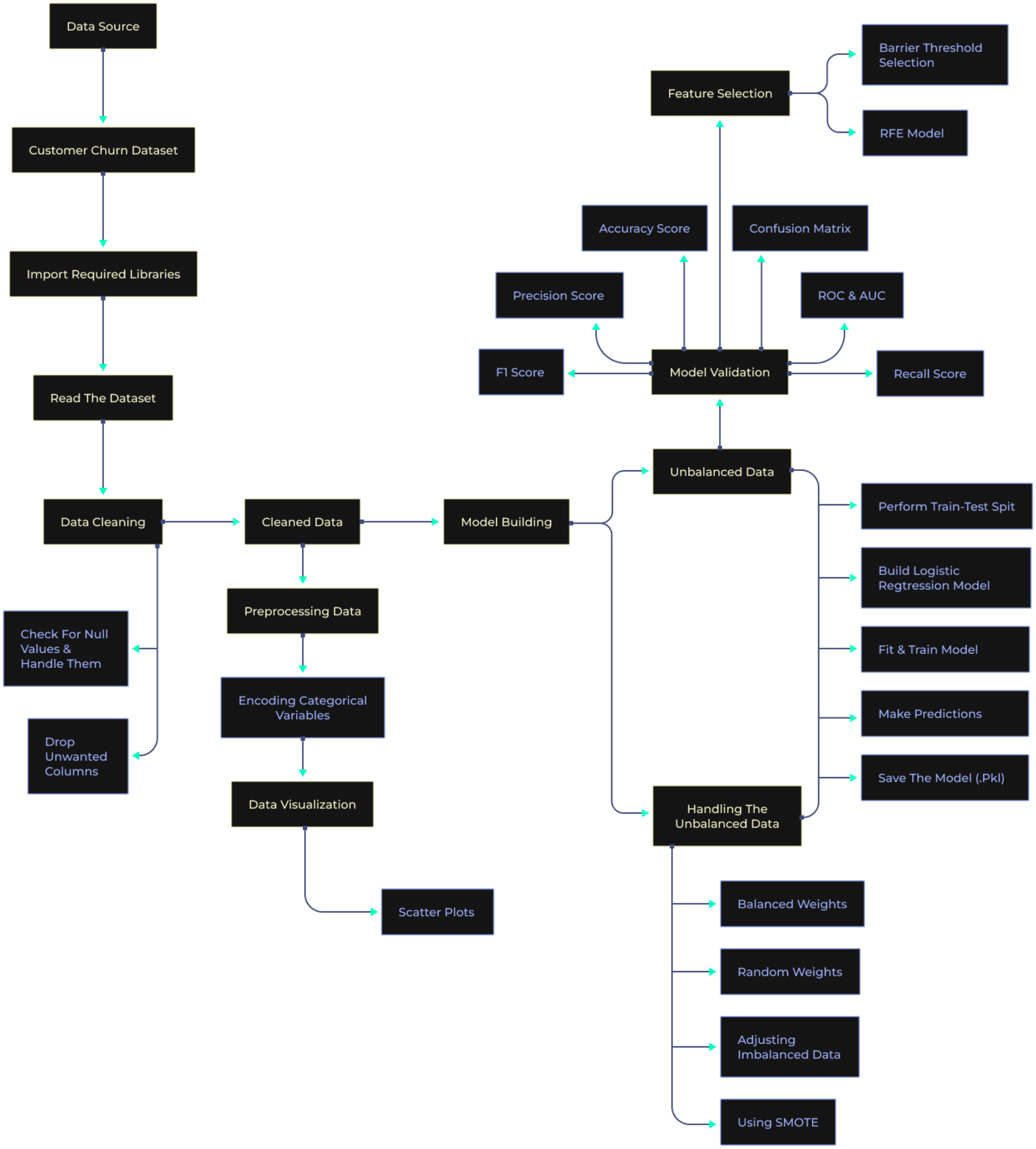

Data Source
Customer Churn Dataset
Import Required Libraries
Read The Dataset
Data Cleaning
Check For Null Values & Handle Them
Drop Unwanted Columns
Cleaned Data
Preprocessing Data
Encoding Categorical Variables
Data Visualization
Scatter Plots
Model Building
Unbalanced Data
Perform Train-Test Spit
Build Logistic Regtression Model
Fit & Train Model
Make Predictions
Save The Model (.Pkl)
Handling The Unbalanced Data
Balanced Weights
Random Weights
Adjusting Imbalanced Data
Using SMOTE
Model Validation
F1 Score
Precision Score
Accuracy Score
Confusion Matrix
ROC & AUC
Recall Score
Feature Selection
Barrier Threshold Selection
RFE Model

Regression project to implement logistic regression in python from scratch on streaming app data.

What will you learn?

- Understanding the basics of classification
- Introduction to Logistics regression
- Understanding the logit function
- Coefficients in logistics regression
- Concept of maximum log-likelihood
- Performance metrics like confusion metric, recall, accuracy, precision, f1-score, AUC, and ROC
- Importing the dataset and required libraries.
- Performing basic Exploratory Data Analysis (EDA).
- Using python libraries such as matplotlib and seaborn for data interpretation and advanced visualizations.
- Data inspection and cleaning
- Using statsmodel and sklearn libraries to build the model
- Splitting Dataset into Train and Test using sklearn.
- Training a model using Classification techniques like Logistics Regression,
- Making predictions using the trained model.
- Gaining confidence in the model using metrics such as accuracy score, confusion matrix, recall, precision, and f1 score
- Handling the unbalanced data using various methods.
- Performing feature selection with multiple methods
- Saving the best model in pickle format for future use.

Project Description

Business Objective

Predicting a qualitative response for observation can be referred to as classifying that observation since it involves assigning the observation to a category or class. Classification forms the basis for Logistic Regression. Logistic Regression is a supervised algorithm used to predict a dependent variable that is categorical or discrete. Logistic regression models the data using the sigmoid function.

Churned Customers are those who have decided to end their relationship with their existing company. In our case study, we will be working on a churn dataset.

XYZ is a service-providing company that provides customers with a one-year subscription plan for their product. The company wants to know if the customers will renew the subscription for the coming year or not.

Data Description

This data provides information about a video streaming service company, where they want to predict if the customer will churn or not. The CSV consists of around 2000 rows and 16 columns.

Aim

Build a logistics regression learning model on the given dataset to determine whether the customer will churn or not.

