SYRIATEL CUSTOMER CHURN PREDICTION

Machine learning project

PROJECT OVERVIEW

Goal:

Predict which customers are likely to churn (i.e., leave the service)

Why It Matters:

- Churn impacts revenue and customer lifetime value.
- Early prediction allows targeted retention efforts.

Solution:

Develop a machine learning model using customer data to predict churn.

BUSINESS UNDERSTANDING

Company:

Syriatel – Telecom provider experiencing churn issues.

Business Objective:

- ✓ Reduce churn by understanding key drivers.
- ✓ Deploy a predictive model to support proactive retention.

Success Metric: High model recall and actionable insights for the business team.

DATASET OVERVIEW

- **Source**: Public dataset simulating telecom customer records.
- **Rows:** ~3,333 **Columns:** 21 features
- **Key Features :** Account length, plan types, charges, service calls
- **Target:** churn (Yes/No)
- **Class Balance:** Slightly imbalanced (~14.5% churned)

DATA PREPARATION

- **Steps Taken**:
  - Handled missing data
  - Converted categorical variables using encoding
  - Normalized numerical features
  - Feature engineering: total charges, interaction terms
- **Train-Test Split**: 80:20 stratified split

Exploratory Data Analysis(EDA)

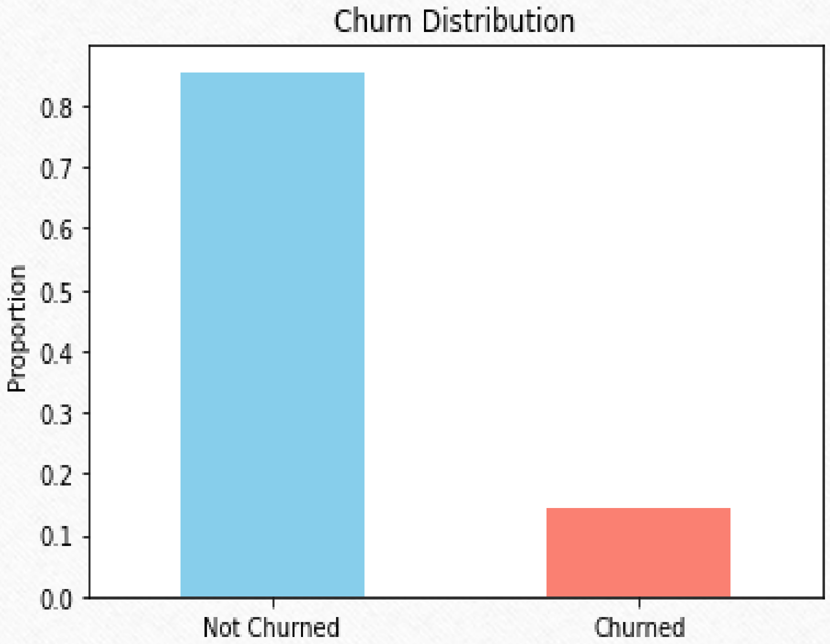
Visualized feature distributions

- Compared churn vs. non-churn groups

Found churners:

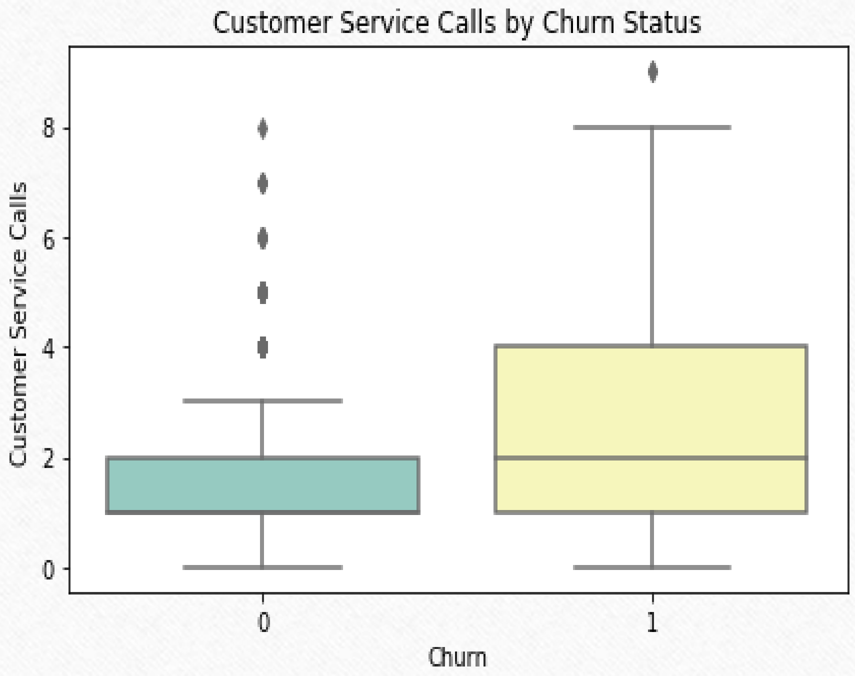
- Use customer service more
- Often lack international plans
- Have higher day-time call minutes

