Project Design Phase-II
Data Flow Diagram

Date	20 October 2023
Team ID	Team-593030
Project Name	Diabetes prediction using Machine learning
Maximum Marks	4 Marks

Data Flow:

Data Collection: Patient data, including health and lifestyle information, is collected from various sources such as electronic health records (EHRs), wearables, patient input, or other healthcare systems.

Data Preprocessing:Data preprocessing is essential to ensure data quality. This stage involves handling missing values, outliers, and standardizing data formats. Additionally, data privacy measures may be applied to protect patient information.

Feature Extraction/Engineering:Relevant features (variables) are identified and extracted from the preprocessed data. New features may also be created or existing ones transformed to improve the model's predictive performance.

Machine Learning Model Development:In this phase, machine learning models are trained using historical patient data. Various algorithms (e.g., logistic regression, decision trees, neural networks) are used to develop the predictive models.

Prediction Engine:The trained machine learning model serves as a prediction engine. It takes input data from patients or healthcare providers and produces diabetes risk assessments based on the model's learned patterns.

Explanations:To enhance transparency and trust, the system may provide explanations for its predictions. These explanations help healthcare providers and patients understand why a particular risk assessment was made.

User Interface:A user-friendly interface is created for healthcare providers to input patient data and access the risk assessments and explanations. It also allows users to interact with the system easily.

Data Storage:Clean and structured data is stored in databases for model training and usage. This data serves as a valuable resource for maintaining and updating the system.

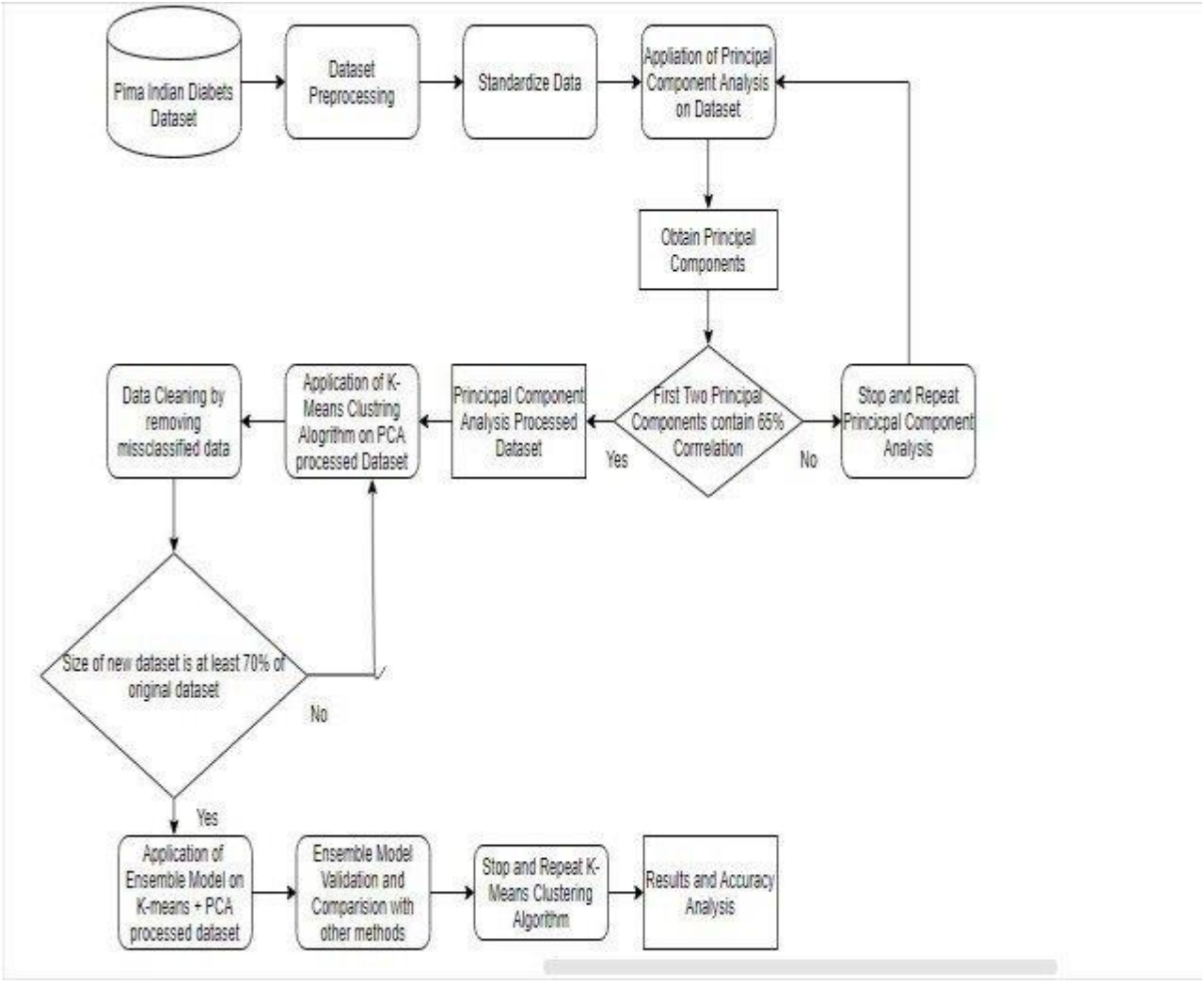
Monitoring and Continuous Learning:Continuous monitoring mechanisms are set up to track the model's performance over time. Feedback and new data are incorporated to ensure that the model remains accurate and relevant.