Tech stack

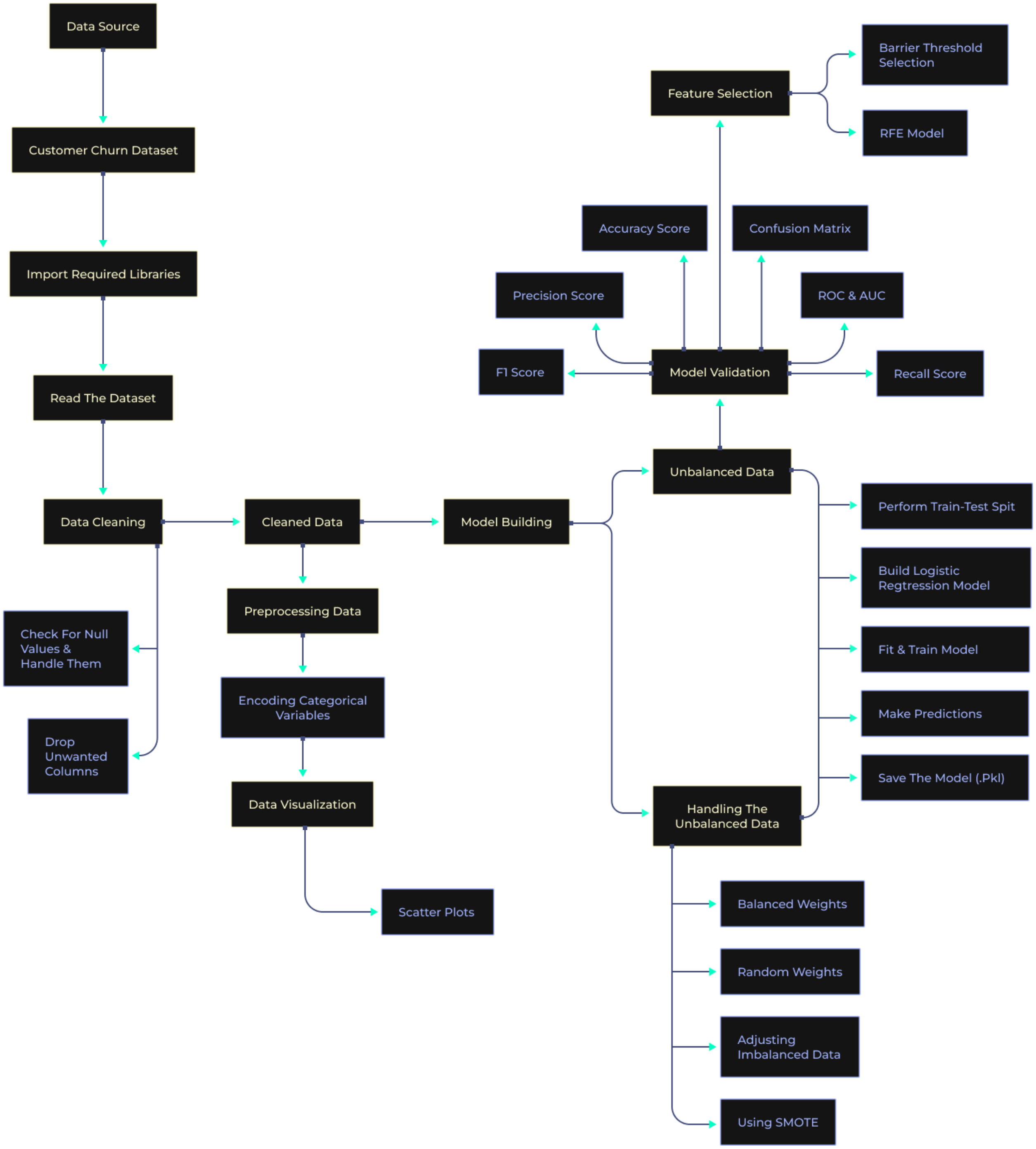
- Language - Python

- Libraries - numpy, pandas, matplotlib, seaborn, sklearn, pickle, imblearn, statsmodel

Approach

1. Importing the required libraries and reading the dataset.
2. Inspecting and cleaning up the data
3. Perform data encoding on categorical variables
4. Exploratory Data Analysis (EDA)
    - Data Visualization
5. Feature Engineering
    - Dropping of unwanted columns
6. Model Building
    - Using the statsmodel library
7. Model Building
    - Performing train test split
    - Logistic Regression Model
8. Model Validation (predictions)
    - Accuracy score
    - Confusion matrix
    - ROC and AUC
    - Recall score
    - Precision score
    - F1-score
9. Handling the unbalanced data
    - With balanced weights
    - Random weights
    - Adjusting imbalanced data
    - Using SMOTE
10. Feature Selection
    - Barrier threshold selection
    - RFE method
11. Save the model in the form of a pickle file

Architecture Diagram


Data Source
Customer Churn Dataset
Import Required Libraries
Read The Dataset
Data Cleaning
Check For Null Values & Handle Them
Drop Unwanted Columns
Cleaned Data
Preprocessing Data
Encoding Categorical Variables
Data Visualization
Scatter Plots
Model Building
Unbalanced Data
Perform Train-Test Spit
Build Logistic Regtression Model
Fit & Train Model
Make Predictions
Save The Model (.Pkl)
Handling The Unbalanced Data
Balanced Weights
Random Weights
Adjusting Imbalanced Data
Using SMOTE
Model Validation
F1 Score
Precision Score
Accuracy Score
Confusion Matrix
ROC & AUC
Recall Score
Feature Selection
Barrier Threshold Selection
RFE Model