**Report**

* **Read me.**  
  • I have created two files for this project one is Naïve\_Bayes\_Algo(2).ipynb and other is KNN\_Algorithm(2). ipynb and three text file.  
  • In order to implement new dataset, we need to create a text file with same format and run code. Code will ask user to input the file name, along with type, that is example: ’1\_2.txt’ .
* **QUESTION 1**  
  • KNN\_Algorithm(2) implements the KNN. First we implement it for smaller data set and then for bigger that is for 120 rows.  
  • First I have implemented all the metrices for all given value of k.  
  • Then I ask for user to give the name of training data of file, then test data of file. And ask for which metrics they want to implement for.  
  • Latter part of code, implements leave one out  
  • Last part of code, I drop the age column and find the accuracy and compare with each other .
* **QUESTION 2**  
  • Naïve\_Bayes\_Algo(2).ipynb implements the naïve gaussian. First we implement it for smaller data set and then for bigger that is for 120 rows.  
  • First I have implemented all the metrices for all given value of k.  
  • Latter part of code, implements leave one out  
  • Last part of code, I drop the age column and find the accuracy and compare with each other .

**Question – Answers.**

* Using Cartesian distance, Manhattan distance and Minkowski distance of order 3 as the similarity measurements show the results of the gender prediction for the Evaluation data that is listed below generated training data for values of K of 1, 3, and 7 .

Text

Description automatically generated

Text

Description automatically generated with medium confidence

Text

Description automatically generated with medium confidence

* for values for K of 1, 3, 5, 7, 9, and 11 and report the results. For which value of K do you get the best performance?

Text, letter

Description automatically generated

For K=1 the accuracy is 62.5 , For K=3 the accuracy is 58.33 , For K=5 the accuracy is 58.33

For K=7 , the accuracy is 60.83 , For K=9 the accuracy is 59.1666 , For K=11 the accuracy is 55.00 .

* So we can state that the K=1 gives the best performance in terms accuracy and best label prediction.

After dropping the age column the Accuracy of the Prediction will get down . It will be near 58.33% . And the Predicted output will be ['W', 'M', 'W', 'W'] . Which is different while predicting with the age column.

Thank you .

Shashwat Chauhan