CHURN DISTRIBUTION



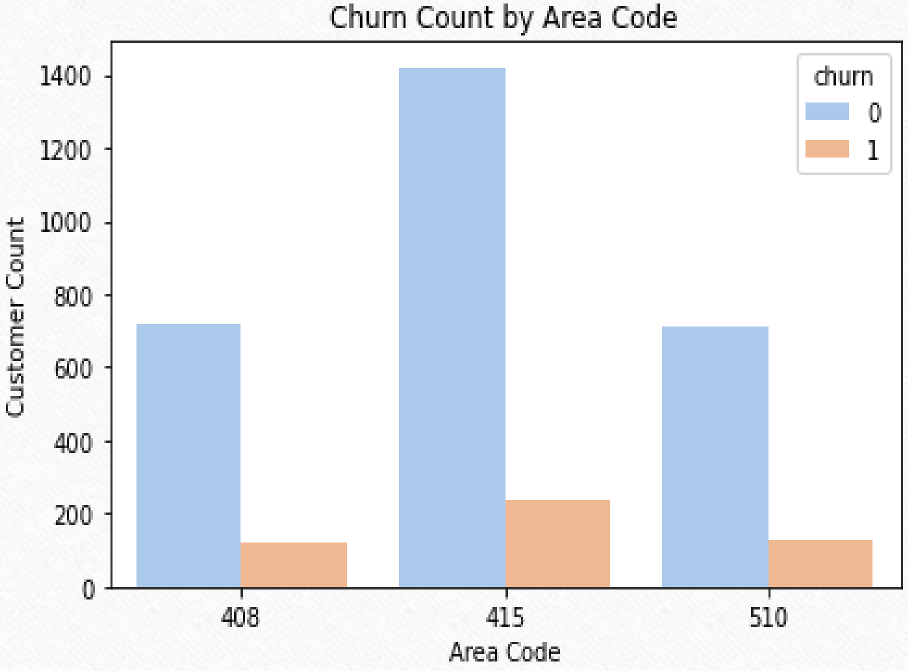
- **Observations**
- - The dataset is highly imbalanced, with significantly more non-churned customers.
- - This imbalance could affect model performance, especially in metrics like accuracy.

CUSTOMER SERVICE CALLS VS CHURN



- Customers who churned (1) appear to have made more customer service calls on average
- This suggests that higher customer service call volume may correlate with increased churn risk
- Could indicate service issues or customer dissatisfaction driving both more calls and eventual churn

CHURN BY AREA CODE



- Area code 415 appears to have the highest total number of customers
- The relative proportion of churned vs retained customers varies by area code
- This helps identify geographic areas that may need targeted retention efforts

STATISTICAL ANALYSIS

- High churn (~14%) observed in customers with:An international planFrequent customer service calls
- Strong correlations between:total_*_minutes and total_*_charge
- ($r \approx 0.99$)Skewed distributions in features like: number_vmail_messages, customer_service_calls
- T-tests show features like total_day_charge and customer_service_calls are statistically different between churned and non-churned groups ($p < 0.05$)
- Chi-square tests confirm strong relationship between international_plan and churn

MODELING

- Used 6 models:
  - **Logistic Regression**
  - **Decision Tree**
  - **Random Forest**
  - **SVM**
  - **XGBoost**
  - **K-Nearest Neighbors**
- Best Model: **XGBoost**
- Good balance of performance and interpretability
- Tuned for best threshold at 0.33

MODEL PERFORMANCE METRICS

Model	Accuracy	Precision (Churn)	Recall (Churn)	F1-Score (Churn)	ROC AUC
Logistic Regression	85.6%	0.51	0.22	0.30	0.80
Decision Tree	90.0%	0.66	0.64	0.65	0.79
Random Forest	93.2%	0.92	0.59	0.72	0.89
SVM (RBF)	85.5%	0.00	0.00	0.00	0.76
XGBoost	94.5%	0.88	0.71	0.79	0.88
K-Nearest Neighbors	87.4%	0.53	0.27	0.36	0.65

KEY RECOMMENDATIONS

- **1. Improve Customer Service Experience:**
  Analyze and reduce high customer service call frequency — a key churn trigger.
- **2. Targeted Retention Offers:**
  Focus on customers with international plans or high day-time usage.
- **3. Monitor Model Continuously:**
  Track model accuracy post-deployment and retrain regularly.
- **4. Enrich Dataset:**
  Add features like billing history, customer feedback, and service sentiment.

NEXT STEPS

- Model Deployment:
- Integrate into CRM system for live predictions
  Dashboarding:
- Visualize churn risk by customer segment
  Continuous Improvement:
- Collect more behavioral data
- Explore time-based churn trends

QUESTIONS?

THANKYOU

HELLEN DIANA NJERI MACHARIA

hellendiana091@gmail.com