**Student Name:** Ajai Balaji R  
**Register Number:** 510923243301  
**Institution:** Global institute of engineering and technology  
**Department:** B.tech[ Artificial intelligence and data science]  
**Date of Submission:** [02/05/2025]

GitHub link: <https://github.com/Aji2206/AJ.NM_course.git>

**Title:**

**Customer Churn Prediction Using Machine Learning to Uncov**er Hidden Patterns

1. **Problem Statement**

Customer churn poses a significant threat to business profitability, especially in subscription-based services. This project seeks to utilize machine learning techniques to predict whether a customer is likely to churn, based on historical and behavioral data, with the goal of uncovering hidden patterns that contribute to customer attrition.

1. **Objective of the Project**

Predict the likelihood of customer churn using machine learning models.

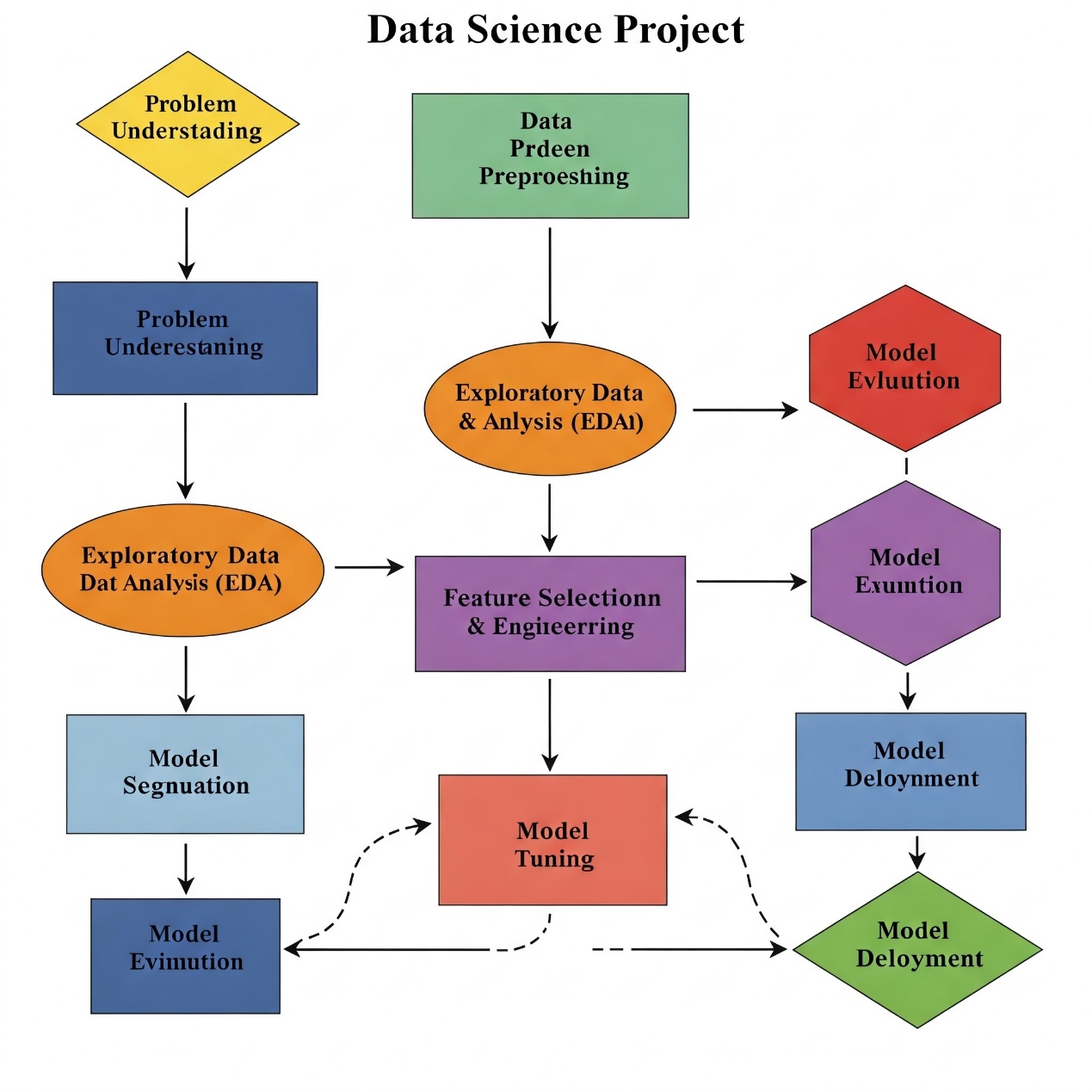
Identify key factors that contribute to churn.

Analyze customer behavior trends and derive actionable insights.

Help businesses improve retention strategies through data-driven decision-making.

