**TABLE OF CONTENT**

(Project Overview).(#project-overview)

(Project Objective).(#project-objective)

(Data Sources).(#data-sources)

(Data Preprocessing).(#data-preprocessing)

(Machine Learning Model).(#machine-learning-model)

(Evaluation of Metrics).(#evaluation-of-metrics)

(Potential Benefits and Recommendation). (#potential-Benefits-and-Recommendation)

(Conclusion).(#conclusion)

## **Project Overview**

* **ConnectTel Telecom Company** faces the pressing need to address customer churn, which poses a significant threat to its business sustainability and growth.
* The company's current customer retention strategies lack precision and effectiveness, resulting in the loss of valuable customers to competitors.

## **Project Objective**

Develop a predictive analytics solution to identify high-risk customers and implement targeted retention strategies to reduce customer churn, enhancing ConnectTel's business sustainability and growth.

## **Data Source**

From 10Alytics

EDA and Feature Engineering were conducted.

## **Data Preprocessing**

This involved the process of cleaning, transforming, and preparing raw data for analysis to enable a high-quality data for building accurate models and obtaining reliable insights.

Steps taken:

* Replacing the missing values with median in Total Charges and converting the objects to strings.
* Removing duplicates: Eliminating duplicate data points to prevent skewing the analysis.
* Encode the categorical features to a numerical feature using a label encoder.
* Merged both the numerical features and the encoded categorical features.
* Dropped off redundant features.
* Segment dataset into data and target label
* Scale dataset features using the MinMaxScaler by using the RandomForestClassifier.
* Plotted out the feature importance chart

## **Machine Learning Model building**

Splinted data into training and evaluation datasets with test\_size=0.33, random\_state=42

Oversampled because the dataset is imbalance by importing SMOTE

7 Machine learning Algorithms were applied to train the dataset which includes:

* Random Forest Classifier ()
* Support Vector Classifier ()
* Logistic Regression ()
* Extreme Gradient Boosting Classifier ()
* Stochastic Gradient Descent Classifier ()
* Gaussian Naïve Base ()
* Decision Tree Classifier ()

Hyper Parameter Optimization Tunning of the 7 models.

## **Evaluation of Metrics**

Based on the analysis, the top-performing models for predicting churn in an imbalanced dataset after ranking the metrics FN>Recall>AUC/ROC>F1-score are:

1. XGBClassifier (XGBC)
2. RandomForestClassifier (RFC)
3. DecisionTreeClassifier (DTC)

These models demonstrate a good balance of metrics, with:

* Low FN rates (9.12%, 9.33%, 10,67%) indicating better identification of true positives (churners).
* High recall rates (0.84, 0.80 ) indicating good identification of true positives (churners).
* High AUC/ROC scores (0.91, 0.85) indicating excellent discrimination between churners and non-churners.
* High F1-scores (0.85, 0.82) indicating a good balance between precision and recall.
* High precision rates (0.87, 0.85) indicating good identification of true positives without flagging too many false positives.

## **Potential Benefits and Recommendation**

1. Upgrading internet and network quality
2. Enhancing customer service relationships
3. Implementing device protection
4. Offering competitive pricing and discounts
5. Introducing incentives and subscription plans
6. Implementing a loyalty program
7. Conducting regular customer feedback surveys
8. Offering personalized plans and services
9. Investing in employee training
10. Monitoring industry trends and competitors

By implementing these strategies, Connecttel can:

* Improve customer satisfaction
* Reduce churn
* Increase retention
* Drive business growth and success

## **Conclusion**

* XGBC, RFC, and DTC are the top-performing models for predicting churn in an imbalanced dataset.
* They demonstrate a good balance of metrics, including low FN rates, high recall, high AUC/ROC scores, high F1-scores, and high precision rates.
* These models are suitable for identifying true positives (churners) while minimizing errors.