Machine learning foundation program

Dr. Sumudu Tennakoon

Diabetes Prediction using Machine Learning

Name: Thehani Fernandopulle

Problem Background

Diabetes, a pervasive metabolic disorder characterized by elevated blood glucose levels, poses a formidable global health challenge. The World Health Organization (WHO) estimates that 422 million adults were living with diabetes in 2014, a number projected to surge to 642 million by 2040 if current trends persist. This escalating prevalence not only impacts individual health but also places an overwhelming burden on healthcare systems and economies worldwide.

Early detection is pivotal in managing diabetes effectively. Timely diagnosis facilitates the implementation of interventions, lifestyle modifications, and medical treatments, thereby mitigating disease progression and reducing the risk of complications. However, conventional methods of diabetes detection often involve invasive procedures, expensive laboratory tests, and are not easily accessible, particularly in resource-constrained regions.

In response to these challenges, a machine learning system has been developed using the Support Vector Machine (SVM) algorithm. SVM is a powerful supervised learning algorithm that, in the realm of healthcare, can contribute significantly to predicting whether an individual has diabetes based on key medical indicators.

Methodology: Utilizing Support Vector Machine for Diabetes Prediction

The SVM algorithm is employed as the cornerstone of this predictive model. In the realm of supervised learning, the algorithm is trained with a dataset comprising essential medical information such as blood glucose levels, BMI, and insulin levels of patients, paired with records indicating the presence or absence of diabetes. This labeled data is crucial for the model to discern patterns and relationships between different parameters.

The SVM model operates by plotting the data in a multi-dimensional space and identifying a hyperplane that effectively separates the data points into two distinct groups – one representing individuals with diabetes and the other without. This hyperplane becomes the basis for classifying new, unseen data into these two categories, enabling the model to predict whether a person has diabetes or not.

Imported libraries.

- Numpy('np')
- Pandas('pd')
- scikit-learn (sklearn) modules.

Work Flow:

Diabetes Data Collection:

Source: Kaggle

https://www.kaggle.com/datasets/uciml/pima-indians-diabetes-database

Data Pre-processing:

The collected data undergoes pre-processing to standardize it, ensuring that all parameters are within the same range for effective training.

Data Splitting:

The dataset is divided into training and testing sets. The training set is used to teach the SVM model, while the testing set evaluates the model's accuracy.

Training the SVM Model:

The SVM model is trained on the labeled dataset, learning the patterns and relationships between medical parameters and diabetes status.

Prediction:

New, unseen data is fed into the trained SVM model, which classifies the data into diabetes or non-diabetes categories.

By implementing this machine learning approach, the goal is to create a reliable, non-invasive, and accessible tool for diabetes prediction. This has the potential to revolutionize early detection efforts, enabling individuals and healthcare professionals to take proactive measures and mitigate the impact of diabetes on a global scale.

Saving the trained model to deploy

Imported libraries

- Pickle

Model accuracy

```
print('Accuracy score of the training data : ', training_data_accuracy)
```

```
Accuracy score of the training data :  0.7866449511400652
```

```
[112] print('Accuracy score of the test data : ', test_data_accuracy)
```

```
Accuracy score of the test data :  0.7727272727272727
```

Deploying the model

IDE-Spyder

Library -Streamlit (‘st’), Numpy (‘np’), Pickle