Assignment

Objective: to build a customer churn prediction model that will predict if a customer will unsubscribe or stop using a service. Use Gradient Boosting Classifier, Decision Tree Classifier, and Random Forest Classifier.

Dataset

<https://www.kaggle.com/datasets/barun2104/telecom-churn>

Step 1:

State the objective of the project you're building.

Step 2: Explain your dataset. What are the features of the dataset, and what does each column represent or mean.

Step 3:

Load all the needed libraries.

Step 4:

Lead your dataset

Step 5:

Explore your dataset. Check the following and state your observations in each of them.

df.shape

df.info()

df.head()

df.tail()

df.isnull().sum()

df.dtypes

df.describe()

df.describe(include='object')

df.duplicated().sum()

Check the value count for all the columns: hint;

for col in df.columns:

print(f"\nColumn: {col}")

print(df[col].value\_counts(dropna=False))

Step 6:

If there are missing values, duplicate rows, incorrect datatype, characters in values (e.g @ -<>+), correct them.

Step 7:

Transform categorical columns to numerical

Step 8:

Drop the target column and any other column that is not relevant

Step 9:

Split your dataset into training and testing

Step 10:

Scale your dataset using standard scaler

Step 11:

Build your models

Step 12:

Evaluate each model performance

Step 13:

Explain your observations of each model performance and state your best performing model