SRM INSTITUTE OF SCIENCE AND TECHNOLOGY

18CSE392T : MACHINE LEARNING - I

Diabetes Prediction using Machine Learning

Meet the Team

RA2111027010002	RA2111027010049	RA2111027010065
Chintapalli Dinesh	Snehal Sukundari	Ratan Priya Singh

TABLE OF CONTENTS

INTRODUCTION

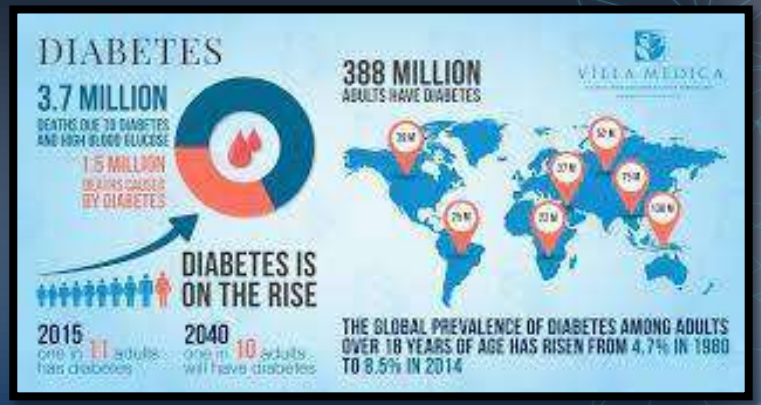
- Diabetes prediction involves using historical patient data to develop models that can forecast the likelihood of an individual developing diabetes in the future.

- Early prediction of diabetes can lead to timely interventions, lifestyle adjustments, and personalized medical care, which are crucial for managing the disease effectively.

- There are two main types of diabetes:
  - Type 1: Diagnosed in childhood
  - Type 2: Diagnosed in adulthood

MOTIVATION

- Predicting diabetes is a crucial area of research because early detection can lead to better management and improved quality of life for individuals at risk.
- By developing accurate prediction models, we can empower both patients and healthcare professionals to take proactive measures and make informed decisions to prevent or manage diabetes effectively.
- According to the World health Organization there were around 70 million people found diabetic in India in 2015.
- As of September 2021, it was estimated that over 463 million adults (20-79 years) were living with diabetes worldwide.
- The world health Organization also believes that diabetes will be the 7th leading cause of death by 2030.



PROBLEM FORMULATION

- To Develop a machine learning model that predicts the likelihood of an individual developing diabetes based on a set of clinical and lifestyle attributes.
- The goal is to create an accurate and interpretable predictive tool that aids healthcare professionals in identifying individuals at risk of diabetes and providing timely interventions.

LITERATURE REVIEW

S. No.	Title of the Work/Authors	Techniques	Results/Limitations
1.	Tao et al.[2]	KNN, Naïve Bayes, Decision Tree, Random Forest, SVM and Logistic Regression	Concentrated on the accuracy of recall and got better result. Filtering criteria can be improved
2.	Loannis et al..[1]	Naïve Bayes, Logistic regression, and Svm	From the three algorithm Svm provided high accuracy of 84%
3.	Weifeng Xu et al.[3]	ID3Naïv Bayes, Randomforest, Adaboost	Random forest classifier method better relative to other in contrast ID3 provided the least accuracy than others.

LITERATURE REVIEW

S. No.	Title of the Work/Authors	Techniques	Results/Limitations
4.	Yunsheng et al. [4]	DISKR and KNN	Accuracy increase can be increase by removing outliers. Space complexity decreased.
5.	Messan et al.[5]	GMM, ELM, ANN LR, and SVM	Comparison of algorithm were done from those method artificial neural network provide better accuracy than other classifier.
6.	Ramiro et al.[6]	Fuzzy rule	Wrong treatment was reduced using fuzzy rule and recommendation system was developed for doctor.

LITERATURE REVIEW

S. No.	Title of the Work/Authors	Techniques	Results/Limitations
7.	Swarupa et al.[7]	KNN,J48, ANN, zeroR, NB, evparameter selection, Filtered classifier and simple cart	Various dataset applied containing diabetes dataset. Cross validation not applied. NBshown high accuracy by providing accuracy of 77,01%.
8.	Pradeep & Dr.Naveen [8]	Decision tree(J48)	J48 is noted as good accuracy provider algorithm. Feature selection has high role in the prediction area.
9.	Sajida et al.[9]	Adaboost, j48,and Bagging	Adaboost was shown improved accuracy than other method.

