Project Report

On

DIABETES PREDECTIVE SYSTEM THROUGH MACHINE LEARNING MODEL

Submitted by

T.Anusha - R170566
S.Madhavi - R170566

Under the guidance of

Ms.Shaik Shabana

Department of Computer Science and Engineering

Rajiv Gandhi University of Knowledge and Technologies(RGUKT),

R.K.Valley, Kadapa, Andra Pradesh.

Rajiv Gandhi University of Knowledge Techonolgies

RK Valley, Kadapa (Dist), Andhra Pradesh, 516330

CERTIFICATE

This is to certify that the project work titled **"DIABETES PREDECTIVE SYSTEM THROUGH MACHINE LEARNING MODEL"** is a bonafied project work submitted by **T.Anusha and S.Madhavi** in **COMPUTER SCIENCE AND ENGINEERING** in partial fulfillment of requirements for the award of degree of **Bachelor of Technology** for the year **2021-2022** carried out the work under the supervision.

INTERNAL GUIDE
(S SHABANA)

HEAD OF THE DEPARTMENT
(HARINADH)

ACKNOWLEDGEMENT

The satisfaction that accompanies the successful completion of any task would be incomplete without the mention of the people who made it possible and whose constant guidance and encouragement crown all the efforts success.

I am extremely grateful to our respected Director,Prof. K. SANDHYA RANI for fostering an excellent academic climate in our institution.

I also express my sincere gratitude to our respected Head of the Department Mr HARINADH for his encouragement, overall guidance in viewing this project a good asset and effort in bringing out this project.

I would like to convey thanks to our guide at college S SHABANA for his guidance, encouragement, co-operation and kindness during the entire duration of the course and academics.

My sincere thanks to all the members who helped me directly and indirectly in the completion of project work. I express my profound gratitude to all our friends and family members for their encouragement.

•

INDEX

ABSTRACT:

Diabetes is an illness caused because of high glucose level in a human body. Diabetes should not be ignored if it is untreated then Diabetes may cause some major issues in a person like: heart related problems, kidney problem, blood pressure, eye damage and it can also affects other organs of human body. Diabetes can be controlled if it is predicted earlier. To achieve this goal this project work we will do early prediction of Diabetes in a human body or a patient for a higher accuracy through applying, Various Machine Learning Techniques. Machine learning techniques Provide better result for prediction by con- structing models from datasets collected from patients. In this work we will use Machine Learning Classification and ensemble techniques on a dataset to predict diabetes. Which are K-Nearest Neighbor (KNN), Support Vector Machine (SVM). The accuracy is different for every model when compared to other models. The Project work gives the accurate or higher accuracy model shows that the model is capa- ble of predicting diabetes effectively.

Introduction

Diabetes is noxious diseases in the world. Diabetes caused because of obesity or high blood glucose level, and so forth. It affects the hormone insulin, resulting in abnormal metabolism of crabs and improves level of sugar in the blood. Early prediction of disease like diabetes can be controlled and save the human life. To accomplish this, this work explores prediction of diabetes by taking various attributes related to diabetes disease. For this purpose we use the Pima Indian Diabetes Dataset, we apply various Machine Learning classification and ensemble Techniques to predict diabetes. Machine Learning Is a method that is used to train computers or machines explicitly. Various techniques of Machine Learning can capable to do prediction, however its tough to choose best technique. Thus for this purpose we apply popular classification and ensemble methods on dataset for prediction.

Purpose

The main aim of this project was to design and implement Diabetes Prediction Using Machine Learning Methods and Performance Analysis of that methods and it has been achieved successfully.We experimented with different classification and ensemble algorithms to predict diabetes.To achieve this goal this project work we will do early prediction of Diabetes in a human body or a patient for a higher accuracy through applying, Various Machine Learning Techniques. Machine learning techniques Provide better result for prediction by constructing models from datasets collected from patients.

Scope

I

Requirement Specification

HARDWARE:

Ram	:	4GB
Hardisk	:	512GB
Processor	:	2GHz

SOFTWARE:

Language	:	Python
Tools	:	Jupyter-Notebook/google collab
GUI	:	Tkinker
Additional Modules	:	Numpy Matplotlib Seaborn Sklearn Pandas

Analysis and Design

Data Collection:

In data collection, we collect the data.

Data Collection and Analysis

PIMA Diabetes Dataset

```
[ ] # loading the diabetes dataset to a pandas DataFrame
    diabetes_dataset = pd.read_csv('/content/diabetes.csv')
```

```
[ ] pd.read_csv?
```

```
[ ] # printing the first 5 rows of the dataset
    diabetes_dataset.head()
```

	Pregnancies	Glucose	BloodPressure	SkinThickness	Insulin	BMI	DiabetesPedigreeFunction	Age	Outcome
0	6	148	72	35	0	33.6	0.627	50	1
1	1	85	66	29	0	26.6	0.351	31	0
2	8	183	64	0	0	23.3	0.672	32	1
3	1	89	66	23	94	28.1	0.167	21	0
4	0	137	40	35	168	43.1	2.288	33	1

Standardized data:

Data Standardization

```
[ ] scaler = StandardScaler()
```

```
[ ] scaler.fit(X)

    StandardScaler(copy=True, with_mean=True, with_std=True)
```

```
[ ] standardized_data = scaler.transform(X)
```

```
[ ] print(standardized_data)

    [[ 0.63994726  0.84832379  0.14964075 ...  0.20401277  0.46849198
       1.4259954 ]
     [-0.84488505 -1.12339636 -0.16054575 ... -0.68442195 -0.36506078
      -0.19067191]
     [ 1.23388019  1.94372388 -0.26394125 ... -1.10325546  0.60439732
      -0.10558415]
```

