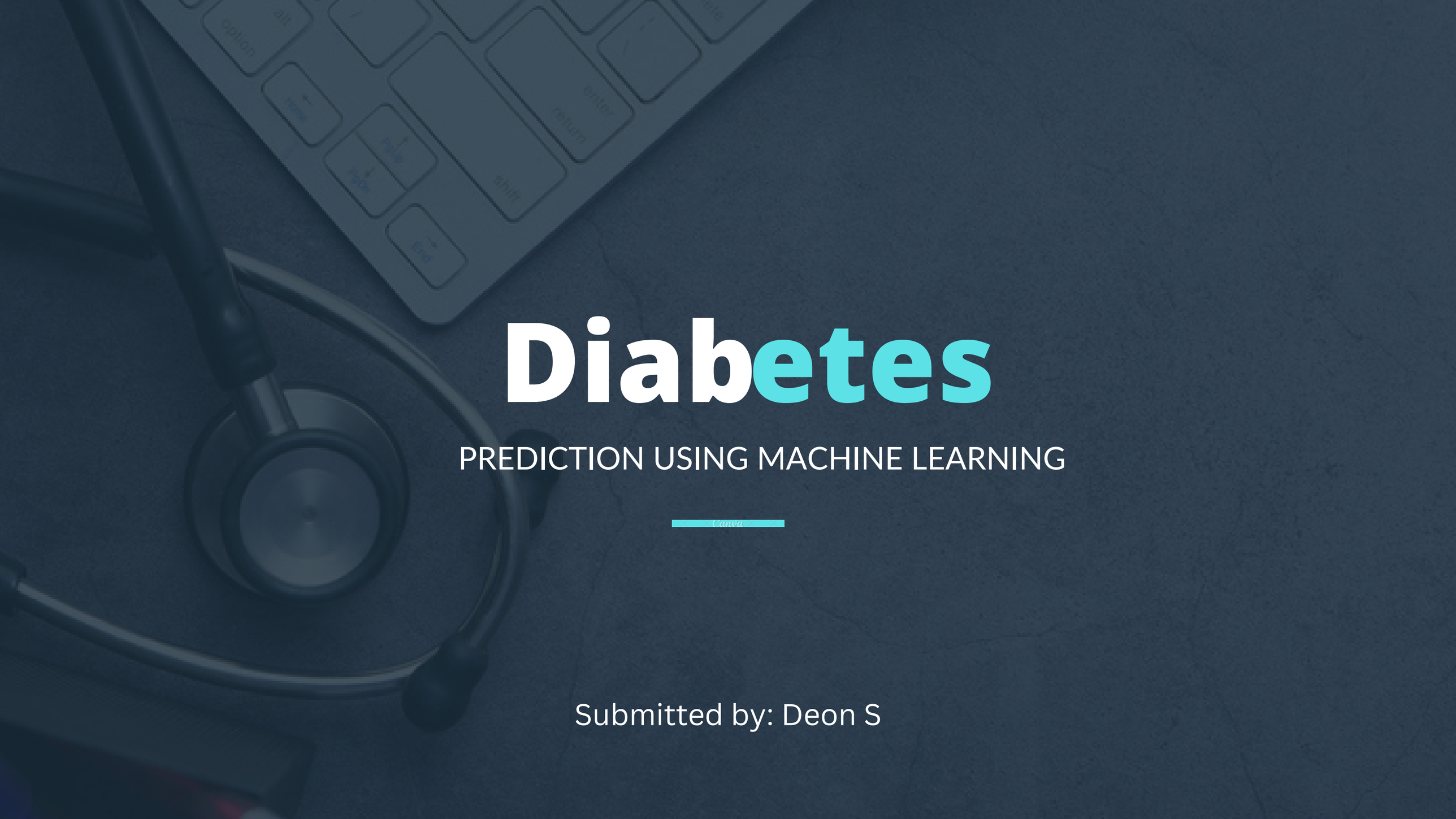


Diabetes

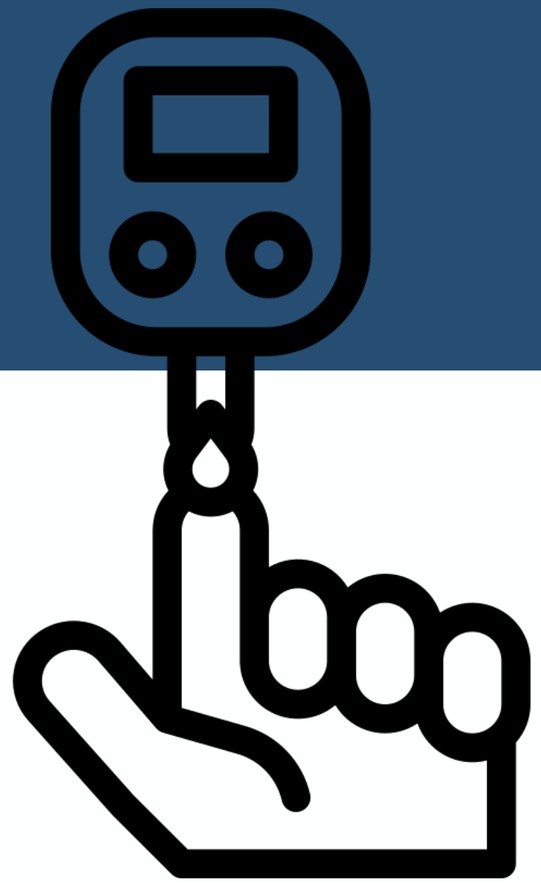
PREDICTION USING MACHINE LEARNING

Submitted by: Deon S


Hey,
I'll be predicting
if you have diabetes
or not.

