**Diabetes prediction using Machine learning**

au622021243025 : JOHN J JASHO

Phase 5 submission Document

In this section we will document the complete project and prepare it for submission



**Overview**

| Group: | Artificial intelligence |
| --- | --- |
| Project Name: | Diabetes prediction using Machine learning |
| Language Used: | Python |
| IDE: | GoogleCollab/jupyter notebook |
| Python version: | Python 3.7 |
| Database: | Required |
| Abstract | It’s an ML-based project which  Involves most of the ML steps  Like collection of data, Exploring the data |

**INTRODUCTION**

Diabetes is a prevalent chronic medical condition that affects millions of people worldwide. It is characterized by the body's inability to regulate blood sugar levels properly, leading to a variety of health complications. Early diagnosis and prediction of diabetes are crucial for effective management and prevention of related complications. Machine learning, a subset of artificial intelligence, has emerged as a powerful tool for diabetes prediction and risk assessment. By leveraging machine learning techniques, healthcare providers and researchers can develop accurate predictive models that analyze various patient attributes to identify individuals at risk of developing diabetes.

This introduction provides an overview of the concept and process of diabetes prediction using machine learning. It outlines the significance of such predictions, the key steps involved, and the potential impact on healthcare and patient outcomes.

**Significance of Diabetes Prediction:**

**1. Early Intervention:** Predicting diabetes at an early stage allows for timely interventions, lifestyle modifications, and medical treatments, which can help prevent or manage the condition more effectively.

**2. Personalized Medicine:** Machine learning models can tailor interventions to individual risk profiles, ensuring that patients receive the most appropriate care and guidance.

**3. Healthcare Resource Optimization:** Predictive models can help healthcare systems allocate resources more efficiently by targeting high-risk individuals for intensive monitoring and care, ultimately reducing the burden on the healthcare system.

**4. Prevention and Cost Reduction:** Early prediction and management of diabetes can lead to a reduction in long-term healthcare costs and improve overall patient well-being.

**Key Steps in Diabetes Prediction Using Machine Learning:**

**1. Data Collection:** Gather a comprehensive dataset containing patient information, including demographics, medical history, lifestyle factors, and relevant biomarkers.

**2. Data Preprocessing:** Clean and preprocess the data by handling missing values, outliers, and ensuring data consistency.

**3. Feature Selection/Engineering:** Identify and select relevant features that contribute to diabetes risk. This may involve creating new features or transforming existing ones.

**4. Model Selection:** Choose an appropriate machine learning algorithm for the prediction task, such as logistic regression, decision trees, support vector machines, or neural networks.

**5. Model Training:** Train the selected model on a portion of the data, allowing it to learn patterns and associations between features and diabetes outcomes.

**6. Model Evaluation:** Assess the model's performance using appropriate evaluation metrics to determine its accuracy and reliability.

**7. Hyper parameter Tuning:** Fine-tune the model's hyper parameters to optimize its predictive power.

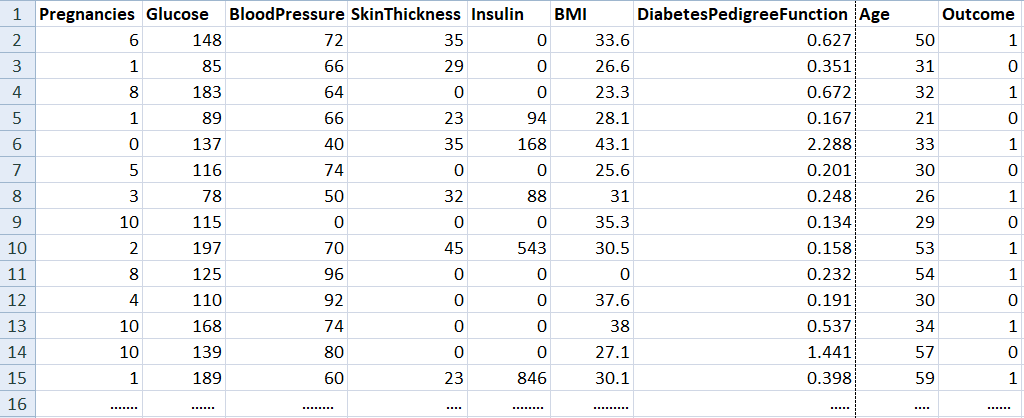
**8. Deployment:** Deploy the trained model for real-world use, such as within healthcare systems or as part of a mobile application.

