

The background features abstract, overlapping green geometric shapes in various shades of green, creating a modern and dynamic feel. The shapes are primarily located on the left and right sides of the slide, framing the central text.

CHURN IN TELECOMUNICATION'S DATASET

PHASE 3 PROJECT

Overview

- ▶ SyriaTel, a telecommunications company is interested in knowing whether a customer will stop doing business with the company. This will help the business in reducing the on money lost because of customers who do not stick around very long.
- ▶ We will use a classifier model to get insights from the dataset. From here we will be able to see whether there are any predicable patterns which in turn will assist Syriatel to have a clear picture of the churn rate.

Business and Data Understanding

- ▶ The dataset originates from SyriaTel Telecommunications (source: <https://www.kaggle.com/datasets/becksddf/churn-in-telecoms-dataset>)
- ▶ Questions to Consider:
 - Which features are more likely to predict churn.
 - What is the correlation between churn and other feature variables.

Our goal will be to build a model that will help us answer these questions.

Modelling

- ▶ We will be using classification models because we have observed that the target variable(churn column) is categorical.
- ▶ Our first model will be a vanilla logistic regression model followed by another logistic regression model that incorporate SMOTE. We are using SMOTE to help address the class imbalance in the target variable.
- ▶ The second model we will use a Decision tree and afterwards we will factor in the hyperparameter tuning and pruning to see whether the model will improve.

Evaluation-Logistic Regression

We evaluated the logistic regression models using the classification report (where we were interested in the accuracy score) and we also added the AUC for comparison but this being an imbalance dataset, the AUC will be more suitable as measure.

Vanilla Logistic regression(without SMOTE)

AUC - 0.81

Logistic regression with SMOTE

AUC - 0.82

Based on the AUC results above we can see there was a slight increase which means the model performance improved with SMOTE

Evaluation-Decision Tree

We have the decision tree using AUC both for the vanilla decision tree and decision tree with hyperparameter tuning and pruning.

Vanilla Decision tree

AUC - 0.85

Decision tree with hyperparameter tuning and pruning

AUC - 0.72

Based on the AUC results above, there was a reduction in AUC which means the decision tree did not perform better with hyperparameter tuning and pruning

Deployment/Conclusion

- Overall, the Logistic regression with SMOTE performance tells us that the model can be used as good predictor of the churn rate for SyriaTel telecommunications company.