The WHO Definition for Diabetes

Diabetes is a chronic disease that occurs either when the pancreas does not produce enough insulin or when the body cannot effectively use the insulin it produces. Insulin is a hormone that regulates blood glucose. Hyperglycaemia, also called raised blood glucose or raised blood sugar, is a common effect of uncontrolled diabetes and over time leads to serious damage to many of the body's systems, especially the nerves and blood vessels.



PROBLEM STATEMENT

To develop a system which can perform early prediction of diabetes for a patient with a higher accuracy by combining the results of different machine learning techniques.

Dataset

This dataset is originally from the National Institute of Diabetes and Digestive and Kidney Diseases.

Variables:

- Pregnancies: Number of times pregnant
- Glucose: Plasma glucose concentration a 2 hours in an oral glucose tolerance test
- BloodPressure: Diastolic blood pressure (mm Hg)
- SkinThickness: Triceps skin fold thickness (mm)
- Insulin: 2-Hour serum insulin (mu U/ml)
- BMI: Body mass index (weight in kg/(height in m)^2)
- DiabetesPedigreeFunction: Diabetes pedigree function
- Age: Age (years)
- Outcome: Class variable (0 or 1)

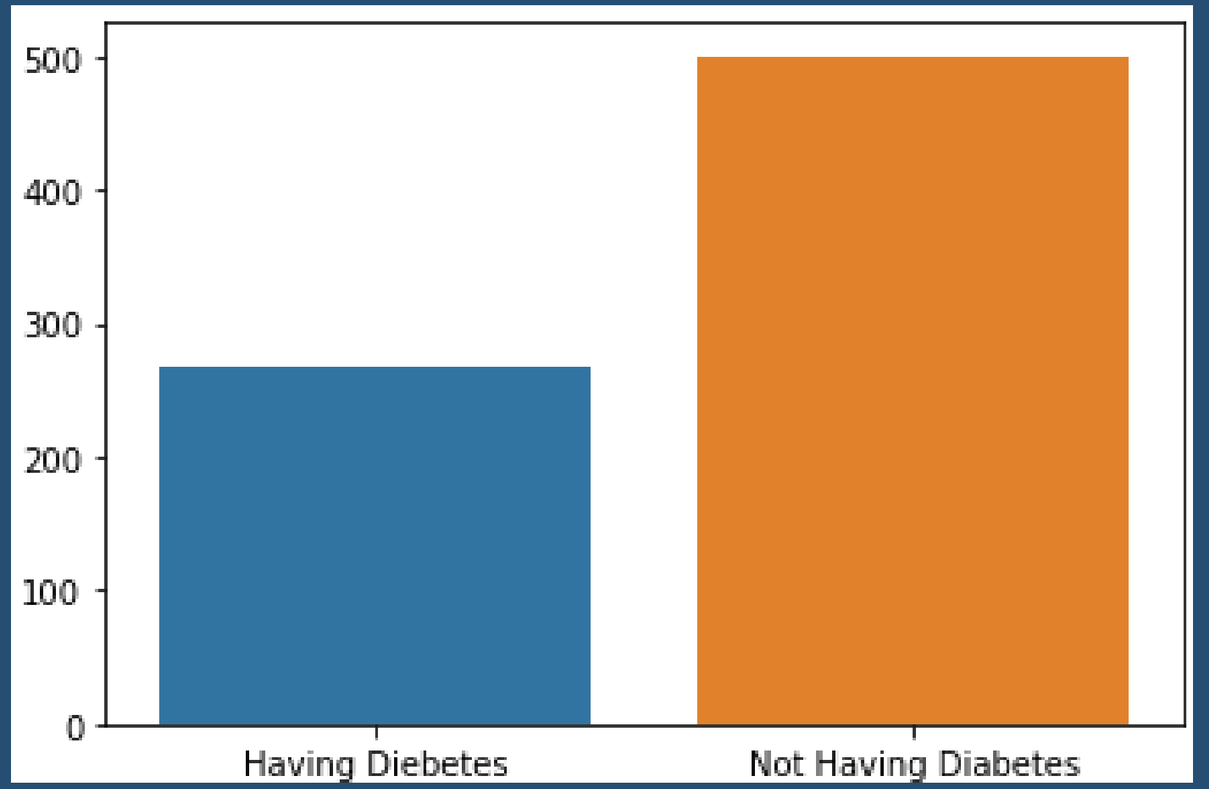
Sample Dataset

	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

Models Used

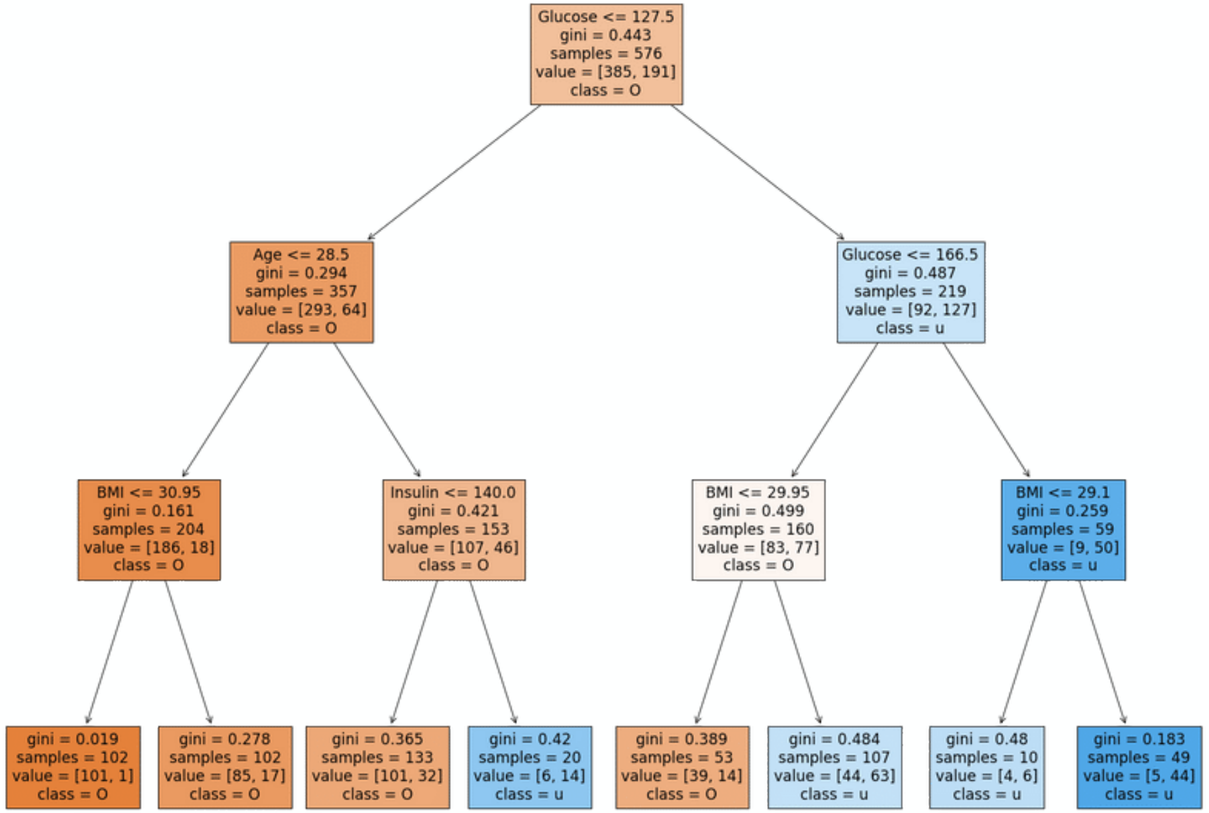
LOGISTIC REGRESSION

