Problem Definition:

The objective is to develop an AI-driven diabetes prediction system that leverages machine learning techniques to analyse medical information and estimate the probability of an individual developing diabetes. The system's goal is to offer early risk evaluation and tailored preventive strategies, enabling individuals to proactively manage their well-being.

Design Thinking:

1. Data Collection:

To proceed with building the AI-powered diabetes prediction system, we require a comprehensive dataset comprising medical attributes like glucose levels, blood pressure, BMI, and other relevant factors. Additionally, the dataset should include corresponding labels indicating whether the individuals have been diagnosed with diabetes or not. This dataset will serve as the foundation for training the machine learning algorithms and enabling accurate predictions.

2. Data Preprocessing:

Before training machine learning models, it is crucial to perform data preprocessing tasks such as cleaning, normalization, and preparation of the medical data. This involves removing any inconsistencies, errors, or missing values in the dataset, as well as standardizing the numerical features to a common scale. By undertaking these preprocessing steps, we ensure the data is in a suitable format for model training, enhancing the accuracy and reliability of the AI-powered diabetes prediction system.

3. Feature Selection:

As part of developing the AI-powered diabetes prediction system, we will conduct a feature selection process to identify the most relevant attributes that can significantly impact the prediction of diabetes risk. This involves analysing the dataset and employing techniques such as statistical analysis, correlation analysis, or machine learning algorithms to determine which features have the strongest associations with the target variable (diabetes). By selecting the most

influential features, we can optimize the model's performance and ensure that it focuses on the most important factors in predicting diabetes risk.

4. Model Selection:

There are several machine learning algorithms that can be explored for experimentation, including Logistic Regression, Random Forest, and Gradient Boosting.

5. Evaluation:

To assess the model's performance, we can utilize various metrics such as accuracy, precision, recall, F1-score, and ROC-AUC. These metrics provide valuable insights into different aspects of the model's predictive capability, allowing us to comprehensively evaluate its effectiveness.

6. Iterative Improvement:

In order to improve the prediction accuracy, we can fine-tune the model parameters and explore techniques like feature engineering. By carefully adjusting the model's hyperparameters and engineering new features that capture relevant information, we can enhance the model's ability to make accurate predictions. This iterative process of parameter tuning and feature engineering enables us to optimize the model's performance for the specific problem at hand.