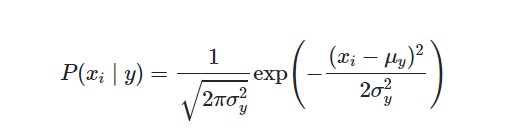
|  |  |
| --- | --- |
| Algorithm: Gaussian Naive Bayes | |
| USN: 1MS18CS024 | NAME: Aravind S |
| USN: 1MS18CS029 | NAME: Arun J Kennedy |

**Description of the Algorithm: <<Write 2-3 Paragraphs about the Algorithm>>**

When dealing with continuous data, an assumption often taken is that the continuous values associated with each class are distributed according to a normal (or Gaussian) distribution. The likelihood of the features is assumed by the formula:



Sometimes assume variance

* is independent of Y (i.e., σi),
* or independent of Xi (i.e., σk)
* or both (i.e., σ)

An approach to create a simple model is to assume that the data is described by a Gaussian distribution with no co-variance between dimensions.

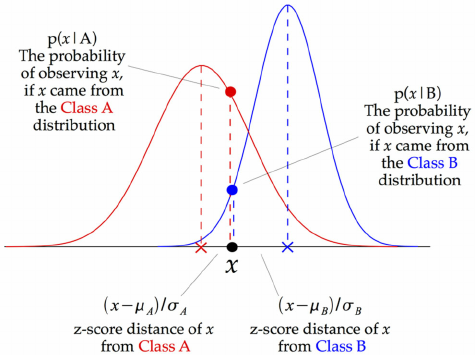


Figure 1: Gaussian Naïve Bayes

The above illustration indicates how a Gaussian Naive Bayes classifier works. At every data point, the z-score distance between that point and each class-mean is calculated, namely the distance from the class mean divided by the standard deviation of that class.

