



EXPOSYS DATA LABS

PREDICTION OF DIABETES

Project by

SHOBHAN AKSHAY GIRIDHARAN
SAMUEL WILLIAM ROBERT

SNEHA JAYWANT SABLE
DHANRAJ JAYA SHETTY

AGENDA



- Abstract
- Types of Diabetes
- Introduction
- Diabetes Infographics
- Flowchart for predicting diabetes using Machine Learning
- Dataset Description
- Data Exploration and Visualization
- Library and IDE used in the Project
- Conclusion

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. But with the growth of machine learning methods we have got the flexibility to 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. In this work we will use Machine Learning Classification and ensemble techniques on a dataset to predict diabetes. Which are XG Boost (XGB), Decision Tree (DT), Support Vector Machine (SVM) and Random Forest (RF).
- 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 capable of predicting diabetes effectively. Our Result shows that Random Forest achieved higher accuracy compared to other machine learning techniques.

TYPES OF DIABETES

TYPE 1 DIABETES

Typically Associated with:
Children to young adults

Risk Factors:
Generic or environmental factors

Treated with:
Insulin

Preventable:
No

Accounts for:
5% of all cases



TYPE 2 DIABETES

Typically Associated with:
Adult onset

Risk Factors:
Obesity, Older age, Family or personal history of the disease

Treated with:
Diet, exercise and sometimes oral drugs or insulin

Preventable: Yes.

Accounts for:
90 - 95% of all cases



GESTATIONAL DIABETES

Typically Associated with:
Pregnant women

Risk Factors:
Obesity, or family history of diabetes.

