**MITWPU**

**School of Computer Science Academic Year 2020-21**

**MCA Tri IV**

**Mini Project on Open Source Development Synopsis**

|  |
| --- |
| **2. Details of Group Member:** |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Roll. No.** | **Class** | **Batch** | **Name of Student** | **Mobile Number** | **Email Id** |
| 25 | SYM | 2024- | Kalpesh Rathod | 8600457836 | [rathodkalpesh3101@gmail.com](mailto:rathodkalpesh3101@gmail.com) |
|  | CA | 2023 |  |  |  |
| 30 | SYM | 2023- | Prinsi Kakadiya | 7624068530 | [prinsikakadiya@gmail.com](mailto:prinsikakadiya@gmail.com) |
|  | CA | 2024 |  |  |  |

**3. Project Overview / Background:**

* **Introduction :**

Diabetes is a disease that occurs when your blood glucose, also called blood sugar, is too high. Glucose is your body’s main source of energy. Your body can make glucose, but glucose also comes from the food you eat.

Diabetes is a chronic medical condition characterized by elevated levels of glucose (sugar) in the blood. It occurs when the body's ability to regulate blood sugar is impaired, leading to either insufficient production of insulin (a hormone that helps regulate blood sugar) or ineffective utilization of insulin.

* **Need for Work / Reasons for Project Selection :**
* Diabetes is an extremely common disease, affecting a diverse age range of people across the world. Those who are diagnosed with diabetes experience significant health concerns because the disease itself has proven to be the catalyst for other health problems.
* Many individuals who struggle with obesity develop diabetes. The disease also kills more people every year than breast cancer and AIDS combined.
* Complications from diabetes can vary. However, the most prevalent co-morbid conditions include kidney disease, amputations, blindness, cardiovascular disease, obesity, hypertension, hypoglycemia, dyslipidemia, and risk of heart attack or stroke.
* **Objectives :**

These projects aim to improve early detection and management of diabetes, which can lead to better healthcare outcomes. Identify individuals at risk of developing diabetes before clinical symptoms manifest, allowing for early intervention and lifestyle modifications.

* **Problem Statement :**

According to the American Diabetes Association, diabetes affects nearly 29 million Americans;

According to the World Health Organization, 347 million people worldwide have diabetes;

Diabetes remains the 7thleading cause of death in America. Additionally, many individuals’ death certificates cite diabetes as the underlying cause of death.

* **Block diagram:**



* **Benefits to Society :**

Creating a one-page website for diabetes prediction can offer several benefits to the surrounding society and individuals:

Early Detection and Prevention

Health Awareness

Reduced Healthcare Burden

Personalized Recommendations

Collaboration with Healthcare Professionals

**4. Methodology :**

* **Data Collection:**

Obtain a dataset containing relevant patient attributes such as age, gender, BMI, blood pressure, glucose levels, insulin levels, and family history of diabetes.

Ensure data integrity by identifying and addressing missing values, outliers, and inconsistencies**.**

* **Data Preprocessing:**

Normalize or standardize numerical features to bring them to a consistent scale.

Encode categorical variables using techniques like one-hot encoding.

Split the dataset into features (X) and target (y) variables.

* **Feature Selection and Engineering:**

Conduct exploratory data analysis to understand the distribution and relationships between variables.

Use techniques like correlation analysis and feature importance scores to select relevant features.

Engineer new features if applicable, such as creating an insulin resistance index.

* **Model Selection and Development:**

Four ML Algorithms will be applied on the dataset and their performance will be evaluated. The models being applied are- Random Forest Classifier, Decision Tree, Support Vector Machine, Xgboost Classifier.

Train multiple models on the training data and tune hyperparameters to optimize their performance.

* **Model Evaluation and Validation:**

Evaluate model performance using appropriate metrics like accuracy, precision, recall, F1-score, and ROC-AUC.

Perform validation on the testing subset to assess how well the model generalizes to new, unseen data.

Address issues of overfitting by adjusting model complexity and regularization.

* **Hyperparameter Tuning:**

Use techniques like grid search or random search to fine-tune hyperparameters of selected models.

Balance model performance with computational resources and time constraints.

* **Results Interpretation and Reporting:**

Analyze the final model's performance and interpret its predictions.

Identify which features contribute most to predictions and explore any insights gained.

Present the findings in a comprehensive report, detailing the methodology, results, and implications.

**Project Flowchart :**



* **Advantages :**
* Identifies diabetes risk factors early, enabling timely intervention.
* Offers more accurate predictions than traditional methods.
* Tailors predictions to individual patient profiles.
* Allocates healthcare resources more effectively.
* Guides patients with targeted lifestyle changes.
* Reduces long-term treatment costs by preventing diabetes.
* **Limitations :**
* Diabetes is a complex condition with multifaceted causes and progression, making it challenging to capture all factors accurately in a model.
* Long-term patient data is essential to assess the model's effectiveness over time, but such data might not be readily available.
* Handling patient health data requires strict adherence to privacy regulations, which can limit data sharing and model development.
* **Applications :**
* Clinical Settings (Hospitals and clinics).
* Home Healthcare.
* Education and Awareness Campaigns.

**5. Project requirements:**

* **Equipment** : Laptop.
* **Software:** Machine learning, Flask , Html, Css, Python.

* **References:**

<https://www.kaggle.com/datasets/iammustafatz/diabetes-prediction-dataset>

* **Name and Signature of Mini Project Team Members :**

**KALPESH RATHOD**

**SIGNATURE:**



**PRINSI KAKADIYA**

**SIGNATURE**:

