**AI BASED:DIABETES PREDICTION SYSTEM**

**Phase 5:Project documentation and submission**

1. **Problem Statement:**

Problem: The project aims to predict the likelihood of an individual developing diabetes based on various health-related features.

Significance: Early diabetes prediction can lead to timely intervention and improved patient outcomes.

**2.phases of development:**

**Phase 1:**

* We collected and explored a given dataset and explain about the project.
* We can discussed the project definition,planning,outcomes,report.

**Phase 2:**

* we can explore innovative techniques such as ensemble methods and deep learning architectures to improve the prediction system's accuracy and robustness.
* we'll explore innovative techniques and approaches to building our spam classifier.

**Phase 3:**

* you will begin building your project by loading and preprocessing the dataset.
* In this phase begin developing the diabetes prediction system by preparing the data and selecting relevant features.

**Phase 4:**

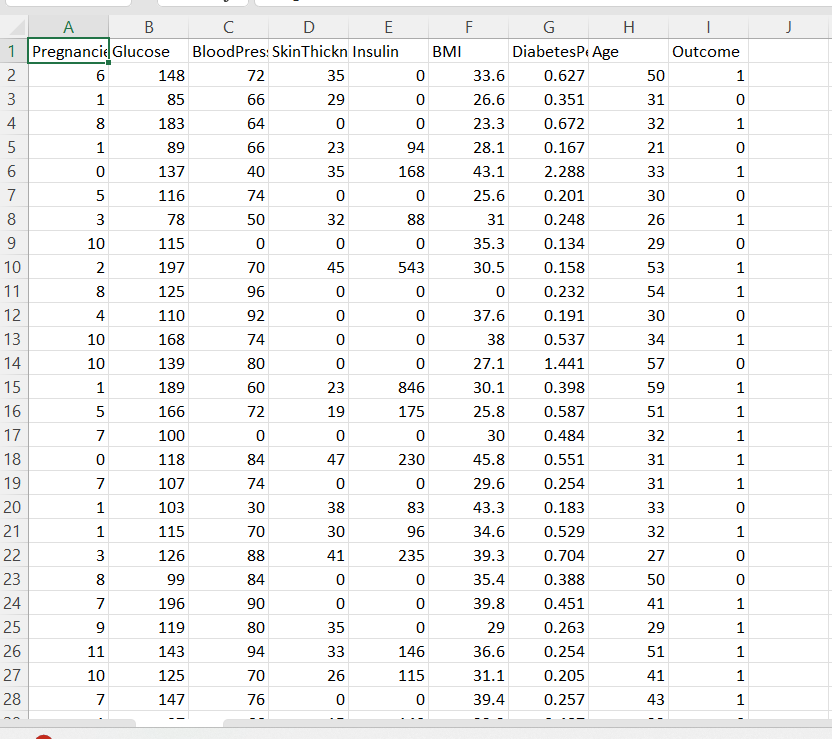
* you will continue building your project.

this phase, we'll continue building the diabetes prediction system by:

* Selecting a machine learning algorithm (SVM)
* Training the model
* Evaluating its performance.

**Dataset description:**

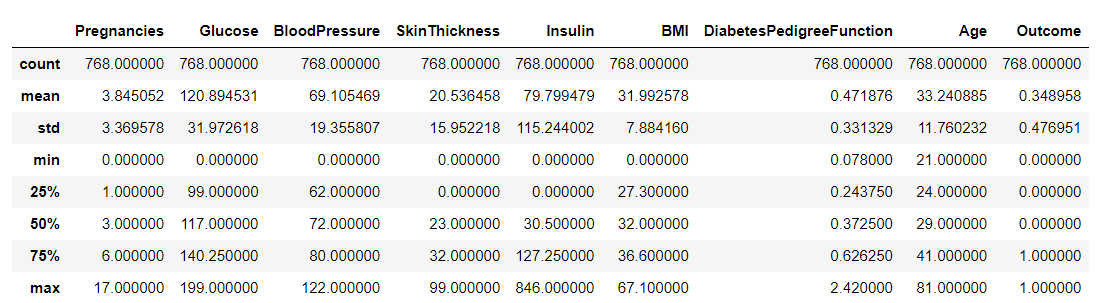
* Dataset: [Diabetes prediction and 9 kb].



