

## **Detailed project proposal (DPP)**

**Student Number:20068347**

**Student Name: Bala Raghavendra Manikanta Mendi**

**Course: MSc Data Science and Analytics with Advanced Research**

**Supervised by: Clive Chandler**

**Project Title: Predicting Diabetes using Machine Learning Algorithm based on Patients' Symptoms**

### **Aim of the project:**

Diabetes is a chronic condition that affects millions of people throughout the world. Early detection and treatment are critical for avoiding complications and improving patient outcomes. In recent years, machine learning algorithms have demonstrated significant promise in predicting and diagnosing diabetes based on patient symptoms. We hope to create a machine-learning model that can accurately predict diabetes based on the patient's symptoms, such as blood glucose level, weight gain or loss, and insulin level, in this project.

### **Research Question**

**The research question of this project is:** Can machine learning algorithms predict diabetes disease based on symptoms such as blood glucose level, weight growth or loss, and insulin level?

### **Hypothesis**

The hypothesis of this project is: Based on the patient's symptoms, such as blood glucose levels, weight increase or loss, and insulin levels, machine learning algorithms can reliably predict diabetes disease.

### **Objectives**

The primary goal of this project is to create a machine-learning algorithm that can effectively predict diabetes based on patient symptoms. The project's precise goals are as follows:

To collect information from diabetic patients.

To delete any missing or incorrect data points from the gathered data.

To identify the characteristics that are highly associated with diabetes.

With the selected features, create a machine-learning model.

To assess the effectiveness of the developed machine learning model.

## Methodology

The methodology for this project includes the following steps:

**Data Collection:** For this experiment, data will be gathered from diabetic patients. Their blood glucose levels, weight growth or reduction, insulin levels, and any relevant symptoms will all be recorded. Medical records and patient interviews will be used to collect data.

**Data Pre-processing:** Any missing or incorrect data points will be removed from the acquired data. To guarantee that the features are on the same scale, the data will be standardized.

**Feature Selection:** For the machine learning model, features that are highly connected with diabetes will be chosen. Statistical analysis and machine learning approaches will be used to choose the features.

**Model Development:** Using the selected features, we will create a machine-learning model. We will utilize many machine-learning techniques to choose the best algorithm for this project. Cross-validation techniques will be used to train the model on the collected data.

**Model Evaluation:** We will analyze the generated machine learning model's performance using several performance metrics such as accuracy, precision, recall, and F1-score. We will compare the performance of the developed model to that of other cutting-edge models.

## Expected Outcomes

The expected outcomes of this project are:

A machine learning model that can accurately predict diabetes based on patient symptoms.

A dataset of diabetes patients can be used for future studies.

A set of performance metrics for evaluating the developed machine learning model's performance.