**Assignment 6:**

2) Compare the performance with the classifier in (2) and write your conclusion.:

* For Classifier 1 the best F-Score value is 0.76118742 for the 6th fold which is Naïve Bayes Model without grid search.
* And for Classifier 2 which uses Grid Search method the best f1 score: 0.732646048658.
* So, by comparing the results, Classifier 1 performed better than Classifier 2.

3) How many samples are enough? Show the impact of sample size on classifier performance?

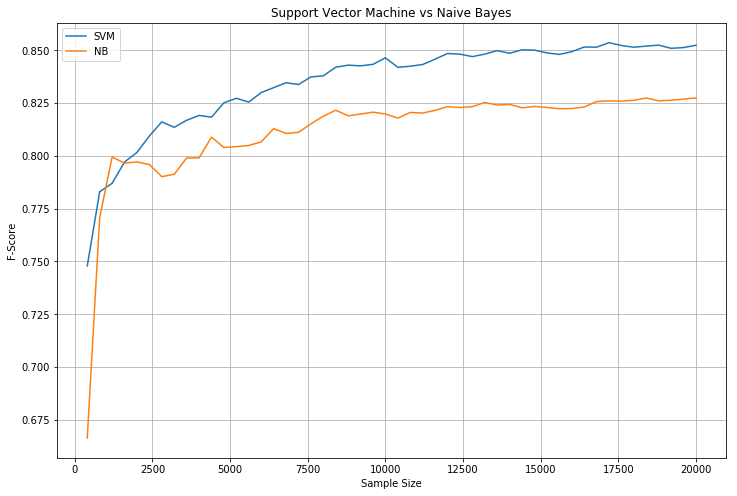
* PFA herewith the diagrams.
* A close up of a screen

  Description generated with high confidence

A picture containing shoji, indoor

Description generated with high confidence

Draw a line chart show the relationship between sample size and average macro f1 score.



Write your analysis on the following:

How sample size affects each classifier’s performance?

* For both the models, f-score increases as the sample size increases.
* For Naïve Bayes, the f-score value increases up to sample size 8000, after that the score value remains fairly constant.
* For SVM model, the f-score value increases up to sample size 12500, after that the score value remains fairly constant.

How many samples do you think would be needed for each model for good?

* Sample size is different for both the models.
* For Support Vector Machine model, the size 12500, would be ideal.
* Because there is not much change in the variation of F-score after that value.
* For Naïve Bayes Model, the sample size between 7500-10000 would be ideal.

performance?

How is performance of SVM classifier compared with Naïve Bayes classifier?

* SVM model performed better than Naïve Bayes model.
* The highest f-score value was 0.85236054358959468 for SVM classifier.
* The highest f-score value was 0.8274205653398834 for Naïve Bayes classifier.