

# 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:**
  - Collect and preprocess healthcare datasets.
  - Perform exploratory and statistical data analysis.
  - Train and optimize predictive models.
  - Deploy a cloud-based web service for predictions.
  - 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**