Customer Churn Prediction Report

Team Members

- Nourhan Khaled
- Fatma Ehab
- Hend Labib
- Mayar Mustafa
- Mona Elgoba

Dataset Description

- Source: Telco Customer Churn Dataset
- Records: ~7,043 rows
- Features: Customer attributes like gender, tenure, billing, and services
- Target Variable: Churn (Yes/No)

Pre-processing and Visualization

- Handled nulls in TotalCharges
- Encoded binary and categorical features
- Scaled numerical features using RobustScaler and MinMaxScaler
- Visualized distributions and churn patterns

Methods (Machine Learning Models)

- Logistic Regression, XGBoost , LightGBM, Stacking Classifier.
- Used ColumnTransformer and Pipeline.
- Hyperparameter tuning with GridSearchCV.
- Handled imbalance using class_weight, SMOTE, and scale_pos_weight.

Evaluation

Models were assessed with emphasis on recall for Churn = 1 due to the business goal of minimizing lost customers. Logistic Regression with SMOTE achieved the best balance with a recall of 0.79 and precision of 0.50. This model is recommended for deployment.

Deployment

- Deployed via Gradio.
- Gradio interface with interactive form and styled prediction display.