SOFTWARE REQUIREMENTS SPECIFICATIONS (SRS)

ON

"DIABETES PREDICTION USING SUPERVISED MACHINE LEARNING"

Submitted to

KIIT Deemed to be University

In Partial Fulfillment of the Requirement for the Award of

BACHELOR'S DEGREE IN INFORMATION TECHNOLOGY

BY

Al Nafis Fuad Shuvo	**20051717**
Ayman Bin Alam	**20051719**
Ferdous Mahmud Fahad	**20051727**
Md Roman Talukdar	**20051734**
Prince Sarker Deganta	**20051745**

UNDER THE GUIDANCE OF

Mrs. Jayanti Dansana

SCHOOL OF COMPUTER ENGINEERING

KALINGA INSTITUTE OF INDUSTRIAL TECHNOLOGY

BHUBANESWAR, ODISHA - 751024

SEPTEMBER 2023

1. Objective

The purpose of this system is to construct a supervised machine learning model to predict whether or not a patient has diabetes. The system will be trained on a dataset of individuals with known diabetes status and different medical parameters, such as blood glucose levels, age, and weight. Once trained, the algorithm will be able to predict the diabetes status of new patients based on their medical characteristics.

2. Requirements

The system should fulfil the following requirements:

Functional requirements:

- The system must be able to predict the diabetes status of new patients with an accuracy of at least 90%.
- The system must be able to manage a dataset of at least 10,000 patients.
- The system must be able to generate forecasts in real time.

The system must be able to give users with a full explanation of how the prediction was produced.

Non-functional requirements:

- Security: The system should secure patient data against unauthorized access, use, disclosure, interruption, modification, or destruction.
- Reliability: The system must be available 99.9% of the time.
- Performance: The system must be able to generate predictions within 1 second of receiving a request.
- Scalability: The system must be able to scale to handle an increasing number of users.
- Usability: The system must be easy to use and navigate for both healthcare professionals and patients.

3. Use Cases

- The system will be utilised by the following users:
- Healthcare professionals: Healthcare providers may utilise the system to forecast the diabetes status of their patients and to make educated decisions regarding their care.
- Patients: Patients can utilise the system to learn more about their risk of having diabetes and to take efforts to avoid the condition.

4. Assumptions and Constraints

- The system makes the following assumptions:
- The training data is indicative of the population of patients that will be utilising the system.
- The medical characteristics utilised to train the model are significant to the prediction of diabetes.
- The system has access to a dependable internet connection.
- The system has the following constraints:
- The system must be able to run on a normal PC or mobile device.
- The system must be able to provide predictions within an acceptable length of time.

5. Design

- The system will be designed utilising a modular architecture. This will make the system easier to maintain and extend. The system will consist of the following components:
- Data preparation component: This component will be responsible for cleaning and preparing the training data.
- Feature engineering component: This component will be responsible for building new features from the current features.
- Model training component: This component will be responsible for training the machine learning model.
- Prediction component: This component will be responsible for creating predictions for new patients.
- Explanation component: This component will be responsible for giving users with a full explanation of how the prediction was made.

6. Implementation

- The system will be developed using the Python programming language. The following machine learning libraries will be used:
- scikit-learn: This package contains a number of machine learning methods.
- pandas: This library provides data structures and analytical tools for dealing with Python.
- NumPy: This library contains a range of mathematical functions and data formats.

7. Testing

- The system will be tested using the following methods:
- Unit testing: Unit tests will be used to test the separate components of the system.
- Integration testing: Integration tests will be used to test the interactions between the different components of the system.
- System testing: System tests will be used to test the overall functionality of the system.
- User acceptability testing: User acceptance testing will be undertaken to guarantee that the system satisfies the demands of its users.

8. Deployment

The solution will be put on a cloud-based platform. This will guarantee that the system is scalable and dependable.

9. Maintenance and Support

The system will be maintained and supported by a team of professional software developers. The team will be responsible for addressing issues and bringing new features to the system.

10. Future Work

- The following are some areas for further work:
- Improve the accuracy of the model: The accuracy of the model may be enhanced by utilising more data and by employing more complex machine learning methods.
- Add additional features to the model: More characteristics can be added to the model to increase its forecasting accuracy. For example, the system might contain data such as family history of diabetes and lifestyle variables.