Treated with:
Diet and exercise or insulin

Preventable:
Recent studies suggest there may be ways to prevent some cases

Accounts for:
Up to 9 of pregnant women



DIABETES INFOGRAPHICS

DIABETES INFOGRAPHICS

Diabetes in numbers

422
MILLION

422 million
people have
diabetes in
the world



that's about 1 out of every 11 people

1 OUT
OF **4**

do not know they
have diabetes

Main types of diabetes



TYPE 1 DIABETES

Body produces insulin
but can't use it well



TYPE 2 DIABETES

Body produces insulin
but can't use it well



GESTATIONAL DIABETES

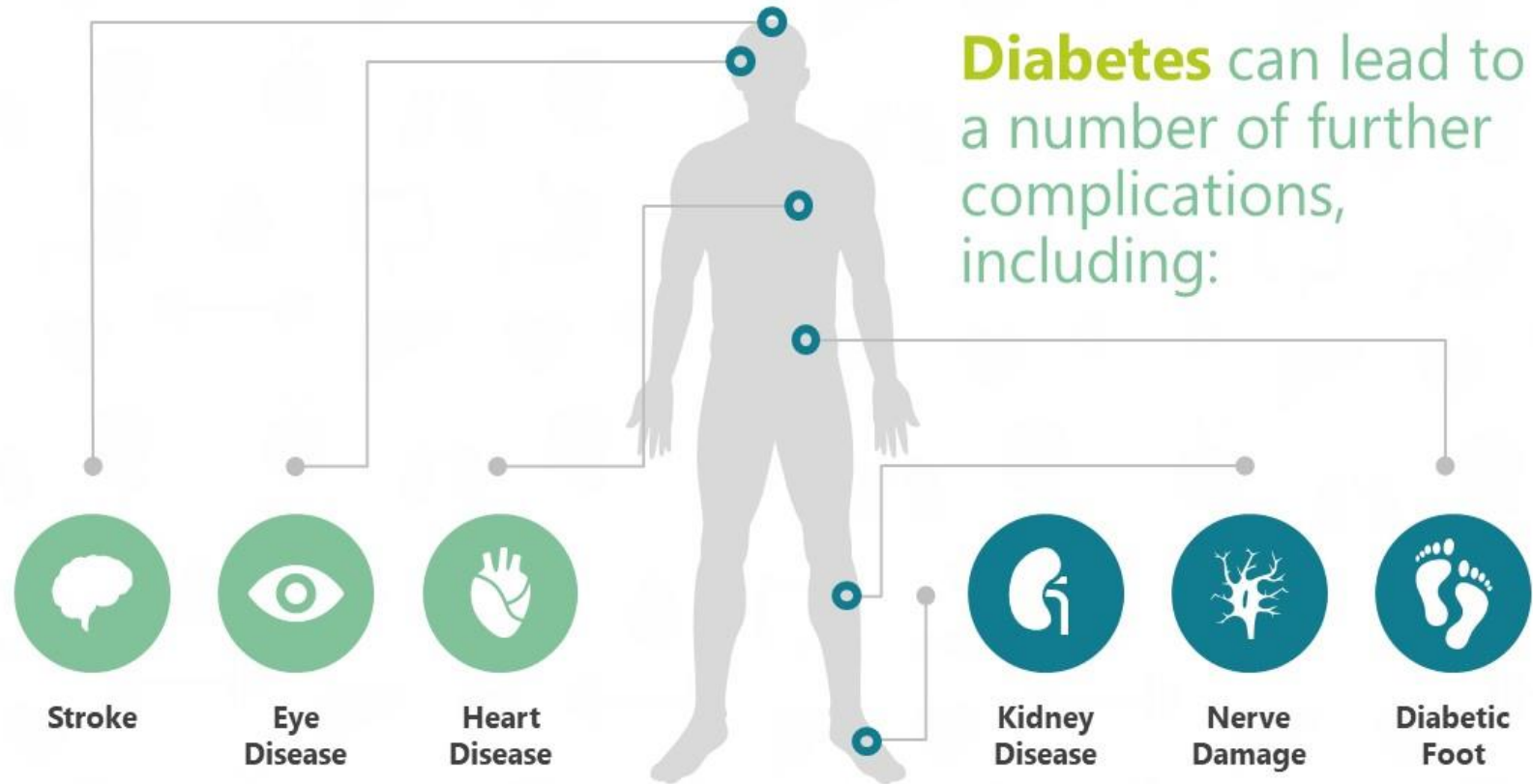
A temporary condition
in pregnancy (GDM)

