Predicting Customer Churn for SyriaTel

Built Machine Learning Model

By: BRIAN KIPNGETICH

Date: 1 September 2024

Overview

- **Project Goal:** Predict customer churn to help SyriaTel retain customers and improve business outcomes.
- **Importance:** Retaining customers is more cost-effective than acquiring new ones. Predictive models help identify customers at risk of leaving.
- **Approach:** We built a machine learning model to predict churn using customer data.

Business Understanding

- Churn impacts revenue and customer lifetime value.
- Understanding customer behavior is key to improving retention.
- **E.g.**: Customers with frequent service issues may be more likely to churn.

Why Classification?

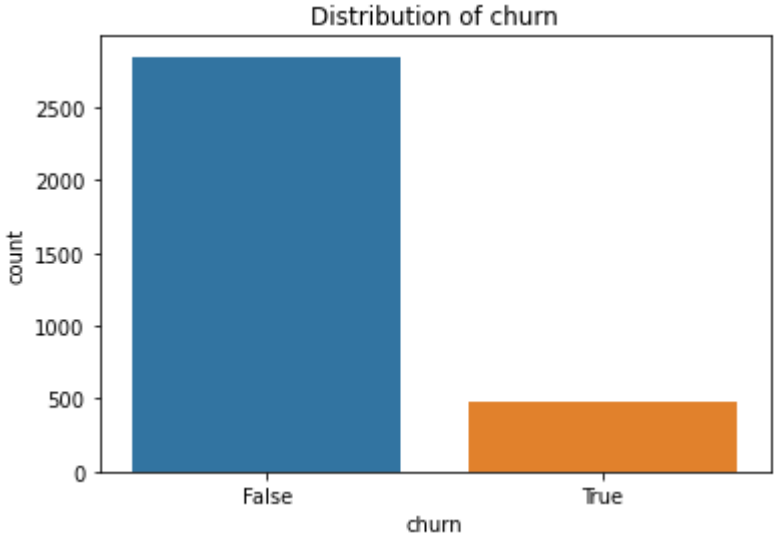
- **What is Classification?**
  - A method used to categorize data into groups (e.g., churn vs. no churn).
- **Purpose:**
  - Helps identify which customers are likely to leave SyriaTel.
- **Benefit:**
  - Enables targeted retention efforts, saving costs, and improving efficiency.

Data Overview

- Dataset provided by SyriaTel on url:'https://www.kaggle.com/datasets/becksddf/churn-in-telecoms-dataset'
- Dataset includes customer demographics, account details, and usage metrics.
- **Key Features**: Age, contract type, monthly charges, and customer service calls.
- **Target Variable**: Churn (Yes/No)

Exploratory Data Analysis

- Churn rate in the dataset is around 14.5%.

