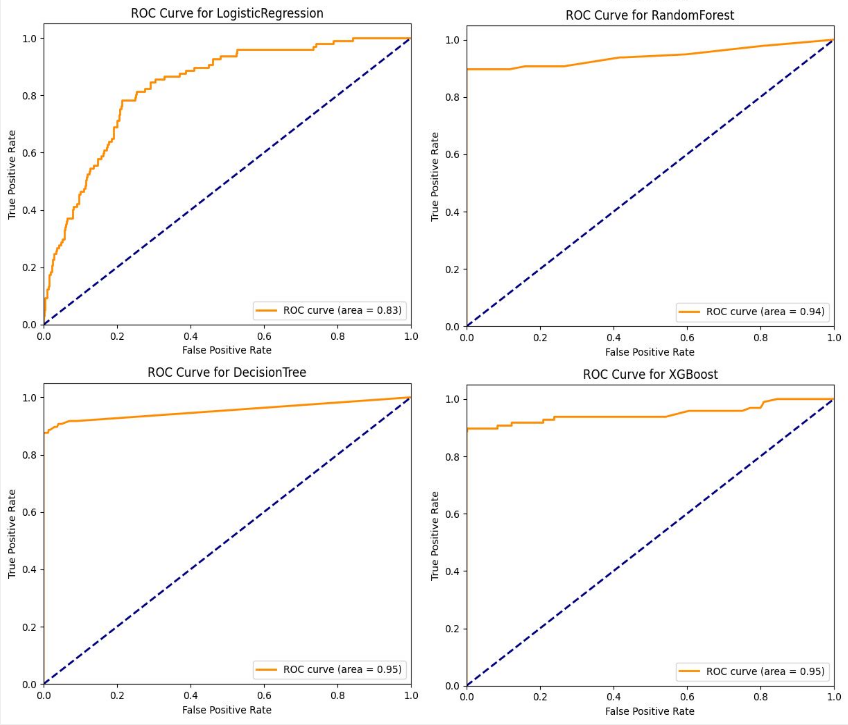
	Model	Accuracy	Precision	Recall	F1 Score	ROC AUC
3	XGBoost	0.962519	0.961916	0.962519	0.961192	0.939935
2	RandomForest	0.943028	0.940876	0.943028	0.940862	0.944583
1	DecisionTree	0.910045	0.921348	0.910045	0.914029	0.894574
0	LogisticRegression	0.785607	0.866687	0.785607	0.811100	0.831868

```
--------------------------------------------------
Balanced Dataset (SMOTE) Results:
```

	Model	Accuracy	Precision	Recall	F1 Score	ROC AUC
3	XGBoost	0.935532	0.932988	0.935532	0.933581	0.917544
2	RandomForest	0.916042	0.911989	0.916042	0.913286	0.923214
1	DecisionTree	0.868066	0.883909	0.868066	0.874322	0.824941
0	LogisticRegression	0.812594	0.833751	0.812594	0.821768	0.810020

Evaluation

ROC Curve



The larger the area under the curve (AUC), the more effectively the model can distinguish churners from non-churners.

Logistic Regression
- 0.83

Decision Tree
- 0.95

Random Forest
- 0.94

XGBoost
- 0.95

Evaluation

Logistic Regression

- False Positives (FP): 119
- False Negatives (FN): 24

Decision Tree

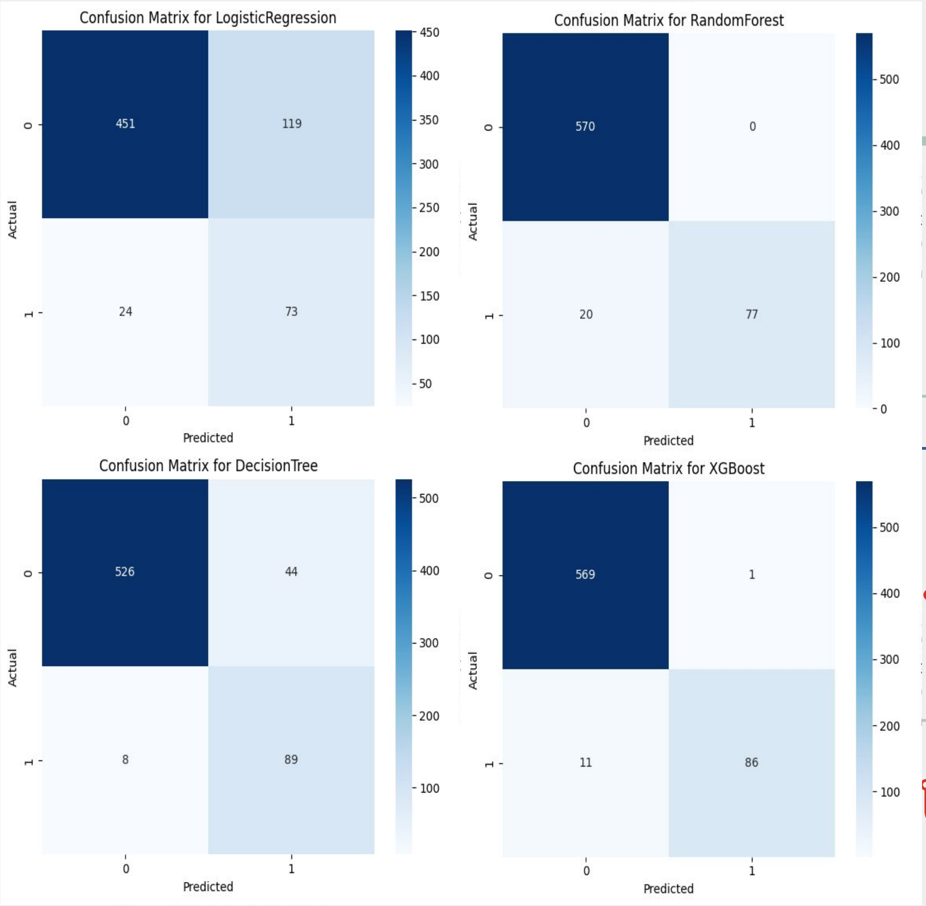
- False Positives (FP): 42
- False Negatives (FN): 18

