

## **Project proposal: Prediction of diabetes based on health indicators**

Diabetes is a persistent condition that affects millions of people worldwide. It is crucial to detect and manage diabetes appropriately in order to avoid serious complications. The aim of this project is to create a machine learning model that can predict the risk of developing diabetes based on common health indicators, in order to facilitate early intervention and improve patients' quality of life.

**Problem:** Diabetes can develop late in life, particularly in people who have no symptoms. A prevention system would enable people at risk to be identified and tailored prevention measures to be implemented.

**Objective:** To create a machine learning model that can predict whether a person will develop diabetes using readily available clinical and biological data (age, gender, BMI, fasting blood glucose levels, etc.).

### **Characteristics of this proposal :**

- Relevance: The project meets a crucial public health need.
- Feasibility depends on the availability of the necessary data and the adaptation of machine learning techniques to this type of problem.
- Consequences: The results of the project could help to improve the management of patients suffering from diabetes.

**Relevance to the Sustainable Development Goals (SDGs):** This project contributes directly to achieving Sustainable Development Goal 3: 'Ensure health and well-being for all, at all ages'. This project will help to reduce the morbidity and mortality associated with diabetes by improving the prevention and detection of this disease.

### **Examples of written texts:**

Article 1: *'Early prediction of diabetes mellitus using machine learning methods'*

- This article examines different machine learning methods for predicting diabetes with reference to clinical features such as blood glucose and body mass index.

Article 2: *'A comparative study on the prediction of diabetes mellitus using machine learning algorithms'*

- This article examines the effectiveness of different machine learning algorithms in predicting diabetes, looking at their accuracy, sensitivity and specificity.

### **Data description:**

- References: Medical databases, health studies, medical records.
- The format used will mainly be CSV or SQL for the data.
- The amount of data required to train the model (several thousand patients) will be considerable.
- Preliminary preparation: The data will be processed, enriched and harmonised to remove missing values and inconsistencies.

### **Approach (Machine Learning or Deep Learning):**

- The model will be trained to predict the presence or absence of diabetes using a supervised learning method.
- Methods such as random forests, artificial neural networks or support vector machines will be considered.
- Evaluation: The results of the model will be measured using indicators such as precision, recall, F1 score and ROC curve.