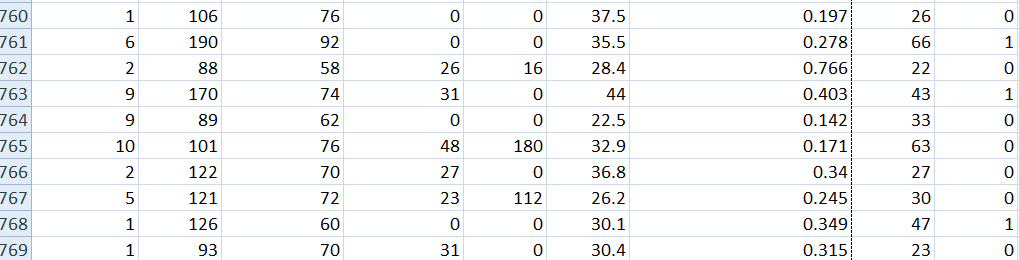
**9. Monitoring and Maintenance:** Continuously monitor the model's performance, update it with new data, and make improvements as needed to ensure ongoing accuracy.

**Data source**

The datasets consists of several medical predictor variables and one target variable, outcome Predictor variables include the number of pregnancies the patient has had, their BMI, insulin level, age, and so on.

**Dataset link:** (<https://www.kaggle.com/datasets/uciml/pima-indians-diabetes-database>)



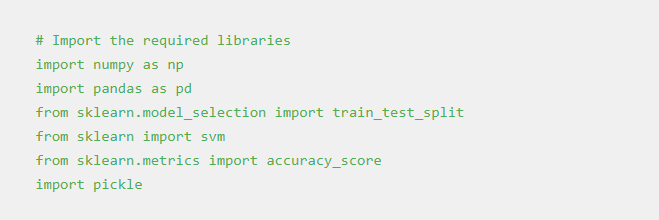


**769 rows X 9 columns**

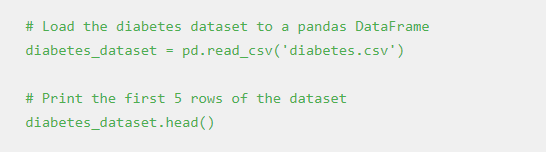
### Exploring the Data

Now we have to set the development environment to build our project. For this project, we are going to build this Diabetes prediction using Machine Learningin [Google Colab](http://colab.research.google.com/). You can also use Jupyter Notebook.

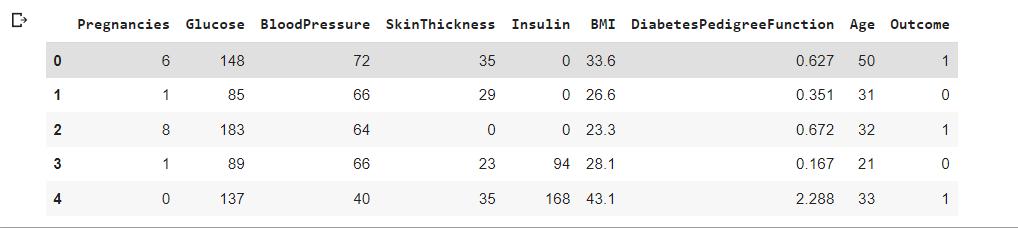
After downloading the dataset, import the necessary libraries to build the model.