RandomForest:

- False Positives (FP): 11
- False Negatives (FN): 27

<u>XGBoost:</u>

- False Positives (FP): 5
- False Negatives (FN): 20



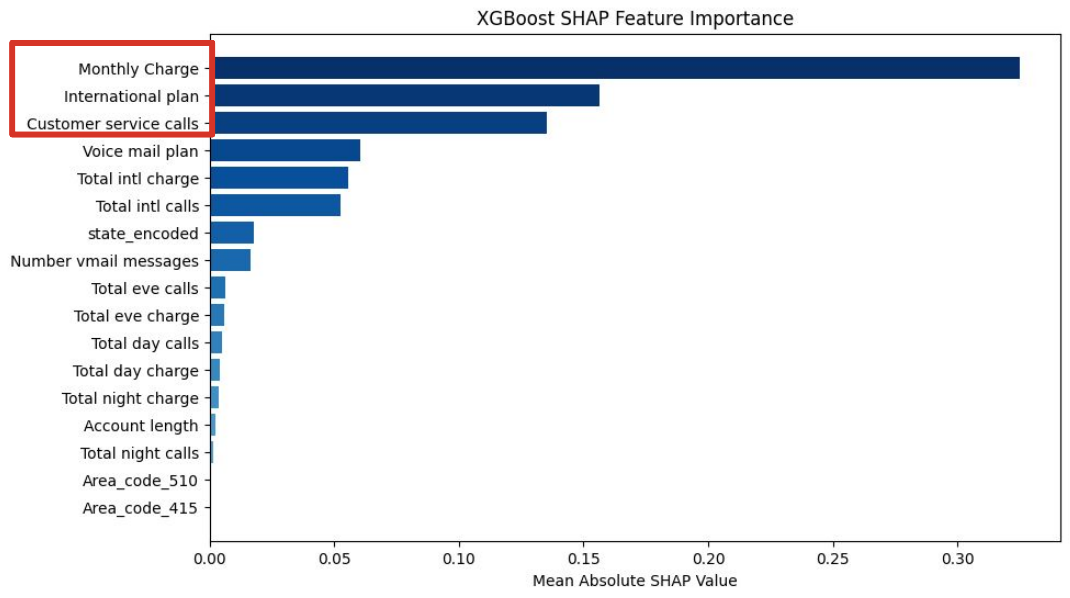
Insight

SHAP Feature Importance

Monthly Charge
- Personalized Pricing Plans
- Loyalty Discounts
- Bundled Services

International Plan
- Discounts
- Seasonal Promotions

Customer Service Calls
- Early Intervention
- Improve Self-Service Channels



Insight

Customer Segmentation

Low-Risk

Medium-Risk

High-Risk

Churn Prediction

0

33%

66%

100%

Potential Strategy

Loyalty Rewards

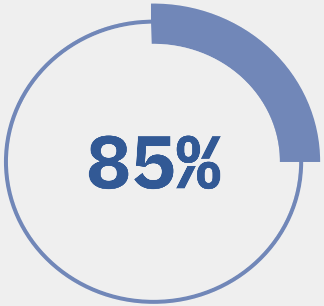
Targeted Discounts

Personalized Outreach

Insight

Test Case

```python
ample_customer = {
    'Account length': 120,
    'State': 'OH',
    'Area code': 415,
    'International plan': 'No',
    'Voice mail plan': 'Yes',
    'Number vmail messages': 30,
    'Total day minutes': 250.0,
    'Total day calls': 105,
    'Total day charge': 38.0,
    'Total eve minutes': 160.0,
    'Total eve calls': 100,
    'Total eve charge': 22.0,
    'Total night minutes': 220.0,
    'Total night calls': 95,
    'Total night charge': 10.0,
    'Total intl minutes': 8.0,
    'Total intl calls': 2,
    'Total intl charge': 2.0,
    'Customer service calls': 1
```



Churn Probability

Predicted churn probability for the sample customer

Customer Segment

High-Risk

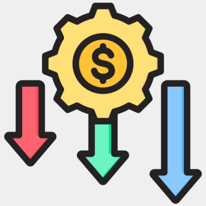
Conclusion

Objective: Predict customer churn in the telecom industry
Model Performance: Among the tested models, **XGBoost** has the best performance, demonstrating its effectiveness in handling class imbalance and complex relationships
Feature Importance: The most influential predictors of churn are **Monthly Charge, International Plan, and Customer Service Calls**, highlighting the impact of usage patterns and customer dissatisfaction.
Customer Segmentation: Customers are divided into **low-risk, medium-risk, and high-risk** categories based on churn probability, allowing for tailored intervention strategies.

Business Value

Prediction

Intervention



Cost

Thanks!

Any Questions?