**AI Based Diabetes Prediction System**

**Introduction**

This project aims to predict the likelihood of diabetes based on various features using machine learning techniques. The project consists of four main scripts: data preprocessing, feature selection, model training, and model evaluation.

**Getting Started**

**Installation**

To run the project, you will need Python 3 and the following libraries:

* pandas
* numpy
* scikit-learn

You can install the required libraries using the following command:

-----------------------------------------------------------------------------------------------------------------

pip install pandas numpy scikit-learn

-----------------------------------------------------------------------------------------------------------------

**Running the Code**

To run the code, navigate to the project directory and execute the following commands in order:

**Run the data preprocessing script:**

-----------------------------------------------------------------------------------------------------------------python data\_preprocessing.py

-----------------------------------------------------------------------------------------------------------------

**Run the feature selection script:**

-----------------------------------------------------------------------------------------------------------------python feature\_selection.py

-----------------------------------------------------------------------------------------------------------------

**Run the model training script:**

-----------------------------------------------------------------------------------------------------------------python model\_training.py

-----------------------------------------------------------------------------------------------------------------

**Run the model evaluation script:**

-----------------------------------------------------------------------------------------------------------------python model\_evaluation.py

-----------------------------------------------------------------------------------------------------------------

**Code Files**

**data\_preprocessing.py**

This script performs data preprocessing tasks such as handling missing values, removing duplicate rows, and scaling the data. The output of this script is a clean and preprocessed dataset that is ready for feature selection and model training.

**feature\_selection.py**

This script selects the most important features from the dataset using a feature selection technique. The selected features are then used for model training.

**model\_training.py**

This script trains a machine learning model using the selected features from the feature selection script. The model is trained using a training dataset, and the training accuracy is reported.

**model\_evaluation.py**

This script evaluates the performance of the trained model using a testing dataset. The evaluation metrics such as accuracy, precision, recall, and f1-score are reported.

**Dataset Source**

The dataset used in this project is the Pima Indians Diabetes Database, which can be found at UCI Machine Learning Repository.

**Conclusion**

This project demonstrates how to use machine learning techniques to predict the likelihood of diabetes based on various features. The code is modular and can be easily adapted to different datasets and machine learning algorithms.