INFRENCE FROM LITERATURE REVIEW

- SVM has the highest accuracy(84%) when compared to algorithms Naïve Bayes & Logistic regression.
- SVM works comparably well when there is an understandable margin of dissociation between classes for Diabetic and Non Diabetic.
- SVMs are effective in handling high-dimensional data.
- SVMs can be regularized, which means that the algorithm can be modified to avoid overfitting.

DATA SET

	A	B	C	D	E	F	G	H	I
1	Pregnancies	Glucose	BloodPressure	SkinThickness	Insulin	BMI	DiabetesPedigreeFunction	Age	Outcome
2	6	148	72	35	0	33.6	0.627	50	1
3	1	85	66	29	0	26.6	0.351	31	0
4	8	183	64	0	0	23.3	0.672	32	1
5	1	89	66	23	94	28.1	0.167	21	0
6	0	137	40	35	168	43.1	2.288	33	1
7	5	116	74	0	0	25.6	0.201	30	0
8	3	78	50	32	88	31	0.248	26	1
9	10	115	0	0	0	35.3	0.134	29	0
10	2	197	70	45	543	30.5	0.158	53	1

Number of row: 769
Number of column: 09

FEATURES/ATRRIBUTES :

1.Pregancies
2.Glucose
3.Blood Pressure
4. Skin Thickness
5. Insulin
6.BMI
7. Diabetes Pedigree Function
8.Age
9.Outcome

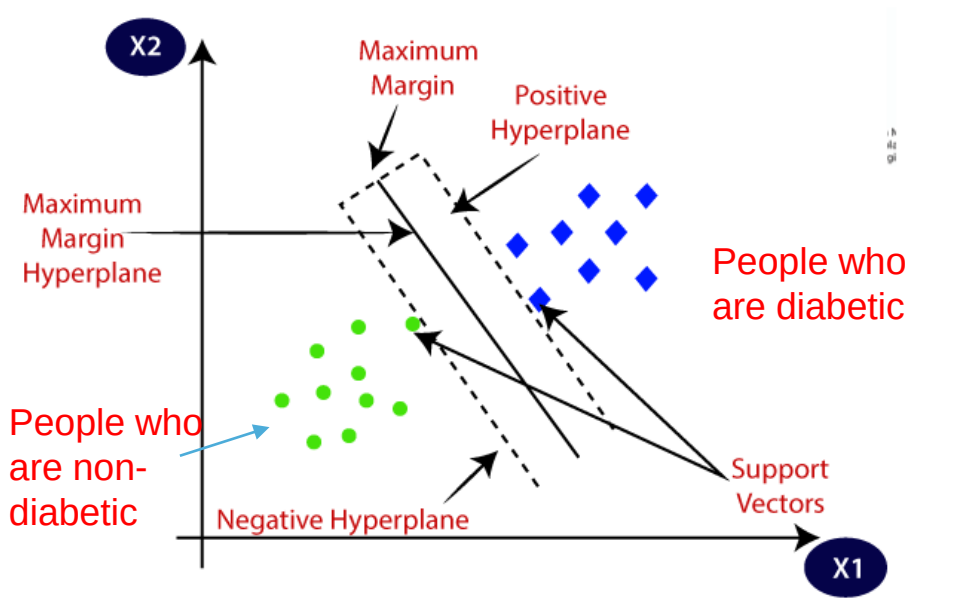
ALGORITHM

Algorithm : Support Vector Machine

- Diabetes prediction often involves complex, non-linear relationships between various factors like glucose levels, BMI, family history, etc. SVMs can model these non-linear relationships effectively using kernel functions.
- Medical datasets used for diabetic prediction can be high-dimensional due to the numerous factors taken into account. SVMs perform well in high-dimensional spaces
- The classes (diabetic and non-diabetic) might be imbalanced, i.e., one class might have significantly fewer instances. SVMs can handle class imbalance well, ensuring that both classes are treated with equal importance

