**GROUP6 LAB WEEK 5: NAÏVE BAYES ALGORITHMS**

Table

Description automatically generated

* In the table, highlight with blue are the best performing algorithms in each category.
* Gaussian Naïve seems to be preforming better than other algorithms in most categories
* When we are interested in correctly predicting PE values lower than 34.5 (i.e., True Positive), Gaussian algorithm offers a better chance of doing that as it has high TP and specificity measures
* When the aim is to prioritise correctly predicting PE higher than 34.5 (i.e., True negative), Complement naïve bayes is the ideal bayes algorithm to use since it has a good combination of high accuracy, TN and sensitivity measures.
* Even though Gaussian and Complement have the joint highest accuracy scores, Gaussian outperforms Complement Naïve Bayes at AUC. AUC is **scale-invariant**. It measures how well predictions are ranked, rather than their absolute values. Meaning that Gaussian ranks predictions better than all other algorithms.
* Categorical Naïve bayes runs perfectly fine when we fit the algorithm using a pipeline. However, when we try to run “predict(X\_test)”, the algorithm gives an index out of bounds error. Upon researching we noticed that this is a problem that hasn’t been fixed yet in the SkLearn community. It requires very careful train, test split to ensure that all classes appear on both test and train data. This is because the error is a result of the test data containing classes that have not been trained.