INTRODUCTION

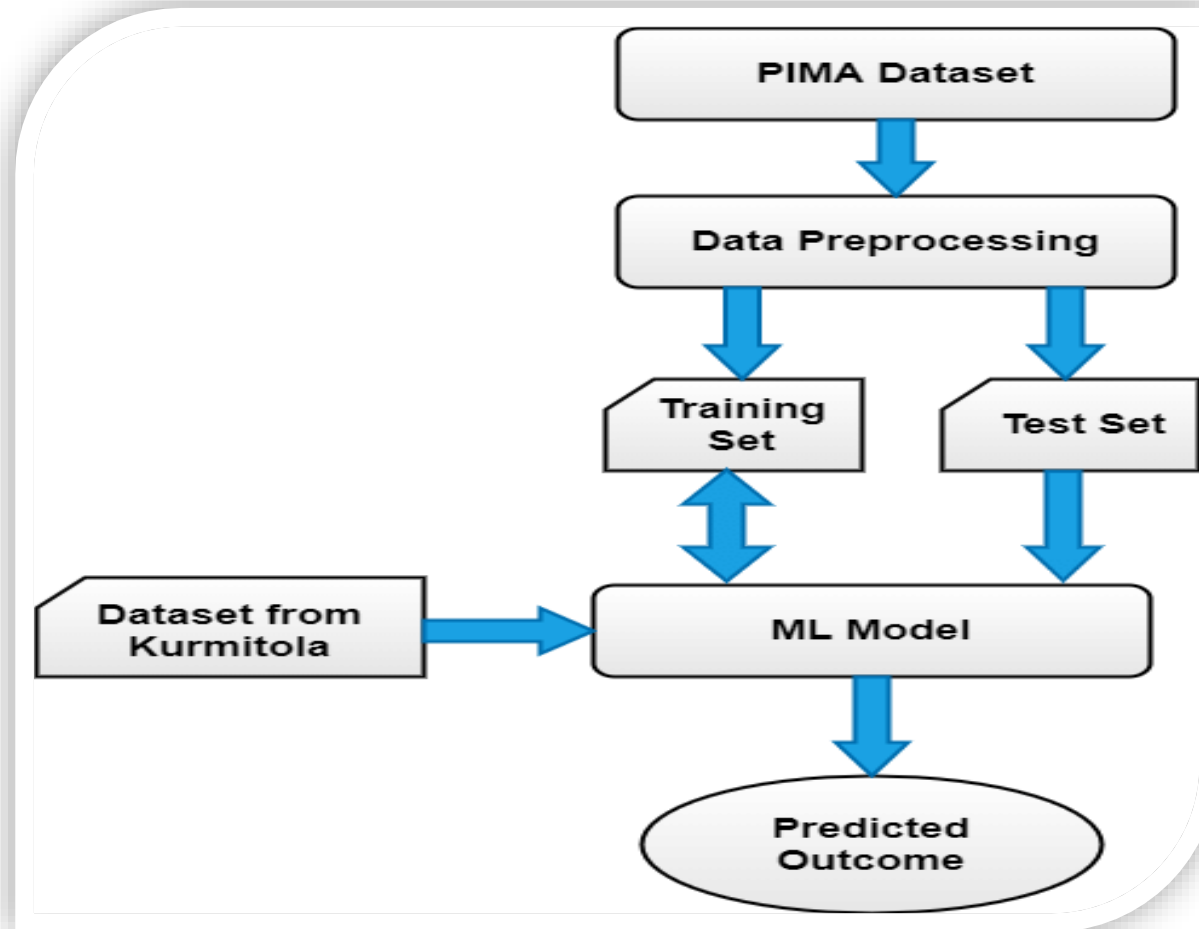
- Around 400 million people suffer from diabetes around the world. Diabetes prediction is challenging as it involves complex interactions or interdependencies between various human organs like eye, kidney, heart, etc.
- The machine learning (ML) algorithms provide an efficient way of predicting the diabetes. The objective of this work is to build a system using ML techniques for the accurate forecast of diabetes in a patient. The decision tree (DT) algorithms are well suited for this.
- In this work, we have applied the DT algorithm to forecast type 2 diabetes mellitus (T2DM). Extensive experiments were performed on the Pima Indian Diabetes Dataset (PIDDD) obtained from Kaggle. Based on the results, we observed that the decision tree was able to forecast accurately when compared to the SVM algorithm on the diabetes data. Other models such as Random Forest Classifier(RFC), XGBoost (XGB) were used which resulted to best accuracy for prediction of diabetes.

