DIABETES-CLASSIFICATION

Project Explanation:

As the data is labeled it falls under Supervised learning.

Here the targets are divided into two classes(0 and 1) thus it falls under Classification.

In this dataset we have the info of 'Pregnancies', 'Glucose', 'BloodPressure', 'SkinThickness', 'Insulin','BMI', 'DiabetesPedigreeFunction', 'Age' of a person based on this they are classified as they have diabetes or not.

Tools Used:

Python 3.9.5

Jupyter Notebook

Pandas 1.2.4 for data analysis

Seaborn 0.11.1 for data visualiztion

Scikit-learn 0.24.2 for machine learning

Algorithms used:

RandomForestClassifier: RandomForests are an ensemble learning method for classificarion,regression and other tasks that operates by constructing a multitude of decision trees at training time.

LogisticRegression: A linear model with a process of modeling the probability of a discrete outcome given an input variable.

Conclusion:

RandomForestClassifier has an accuracy of 81.81% and LogisticRegression has an accuracy of 81.16%.