**IBM NAAN MUDHALVAN**

**DOMAIN – DATA ANALYTICS WITH COGNOS**

|  |  |
| --- | --- |
| **DATE** | 11.10.2023 |
| **TEAM ID** | PROJ\_224694\_TEAM\_2 |
| **PROJECT TITLE** | CUSTOMER CHURN PREDICTION-TELECOMMUNICATION CUSTOMER CHURN PREDICTION |

**PROBLEM:**

To predict the customers who are likely to churn in the telecom industry.

**PROBLEM DEFINITION:**

* Telecommunication companies (telco’s) have a significant problem with customer churn, which is the loss of customers who stop using their services.
* To solve this problem, telco’s need to identify the customers who are likely to churn and take preemptive measures to retain them.

**SOLUTION:**

* Import necessary libraries and then understand the data.
* Remove inconsistent or erroneous records from the dataset – Data Cleaning.
* Perform Exploratory data Analysis (EDA) using visualization techniques.
* Predict the outcome using machine learning models like Logistic Regression, Random forest, Gradient Boost XGBoost and CatBoost.
* CatBoost(Categorical Boost) can predict extraordinally faster than other models.

**STEPS:**

1. Data Exploration
2. Exploratory Data Analysis(EDA) and Data Cleaning
3. Data Visualization
4. Feature Engineering
5. Data splitting
6. Comparing different ML Models
7. Model Selection : CatBoost
8. Hyperparameter tuning
9. Check model’s performance

**FLOW CHART:**

**Data Exploration:**

Get data from dataset <https://www.kaggle.com/datasets/blastchar/telco-customer-churn> & store data in panda DataFrame by pd.read\_csv() function.

**EDA & Data Cleaning:**

Understand the data in the dataset related to Demographics, Account details, usage patterns. Also clean the data.

**Data Visualization:**

Visualize demographic service & customer account information as bar graphs, histogram & stacked plots

**Feature Engineering:**

Extract features from raw data. Covert categorical data into numerical data

**Data Splitting:**

Split the dataset as training data and testing data

**Comparing different ML Models:**

Compare the machine learning models like Logistic regression ,Random forest, Gradient boost, XGBoost and CatBoost.

And CatBoost

**Model Selection :CatBoost**

CatBoost performs well compared to other models

**Hyperparameter tuning :**

This is done to get better performance of model (CatBoost)

**Check model’s performance:**

Evaluate CatBoost model’s performance

**Data Exploration:**

* Import pandas , numpy for data handling and matplotlib , seaborn for data visualization.
* Dataset has 21 columns out of which 19 are independent variables.
* “Churn” is a dependent variable

**EDA & Data cleaning:**

* Understand Customer Demographic , Customer Account , Customer Service informations
* Clean data ,identify patterns & inconsistencies.

**Data visualization:**

* Understand and analyze data by various visualization methods like bar graphs , histogram & stacked plots.
* Use gender , SeniorCitizen , partner features for it.

**Feature Engineering:**

* Replace the categorical labels gender , partner , dependents , paperlessBilling , phone service , churn with a numeric value.
* Rescale the categorical features between[0,1]

**Data Splitting:**

* Dataset is splited into training & test data
* Training data is used for model building
* Test data is used for evaluating model’s performance.

**Comparing different ML Model :**

* Compare algorithm like

i Logistic regresssion

ii. CatBoost

iii. Random forest

iv.Gradient Boost

v. XGBoost

* And then predict their accuracies

**Model Selection**:

* Catboost can predict extraordinally aster than XGBoost.
* It also reduces the prediction time.

**Hyperparameter tuning:**

* Hyperparameter tuning is done to get better performance of model
* Techniques to find best hyperparameters for model.

i.Grid search

ii.Random search

iii. Bayesian optimization.

**Check model’s performance:**

* Evaluate model’s performance by two ways

i. Confusion matrix.

ii. evaluation matrics.

|  |  |
| --- | --- |
| **TEAM ID** | **PROJ\_224694\_TEAM\_2** |
| **TEAM MEMBERS** | **S.RAJARAJESHWARI**  **SANDHIYA.S**  **PRIYADHARSHINI .S**  **KOKILA.U**  **VISHNUPRIYA P** |
| **NAME** | **PRIYADHARSHINI S** |
| **REGISTER NUMBER** | **420421205015** |