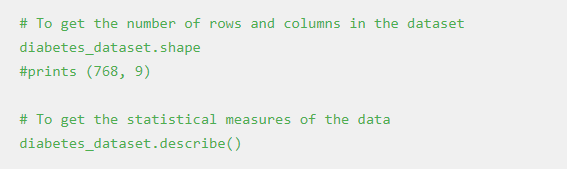


Load the data using the *read\_csv* method in the pandas library. Then the *head()* method in the pandas library is used to print the rows up to the limit we specify. The default number of rows is five.

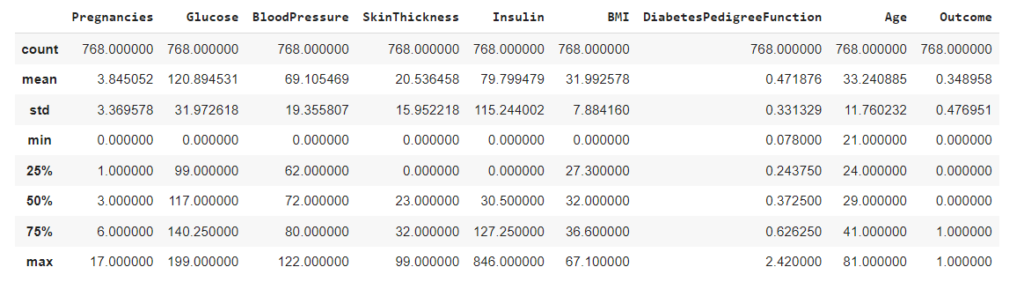


**Output:**

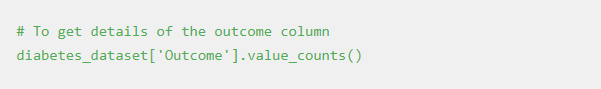




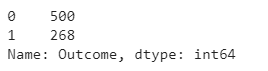
**Output:**



And, it is clear that the Outcome column is the output variable. So let us explore more details about that column.

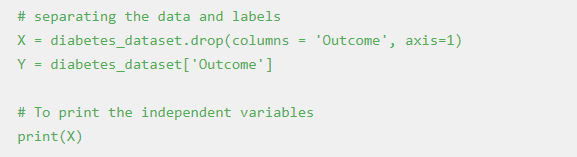


In the output, the value 1 means the person is having Diabetes, and 0 means the person is not having Diabetes. We can see the total count of people with and without Diabetes.

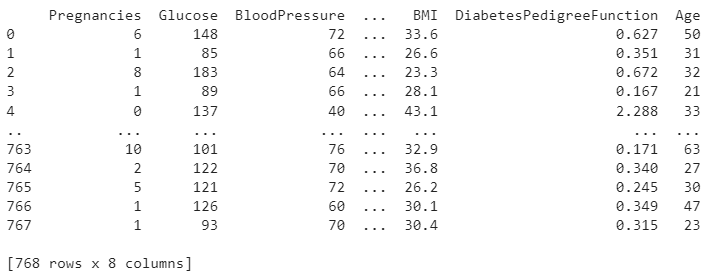


### Splitting the data

The next step in the building of the Machine learning model is splitting the data into training and testing sets. The training and testing data should be split in a ratio of 3:1 for better prediction results.



**Output:**



#Split the data into train and test

X\_train, X\_test, Y\_train, Y\_test = train\_test\_split(X,Y, test\_size = 0.2, stratify=Y, random\_state=2)

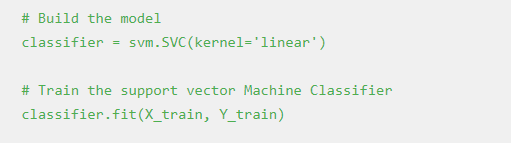
print(X.shape, X\_train.shape, X\_test.shape)