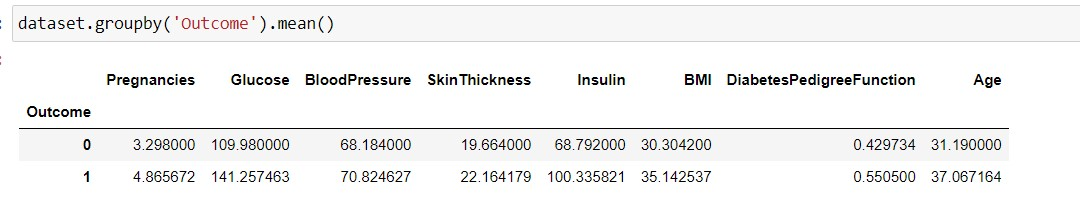
* Features: [List key features such as age, BMI, blood pressure, glucose,skin thickness,insulin,etc.]

**Data Preprocessing:**

* Handle Missing Values: Describe how missing data was handled (e.g., imputation or removal).

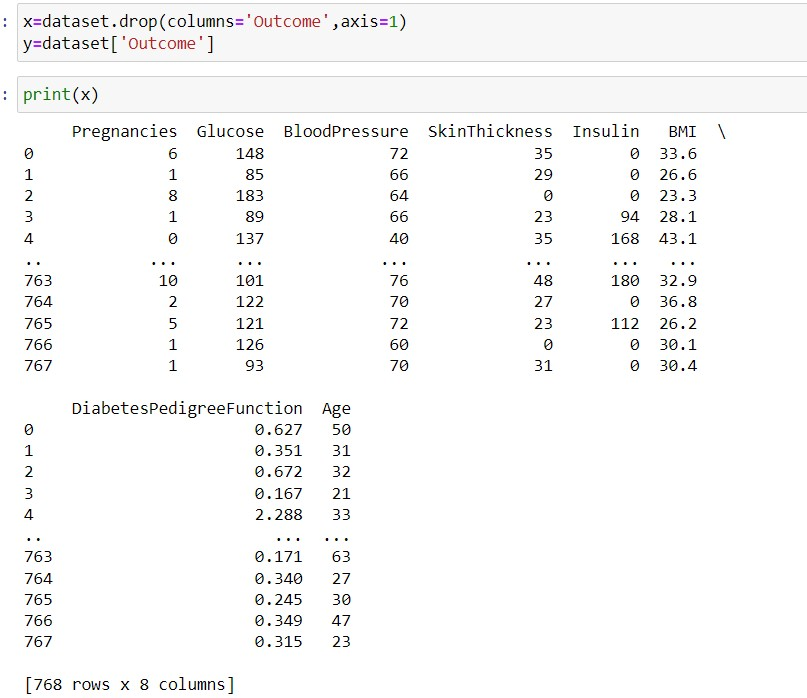


* Outlier Detection and Treatment: Explain the approach to identifying and addressing outliers.



* Data Scaling: Detail the normalization or standardization methods used.
* Encoding Categorical Variables: Explain how categorical data was encoded (e.g., one-hot encoding).

