# PROJECT-2 CONCLUSION

Upon applying PCA, a maximum accuracy of 97% for 7 components was achieved, which was made possible by tweaking the parameters of MLP Classifier and iterating through the number of components. This means that there are at least 97% chances that the submarine will survive a mine field. The calculated Confusion matrix lets us know that a rock was wrongly identified as a mine and a mine was wrongly identified as a rock. Hence, out of the 30% test set i.e., 63 samples, 61 were correctly identified and 2 were wrongly identified. A rock being wrongly identified as a mine won't actually create highly undesirable circumstances. The submarine may get into alert mode or may change its direction but eventually it won't end up getting destroyed. The other case where mine was wrongly identified as rock will lead to explosions which is definitely undesirable. Hence, if we consider the case of a rock being wrongly identified as a mine to be not so harmful then the accuracy is definitely even better! The highest accuracies are observed when the number of components is ranging from 5 to 15. This is because as the number of components increases, the model becomes more complex due to increased dimensions (overfitting) and the test data doesn't give great results. With more components considered, components that are noise or have hardly any correlation with the class also get included. Hence, we could achieve a great value of accuracy with just 7 components. Therefore, the plot of accuracy vs number of components shows a peak initially when the components are below 15 and then remains in the same bracket of accuracy between 75% to 90%. Because of some overall randomness in parameters, we can observe peaks and drops at various points. The best combination of parameters was found by tweaking all parameter values. The parameter 'solver' was kept 'adam' because 'lbfgs' works better for small datasets and 'sgd' depends a lot on learning rate. The parameter 'activation' was kept as 'tanh' because 'logistic' could provide the highest value of 95% and 'relu' and 'identity' were giving values between 85% to 95% even after tweaking the parameters as much as possible. When the 'random_state' value is set to 1, highest accuracy was achieved for 7 components. For other values of 'random_state' or while keeping it different for each run, the highest accuracy observed was below 97% but not significantly less. If we don't set 'random_state' value, it will give us different results each time because a different random sequence will be generated each time. For repeatability purposes, it's better to specify 'random_state'. The parameter 'hidden_layer_sizes' gave comparatively less accuracy if the value was too high (>100) or too low (<20) and 'max_iter' gave less accuracy if its value started decreasing and went below 1000.