**Output:**

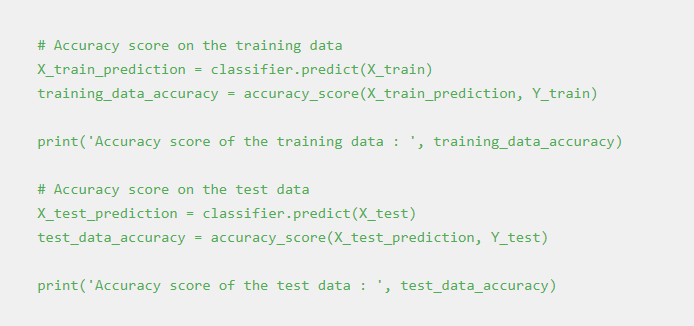
*(768, 8) (614, 8) (154, 8)*

### Training the model

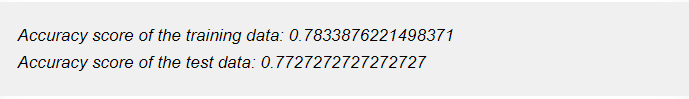
The next step is to build and train our model. We are going to use a Support vector classifier algorithm to build our model.



After building the model, the model has to predict output with test data. After the prediction of the outcome with test data, we can calculate the accuracy score of the prediction results by the model.

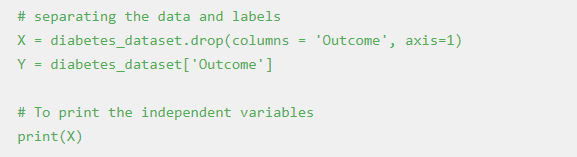


**Output:**

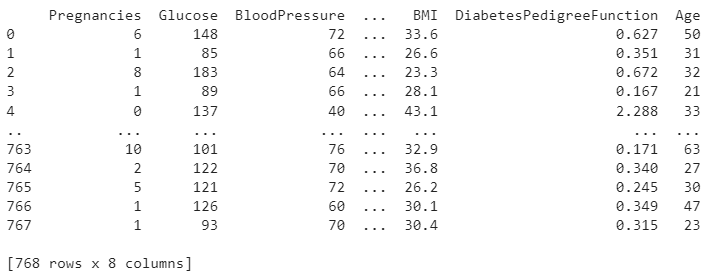


### Splitting the data

The next step in the building of the Machine learning model is splitting the data into training and testing sets. The training and testing data should be split in a ratio of 3:1 for better prediction results.



**Output:**



#Split the data into train and test

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)

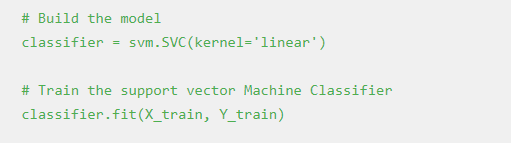
**Output:**

*(768, 8) (614, 8) (154, 8)*

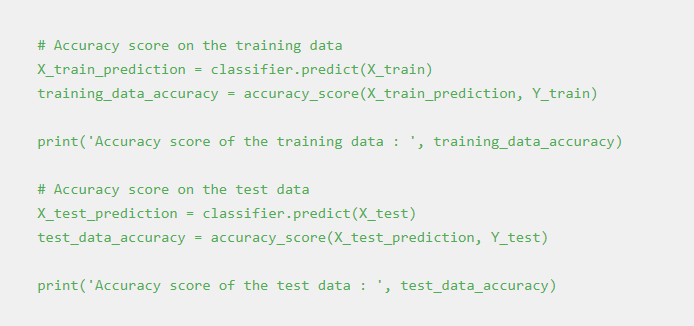
### Training the model

### 2.png

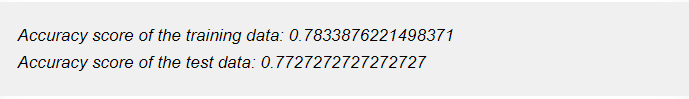
The next step is to build and train our model. We are going to use a Support vector classifier algorithm to build our model.