**Algorithm Pseudocode:**

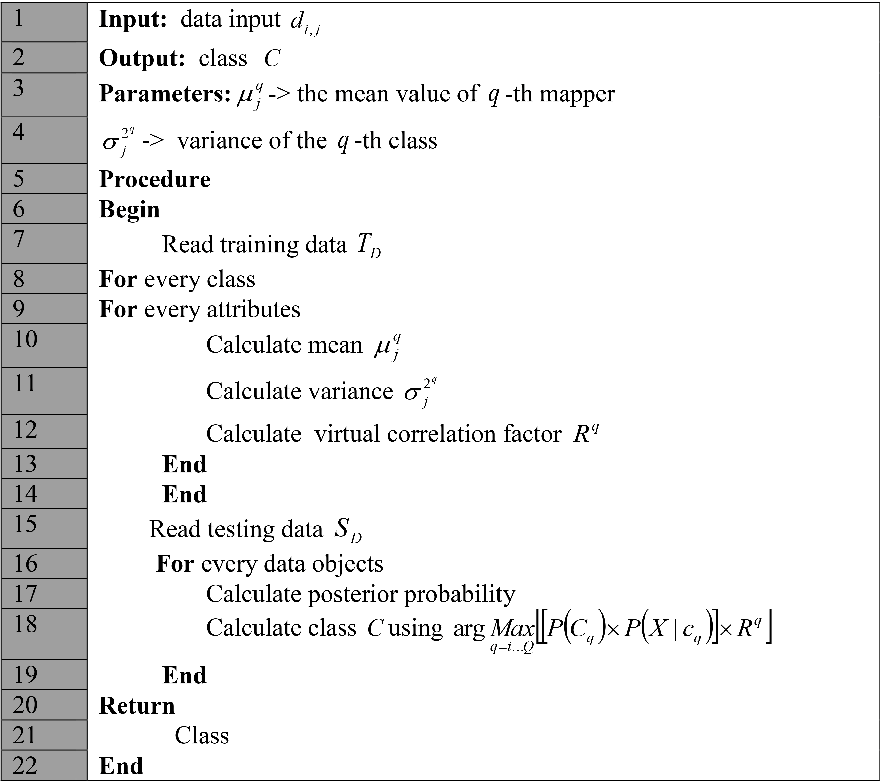
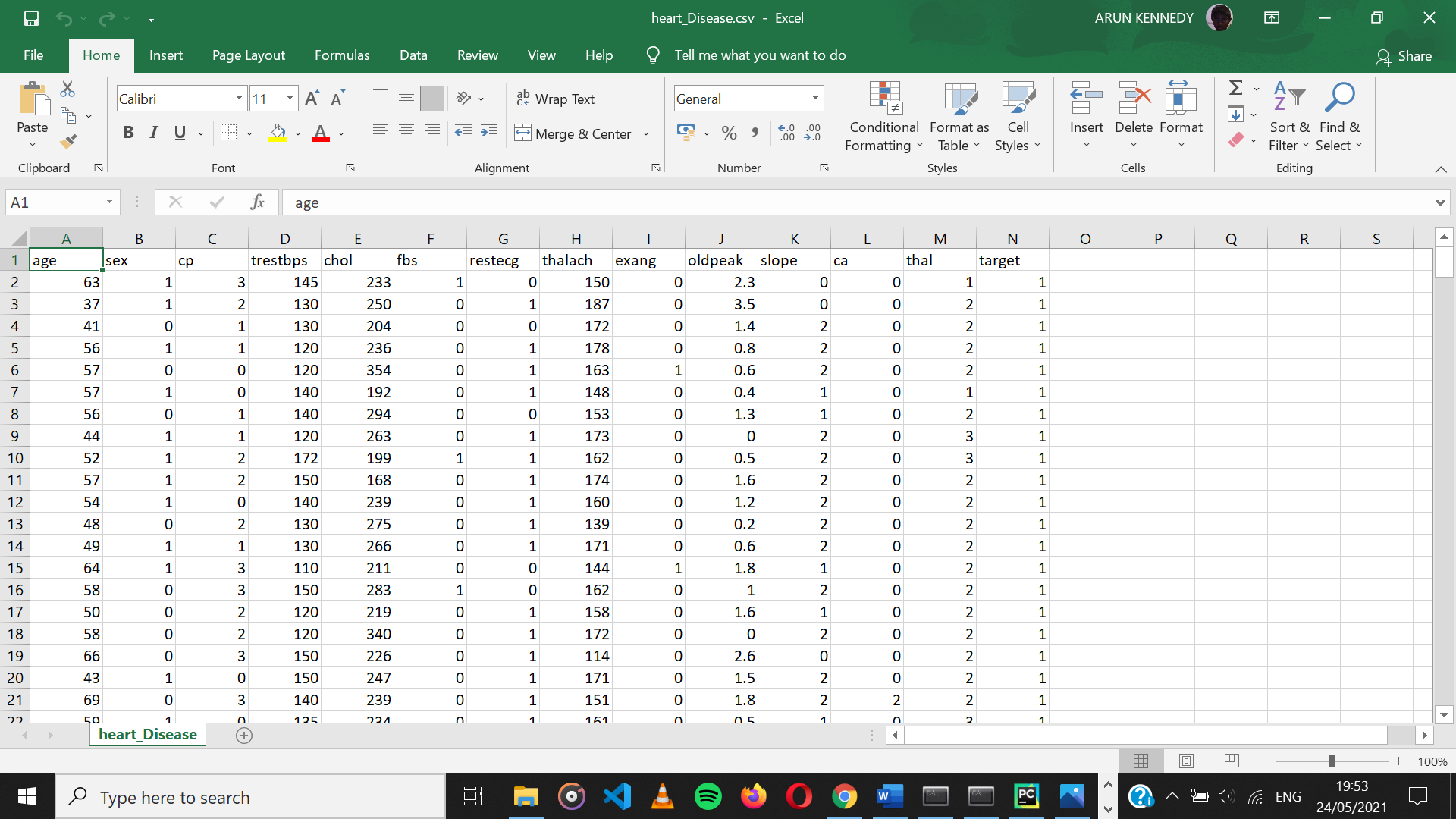


Figure 2: Gaussian Naïve Bayes Pseudocode

**Data set Used: (Attach Screen shot of the few rows and also the Kaggle/Dataset link)**



Dataset link: <https://www.kaggle.com/ronitf/heart-disease-uci>

**Challenges faced during the implementation of the program:**

Python’s sklearn has the library for Gaussian Naïve Bayes model so no particular challenges were faced.