ALGORITHM

Algorithm : Support Vector Machine

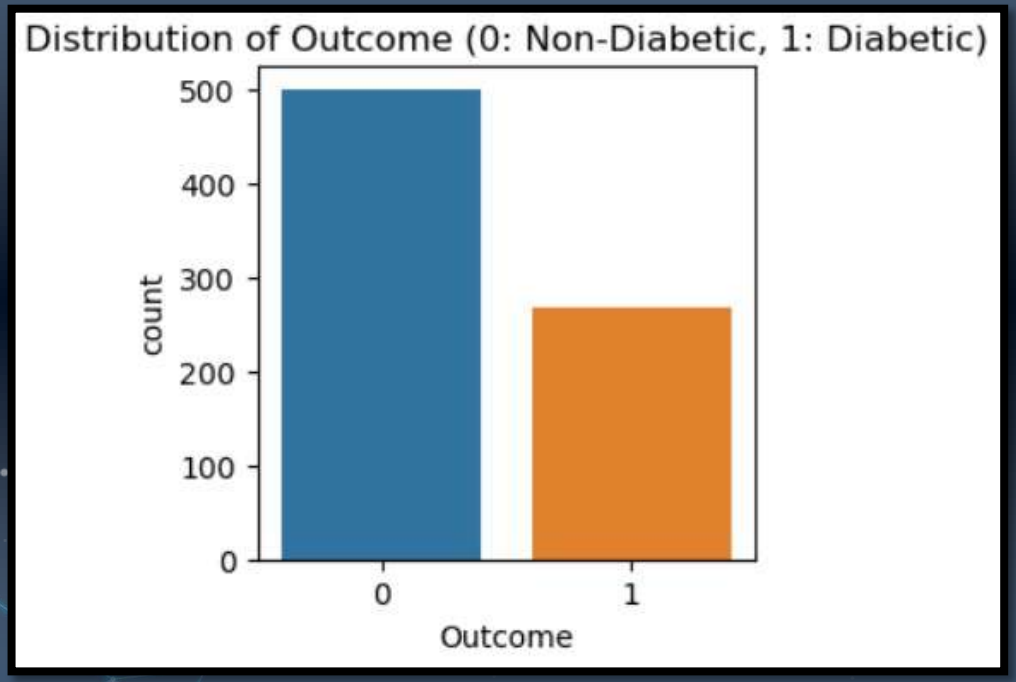


WHY SVM

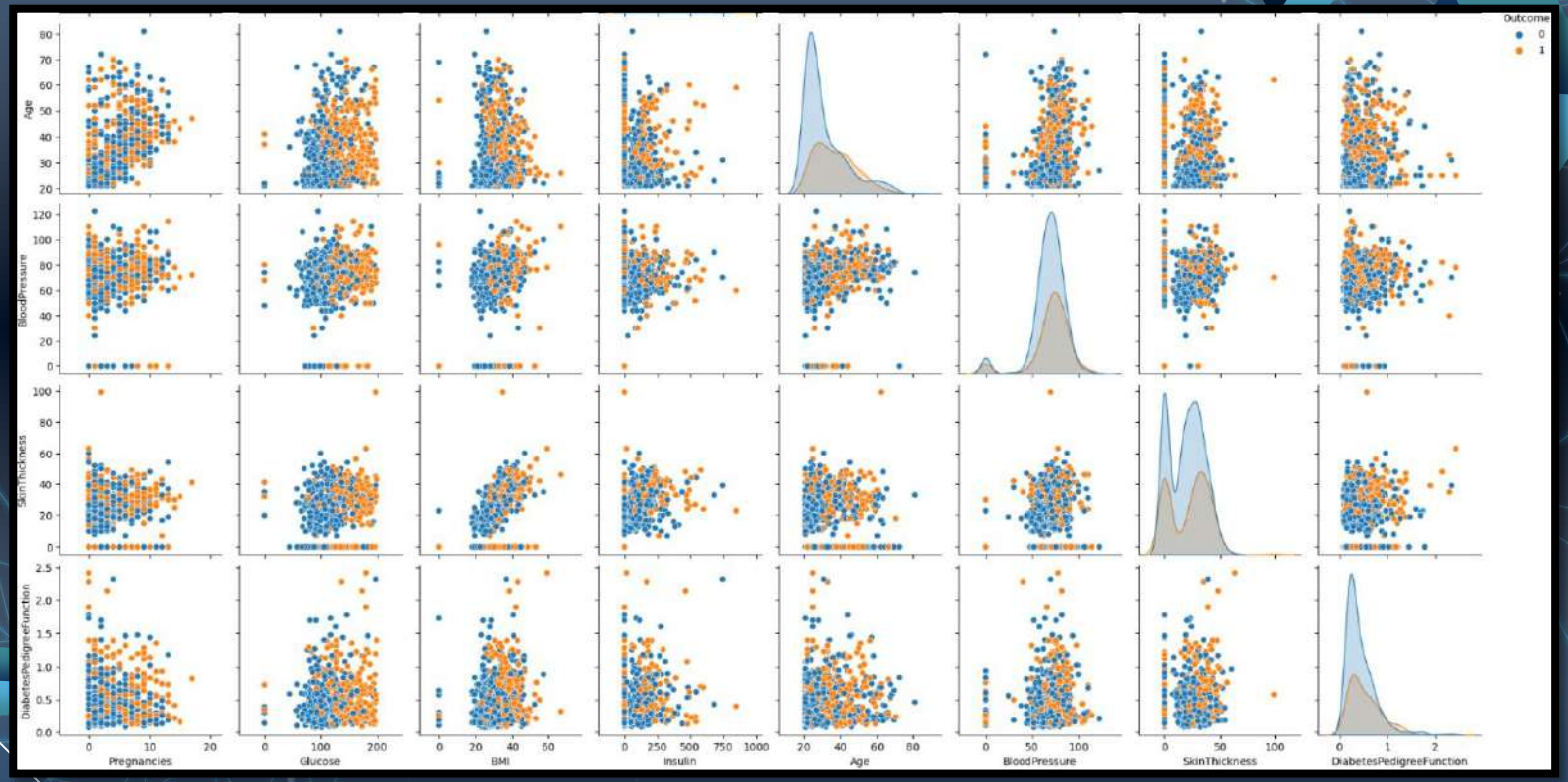
National Library of Medicine
National Center for Biotechnology Information