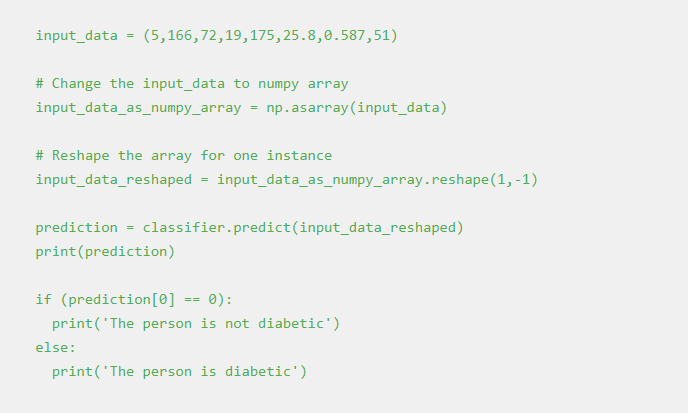
After building the model, the model has to predict output with test data. After the prediction of the outcome with test data, we can calculate the accuracy score of the prediction results by the model.



**Output:**



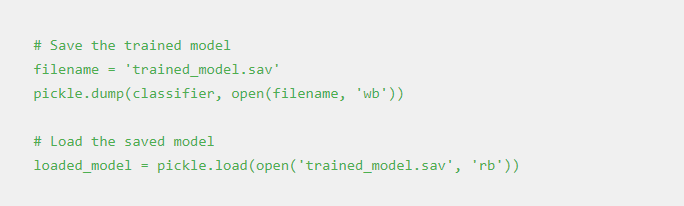
**5. Evaluating the model**



**Output**:

*The person is diabetic*

**Saving the file**



Once you run this code a new file named trained\_model.sav will be saved in the project folder.

## Deploying the model

One of the most important and final steps in building a Machine Learning project is Model deployment. There are many frameworks available for deploying the Machine learning model on the web. Some of the most used Python frameworks are Django and Flask. But these frameworks require a little knowledge of languages such as HTML, CSS, and JavaScript.

So, a new framework known as Streamlit was introduced to deploy the Machine Learning model without the need to have the knowledge of Front End Languages. It is quite easy to deploy using Streamlit. So, we will use the [Streamlit](https://streamlit.io/) framework to deploy our model. Although Streamlit has many advantages over the other frameworks, lot more features are under development. If you are getting started in Machine Learning then this framework will be a perfect start to deploy your machine learning model on the web.

### Python Code to Deploy ML model using Streamlit

To install Streamlit run the following command in the command prompt or terminal.



Open a new Python file and put the following code.

***App.py***

**import numpy as np**

**import pickle**

**import streamlit as st**

**# Load the saved model**

**loaded\_model = pickle.load(open('C:/Users/ELCOT/Downloads/trained\_model.sav', 'rb'))**

**# Create a function for Prediction**

**def diabetes\_prediction(input\_data):**

**# Change 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):**

**return 'The person is not diabetic'**

**else:**

**return 'The person is diabetic'**

**def main():**

**# Give a title**

**st.title('Diabetes Prediction Web App')**

**# To get the input data from the user**

**Pregnancies = st.text\_input('Number of Pregnancies')**

**Glucose = st.text\_input('Glucose Level')**

**BloodPressure = st.text\_input('Blood Pressure value')**

**SkinThickness = st.text\_input('Skin Thickness value')**

**Insulin = st.text\_input('Insulin Level')**

**BMI = st.text\_input('BMI value')**

**DiabetesPedigreeFunction = st.text\_input('Diabetes Pedigree Function value')**

**Age = st.text\_input('Age of the Person')**

**# Code for Prediction**

**diagnosis = ‘’**

**# Create a button for Prediction**

**if st.button('Diabetes Test Result'):**

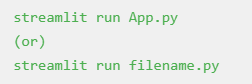
**diagnosis = diabetes\_prediction([Pregnancies, Glucose, BloodPressure, SkinThickness, Insulin, BMI, DiabetesPedigreeFunction, Age])**

