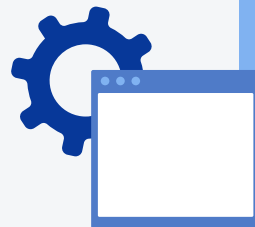# TELECOM CUSTOMER CHURN

TELECOM CUSTOMER CHURN Defination

Telco customer churn refers to the loss of subscribers in telecommunications, either through service cancellations or switching to other providers, posing a significant challenge by leading to revenue loss and impeding potential growth

OBJECTIVE

❑ Create visualizations and extract a set of insights to understand the reasons that led users to leave the company.

❑ Address the problem of imbalanced datasets using different techniques.

❑ Implement a set of machine and deep learning models, compare the results, and choose the best model.
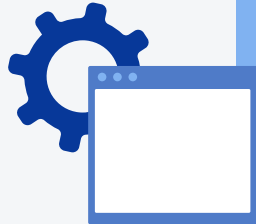
DATA DESCRIPTION

The dataset contains twenty one columns (independent variables) that indicate the characteristics of the clients of a fictional telecommunications corporation. The Churn column (response variable) indicates whether the customer departed within the last month or not. The class No includes the clients that did not leave the company last month, while the class Yes contains the clients that decided to terminate their relations with the company.

7043 Samples	21 Features	26.5% Churn

DATA EXPLORATION

❑ The percentage of male in data is :50.48% and The percentage of female in data is :49.52%

❑ The percentage of Customers who their age less than 65 is :83.79% and The percentage of Customers who their age less than 65 is :16.21%

❑ The percentage of customers that have Dependents :29.96% , The percentage of customers that have no Dependents :70.04%

❑ The percentage of customers using phone services is : 90.32% and The percentage of customers not using phone services is : 9.68%

❑ The percentage of customers have left the company after one year of usage is : 31.04 % and The percentage of customers have left the company after two year of usage is : 45.58 %

❑ The percentage of customers have left the company after three year of usage is : 57.39 % and The percentage of customers have left the company after four year of usage is : 68.21 %

DATA EXPLORATION(CON.)

❑ The percentage of customers have left the company after five year of usage is : 80.02 % and The percentage of customers have left the company after six year of usage is : 100.0 %

❑ Most customers prefer to use 'Fiber optic' then 'DSL'

❑ The percentage of customers who not using the OnlineSecurity because no internet service is :21.67% and The percentage of customers who not using the OnlineSecurity despite of using internet service is :49.67%

❑ The percentage of customers who not using the TechSupport because no internet service is :21.67% and The percentage of customers who not using the TechSupport of using internet service is :49.31%

❑ The percentage of customers who leaving the company is : 26.54% and The percentage of customers who not leaving the company is : 73.46%

WHY THEY LEFT..
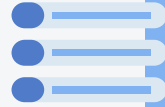
- ❑ suggests that having a partner may contribute to greater customer stability within the company because 17% of customers who do not have partners left the company, out of the total 26% of customers who left

- ❑ The company should review its phone services as 24% of users who utilize these services have left out of the total 26% of customers who left

- ❑ The company should investigate its internet services, especially for customers using 'Fiber optic', as 18.4% of users with this service left the company out of the total 26% who left

- ❑ The company should emphasize the importance of security services to customers, as I've observed that 21% of customers who do not use security services left the company

- ❑ The company should emphasize the importance of device protection services to customers, as I've observed that 17.2% of customers who do not use device protection services left the company

WHY THEY LEFT(CON.)

❑ The company should emphasize the importance of device protection services to customers, as I've observed that 17.2% of customers who do not use device protection services left the company

❑ The company should emphasize the importance of tech support services to customers, as I've observed that 20.5% of customers who do not use tech support services left the company

❑ The company should take note that there may be an issue with the 'Month-to-month' contract, as 24.5% out of the total 26% of users who left the company were on this type of contract.

❑ The suggestion is to review and address potential issues in the contract structure. It emphasizes the importance of thorough testing for new customers and recommends improving existing services and introducing attractive offerings to enhance customer satisfaction, thereby reducing the likelihood of customer churn.

DATA Preprocessing

❑ Drop the customerID column because not important

❑ Replace the values of 'No internet service' with 'No' in multiple columns

❑ Convert values in certain columns from categorical to integer

❑ Perform one-hot encoding on 'InternetService,' 'Contract,' and 'PaymentMethod'.