- For predicting blood pressure status, they used conditional decision making and for predicting diabetes, they used SVM, KNN, and decision tree. Among these models, SVM worked better as they got 75% accuracy which is better than other classifier algorithms.
- Maximizes the distance from the nearest data points.
- They can be used to avoid the difficulties of using linear functions in the high-dimensional feature space.
- Robustness to overfitting.

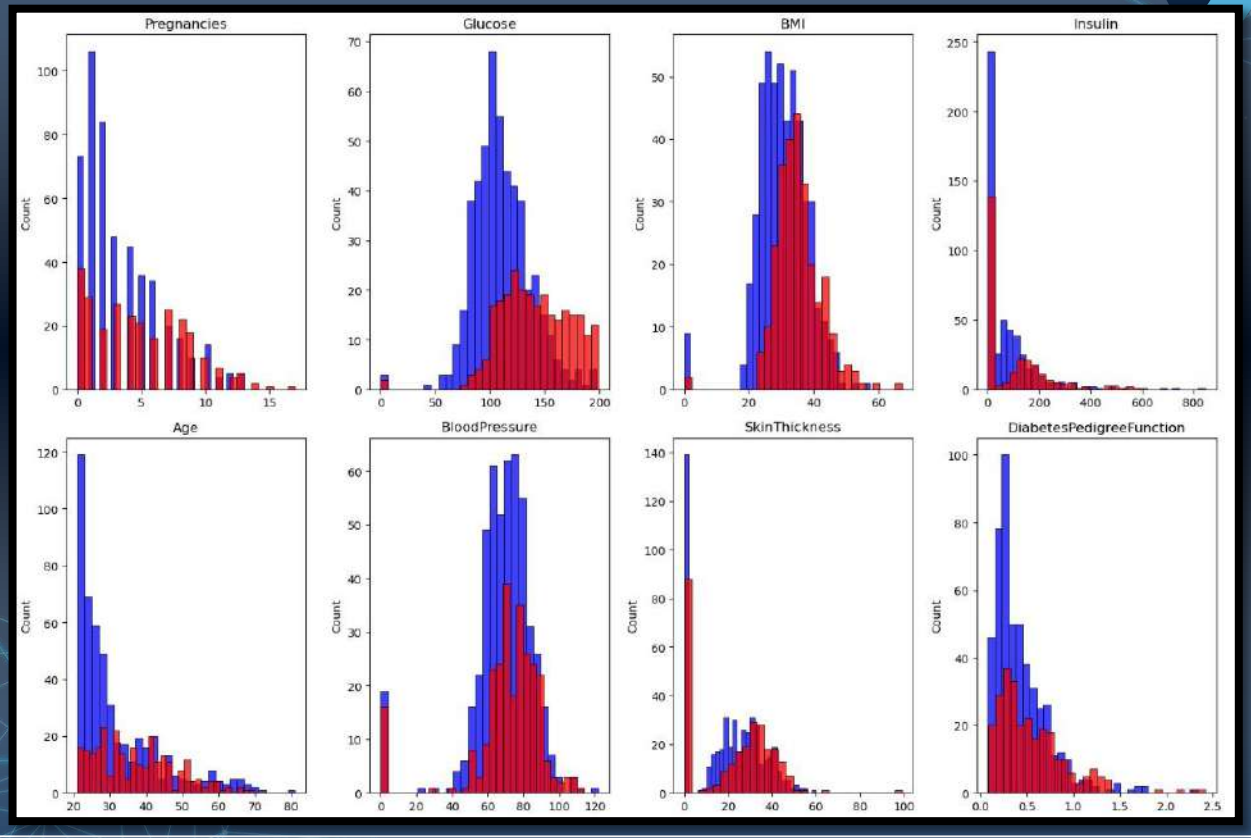
BAR GRAPH


Distribution of Outcome (0: Non-Diabetic, 1: Diabetic)
500
400
300
200
100
0
count
0
1
Outcome

