

# **AI-Based Diabetes Prediction System**

## **APPROACH**

The approach for an AI-Based Diabetes Prediction System involves collecting diverse health data while prioritizing privacy. This data is preprocessed to handle missing values and outliers, and features are carefully engineered. Machine learning models like Logistic Regression, Random Forest, and Deep Learning are considered. Model training and hyperparameter tuning follow, with emphasis on interpretability. The system is deployed based on the intended use case, and thorough testing ensures compliance with standards. Involving healthcare experts is crucial for domain-specific validation.

## **Tech Stack:**

### **1. Programming Languages:**

- Python (for data preprocessing, modeling, and deployment)
- HTML/CSS/JavaScript (for web-based interfaces)

### **2. Libraries and Frameworks:**

- Pandas, NumPy, Scikit-learn (for data manipulation, preprocessing, and modeling)
- TensorFlow, PyTorch (for deep learning, if applicable)
- Flask, Django (for web application development, if required)
- Docker (for containerization of the application)

### **3. Database (Optional):**

- Depending on the scale of the application, you might need a database for storing and managing user data.

## **Algorithms and Techniques:**

### **1. Machine Learning Algorithms:**

- Logistic Regression
- Random Forest

- Support Vector Machines
- Gradient Boosting algorithms (e.g., XGBoost)
- Neural Networks (if deep learning is applicable)

## 2. Feature Selection/Importance:

- Techniques like Recursive Feature Elimination (RFE), feature importance from tree-based models.

## 3. Hyperparameter Tuning:

- Grid search, random search, Bayesian optimization, etc., to find the optimal hyperparameters for the chosen models.

## 4. Data Imputation:

- Techniques like mean imputation, median imputation, K-nearest neighbors imputation for handling missing values.