RISK FACTOR AND FURTHER COMPLICATIONS

RISK AND FURTHER COMPLICATIONS



FLOWCHART FOR PREDICTING DIABETES USING MACHINE LEARNING

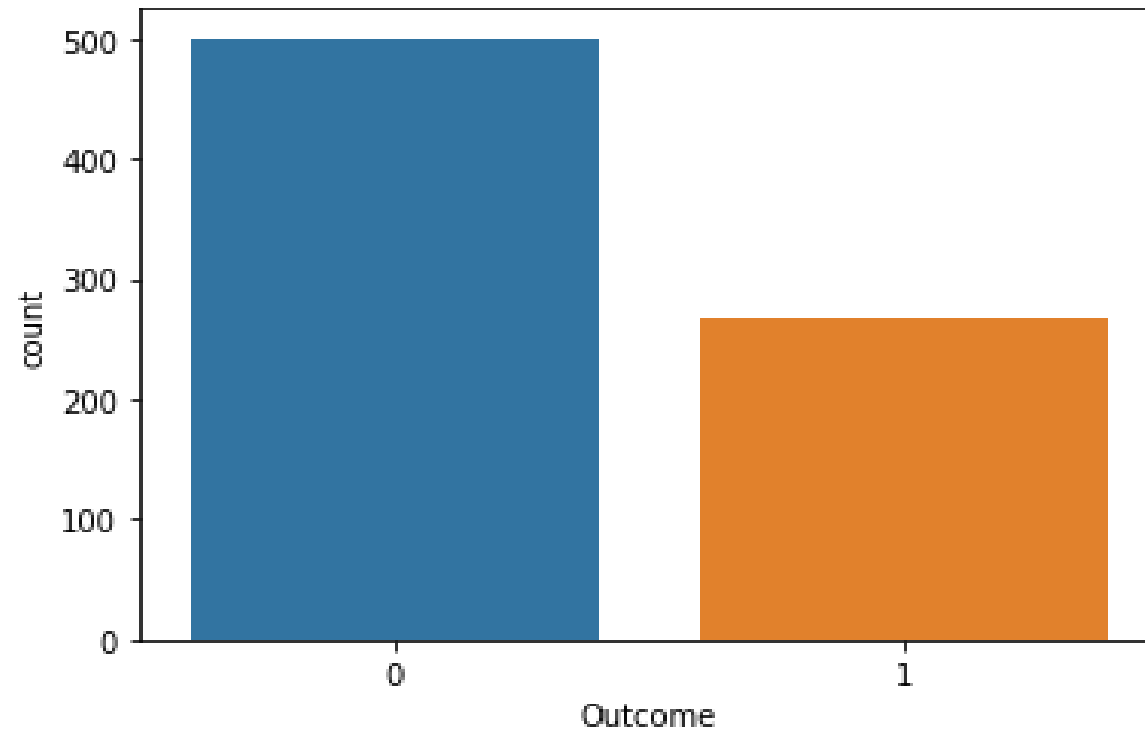


Dataset Description

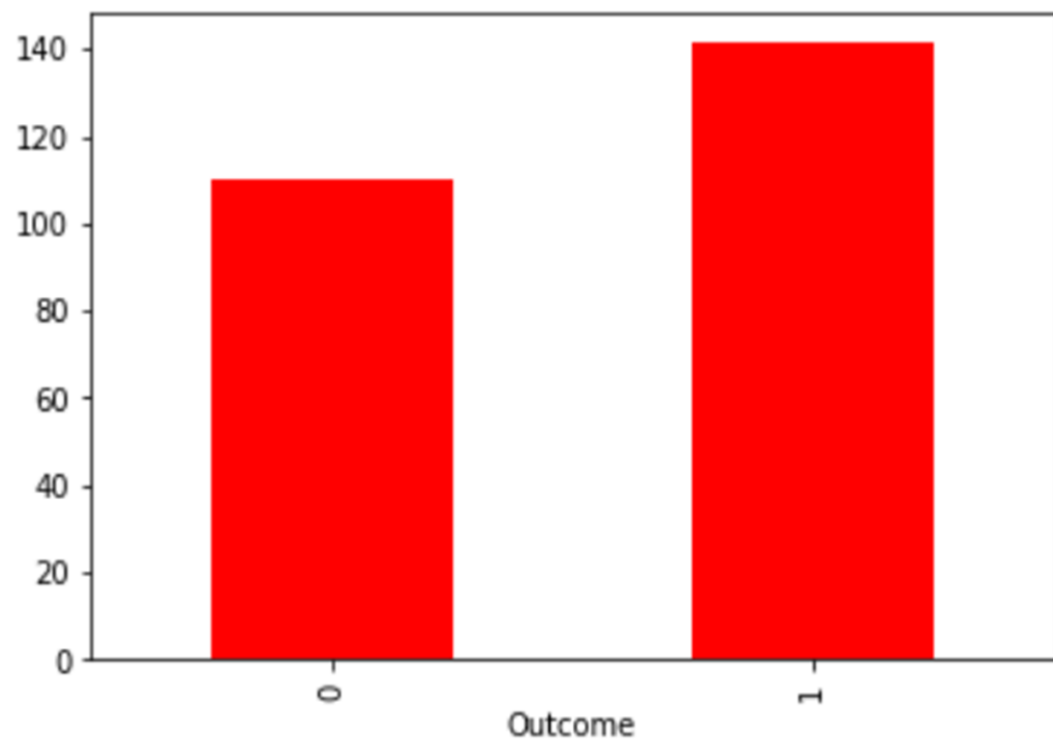
	Pregnancies	Glucose	BloodPressure	SkinThickness	Insulin	BMI	DiabetesPedigreeFunction	Age	Outcome
0	6	148	72	35	0	33.6	0.627	50	1
1	1	85	66	29	0	26.6	0.351	31	0
2	8	183	64	0	0	23.3	0.672	32	1
3	1	89	66	23	94	28.1	0.167	21	0
4	0	137	40	35	168	43.1	2.288	33	1
5	5	116	74	0	0	25.6	0.201	30	0
6	3	78	50	32	88	31.0	0.248	26	1
7	10	115	0	0	0	35.3	0.134	29	0
8	2	197	70	45	543	30.5	0.158	53	1
9	8	125	96	0	0	0.0	0.232	54	1