❑ Apply min-max scaling to the columns: 'tenure,' 'MonthlyCharges,' and 'TotalChar'

❑ Split the data into "X" and "Y"

❑ Split 'X','Y' into x_train , y_train,x_test,y_test

MODELS

❑ **Deep Learning Model :** I built a deep neural network with an input layer of 26 nodes and ReLU activation, followed by two hidden layers with 30 and 15 nodes, both using ReLU activation, and an output layer with 1 node and sigmoid activation for binary classification .

❑ **Enhancement of Deep Learning Model:** I enhanced the performance of the deep learning model by addressing the imbalance in the dataset using techniques such as oversampling, undersampling, and SMOTE.

❑ **Machine Learning Model :** I constructed a set of machine learning models, including logistic regression, KNN, SVM, Decision Tree, and Random Forest.

RESULT ANALYSIS

78% **Deep Learning model**	**76%** **Deep Learning with UnderSampling**	**87%** **Deep Learning with OverSampling**
85% **Deep Learning with SMOTE**	**74%** **Deep Learning with Ensemble**	**79%** **Decision Tree**
80% **KNN**	**83%** **Random Forest**	**82%** **SVM**

CONCLUSION

❑ The first trial achieved an overall accuracy of 78%, and the model demonstrated good generalization on unseen data. However, precision, recall, and F1-score for class 1 were suboptimal due to dataset imbalance, with class 1 being the minority class. To address this issue, I enhanced the model's performance by employing various imbalance-solving techniques, including Oversampling, Undersampling, SMOTE, and Ensemble methods.

❑ After applying undersampling, we observed improvements in recall and F1-score for class 1; however, the overall accuracy has not improved and decrease to 76%.

❑ After applying Oversampling, we observed improvements in recall , precision and F1-score for class 1; and the overall accuracy improved and increase to 87%.
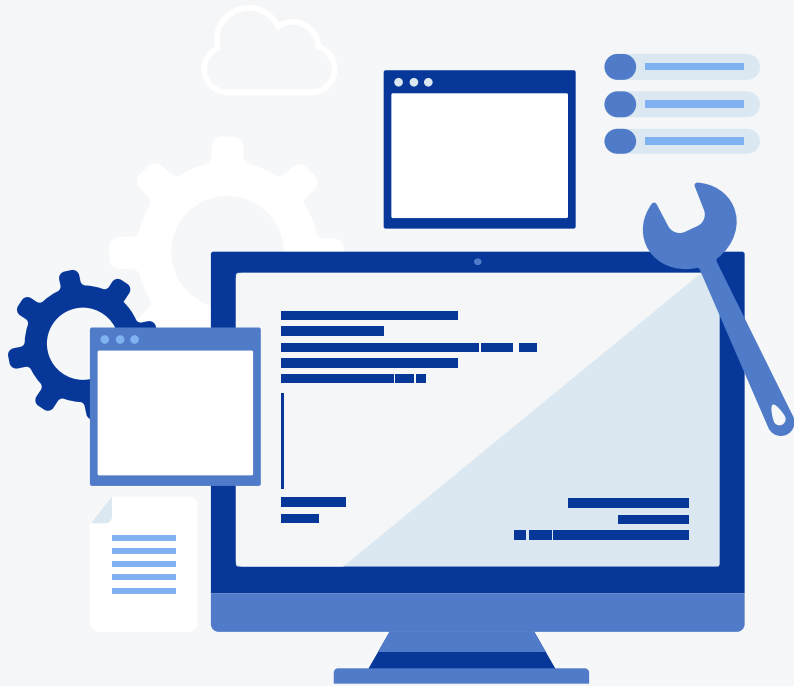
CONCLUSION(CON.)

❑ After applying SMOTE technique, I observed improvements in recall , precision and F1-score for class 1; and the overall accuracy improved and increase to 85%.

❑ After applying Ensemble methods, I found that the model did not improve, and the overall accuracy decreased to below 78%.

❑ I applied a set of machine learning models with tuning ; (DecisionTree: Accuracy - 0.7904) - (RandomForest: Accuracy - 0.8296) - (KNeighbors: Accuracy - 0.8049) - (SVM: Accuracy - 0.8170)

❑ We found that the results improved after addressing the imbalance issue, and the best model is the deep learning model with imbalance resolution using oversampling. The overall accuracy reached 87%.

REFERENCES

❑ Abou dataset :
https://community.ibm.com/community/user/businessanalytics/blogs/steven-macko/2019/07/11/telco-customer-churn-1113

❑ Project URL : https://community.ibm.com/community/user/businessanalytics/blogs/steven-macko/2019/07/11/telco-customer-churn-1113

Thanks