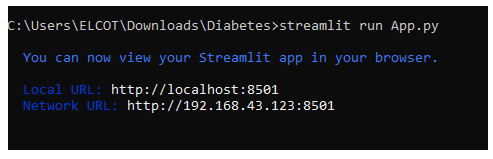
**st.success(diagnosis)**

**if \_\_name\_\_ == '\_\_main\_\_':**

**main()**

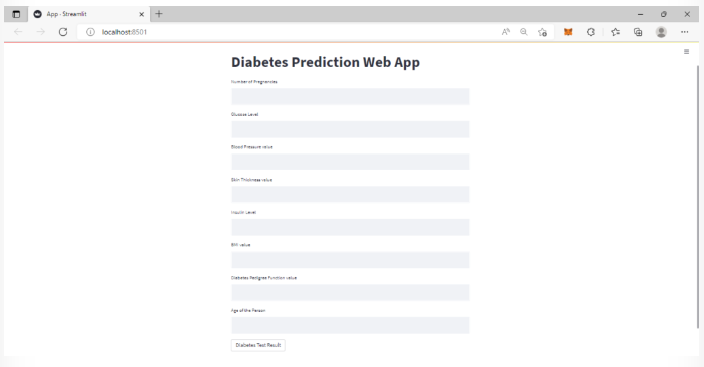
**Save the file after pasting the code. And then to deploy using streamlit go to command prompt and run the following command.**





After running the command the web app will open in the localhost webserver. Otherwise, go to your browser and type *localhost:8501*. The following output will be shown.

**Output**:



Sample Input data for a person does not have diabetes is {1, 85, 66, 29, 0, 26.6, 0.351, 31}. These data as input will generate the following output in the web app.



Sample input data for a person who have diabetes is {6, 148, 72, 35, 0, 33.6, 0.627, 50}. These data as input will generate the following output in the web app.



## Conclusion

Diabetes prediction using machine learning has emerged as a promising approach to address the growing global health concern of diabetes. This technology provides the medical community with valuable tools to identify individuals at risk, enabling early intervention and personalized care. The following key points summarize the significance and impact of diabetes prediction through machine learning:

**Early Intervention and Prevention:** Diabetes prediction models empower healthcare professionals to identify at-risk individuals before symptoms manifest. Early intervention, lifestyle modifications, and targeted treatments can prevent or delay the onset of diabetes, thereby improving patients' quality of life and reducing the long-term health and economic burdens associated with the condition.

**Personalized Healthcare:** Machine learning models enable the delivery of personalized healthcare interventions by tailoring treatment plans to individual risk profiles. This patient-centric approach maximizes the effectiveness of interventions, ensuring that patients receive the most appropriate care and guidance.

**Efficient Resource Allocation:** Healthcare systems can use predictive models to allocate resources more efficiently. By focusing resources on individuals at high risk, healthcare providers can reduce the burden on the system while optimizing patient outcomes.

**Cost Reduction:** Early diagnosis and intervention have the potential to significantly reduce long-term healthcare costs associated with diabetes complications, hospitalizations, and treatments. The economic benefits of prevention and management make machine learning-based prediction systems economically attractive.

**Continuous Monitoring and Improvement:** Machine learning models can be deployed in real-world healthcare settings and continuously monitored to ensure they remain accurate and up-to-date. Regular retraining and updates with new data enhance the models' predictive power and reliability over time.

**Privacy and Ethical Considerations:** The responsible use of sensitive medical data is of paramount importance. Adherence to privacy regulations and ethical guidelines, such as HIPAA, is crucial when handling patient information for predictive modeling.

In conclusion, diabetes prediction using machine learning represents a paradigm shift in healthcare, empowering healthcare professionals with tools to predict, prevent, and manage diabetes more effectively. As the field continues to advance, collaboration between data scientists, healthcare experts, and stakeholders is essential to create robust and ethical predictive models that can have a significant positive impact on public health. The combination of cutting-edge technology and medical expertise promises to drive improved patient outcomes and a more sustainable healthcare system in the fight against diabetes.