DIABETES PREDICTION SYSTEM

ARTIFICIAL INTELLIGENCE

PHASE 2

DATA SOURCES:

Gather diabetes-related data from sources such as medical records, wearable devices, or publicly available datasets like the NHANES.

DATA PREPROCESSING:

Clean the data, handle missing values, normalized and prepared for training machine learning models.

FEATURE SELECTION AND ENGINEERING:

Use techniques like Recursive Feature Elimination (RFE), feature importance from tree-based models, or domain knowledge to select the most relevant features. Create new features if needed, like BMI, blood pressure categories, or glycemic indices.

MACHINE LEARNING ALGORITHMS:

Experiment with various machine learning algorithms, including but not limited to:

* Logistic Regression
* Random Forest
* Support Vector Machines (SVM)
* Gradient Boosting (e.g., XGBoost, LightGBM)
* Neural Networks (Deep Learning)

MODEL EVALUATION AND METRICS:

Choose appropriate evaluation metrics for classification problems, including:

* Accuracy
* Precision
* Recall
* F1-score
* Area Under the Receiver Operating Characteristic (ROC AUC)

MODEL SELECTION:

Choose one or more machine learning algorithms suitable for binary classification (predicting diabetes vs. non-diabetes). Common choices include Logistic Regression, Random Forest, Support Vector Machines (SVM), Gradient Boosting (e.g., XGBoost), or Neural Networks.

MODEL TRAINING:

* Train the selected machine learning models on the training dataset.
* For deep learning models, you can use frameworks like TensorFlow or PyTorch.

MODEL DEVELOPMENT:

* Develop a user-friendly interface for healthcare professionals to input patient data and receive predictions.
* Deploy the selected model(s) using frameworks like Flask, Django, or FastAPI.
* Host the system on a server or cloud platform (e.g., AWS, Google Cloud, or Azure).