PHASE 2 : INNOVATION PHASE

COURSE	Artificial Intelligence
PROJECT	AI Based Diabetes Prediction System
DATE	10-10-2023

OBJECTIVE:

The primary objective of this phase is to implement and deploy the AI-based Diabetes Prediction system, ensuring that it accurately predicts the risk of diabetes based on relevant features.

STATEMENT:

In this phase, we can explore innovative techniques such as ensemble methods and deep learning architectures to improve the prediction system's accuracy and robustness.

DESCRIPTION OF THE DATASET:

This dataset is originally from the National Institute of Diabetes and Digestive and Kidney Diseases. The objective is to predict based on diagnostic measurements whether a patient has diabetes.

Several constraints were placed on the selection of these instances from a larger database. In particular, all patients here are females at least 21 years old of Pima Indian heritage.

- Pregnancies: Number of times pregnant
- Glucose: Plasma glucose concentration a 2 hours in an oral glucose tolerance test
- Blood Pressure: Diastolic blood pressure (mm Hg)
- Skin Thickness: Triceps skin fold thickness (mm)
- Insulin: 2-Hour serum insulin (mu U/ml)
- BMI: Body mass index (weight in kg/(height in m)^2)
- Diabetes Pedigree Function: Diabetes pedigree function
- Age: Age (years)
- Outcome: Class variable (0 or 1)

TRANSFORMATION STEPS

INTRODUCTION :

The transformation phase of the AI-based Diabetes Prediction System is a pivotal stage where we convert our design into a practical and functional solution. This phase encompasses a series of well-defined steps aimed at ensuring the accuracy, robustness, and real-world applicability of our predictive model. The transformation phase focuses on data preprocessing, model selection, evaluation, and ultimately, deployment, making it a critical bridge between conception and practical implementation.

STEPS INVOLVED :

STEP 1: DATA PREPROCESSING AND SPLITTING

- Begin by cleaning and preprocessing diabetes dataset, handling missing values, encoding categorical variables, and scaling or normalizing features as needed.
- Perform feature engineering to create relevant features that may improve prediction accuracy.

STEP 2: MODEL SELECTION

Here's an ensemble approach for diabetes prediction .

This may include:

- Random Forest Classifier
- Logistic Regression
- Support Vector Machine (SVM)

The goal of using an ensemble approach is to leverage the strengths of each individual classifier

to improve the overall prediction accuracy of the system . Here's a brief overview of each of these classifiers:

- Random Forest Classifier:
  Random Forest is an ensemble learning method based on decision trees. It creates multiple decision trees during training and combines their predictions. It is known for its robustness, ability to handle complex data, and resistance to over fitting.
- Logistic Regression:
  Logistic Regression is a simple yet effective linear classification algorithm. It is widely used in binary classification problems, such as predicting the probability

of a binary outcome (e.g., diabetes vs. non-diabetes). It's interpretable and computationally efficient.

- Support Vector Machine (SVM):

  SVM is a powerful classification algorithm that finds a hyper plane that best separates data points belonging to different classes. It can handle both linear and non-linear classification tasks through kernel functions. SVMs are known for their ability to handle high-dimensional data and find optimal decision boundaries.

A Voting Classifier is created, combining the predictions of the individual classifiers using a 'soft' voting strategy. The ensemble aims to improve prediction accuracy by aggregating multiple model predictions.

STEP 3: MODEL EVALUATION

Assess the ensemble model's performance using appropriate evaluation metrics (e.g., accuracy, precision, recall, F1-score) . We can evaluate the model's performance using appropriate evaluation metrics, including accuracy, precision, recall, and F1-score. Additionally, we can conduct cross-validation to ensure robustness and minimize over fitting.

STEP 4: MODEL DEPLOYMENT

- Once we are satisfied with the ensemble's performance, we can use it to make predictions on new, unseen data. To deploy your ensemble model within our AI-based Diabetes Prediction system and integrate it into a web application using Flask.
- To make our web application accessible over the internet, we can deploy it on a web hosting service. Popular choices include Heroku, AWS, and GCP. Each service has its deployment process.
- By following these steps, we can deploy your ensemble model within a Flask-based web application, allowing users to interact with the model and receive diabetes predictions through a user-friendly interface.

CONCLUSION :

Finally, we embark on the exciting journey of model deployment, making our AI-based Diabetes Prediction System accessible to users through a web application powered by Flask. By embracing modern web hosting services like Heroku, AWS, or GCP, we take our system beyond the development environment, making it readily available for healthcare professionals and individuals seeking accurate diabetes risk predictions. The innovation phase is the manifestation of our innovative solution, designed to make a positive impact on healthcare and the lives of those it serves.