Logistic regression is a machine learning technique used when dependent variables are able to categorize. The outputs obtained by using the logistic regression is based on the available features. Logistic regression had an accuracy of 0.786



DECISION TREE

A decision tree is a flowchart-like tree structure, where each internal node denotes a test on an attribute, each branch represents an outcome of the test, and each leaf node (terminal node) holds a class label. Similar to logistic regression, the decision tree also had the same accuracy of 0.786



Since both models had the best accuracy, I decided to proceed with logistic regression.

Saving the Trained Model

Using the pickle library, the trained model was saved into the local files. This trained model was used in the streamlit backend to predict if the patient has diabetes or not.

Saving the trained model

```
import pickle
```

```
filename = 'trained_model_1.sav'
pickle.dump(model, open(filename, 'wb'))
```

```
# loading the saved model
loaded_model = pickle.load(open('trained_model_1.sav', 'rb'))
```

```
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)

prediction = loaded_model.predict(input_data_reshaped)
print(prediction)

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

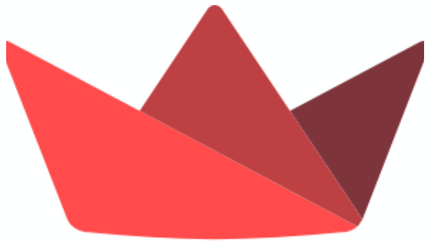
User Interface

UI was created using the Streamlit library

GitHub Repository

Scan to view my GitHub Repo

Tools I Work With



Thank You