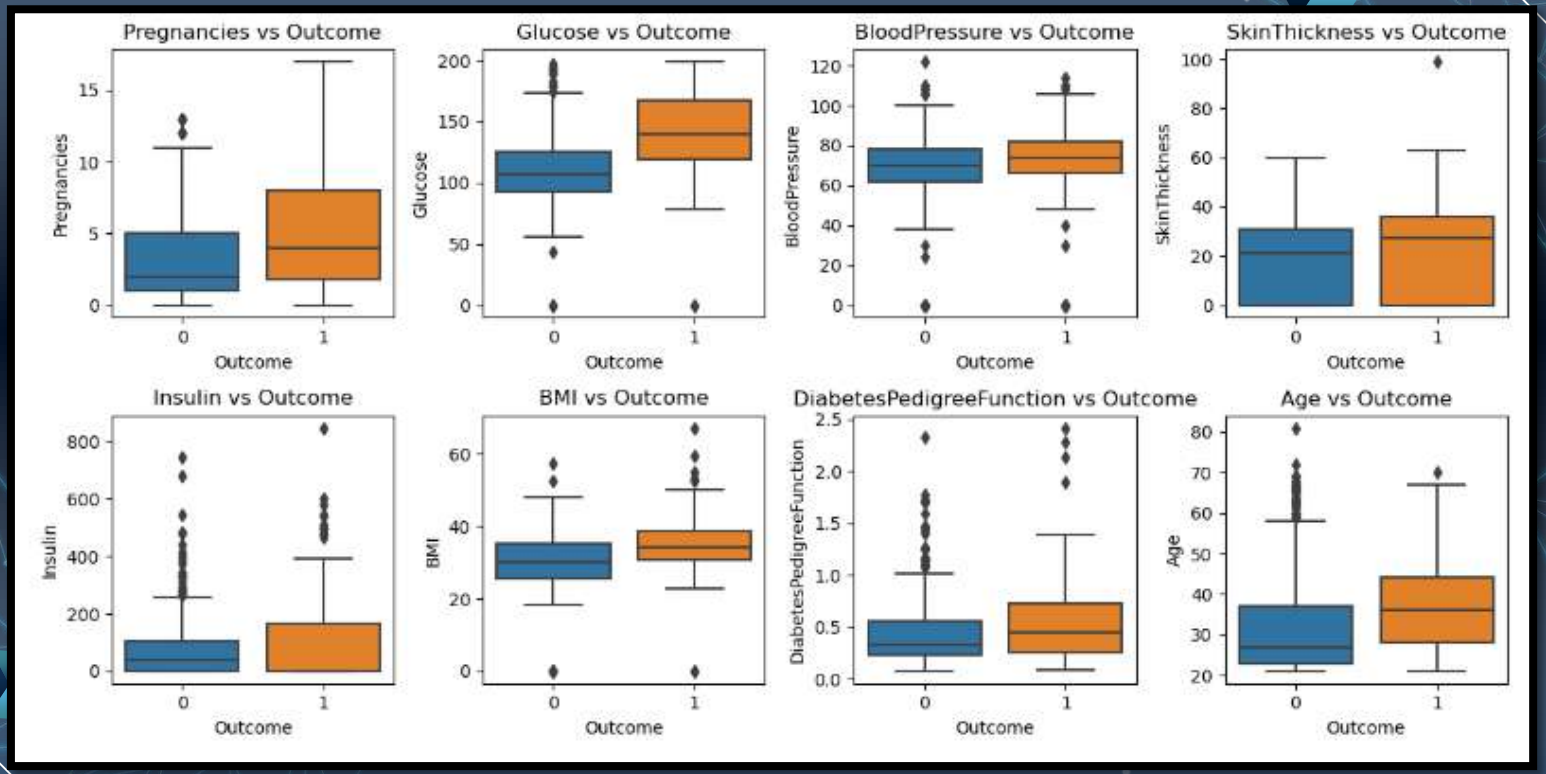
PAIR PLOT OF FEATURES



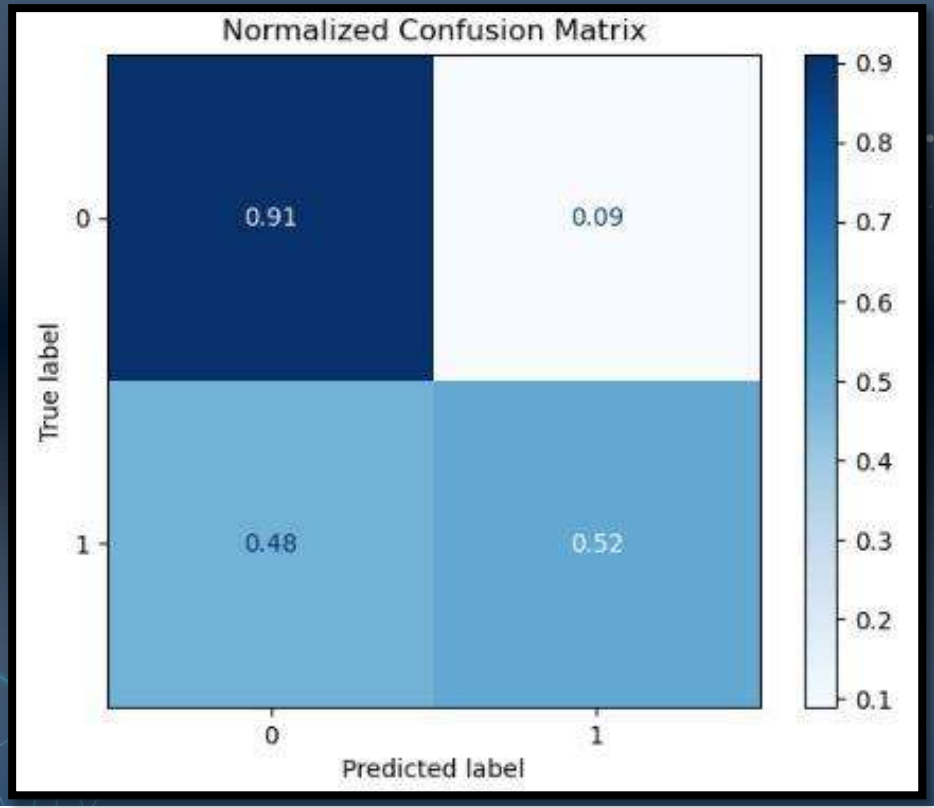
HISTOGRAM OF ATTRIBUTES



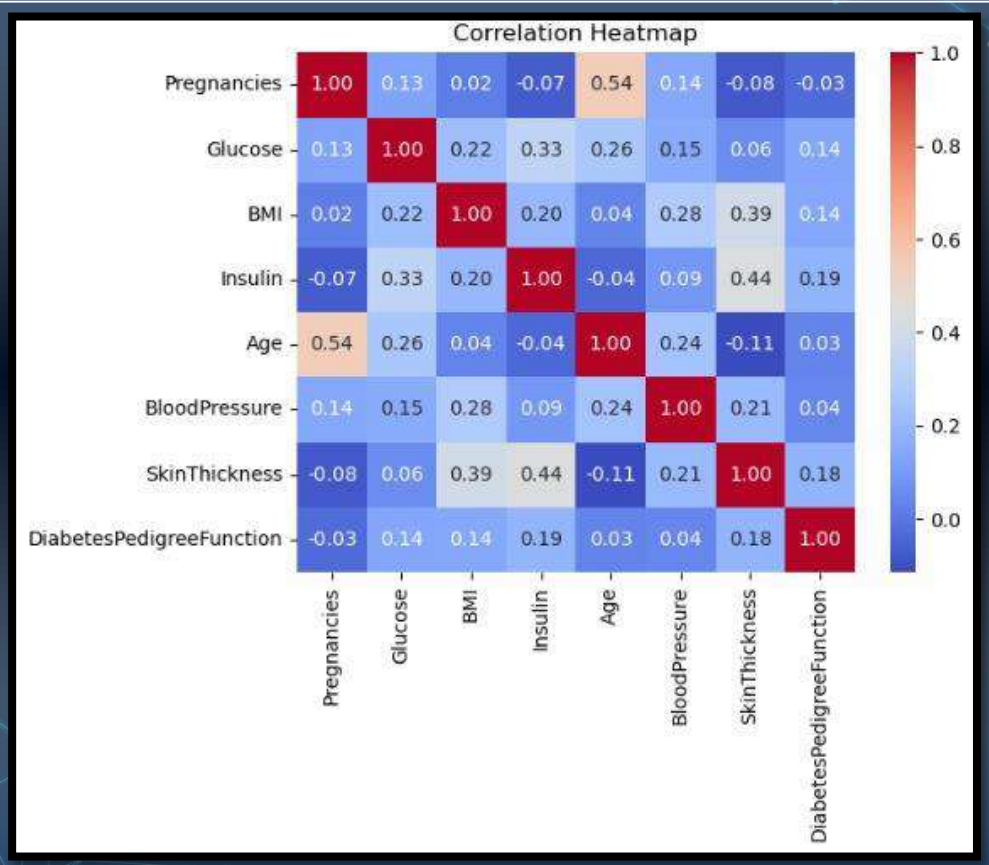
BOX PLOT


Pregnancies vs Outcome
Pregnancies
Outcome
Glucose vs Outcome
Glucose
Outcome
BloodPressure vs Outcome
BloodPressure
Outcome
SkinThickness vs Outcome
SkinThickness
Outcome
Insulin vs Outcome
Insulin
Outcome
BMI vs Outcome
BMI
Outcome
DiabetesPedigreeFunction vs Outcome
DiabetesPedigreeFunction
Outcome
Age vs Outcome
Age
Outcome

CONFUSION MATRIX



CORRELATION MATRIX



RESULT

```
# accuracy score on the test data
print('Accuracy score of test data : ', test_data_accuracy)
```

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

```
#Making a Predictive System

