Answers

6. a) The value of Accuracy is different than the one given for NB in step 5.4 is because of the stochastic nature of the algorithm or evaluation procedure, or differences in numerical precision and also because we are using 2-fold cross validation and 10-fold cross validation is used in 5.4

6. b) The confusion matrix is different because in 6.2, SVM model is used and we are using GaussianNB model. Also, in 6.2, the dataset is split in 80% as training dataset & 20% as test dataset and we are using 50% as training dataset & 50% as test dataset.

6. c) P, R, and F1 scores different b different because in 6.2, SVM model is used and we are using GaussianNB model. Also, in 6.2, the dataset is split in 80% as training dataset & 20% as test dataset and we are using 50% as training dataset & 50% as test dataset.

6. d) Samples in training dataset = 75

6. e) Samples in test dataset = 75