Healthcare Predictive Analytics Project Proposal

Team Members:

The following team members are all designated as Data Scientists:

- Muuhmd Salah Abd El-Bari (Team Leader)
- Nermine Samir Abd El-Aty
- Toqa Hamed Mohammed
- George Emil Sadik
- Mostafa Mohammed Fouad
- Omar Ahmed Mostafa

Overview:

This project aims to develop a healthcare predictive analytics system using data science and machine learning techniques to enhance early risk assessment and support clinical decision-making. By analyzing patient demographic, clinical, and lifestyle data, the system identifies patterns associated with potential health risks.

Within this broader healthcare context, the project focuses on stroke risk prediction as a critical application, enabling the identification of high-risk individuals and supporting preventive healthcare and data-driven treatment strategies.

Project Objectives:

- Analyze healthcare data to identify key risk factors associated with stroke
- Perform exploratory data analysis to understand medical and lifestyle patterns
- Build and evaluate machine learning models for stroke risk prediction
- Deliver an accurate and interpretable healthcare prediction system

Scope:

➢ **Target Audience**

This project is designed for healthcare professionals, hospitals, clinical centers, and medical research institutions seeking to enhance clinical decision-making through data-driven and AI-based stroke risk assessment tools.

➢ **Data Sources**

This project utilizes a publicly available healthcare dataset obtained from Kaggle, which contains anonymized patient demographic, clinical, and lifestyle information. The dataset includes attributes such as age, gender, hypertension status, history of heart disease, average glucose level, body mass index (BMI), smoking habits, and stroke outcome labels.

➢ **Dataset Link**

https://www.kaggle.com/code/bhargavic2025/explainable-stroke-risk-prediction-healthcare-ml

➢ **Key Features**

• Machine learning–based stroke risk classification

• Exploratory data analysis and visualization of clinical risk factors

• Explainable AI techniques to interpret feature importance and model predictions

• Scalable model deployment and experiment tracking using MLOps practices

Tools & Technologies

• Programming Language: Python

• Libraries: Pandas, NumPy, Matplotlib, Seaborn

• Machine Learning Models: Logistic Regression, Decision Trees, Random Forest

• Database: SQL

Final Milestones Summary:

Milestone	Key Deliverables
1. Data Collection, Exploration & Preprocessing	EDA Report, Interactive Visualizations, Cleaned Dataset
2. Data Analysis, Visualization & Feature Engineering	Data Analysis Report, Visualizations of Health Trends, Feature Engineering Summary
3. Model Development & Optimization	Model Evaluation Report, Model Code, Final Model
4. MLOps, Deployment & Monitoring	Deployed Model, MLOps Report, Monitoring Setup
5. Final Documentation & Presentation	Final Project Report, Final Presentation

----

Conclusion:

The Healthcare Predictive Analytics project leverages machine learning to predict patient risks and health outcomes, offering valuable insights that healthcare professionals can use for improving patient care. By focusing on data exploration, predictive modeling, and deployment, this project ensures the development of a functional and scalable system for healthcare decision-making.