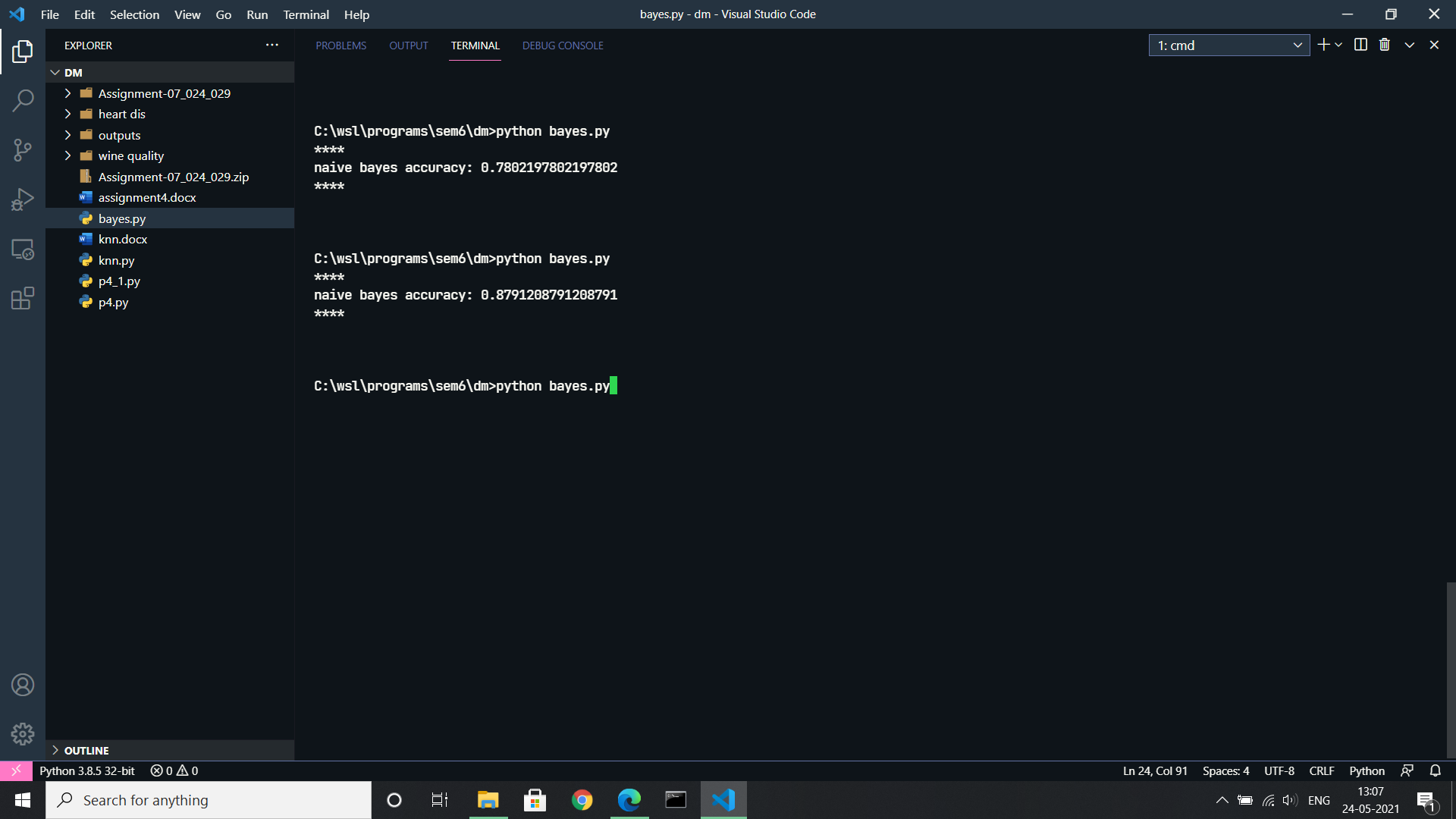
**Advantages & Disadvantages of the Algorithm:**

|  |  |
| --- | --- |
| Advantages | Disadvantages |
| Fast and flexible model gives highly reliable  results. | Large data records are required to achieve a good result. |
| Works well with large data. | Shows lower performance than the other classifiers according to the type of problem. |
| There is no need to spend much time for training. | It assumes that all the features are independent. While it might sound great in theory, in real life, you’ll hardly find a set of independent features. |
| Provides better grading performance by eliminating insignificant specifications. |  |

**Applications of the Algorithm:**

* Text classification can be performed using Gaussian Naïve Bayes model.
* Naive Bayes classifier can be applied on TF-IDF Vectorized Matrix to perform better information retrieval.
* It can be used to predict whether a person will buy a particular product based on his age, income and other parameters.
* Email services (like Gmail) use this algorithm to figure out whether an email is a spam or not. This algorithm is excellent for spam filtering.
* Collaborative Filtering and the Naive Bayes algorithm work together to build recommendation systems. These systems use data mining and machine learning to predict if the user would like a particular resource or not.

**Output: (Screen shots)**



**References:**

<https://www.semanticscholar.org/paper/PGNBC%3A-Pearson-Gaussian-Na%C3%AFve-Bayes-classifier-for-Babu-Ramadevi/4b08a70db00edadbca433b78575cc721cb591ab5>

<https://www.researchgate.net/publication/338950098_Comparison_of_Multi-class_Classification_Algorithms_on_Early_Diagnosis_of_Heart_Diseases/figures?lo=1>

<https://www.saedsayad.com/naive_bayesian.htm>