Splitting data:

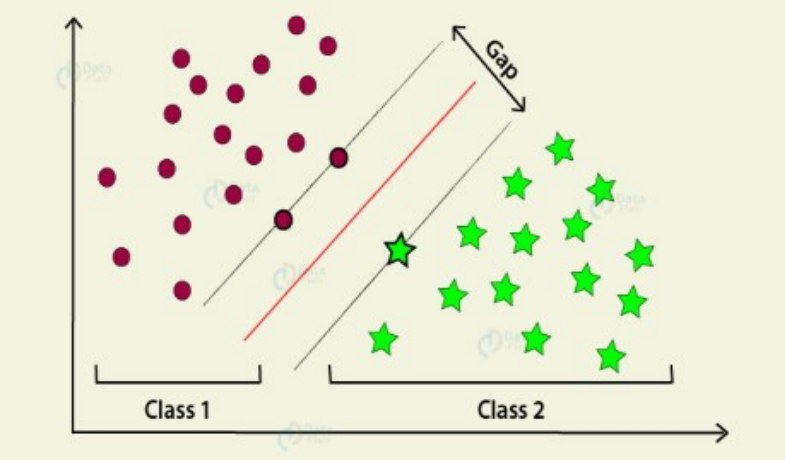
Train Test Split

```
[ ] X_train, X_test, Y_train, Y_test = train_test_split(X,Y, test_size = 0.2, stratify=Y, random_state=2)
```

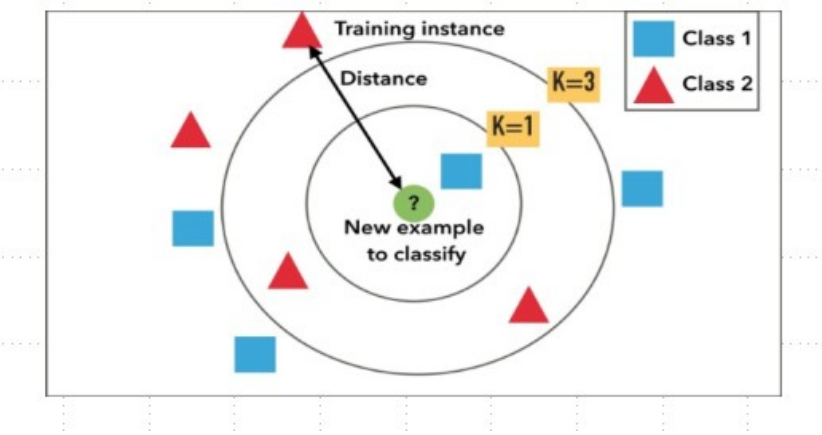
```
[ ] print(X.shape, X_train.shape, X_test.shape)

    (768, 8) (614, 8) (154, 8)
```

Support Vector Machine:



K Nearest Neighbour :



Making a predective system:

Making a Predictive System

```
[ ]  input_data = (5,166,72,19,175,25.8,0.587,51)

     # changing the input_data to numpy array
     input_data_as_numpy_array = np.asarray(input_data)

     # reshape the array as we are predicting for one instance
     input_data_reshaped = input_data_as_numpy_array.reshape(1,-1)

     # standardize the input data
     std_data = scaler.transform(input_data_reshaped)
     print(std_data)

     prediction = classifier.predict(std_data)
     print(prediction)

     if (prediction[0] == 0):
       print('The person is not diabetic')
     else:
       print('The person is diabetic')
```

```
[[ 0.3429808   1.41167241  0.14964075 -0.09637905  0.82661621 -0.78595734
   0.34768723  1.51108316]]
[1]
The person is diabetic
```

Libraries used:

Pandas:

Pandas is **a Python package providing fast, flexible, and expressive data structures designed to make working with "relational" or "labeled" data both easy and intuitive**. It aims to be the fundamental high-level building block for doing practical, real-world data analysis in Python.

Numpy:

NumPy is **a Python library used for working with arrays**. It also has functions for working in domain of linear algebra, fourier transform, and matrices. NumPy was created in 2005 by Travis Oliphant. It is an open source project and you can use it freely. NumPy stands for Numerical Python.

Sklearn:

Scikit-learn is **a key library for the Python programming language that is typically used in machine learning projects**. Scikit-learn is focused on machine learning tools including mathematical, statistical and general purpose algorithms that form the basis for many machine learning technologies.

Seaborn:

Seaborn is a library in Python predominantly used for making statistcal graphs. Seaborn is a data visualization library built on top of matplotlib and closelyintegrated with pandas data structures in python.Visualization is the central part of Seaborn which helps in exploration and understanding of data.

CONCLUSION

The proposed approach uses various classification and ensemble learning method in which SVM, KNN are used. To prevent and cure diabetes and to improve the lives of all people affected by diabetes". To support the lives of the people all over the world, Support vector machine and NB techniques give the accuracy of 77.73% and 73.48% respectively from the existing method and the proposed method improves the accuracy has been achieved. The Exper- imental results can be asst health care to take early predic- tion and make early decision to cure diabetes and save humans life.

REFERENCES

- **https://numpy.org/doc/stable/**
- https://matplotlib.org/
- https://scikit-learn.org/stable/index.html
- https://seaborn.pydata.org/


THANK YOU