- DataSet is Unbalanced with Negative cases having 88 percent of data and Positive cases having 12 percent of total data
- First we tried with logical regression with the Total data but ROC_AUC score is 0.5 which means the prediction is very bad
- Then with dataset Random Forest Classifier was used and the result was same
- The next approach taken was Oversampling or Undersampling
- In Oversampling the data with 1 as response is repeated and size is made equivalent to 0 cases. In this ROC_AUC score has improved with Random Forest

- We are able to almost around 0.8 and the Recall of 1 too has improved but Precision for that has reduced.
- Next step was with AdaClassifier and XGBClassifier with the Oversampled data
- Next step involved Voting and stacking of the three model outputs . Based on all the combinations Voting provided better results
- Grid search was done for all three methods to get the best estimator model and In voting and stacking the best models were used
- In Stacking LogisticRegression was used to come to conclusion on results

- Next on the dataset we are using undersampling method . The reason to first go for oversampling was we would be losing data on prediction of 0 response

- With the undersampling method the predicted 0 Response data values size is reduced to same as predicted 1 response values

- All the three RF ,XGBClassifier and AdaClassifier are used to get results.

- Same Grid Search is used on all three algorithm to get best models

- Voting Stacking was used on undersampling method

- Screenshots and files are attached in the Github

- Using Google Cloud Computing and streamlit model was deployed and link has been shared

- In the video explanation too the video has beens showed