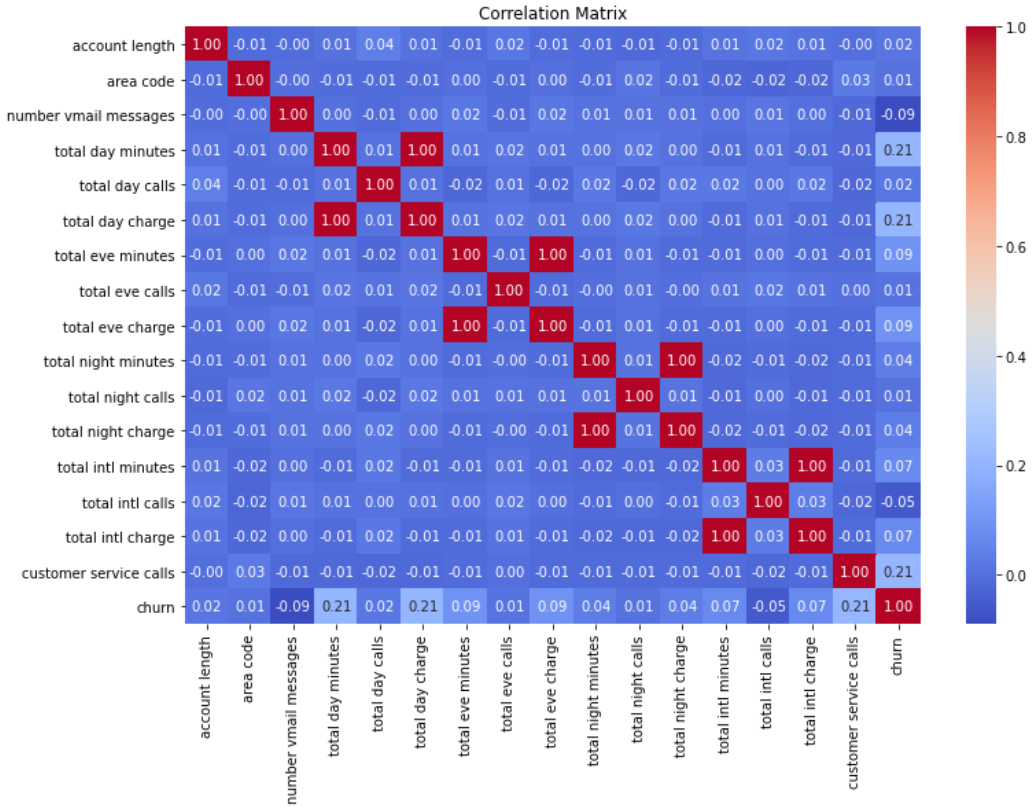


Exploratory Data Analysis Cont’d

- **Correlation**: Features like tenure and service calls are strongly related to churn.

See the correlation matrix in the next slide

Exploratory Data Analysis Cont'd



Modeling Approach

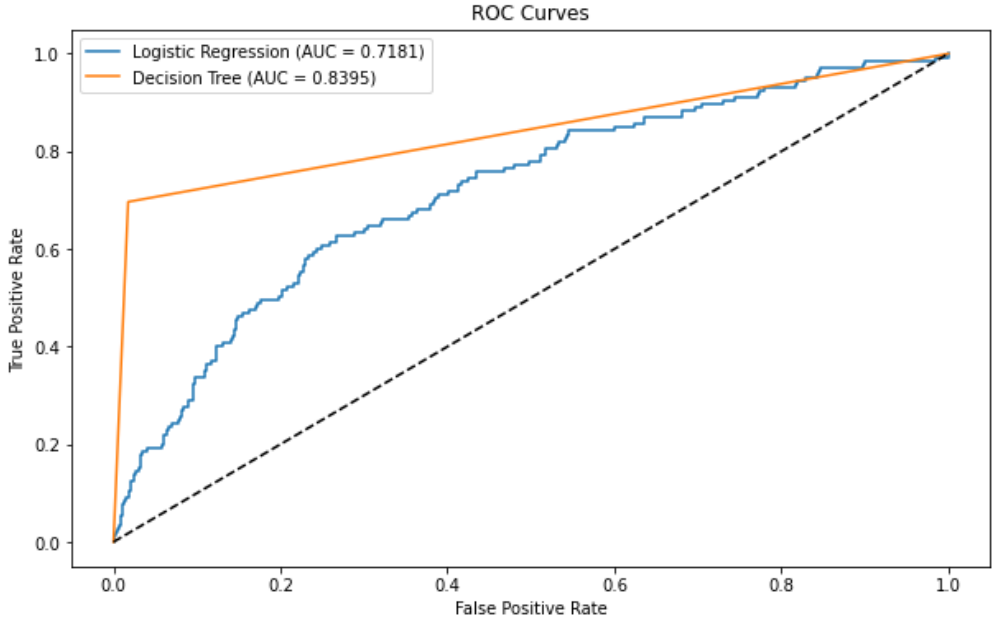
- The modelling approach was done by having three models as below to pick the best for deployment.
- **Baseline Model**: Dummy classifier to benchmark performance.
- **Logistic Regression**: Chosen for its simplicity and effectiveness.
- **Decision Tree**: Offers interpretability and captures non-linear relationships.

Model Evaluation

- **Metrics**: Accuracy, precision, recall, F1 score, and ROC-AUC.
- **Comparison**: Logistic regression outperformed other models.
- **ROC-AUC**: Indicates logistic regression has the best ability to distinguish churners.

ROC Curve Comparison

- ROC curve compares true positive rate vs. false positive rate.
- This is the roc-auc curve for the three models



- Decision tree: Highest AUC (Area Under the Curve), indicating strong performance.
- logistic regression performs well but slightly less effective than Decision tree.

Conclusion

- **Best Model**: Decision tree provides the most balanced and accurate predictions.
- **Business Impact**: Helps SyriaTel identify and retain high-risk customers.
- **Key Insight**: Data-driven decisions can significantly reduce churn.

Recommendations

- **Target High-Risk Customers**: Use model predictions for focused retention efforts.
- **Personalized Offers**: Provide incentives to prevent churn.
- **Monitor and Improve**: Continuously refine the model for better accuracy.

THE END

- Thank you for this opportunity to help you reduce the company churn rate through data analysis insights.
- Q & A