







1.) Objective 1: Review existing literature on diabetes diagnosis and prediction.

Objective 2: Develop a model using machine learning techniques.

Objective 3: Analyze the diabetes dataset and use Support Vector Machine and Random Forest algorithms to develop a prediction engine.

The scope of your project could include:

1)

Project Scope

and Objectives:

2)

Data

Collection

3)

Data

Preprocessing

Developing a user-friendly interface for the prediction system.

Ensuring that the system is accurate and reliable.

Testing the system on a large dataset to ensure that it is effective in predicting diabetes.

Ensuring that the system is secure and that patient data is protected.

2.)

There are several AI-based systems that use data collection to predict diabetes. One such system is a decision support system (DSS) that utilizes bidirectional long/short-term memory (BiLSTM)to accurately predict diabetic illness from patient data . Another example is an AI/ML-based medical device that has been approved by the US Food and Drug Administration for automatic retinal screening, clinical diagnosis support, and patient selfmanagement tool.

1)

Project Scope

and Objectives:

2)

Data

Collection

3)

Data

Preprocessing

data set link: https://www.kaggle.com/datasets/mathchi/diabetes-data-set

1)

Project Scope

and Objectives:

2)

Data

Collection

3)

Data

Preprocessing

Data Preprocessing: Cleaning and preparing the collected data for analysis. This step includes removing missing values, handling outliers, and normalizing the data.

data set link: https://www.kaggle.com/datasets/mathchi/diabetes-data-set

4)

Feature Selection

and Engineering

5)

Model Selection

6)

Model Training

The model contains the use of Boruta feature selection, the extraction of salient features from datasets, the use of the K-Means++ algorithm for unsupervised clustering of data, and stacking of an ensemble learning method for classification. The model was evaluated by accuracy, precision, and F1 index on the PIMA Indian diabetes dataset and achieved an accuracy rate of 98% .Another study focuses on data-driven diabetes risk factor prediction using machine learning techniques. The study applies two-fold feature selection techniques (i.e., principal component analysis, PCA, and information gain, IG) to boost the prediction accuracy. Then, the optimal features are fed into five ML algorithms, namely decision tree, random forest, support vector machine, logistic regression, and KNN ².

4)

Feature Selection

and Engineering

5)

Model Selection

6)

Model Training

The model contains the use of Boruta feature selection, the extraction of salient features from datasets, the use of the KMeans++ algorithm for unsupervised clustering of data, and stacking of an ensemble learning method for classification. The model was evaluated by accuracy, precision, and F1 index on the PIMA Indian diabetes dataset and achieved an accuracy rate of 98% The study applies two-fold feature selection techniques (i.e., principal component analysis, PCA, and information gain, IG) to boost the prediction accuracy. Then, the optimal features are fed into five ML algorithms, namely decision tree, random forest, support vector machine, logistic regression, and KNN 2.







