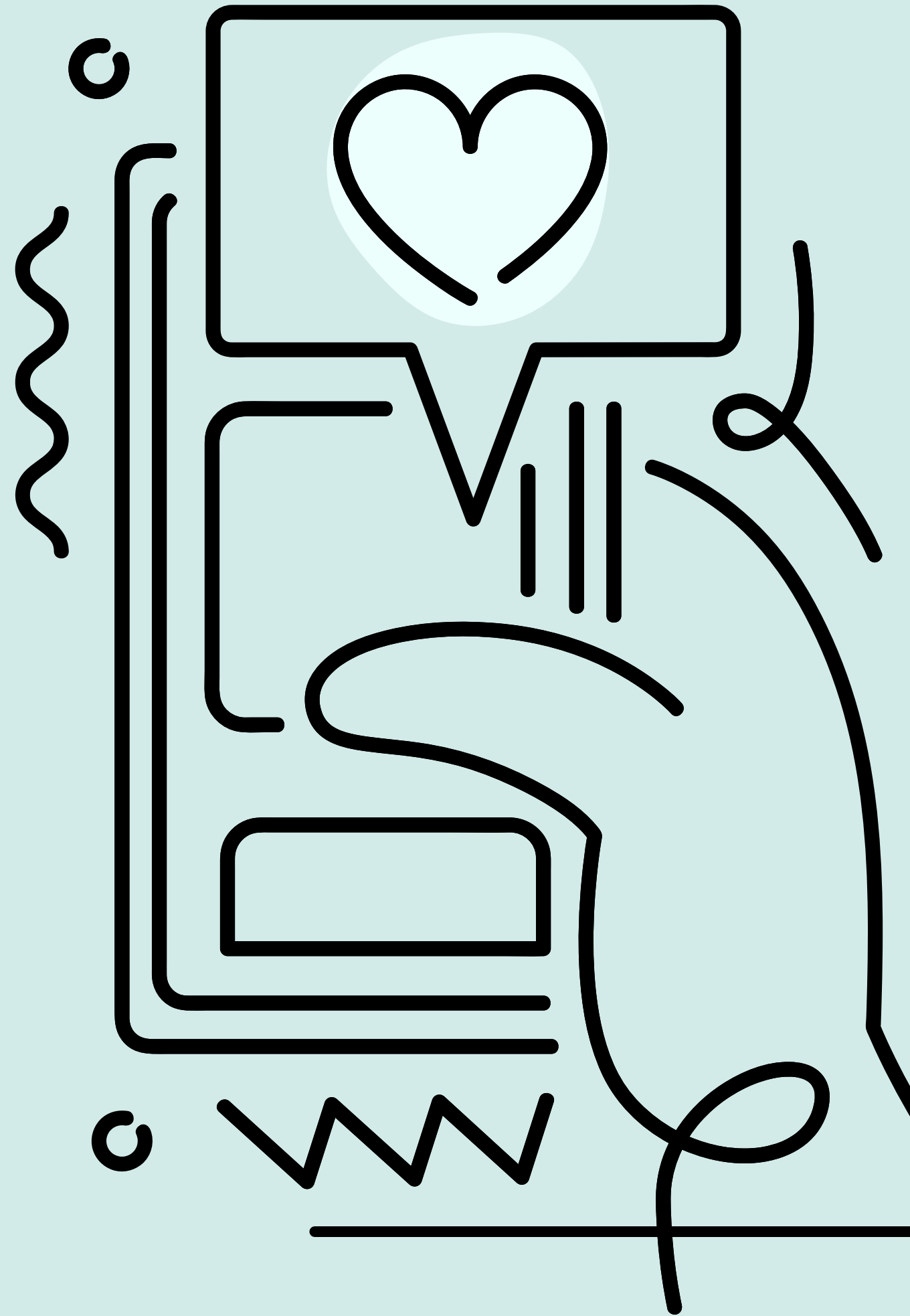


PHASE 3 PROJECT

Author: Lilian Ngige



SYRIATEL CUSTOMER CHURN

The data has a target variable called churn and several features

The data set has numerical data and categorical data

Categorical data are encoded

DATA

The dataset is a binary classification

This dataset has 3,333 rows and 21 columns

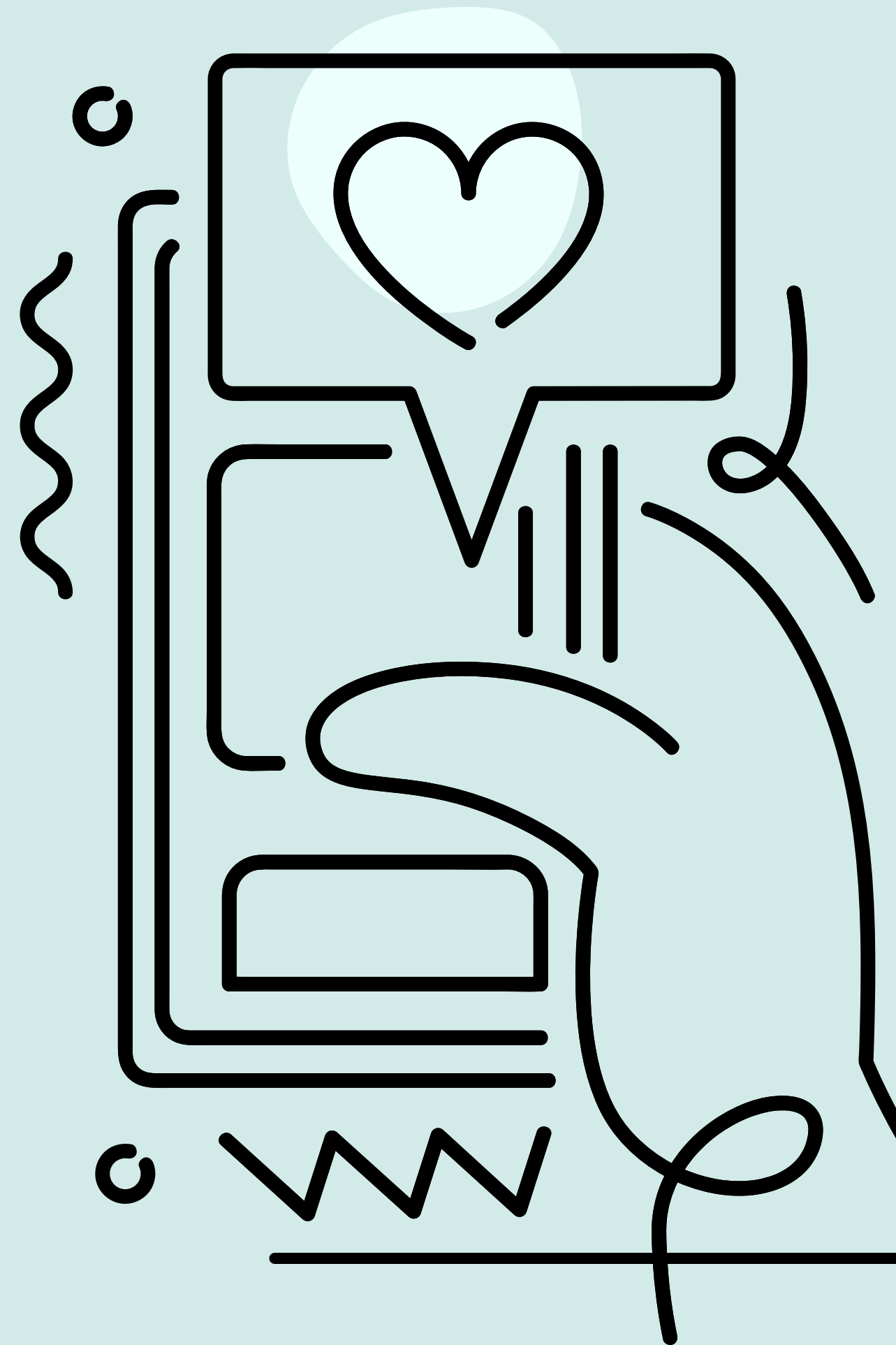


Project Overview:

The project aims at analysing SyriaTel data and use machine learning algorithm in building models that predict whether a customer will churn or not.

Business Problem:

SyriaTel, a telecommunications company, is facing the challenge of customer churn. The company wants to predict whether a customer is likely to stop doing business with them in the near future. Reducing customer churn is essential for maintaining revenue and profitability.