Ethical Considerations:Efforts may be made to address potential biases in the predictions to ensure that the system provides equitable results for all patients.

Flow in the model:

In a diabetes prediction model, data is first collected from patients and preprocessed to ensure data quality. Relevant features are selected, and a machine learning model is developed and trained on historical data. The model is tested and validated for accuracy. It serves as the prediction engine, taking input data and providing risk assessments. The system offers explanations for predictions, enhancing transparency. A user-friendly interface allows users to interact with the model. Data and model parameters are stored in databases. Ethical considerations are addressed to ensure fairness. Continuous monitoring and updates maintain model accuracy and relevance over time.

Data Flow Diagram:



User stories:

User Type	Functional Requirement	User Story Number	User Story/Task	Acceptance criteria	Priority	Release
	Develop a Diabetes Prediction Model	User – 1	As a user, I want to input relevant patient data,including,gender,BMI,family history, etc., so that the machine learning model can make predictions.	The system should provide input fields for relevant patient data.	High	Sprint 1
	User Management	User - 2	As an admin, I want to be able to manage user accounts, including adding, updating, or deleting users.	The admin should be able to add, update, or delete user accounts as needed.	High	Sprint 1
	Data collection	User – 3	Prior to model training, perform pre-processing tasks, including resizing the images, standardizing pixel values, and dividing the dataset into training and test subsets.	Pre-processing and the splitting of the plant village dataset	High	Sprint 2
	Reporting and Visualization	User – 4	As a user, I want to view a report or visualization of the model's predictions and insights.	The system should generate reports or visualizations based on model predictions. Users should be able to view and interpret these reports.	High	Sprint 2

	Security and Privacy	User – 5	As a user, I want assurance that my personal health data is kept secure and private.	The system should comply with relevant data privacy regulations.	High	Sprint 3
	Integration and Deployment	User – 6	As a system administrator, I want an easy deployment process for the diabetes prediction system.	The system should provide clear deployment instructions.	Mediu m	Sprint 4
	Model Deployment	User – 7	Deploy the trained deep learning model as an API or web service and make it accessible for potato classification.	Check the scalability of the model	Medium	Sprint 4
	Testing and Quality Assurance	User – 8	Thoroughly test the model and web interface, identify and report bugs, fine-tune parameters, and optimize performance based on user feedback.	Creating the web application	Medium	Sprint 6