Michael Mohab Soltan Nashed (eng.michaelsoltan@gmail.com)

Data Science Project Proposal

1. Project Description

This project, titled 'Customer Churn Prediction and Analysis', aims to build a machine learning model to predict customer churn using the Telco Customer Churn dataset from Kaggle. The goal is to help telecommunication companies identify customers likely to leave, allowing proactive retention strategies. Through data preprocessing, exploratory analysis, feature engineering, model training, and deployment, this project follows the IBM Data Science Round 3 framework to deliver an accurate, scalable, and business-relevant churn prediction system.

2. Team Members & Roles

Name	Email	Student ID	Role
Michael Mohab Soltan Nashed (Leader)	eng.michaelsoltan@gmail.com	21077122	Project Lead & Data Collection & Preprocessing
Hazem Mohamed Salem	Hazem.mohamedsm@gmail.com	21041541	Model Deployment
Marwan Ahmed Salah Ahmed	marwanshehata328@gmail.com	21083763	EDA & Data Visualization
Kareem Moataz Mostafa Mohamed Elmahdy	Kareemmoataz16@gmail.com	21049214	Feature Engineering & Model Optimization
Ibrahim Emam Mahmoud	Ibrahimelnogomy7@gmail.com	21000650	Model Development & Evaluation
Mohamed Tawfik	Mohammed_20230458@fci.helwan.edu.eg	21093334	Documentation & Reporting

Hossam El
Din

3. Objectives

- Collect and preprocess customer churn data from Kaggle.
- Perform exploratory data analysis (EDA) to understand key churn patterns.
- Engineer meaningful features to improve predictive performance.
- Develop, train, and optimize classification models for churn prediction.
- Evaluate model accuracy using metrics like precision, recall, F1-score, and ROC-AUC.
- Deploy the best-performing model as a web service or API.
- Document the workflow, findings, and business insights effectively.

4. Tools & Technologies

The project utilizes the following tools and technologies:

- Programming Languages: Python
- Libraries: pandas, NumPy, matplotlib, seaborn, scikit-learn
- Development Environment: Google Colab
- Deployment Frameworks: TBC
- Version Control: GitHub
- Dataset Source: Kaggle – Telco Customer Churn (blastchar/telco-customer-churn)

5. Milestones & Deadlines

Milestone	Description	Deadline
Milestone 1	Data Collection, Exploration & Preprocessing	12 Oct 2025
Milestone 2	Advanced Data Analysis & Feature Engineering	18 Oct 2025
Milestone 3	Model Development & Optimization	8 Nov 2025
Milestone 4	MLOps, Deployment & Monitoring	25 Nov 2025
Milestone 5	Final Documentation & Presentation	20 Dec 2025

6. Key Performance Indicators (KPIs)

Data Quality:

• Percentage of missing values handled: 100%
• Data accuracy after preprocessing: 98%
• Dataset diversity (representation of categories): 100%

Model Performance:
• Model accuracy (F1-Score): TBC
• Model prediction speed (Latency): TBC
• Error rate (False Positive/Negative Rate): TBC

Deployment & Scalability:
• API uptime: TBC
• Response time per request: TBC

Business Impact & Practical Use:
• Reduction in manual effort: TBC
• Expected cost savings: TBC
• User satisfaction: TBC

7. References

Dataset: Telco Customer Churn – Kaggle
(https://www.kaggle.com/datasets/blastchar/telco-customer-churn)