Genztechs
Data Science Juniors
Final Project

Project Title: Predictive Analytics for Customer Churn in a Subscription-Based Business

Objective:

Build a predictive model to identify customers who are likely to churn (cancel their subscription).

Project Steps:

1. Data Collection

- Use a publicly available dataset (e.g., Telco Customer Churn on Kaggle).

2. Exploratory Data Analysis (EDA)

- Analyze the dataset to understand patterns and correlations.
- Visualize key insights:
    - Distribution of churn vs. non-churn customers.
    - Correlation between features and churn.
    - Trends in customer behavior (e.g., high churn among low-tenure customers).
- Create visualization using matplotlib and PowerBi.

3. Data Preprocessing

- Handle missing values and outliers.
- Encode categorical variables (e.g., one-hot encoding).
- Normalize/scale numerical features.
- Split data into training and testing sets.

(Perform any of the pre-processing step or technique if necessary)

4. Model Building

- Train and evaluate machine learning models:

    - Logistic Regression.
    - Random Forest.
- Use metrics like accuracy, precision, recall, F1-score, and ROC-AUC to evaluate models.

5. Model Interpretation

- Use SHAP or LIME to explain model predictions.
- Identify key drivers of churn (e.g., high monthly charges, poor customer support).