THE DATA IS GATHERED FROM UCI REPOSITORY WHICH IS NAMED AS PIMA INDIAN DIABETES DATASET. THE DATASET HAVE MANY ATTRIBUTES OF 768 PATIENTS.

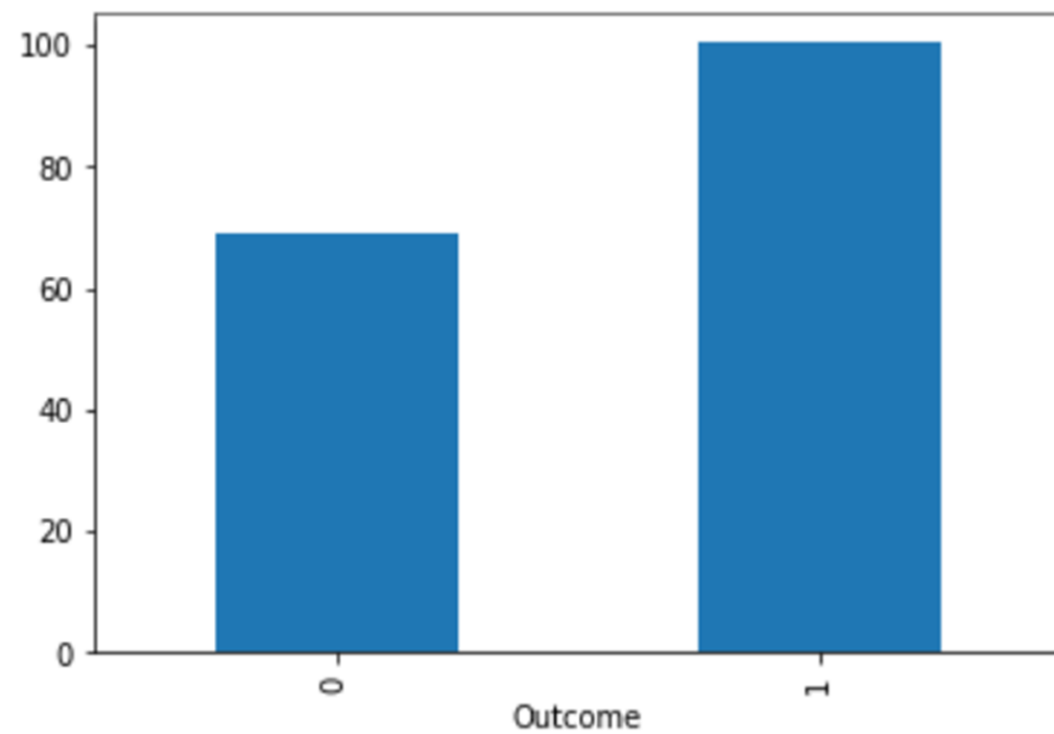
RATIO OF DIABETIC AND NON-DIABETIC PATIENT



DATA EXPLORATION AND VISUALIZATION



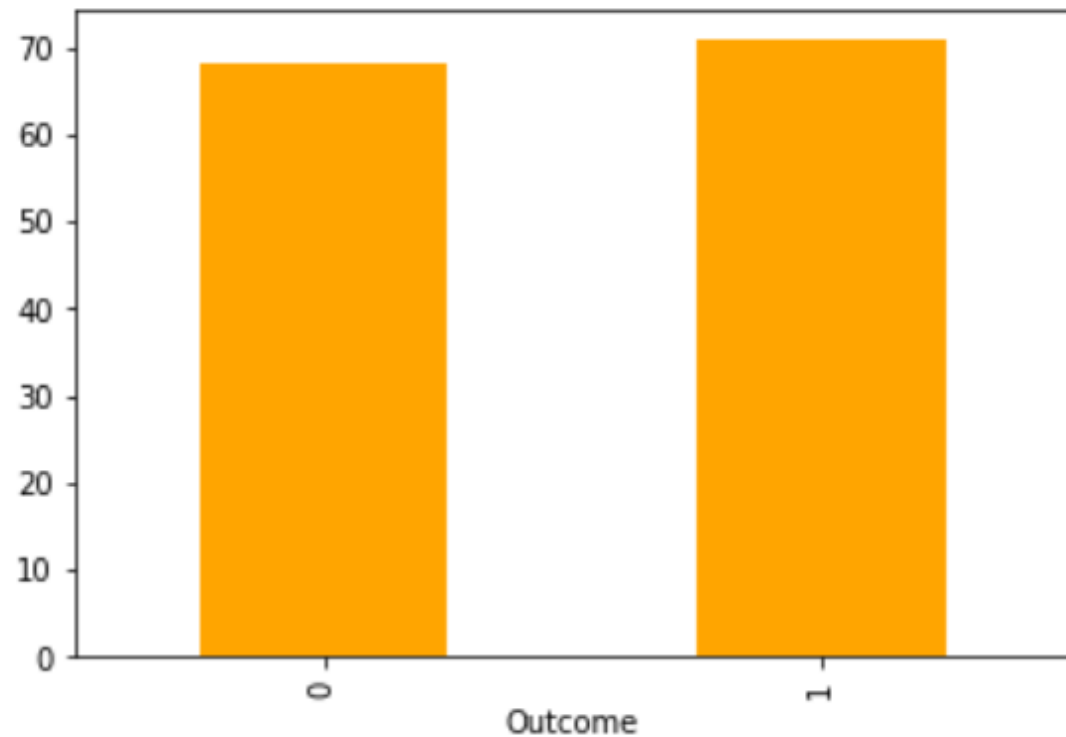
Glucose



