

# EYouth X DEPI Project Proposal



## **Our supervisor:**

Mahmoud Talaat

## **Submitted by:**

Moaaz Elhaiys (Team leader)

Farida Yasser

Malak Abdelmawgoud

Zahraa Hamdy

Mohamed Alaa

Mohamed Hassanein

# Proposal for Healthcare Predictive Analytics Project

## 1. Project Overview

The Healthcare Predictive Analytics Project aims to leverage machine learning models to improve healthcare outcomes by providing data-driven insights. This project focuses on developing a predictive model that assists healthcare professionals in patient risk assessment, trend identification, and data-driven decision-making. The primary goal is to enhance patient care and optimize resource management using predictive analytics.

## 2. Objectives

- Develop a predictive model to forecast healthcare-related outcomes.
- Identify key risk factors affecting patient health.
- Improve decision-making for healthcare professionals through data insights.
- Enhance patient care management using predictive analytics.

## 3. Tools & Technologies

The following tools and technologies will be used throughout the project:

- Programming Language: Python
- Data Exploration & Visualization: Pandas, Matplotlib, Seaborn, Plotly, Tableau
- Machine Learning & Model Development: Scikit-learn, XGBoost
- Version Control & Collaboration: Git, GitHub

## 4. Milestones and Deadlines

Milestone 1: Data Collection, Exploration, and Preprocessing

Deadline: 10/03/2025

This milestone focuses on acquiring and preparing healthcare data for predictive modeling.

Tasks:

1. Data Collection: Gather healthcare data, including patient records, clinical metrics, and lifestyle attributes.
2. Data Exploration: Perform exploratory data analysis (EDA) to identify patterns, trends, and inconsistencies.
3. Data Preprocessing: Handle missing data, normalize/standardize numeric features, and encode categorical variables.

#### Dataset Features:

- HeartDisease (Binary: Yes/No) – Indicates if the patient has heart disease.
- BMI (Numeric) – Body Mass Index, a measure of body fat.
- Smoking (Binary: Yes/No) – Indicates smoking history.
- AlcoholDrinking (Binary: Yes/No) – Heavy alcohol consumption indicator.
- Stroke (Binary: Yes/No) – History of strokes.
- PhysicalHealth (Numeric: 0-30) – Number of days with poor physical health.
- MentalHealth (Numeric: 0-30) – Number of days with poor mental health.
- DiffWalking (Binary: Yes/No) – Difficulty walking or climbing stairs.
- Sex (Categorical: Male/Female) – Biological sex of the patient.
- AgeCategory (Categorical) – Age group classification.
- Race (Categorical) – Ethnicity of the patient (one-hot encoded).
- Diabetic (Categorical: Yes/No/Borderline) – Diabetes diagnosis status.
- PhysicalActivity (Binary: Yes/No) – Engagement in physical activity.
- GenHealth (Categorical: Excellent/Very Good/Good/Fair/Poor) – Self-reported health status.
- SleepTime (Numeric) – Average sleep hours per day.
- Asthma (Binary: Yes/No) – Asthma diagnosis.
- KidneyDisease (Binary: Yes/No) – Kidney disease diagnosis.
- SkinCancer (Binary: Yes/No) – Skin cancer diagnosis.

#### Deliverables:

- Dataset Exploration Report
  - EDA Notebook
  - Cleaned Dataset
-

## Milestone 2: Data Analysis and Visualization

Deadline: 30/03/2025

### Tasks:

1. Data Cleaning: Address remaining missing values, inconsistencies, and outliers.
2. Data Analysis: Identify relationships between health metrics and patient outcomes using statistical methods.
3. Data Visualization: Create compelling visualizations (heatmaps, trend lines, scatter plots) and interactive dashboards.

### Dataset Features:

- Statistical analysis of health conditions and their correlation with disease risk.
- Feature importance analysis to determine influential factors.
- Visual representation of patient health trends.

### Deliverables:

- Cleaned Dataset and Analysis Report
- Visualizations of Health Trends

---

## Milestone 3: Predictive Model Development and Optimization

Deadline: 15/04/2025

### Tasks:

1. Model Selection: Choose appropriate machine learning algorithms for prediction.
2. Model Training: Split data into training and testing sets and train models.
3. Model Evaluation: Assess performance using metrics like accuracy, precision, recall, and ROC-AUC.
4. Model Optimization: Fine-tune hyperparameters to improve model performance.

### Dataset Features:

- Feature selection for machine learning models.
- Model input variables: demographics, health history, and lifestyle factors.

### Deliverables:

- Predictive Model Performance Report
  - Model Code
  - Final Optimized Model
- 

#### Milestone 4: MLOps, Deployment, and Monitoring

Deadline: 25/04/2025

##### Tasks:

1. MLOps Implementation: Use MLflow or Kubeflow for tracking experiments and versioning models.
2. Model Deployment: Deploy the model as a REST API or web application.
3. Model Monitoring: Set up monitoring for performance tracking and model drift detection.
4. Performance Reporting: Generate periodic reports on model performance.

##### Dataset Features:

- Deployment readiness of the model.
- Live performance metrics tracking.

##### Deliverables:

- Deployed Predictive Model
  - MLOps Report
  - Model Monitoring Setup
- 

#### Milestone 5: Final Documentation and Presentation

Deadline: 01/05/2025

##### Tasks:

1. Final Report: Summarize the project, including challenges and key insights.
2. Final Presentation: Create a presentation for healthcare stakeholders showcasing the predictive model's impact.

##### Dataset Features:

- Summary statistics of model performance.
-

- Real-world application case studies.

Deliverables:

- Final Project Report
- Final Presentation

This proposal ensures that the entire project is completed within the stipulated timeframe, with each milestone contributing to the overall success of the Healthcare Predictive Analytics Project.