Updated CDAC Final Project Roadmap - Customer Churn Analysis

Project Goal

Predict whether a customer will churn and analyze key reasons for churn to guide business strategy.

Phase 1: Setup & Data Understanding

- Load dataset using Pandas
- Check data types, structure, and missing values using df.info(), df.head(), df.isnull().sum()

Phase 2: Data Cleaning

- Convert TotalCharges to numeric
- Drop missing values if any
- Encode binary (Yes/No, Male/Female) columns
- One-hot encode multi-category columns
- Normalize numeric columns using StandardScaler

Phase 3: Exploratory Data Analysis (EDA)

- Churn rate overall
- Visualize churn vs: Gender, SeniorCitizen, Contract, InternetService, etc.
- Boxplots for charges
- Correlation heatmap

Phase 4: Model Building (Classification Models)

- Logistic Regression
- Decision Tree
- Random Forest
- XGBoost
- SVM (Linear and RBF)
- Naive Bayes
- K-Nearest Neighbors (KNN)

Updated CDAC Final Project Roadmap - Customer Churn Analysis

Train/test split using train_test_split

Phase 5: Model Evaluation

- Accuracy, Precision, Recall, F1-score
- Confusion Matrix
- ROC-AUC Curve
- Log Loss
- Use classification_report

Phase 6: Unsupervised Learning

- KMeans Clustering
- Elbow method for optimal K
- Cluster visualization
- PCA for dimensionality reduction

Phase 7: Regression (Optional)

- Ridge Regression
- Lasso Regression
- Polynomial Regression

Phase 8: Business Insights

- Identify top churn drivers
- Segment high-risk customers
- Recommend targeted retention strategies

Phase 9: Dashboard / App (Optional)

- Streamlit app or Power BI dashboard
- Upload input to predict churn
- Visual analytics for stakeholders

Updated CDAC Final Project Roadmap - Customer Churn Analysis

Phase 10: Documentation & Submission

- Final report (PDF or Word)
- 8-10 slide PPT for viva
- Cleaned dataset and notebook
- Optional: deployed app link