

**DATA SCIENCE INSTITUTE**

**HACKATHON QUESTION**

**Problem Statement:**Using the telecom customer churn dataset, your task is to develop a machine learning model that accurately predicts customer churn based on demographic, contract, and service usage features.

**Steps to Follow:**

1. Perform data preprocessing: handle missing values, encode categorical variables, and normalize data.
2. Perform exploratory data analysis (EDA) to identify important trends.
3. Build machine learning models (Logistic Regression, Random Forest, and XGBoost).
4. Compare model performance using metrics like accuracy, precision, recall, F1-score, and ROC-AUC.

**Bonus Challenge:**Identify key churn drivers using feature importance analysis and suggest strategies to reduce churn.

**DATASET LINK:**

<https://github.com/Nextonit/Nextonit-Projects/blob/main/Telco_Customer_Churn.csv>

**Upload Your Solution File Here:**

<https://forms.gle/KcSBFadRea25LLzaA>

**Nextonit Data Science Course Details**

<https://github.com/Nextonit/Nextonit-Projects/blob/main/Nextonit%20Course%20Details.pdf>

Here’s an explanation of each column in the dataset:

1. **customerID**: Unique identifier for each customer.
2. **gender**: Gender of the customer (Male/Female).
3. **SeniorCitizen**: Indicator if the customer is a senior citizen (1 = Yes, 0 = No).
4. **Partner**: Indicates if the customer has a partner (Yes/No).
5. **Dependents**: Indicates if the customer has dependents (Yes/No).
6. **tenure**: Number of months the customer has been with the company.
7. **PhoneService**: Indicates if the customer has a phone service (Yes/No).
8. **MultipleLines**: Indicates if the customer has multiple phone lines (Yes/No).
9. **InternetService**: Type of internet service (DSL, Fiber optic, No service).
10. **OnlineSecurity**: Indicates if the customer has online security (Yes/No).
11. **OnlineBackup**: Indicates if the customer has online backup (Yes/No).
12. **DeviceProtection**: Indicates if the customer has device protection (Yes/No).
13. **TechSupport**: Indicates if the customer has tech support (Yes/No).
14. **StreamingTV**: Indicates if the customer has streaming TV services (Yes/No).
15. **StreamingMovies**: Indicates if the customer has streaming movie services (Yes/No).
16. **Contract**: Type of contract (Month-to-month, One year, Two year).
17. **PaperlessBilling**: Indicates if the customer has paperless billing (Yes/No).
18. **PaymentMethod**: The method of payment (e.g., Electronic check, Mailed check).
19. **MonthlyCharges**: Amount charged to the customer each month.
20. **TotalCharges**: Total amount charged to the customer over the tenure.
21. **Churn**: Target variable indicating if the customer has churned (Yes/No).