AI BASED DIABETES PREDICTION SYSTEM

Phase 5 – Project submission

PROJECT TITLE	AI based diabetes prediction system
SKILLS TAKEN AWAY	❖ Python script ❖ EDA ❖ UI deployment
DOMAIN	Medical
TEAM MEMBERS	K. MADHUMITHA M.PAVITHRA M.ARTHI VARSHINI S. ROSHINI V.KANAGAVALLI

INTRODUCTION:

Diabetes is a health issue that affects how your body converts food into energy. The majority of the food you consume is broken down into sugar (also known as glucose) and released into your bloodstream. When your blood sugar rises, your pancreas releases insulin. Diabetes, if not managed carefully and continuously, can cause a buildup of sugars in the blood, increasing the risk of serious consequences such as stroke and heart disease. As a result, I decided to anticipate using Machine Learning in Python.

OBJECTIVES:

- Determine whether or not a person has diabetes.
- Discover the most telling signs of diabetes and experiment with several classification methods to get the best accuracy.

ABSTRACT:

ABSTRACT: Diabetes Mellitus is a serious illness that affects a large number of people. Diabetes mellitus can be brought on by a number of factors, including advanced age, obesity, a poor diet, genetics, high blood pressure, and lack of exercise. Diabetes increases a person's risk of developing heart disease, renal

disease, stroke, vision problems, nerve damage, and other illnesses. Hospitals currently gather the data needed for diabetes diagnosis using a variety of tests, and then treat patients according to their diagnosis. Database volume is high in the healthcare sector. With the use of big data analytics, it is possible to examine enormous datasets, uncover hidden patterns and information, gain knowledge from the data, and forecast results appropriately. The accuracy of categorization and prediction using the current method is not very great. In this work, we have suggested

DETAILS ABOUT THE DATASET:

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

The datasets consist of several medical predictor variables and one target variable, Outcome. Predictor variables includes the number of pregnancies the patient has had, their BMI, insulin level, age, and so on.

- **Pregnancies**: Number of times pregnant
- **Glucose**: Plasma glucose concentration 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)

DATA PREPROCESSING:

- ✓ Handling missing values: Impute missing data using techniques like mean, median, or predictive modelling.
- ✓ Outlier detection: Identify and handle outliers that can skew your model's predictions.
- ✓ Data normalization or scaling: Ensure that all features are on a similar scale, which is crucial for many machine learning algorithms.

SPLITTING THE DATA SET INTO TRAINING & TEST DATA:

Dividing the dataset into two subsets: training data, and test data. Typically, I have used an 80-20 split is used, where the training set is the largest portion.

MODEL SELECTION:

I have Selected the SVM classifier (support vector machines) as my machine learning model. SVM is a suitable choice for binary classification tasks like diabetes prediction.

MODEL TRAINING:

Training our selected model using the training dataset. During training, the model learns to make predictions by adjusting its internal parameters to minimize the prediction error.

MODEL EVALUATION AND TESTING:

Model evaluation is done by checking its accuracy. After that its tested. Thus, AI model has been built and trained using machine learning algorithm and its performance have been evaluated.

CONCLUSION:

In conclusion, developing an AI-based diabetes prediction system using a Support Vector Machine (SVM) classifier involves a series of well-defined steps. This approach can be effective in binary classification tasks, providing interpretable results, and it has the potential to make a positive impact in healthcare. Thus, by using a Support Vector Machine classifier in our diabetes prediction system can provide valuable insights and assist healthcare professionals in making informed decisions.

FUTURE ENHANCEMENT:

Future enhancements for an AI-based diabetes prediction system should focus on refining data quality and quantity, incorporating advanced feature engineering and deep learning techniques, ensuring explainability and real-time monitoring, personalizing predictions and integrating with electronic health records and telehealth services, fostering ethical and regulatory compliance, and supporting continuous improvement through feedback and research, with a global perspective, to enhance patient outcomes and expand the system's impact on healthcare.