Project Planning & Management

Project Proposal

- **Title:** Healthcare Predictive Analytics

- **Objective:** Develop a machine learning model to predict patient health risks, assisting healthcare professionals in decision-making.

- **Scope:**

    o Collect and preprocess healthcare datasets.

    o Perform exploratory and statistical data analysis.

    o Train and optimize predictive models.

    o Deploy a cloud-based web service for predictions.

    o Monitor and refine model performance over time.

Milestones & Timeline

Milestone	Task	Deliverables	Deadline
Data Collection & Preprocessing	Collect datasets, clean & preprocess	EDA report, cleaned dataset, visualizations	1/4/2025
Advanced Data Analysis & Feature Engineering	Perform statistical analysis & create features	Analysis report, engineered dataset	8/4/2025
Model Development & Optimization	Train & optimize ML models	Model evaluation report, final trained model	20/4/2025
Deployment & Monitoring	Deploy model & set up monitoring	Deployed model, MLOps report	30/4/2025
Final Documentation & Presentation	Prepare final report & presentation	Final project report, presentation slides	7/5/2025

Task Assignment & Roles

Task	Team Member
Data Collection & Cleaning	Alaa, Ahmed Osama
Exploratory Data Analysis (EDA)	Esraa, Abanoup
Statistical Analysis & Pattern Recognition	Mohamed, Ahmed Sameh
Feature Engineering & Transformation	Esraa, Alaa
Model Training & Evaluation	Abanoup, Ahmed Osama
Hyperparameter Tuning & Feature Selection	Ahmed Sameh, Esraa
Model Deployment	Alaa, Mohamed
Performance Monitoring & Maintenance	Mohamed, Abanoup
Project Report & Documentation	Ahmed Osama, Ahmed Sameh
Presentation & Stakeholder Communication	Esraa, Alaa

Risk assessment & Mitigation

Risk	Impact	Mitigation Strategy
Low-quality or insufficient data	High	Use multiple datasets from different sources (Kaggle, MIMIC-III, etc.), perform data augmentation if needed.
Data preprocessing challenges	Medium	Automate cleaning steps using Python scripts, document all transformations for reproducibility.
Feature engineering complexity	Medium	Conduct iterative feature selection using domain knowledge & statistical tests.
Model underperformance	High	Experiment with multiple algorithms, apply hyperparameter tuning, and ensemble techniques.
Overfitting on training data	High	Use cross-validation, regularization techniques, and ensure a well-balanced dataset.
Computational resource limitations	Medium	Utilize cloud-based services (Google Colab, AWS, or local GPU) for training heavy models.
Deployment & monitoring issues	Low	If deployment is required, use lightweight tools like Streamlit for interactive results without complex backend setup.
Lack of explainability in model predictions	Medium	Use SHAP, LIME, or other explainability techniques to interpret results for healthcare professionals.

KPIs (Key Performance Indicators)

- Model Accuracy: **≥85%**

- Prediction Response Time: **<2 sec per request**

- User Adoption: **At least 10 test users provide feedback**

Literature Review

1. Disease Prediction Using AI

- AI is used in healthcare to **predict diseases** early.

- Many studies focus on using **Machine Learning (ML)** to analyze medical data.

2. Common Machine Learning Models

- **Random Forest**: Uses multiple decision trees to make better predictions.

- **SVM (Support Vector Machine)**: Good for classifying diseases.

- **Neural Networks**: Mimic the human brain for complex predictions.

3. Medical Datasets Used

- Researchers use datasets like:
    - **Heart Disease Dataset**
    - **Diabetes Dataset**
    - **Cancer Prediction Data**

4. Limitations of Previous Research

- Some models are not easy for doctors to understand.

- Some studies use small datasets, which may not be accurate.

5. How Our Project is Different

- We will try to use **more reliable data**.

- We will focus on making the model easy to understand for doctors.

Requirement Gathering

1. Stakeholder Analysis

Key stakeholders and their needs:

- **Patients:** Want an accurate prediction of disease risks based on their health data.
- **Doctors:** Need a dashboard that displays patient risk analysis to assist in making better medical decisions.
- **Hospitals:** Aim to improve healthcare quality and reduce treatment costs by predicting diseases.
- **Healthcare Researchers:** Can use the system to analyze health data and discover new disease patterns.

2. User Stories & Use Cases

Potential user stories:

- **As a patient,** I want to input my health data (such as age, blood pressure, sugar level, etc.) to receive a prediction of my disease risk.
- **As a doctor,** I want a dashboard that provides a risk analysis of patients based on their health data to help me make accurate medical decisions.
- **As a hospital,** I want to receive analytical reports on patients who are likely to develop chronic diseases so that preventive measures can be taken.

3. Functional Requirements

- Input patient data (manually or via CSV files).
- Process and analyze data using AI algorithms (such as Random Forest, SVM, Neural Networks).
- Display prediction results in a visual format (graphs, tables, charts).
- Provide recommendations to doctors based on the analysis results.

4. Non-functional Requirements

- **Performance:** The system should provide results within **5 seconds** per analysis request.
- **Security:** Patient data should be encrypted and access to sensitive information should be protected.
- **Usability:** The user interface should be intuitive and easy to navigate for both doctors and patients.
- **Scalability:** The system should handle a large number of users and datasets without performance degradation.