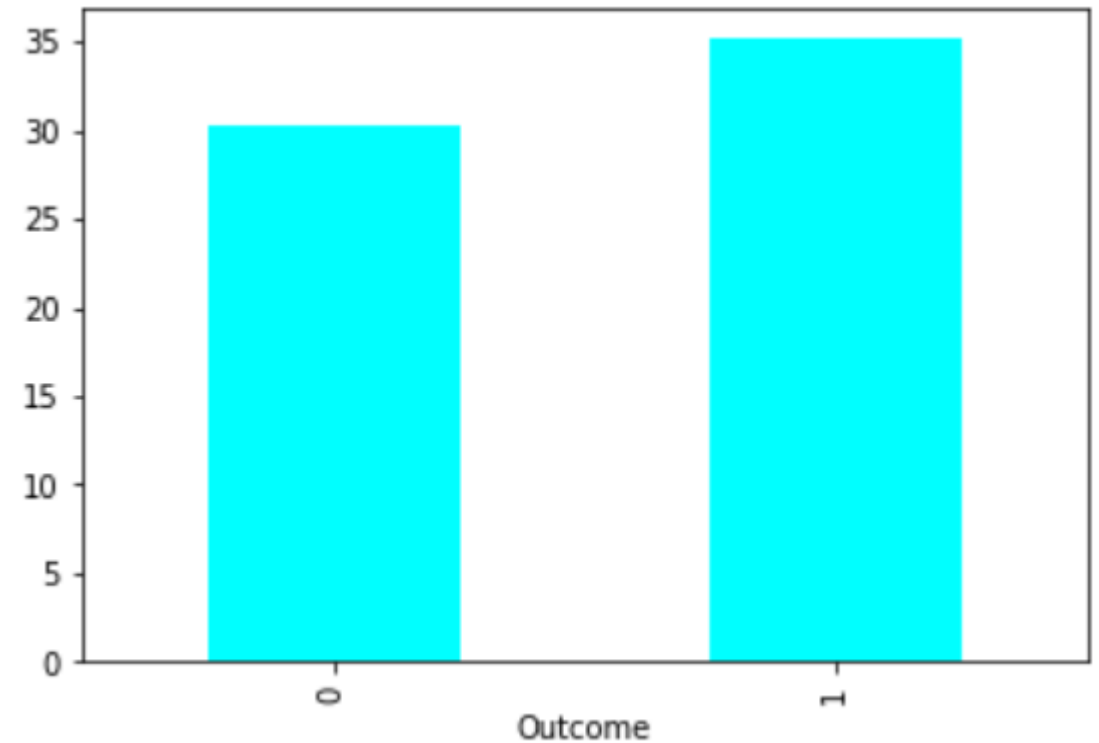
Insulin

DATA EXPLORATION AND VISUALIZATION

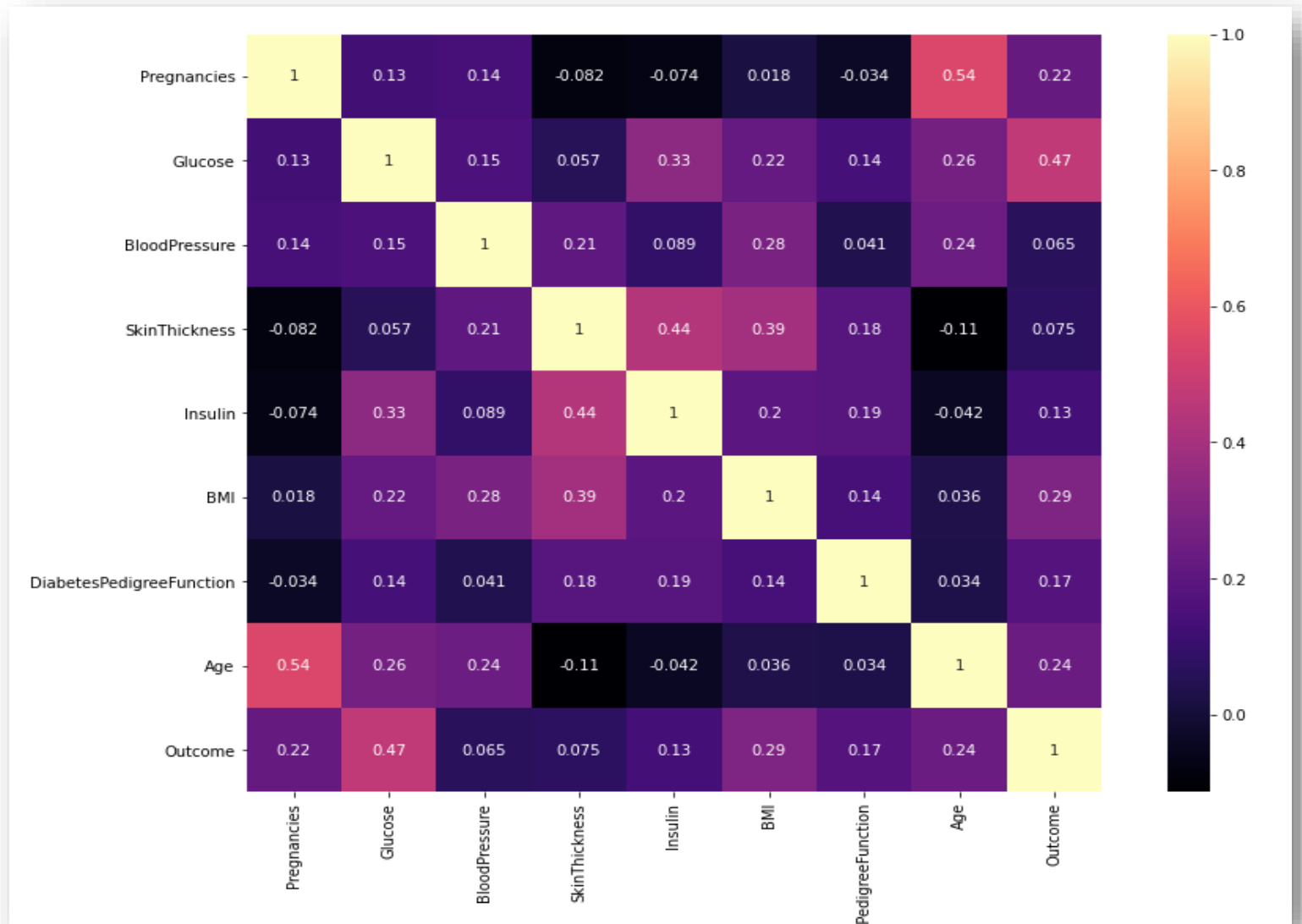
Blood Pressure



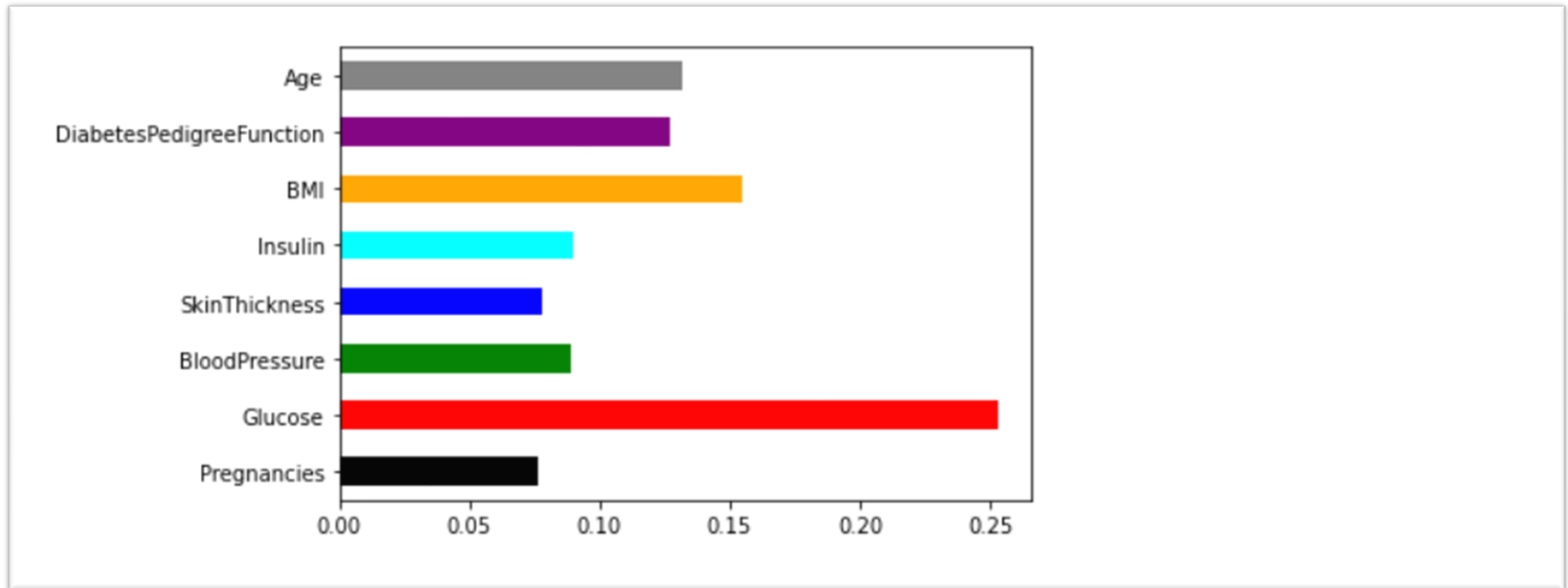
BMI



HEATMAP OF CORRELATED DATA



DIABETES PREDICTION OF GLUCOSE & BMI



The graph shows the most important feature in this diabetes prediction is glucose & BMI.

LIBRARY AND IDE USED IN THE PROJECT

- a. Pandas
- b. NumPy
- c. Seaborn
- d. Matplotlib
- e. Sci-kit learning

IDE Used for compiling:- Jupyter Notebook (Python)

CONCLUSION

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. The proposed approach uses various classification and ensemble learning method in which SVM, XGBoost, Random Forest, Decision Tree classifiers are used. Random Forest Classifier is the best model for this prediction since it has an accuracy score of 75.98% .

The Experimental results can be assisted health care to take early prediction and make early decision to cure diabetes and save humans life.