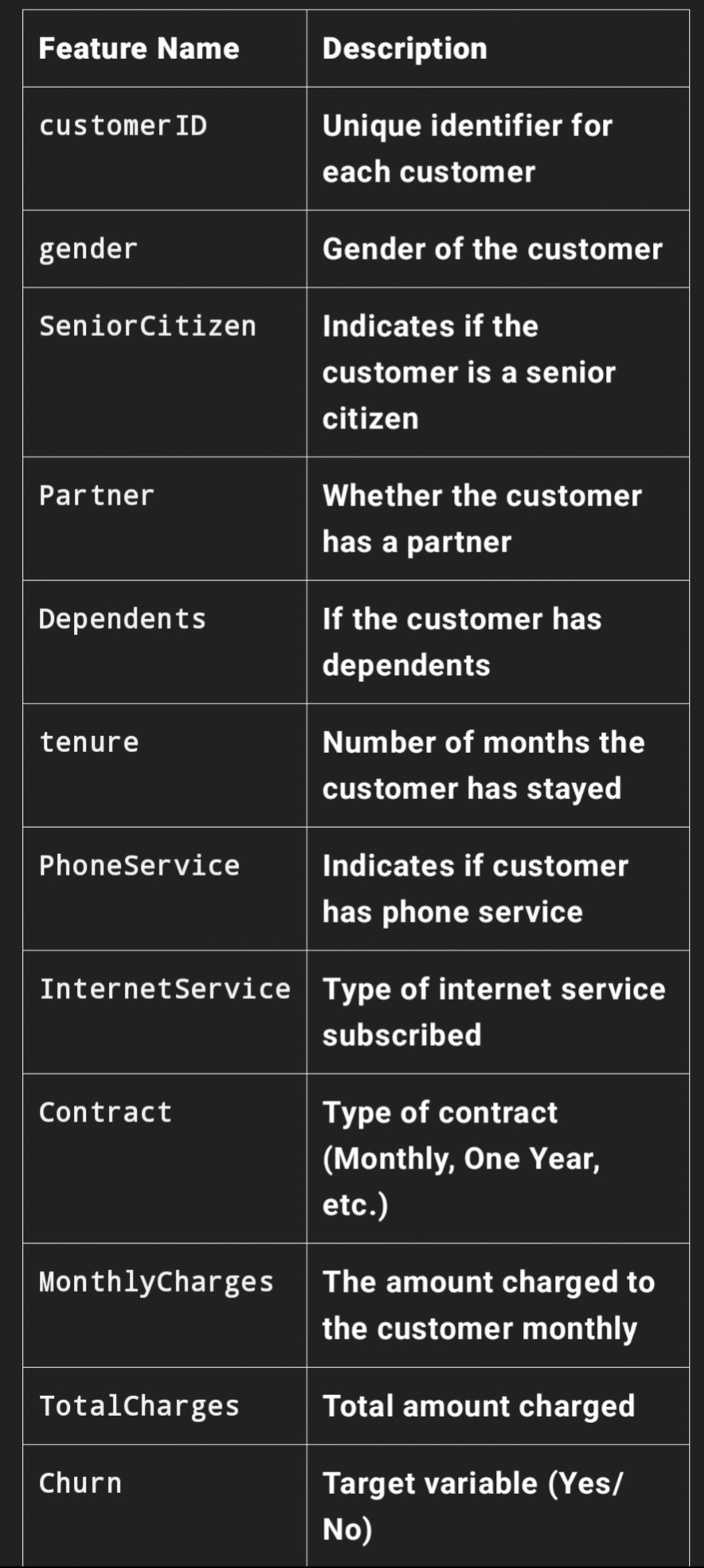
1. **Project Workflow**

Flow of the project:



1. Problem Understanding
2. Data Collection
3. Data Preprocessing
4. Exploratory Data Analysis (EDA)
5. Feature Selection & Engineering
6. Model Selection & Training
7. Model Evaluation
8. Model Tuning
9. Model Deployment
10. **Data Description**

Sample fields in a typical churn dataset:



1. **Data Processing**

Remove or impute missing values (e.g., blank TotalCharges)

Convert categorical data into numeric format using encoding techniques:

Label Encoding for binary columns

One-Hot Encoding for multi-class columns

Normalize numerical features (e.g., tenure, MonthlyCharges)

Split dataset into training and testing sets (e.g., 80/20 split)

1. **Exploratory Data Analysis (EDA)**

Analyze churn distribution (churn vs non-churn)

Visualize feature distributions using histograms, boxplots

Use heatmaps to identify correlation between features

Churn rate by:

Contract type

Internet service

Customer tenure

1. **Tools and Technologies Used**

Programming Language: Python

Data Processing: Pandas, NumPy

Visualization: Matplotlib, Seaborn

Modeling: Scikit-learn, XGBoost, Random Forest

Environment: Jupyter Notebook / Google Colab

Optional Deployment: Flask / Streamlit for web-based dashboard

1. **Team members:**

**Ajai balaji R**

**Bhuvaneshwaran p**

**Nithish R**

**Vignesh baskar s**