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

input_data_as_numpy_array = np.asarray(input_data)

input_data_reshaped = input_data_as_numpy_array.reshape(1,-1)

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
```

CONCLUSION

- In this project, we successfully developed a machine learning model for diabetes prediction using Python. We started by collecting and preprocessing a dataset containing various patient attributes, including age, BMI, and glucose levels, and the binary diabetes label.
- This project demonstrates the effectiveness of machine learning in diabetes prediction. The model can be a valuable tool in helping to identify individuals at risk of diabetes, potentially leading to early interventions and improved patient outcomes.

REFRENCES

- Gauri D. Kalyankar, Shivananda R. Poojara and Nagaraj V. Dharwadkar," Predictive Analysis of Diabetic Patient Data Using Machine Learning and Hadoop", International Conference.
- B. Nithya and Dr. V. Ilango," Predictive Analytics in Health Care Using Machine Learning Tools and Techniques", International Conference on Intelligent Computing and Control Systems.
- P. Suresh Kumar and S. Pranavi "Performance Analysis of Machine Learning Algorithms on Diabetes Dataset using Big Data Analytics", International Conference on Infocom Technologies and Unmanned Systems, 978-1-5386-0514-1, Dec. 18-20, 2017.
- Tejas N. Joshi, Prof. Pramila M. Chawan, "Diabetes Prediction Using Machine Learning Techniques". Int. Journal of Engineering Research and Application, Vol. 8, Issue 1, (Part -II) January 2018.

Thank you !