**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 the Data world of Kaggle of the 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:** This experiment will gather data from Data World like Kaggle which has patients’ data. Their blood glucose levels, weight growth or reduction, insulin levels, and any relevant symptoms will all be recorded.

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

**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:** We will create a machine-learning model using the selected features. We use different algorithms which will give the best results.

**Model Evaluation**

**Expected Outcomes**

The expected outcomes of this project are:

We will get a better prediction from the model.

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.