**Narayanam Naga Vamsi Sandeep**

**WSU id: X638B786**

**BSAN 735 – Project 24 – Step -1**

**Data****set link:** [Telecom Customer data Churn Analysis](https://data.world/bob-wakefield/call-center-data)

The Churn Dataset consists of 22 columns with 12892 rows.

**Dataset Description:**

The dataset contains information about telecom customers, including their demographic details, account information, and service usage. The dependent variable is 'churn,' which indicates whether a customer has left the service provider. Independent variables include features such as customer age, gender, tenure with the company, service usage patterns, charges, and contract details.

**Dependent Variable (Target):**

This is a classification problem. The dependent appears to be churn, which indicates whether a customer has churned (left the service) or not.

**Independent Variables:** There are 19 independent variables.

The independent variables in the telecom churn dataset are as follows:

* state
* account\_length
* area\_code
* international\_plan
* voice\_mail\_plan
* number\_vmail\_messages
* total\_day\_minutes
* total\_day\_calls
* total\_day\_charge
* total\_eve\_minutes
* total\_eve\_calls
* total\_eve\_charge
* total\_night\_minutes
* total\_night\_calls
* total\_night\_charge
* total\_intl\_minutes
* total\_intl\_calls
* total\_intl\_charge
* number\_customer\_service\_calls

**Task of Interest:**

The primary task of interest for this project is predicting customer churn in the telecom industry. This involves developing a classification model that can accurately identify customers who are likely to discontinue their service with the telecom provider. The goal is to use the model's predictions to implement targeted retention strategies, thereby reducing churn rates and increasing customer loyalty and profitability.

**Analysis Plan:**

* **Data Preprocessing:** Clean the data by handling missing values and outliers, encoding categorical variables, and normalizing numerical features.
* **Exploratory Data Analysis (EDA):** Perform an initial analysis to understand the distribution of variables, identify patterns, and explore correlations between features and churn.
* **Feature Engineering:** Enhance the dataset by creating new features and selecting the most relevant ones to improve model performance.
* **Model Building:** Split the data into training and testing sets, experiment with various classification algorithms, and tune hyperparameters to optimize model performance.
* **Model Evaluation:** Assess the model's performance using metrics such as accuracy, precision, recall, F1-score, and ROC-AUC, and compare different models to select the best one.
* **Model Interpretation:** Analyze feature importance to understand the factors contributing to churn and interpret model predictions to derive actionable insights.
* **Implementation and Monitoring:** Deploy the selected model into production, monitor its performance over time, and update it as necessary to maintain accuracy.