**NAÏVE BAYES ALGORITHM.**

The Naïve Bayes classifier is a supervised machine learning algorithm, which is used for classification tasks, like text classification.

Naïve Bayes is a probabilistic machine learning algorithm based on the Bayes Theorem.

Naive Bayes is a classification algorithm for binary (two-class) and multi-class classification problems

EXAMPLE CODE.

This code uses the diabetes dtaset which has been attached to this file.

1. **Loading and Exploring Data.**

import pandas as pd

data = pd.read\_csv("diabetes.csv")

data.head(10)

data.dtypes

data.describe()

- Loads the diabetes dataset into a Pandas DataFrame and displays the first 10 rows, data types of columns, and descriptive statistics of the numerical columns.

2. **Train-Test Split.**

from sklearn.model\_selection import train\_test\_split

X = data.drop("Outcome", axis=1)

y = data[["Outcome"]]

X\_train, X\_test, y\_train, y\_test = train\_test\_split(X, y, test\_size=0.30, random\_state=1)

- Imports the `train\_test\_split` function and splits the data into training and testing sets.

3. **Gaussian Naive Bayes Model.**

from sklearn.naive\_bayes import GaussianNB

model = GaussianNB()

- Imports the Gaussian Naive Bayes model from scikit-learn and initializes a Gaussian classifier.

4. **Flatten the Target Variable.**

y\_train\_flattened = y\_train.values.ravel()

- Flattens the `y\_train` DataFrame to a 1D array. This is necessary for training the model.

5. **Training the Model.**

model.fit(X\_train, y\_train\_flattened)

- Trains the Gaussian Naive Bayes model using the training sets.

6. **Model Prediction and Evaluation.**

y\_pred = model.predict(X\_test)

from sklearn import metrics

print("accuracy:", metrics.accuracy\_score(y\_test, y\_pred))

test\_pred = model.predict(X\_test)

print(metrics.classification\_report(y\_test, test\_pred))

print(metrics.confusion\_matrix(y\_test, test\_pred))

- Predicts the response for the test datasets.

- Calculates and prints the accuracy score of the model.

- Prints the classification report, which includes precision, recall, and F1-score.

- Prints the confusion matrix, a table showing the performance of the classification model.

This code essentially loads a diabetes dataset, splits it into training and testing sets, trains a Gaussian Naive Bayes model, and evaluates its performance on the test set using accuracy, a classification report, and a confusion matrix.