# SGN-41007: Task01 Competition Report Group 13

## Feature extraction

We tried 3 different kinds of feature extraction.

* The first one just reshapes the Vector.
* Second one takes the mean of all samples
* Third one uses the mean and deviation of the distribution.

We see different results on different classifiers for these types of feature extraction used.

## Values from our Testbench

We separated the Available data simply on a 80-20 ratio using train\_test\_split. We used the 20% to get an estimate of our accuracy by using accuracy\_score function. We see that the accuracies in our testset are about 0.2 higher than what we get in the Kaggle leaderboard. This is propably due to our simple split in the testdata without taking the way the data was split in this course into account.

## Submission

For the submission we used a Extremely Randomized RandomForestClassifier with 1000 trees and the Mean and Deviation Feature Extraction. This showed the best results in our testbench and also in Kaggle. We saw that increasing the number of trees gave us better results in the testbench and Kaggle. In our Testbench we could reach a accuracy score of 0.93 while in Kaggle we could reach 0.716.

## Attachments

1. Python Code used
2. Submission file