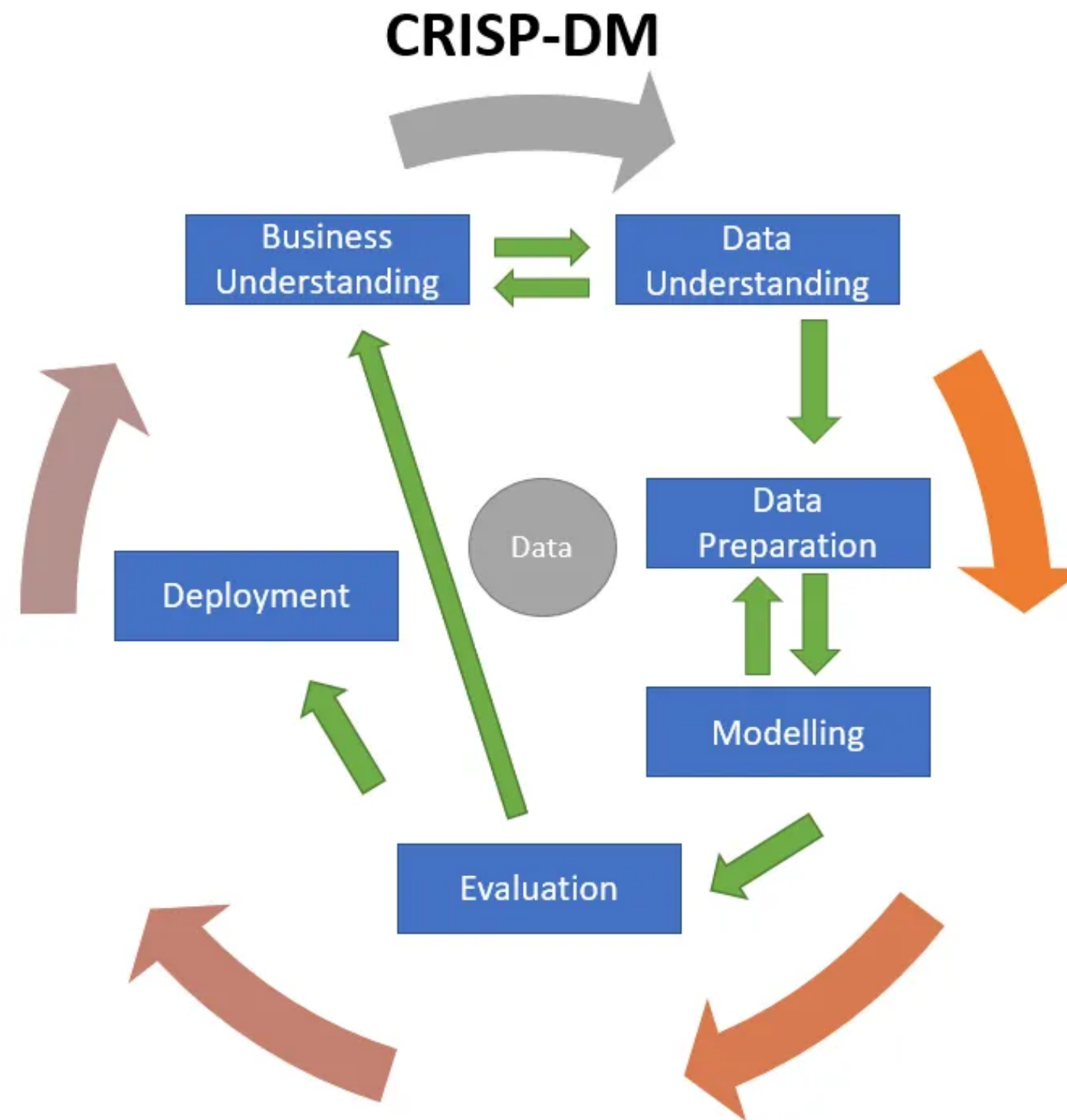




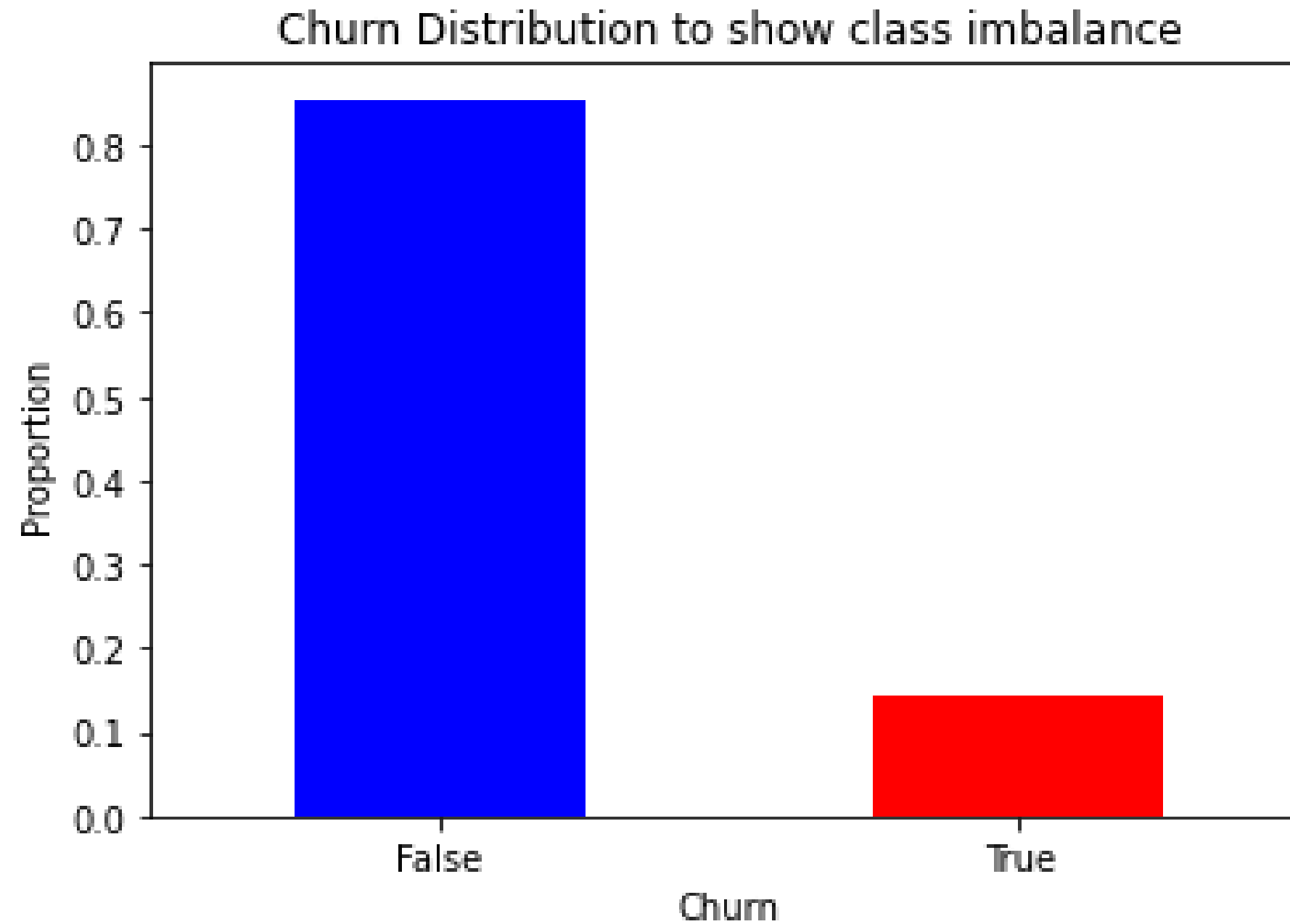
Project Objective

Build a binary classification model to predict customer churn. The model will identify customers who are likely to churn, to enable SyriaTel to take proactive measures to retain them.

Method used



Target Variable



The distributions shows
class imbalance

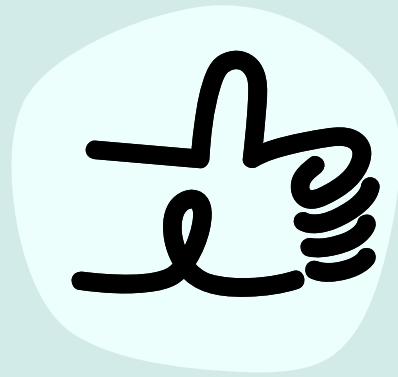
85% of customer have churned
and 14% have not churned

MODELLING:

Machine learning models to use for predicting customer churn in the telecom company.



Logistic Regression



Random Forest



Decision Tree

RESULTS

LOGISTIC REGRESSION

Accuracy: 0.74

RANDOM FOREST MODEL

Accuracy: 0.96

DECISION TREE:

Accuracy: 0.93

Random Forest is the best model with a striking balance in the 3 metrics of evaluation.

Evaluation

Random Forest model is the best performing model as shown in all 4 metrics

Accuracy: 0.98

Precision: 0.99

Recall: 0.96

F1 score: 0.98

CONCLUSION

The Random Forest model outperforms both Logistic Regression and the Decision Tree with tuned hyperparameters in terms of accuracy, precision, recall, and F1-Score. It achieves the highest overall accuracy and a well-balanced F1-Score, indicating robust performance in predicting customer churn.



RECOMMENDATION: CHURN RETENTION STRATEGIES



Utilize the predictions from random forest model to implement customer retention strategies for identified churners.
Develop specific marketing or service offers to reduce churn among high-risk customers.



The high-performance model should be deployed for predicting customer churn within the telecom company.

Next steps

Further improvement-Continue to monitor and fine tune the models over time.

Consider Feature engineering and data augmentation to potentially improve the model performance.



THANK YOU