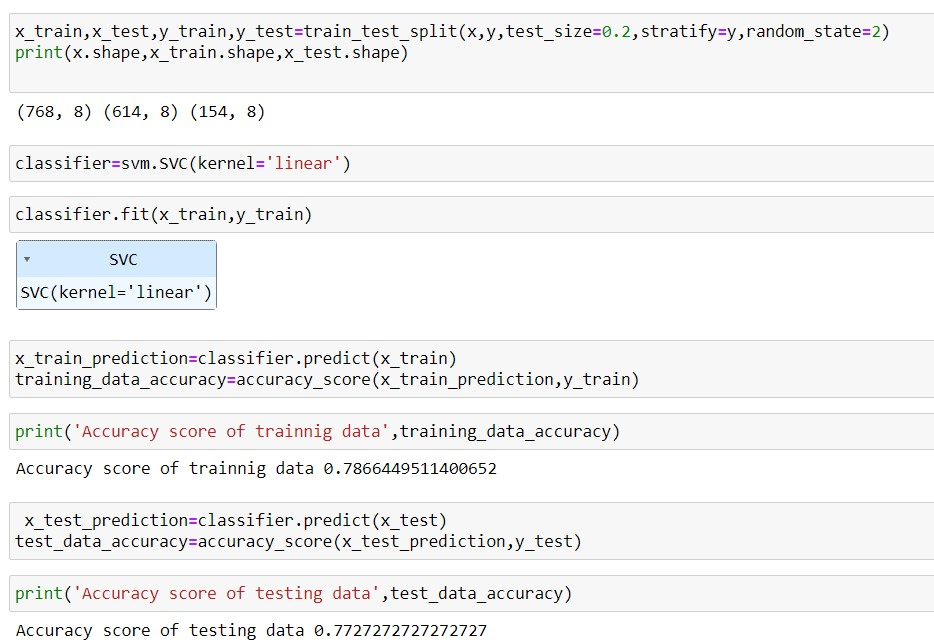
**Feature Selection:**

* Feature Engineering: Describe any engineered features or interactions.
* 
* Feature Selection Techniques: Explain the methods used to select the most relevant features, such as correlation analysis or feature importance ranking.



**Machine Learning Algorithm:**

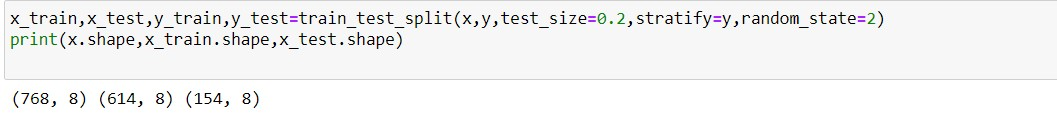
* Chosen Algorithm: State the machine learning algorithm(s) used (e.g., SVM).



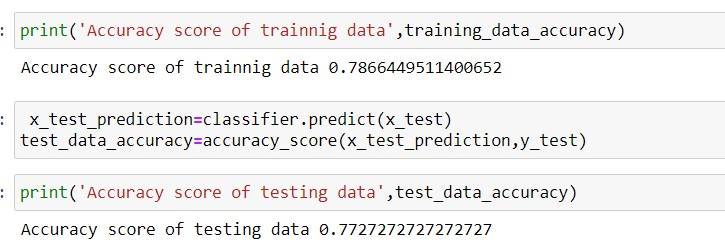
* Rationale: Explain why the chosen algorithm is suitable for this classification task.
* an appropriate SVM variant based on the nature of your problem:
* Linear SVM: Use when the data can be effectively separated by a linear boundary.
* Kernel SVM: Employ when the data requires a non-linear decision boundary.

**Model Training:**

* Select the SVM kernel (e.g., linear, polynomial, radial basis function) based on your data's characteristics.



* Split the dataset into training and testing sets.



* Train the SVM model on the training data using an optimization algorithm.

**Evaluation Metrics:**

* Metrics Used: Define evaluation metrics such as accuracy, precision, recall, F1-score, and ROC AUC.
* Rationale: Explain the choice of metrics and how they align with the project's goals.

**Project Documentation and reporting:**

* Separate from the report, create documentation for your diabetes prediction system, including instructions for usage, system architecture, and any code-related documentation.
* Remember to use clear and concise python language in your documentation and reporting.

**Final outcome:**

* We showed our final outcome for this project:

