**Date**: 23rd March, 2022

**Abraca-data**

**Weekly Report**

**Tasks performed in this week**

- Obtained results after applying logistic regression and KNN on the dataset after resampling the data and feature selection.

- Results were obtained in the form of confusion matrix and performance measures being accuracy, recall, f1-score, and AUC.

- Plotted Receiver operating characteristic (ROC) curve obtaining Area under the curve (AUC) which helps us determine the performance of logistic regression and KNN on test data.

----------------------------------------------------------------------------------------------------------------

**Outcomes of the tasks performed**

- The accuracy, recall, F1 score as well as AUC score of KNN is better than logistic regression, and hence it is a better model.

----------------------------------------------------------------------------------------------------------------

**Tasks to be performed in the upcoming week**

-Modification in the step of Resampling: We will code the ‘Synthetic Minority Overfitting Technique’ for resampling. For which earlier we had done simple oversampling by replicating the data in the minority class

-Study various methods for dimensionality reduction of the dataset (such as ANOVA score function or LDA or PCA). Select and apply a suitable method to the dataset.