

**CS4092D Machine Learning Lab**  
**Module2 (NBC): Exercise**  
**Naïve Bayes Classifier**  
**S4 MCA: Winter 2022-23**

**Date: 25-Jan-2023**

Q. Implement the **Naïve Bayes Classifier(NBC)** from scratch using Python. You must use only the basic libraries in Python to implement this exercise. Use the Dermatology dataset shared in *Module1* to train and test your classifier models (the dependent variable is “Class code”). For the same, **divide your dataset into train and test sets**: you may **randomly** take 70% of the datapoints(rows) in the dataframe as the train-dataset and the remaining 30% of the datapoints goes to the test-dataset. Compare your **model’s accuracy** on these data sets with respect to Python’s Naïve Bayes Classifier function: you may use the GaussianNB() function defined in the sklearn.naive\_bayes library as the reference function for comparing your classifier’s accuracy.

You may refer to Sections 9, 12, 13 and 14 in the following [link](#) for learning to build the reference model using GaussianNB(), and to understand the steps used to split the data-set into train/test and finally the steps used to calculate the model’s accuracy.

**NB:** In your implementation of NBC, **you should not use** any built-in functions that define the Naïve Bayes Classifier(eg: GaussianNB()). However, **you may use** any functions defined in the basic Python libraries, viz., pandas, numpy, re, matplotlib, math, etc.