report

## Dividing the dataset using TEST, TRAIN and split:

I have divided the Dataset by “prefered\_device” column i.e. (Mobile and Laptop) as asked in the requirements.

Later, Divide the Mobile and Laptop datasets in test and train dataset in the ratio of 80:20 for each data dataset. And also divided the datasets into 70:30 ratio as asked in requirements.

People prefers Mobile to visit the airline’s website than Laptop.

Since, Laptop dataset has less data, its train dataset has less data to train the model so the accuracy of the laptop dataset is low as compared to Mobile dataset.

## list of models implemented:

I have implemented total 11 models to train the Mobile and Laptop dataset. In the requirement you asked Linear Discriminant Analysis, Logistic Regression, KNN Classifier, Naïve Bayes Classifier, AdaBoost, Bagging Classifier, Gradient Boost. I have implemented all 7 models with additional hybrid model Gradient Boost with ADA Boost Classifier.

1. Decision Tree
2. Random Forest Classifier
3. XGBoost
4. ADA Boost
5. Gradient Boost Classifier
6. Gradient Boost with ADA Boost Classifier
7. Bagging Classifier
8. Linear Discriminant Analysis
9. KNN Classifier
10. Naïve Bayes Classifier
11. Logistic Regression

## Tuning of models:

I have implemented GridSearch for Hyperparamenter Tunning of the models are:

1. Decision Tree
2. XGBoost
3. Bagging Classifier

## Feature Selection:

I have selected the features according to the correlation matrix of Mobile dataset for Building Mobile Model and correlation matrix of Laptop dataset for Building Laptop Model.

I have selected the preferred\_location\_type, Yearly\_avg\_comment\_on\_travel\_page, week\_since\_last\_outstation\_checkin, following\_company\_page columns to train the Mobile model.

I have selected the preferred\_location\_type, Yearly\_avg\_comment\_on\_travel\_page, week\_since\_last\_outstation\_checkin, following\_company\_page columns to train the Laptop model.

As their correlation w.r.t to target column i.e. Taken product is high.

## Accuracy of models:

For Mobile Model (80:20) :

1. Random Forest Classifier having test accuracy of 91.64% and training accuracy is 99.42% which means model is slightly overfitting.

When the training accuracy is way more than testing accuracy then the model is said to be overfitting.

1. Bagging Classifier having test accuracy of 90.33% and training accuracy is 98.45% which means model is slightly overfitting.
2. Decision Tree having test accuracy of 87.98% and training accuracy is 99.42% which means model is slightly overfitting.

For Mobile Model (70:30) :

1. Random Forest Classifier having test accuracy of 90.08% and training accuracy is 99.54% which means model is slightly overfitting.
2. Decision Tree having test accuracy of 87.35% and training accuracy is 99.54% which means model is slightly overfitting.
3. Bagging Classifier having test accuracy of 89.20% and training accuracy is 98.71% which means model is slightly overfitting.

*In Conclusion, we can see that the 80:20 distribution model worked better than 70:30 distribution models.*

For Laptop Model (80:20) :

1. Logistic Regression having test accuracy of 83.24% and training accuracy is 83.93%.
2. Linear Discriminant Analysis having test accuracy of 81.41% and training accuracy is 82.65%..
3. Naïve bayes classifier having test accuracy of 81.32% and training accuracy is 82.72%..

For Mobile Model (70:30) :

1. Linear Discriminant Analysis having test accuracy of 82.57% and training accuracy is 84.07%.
2. Logistic Regression having test accuracy of 82.47% and training accuracy is 84.72%.
3. Naïve bayes classifier having test accuracy of 80.88% and training accuracy is 82.53%.

*In Conclusion, we can see that the 80:20 distribution and 70:30 distribution models have non-significant difference in accuracy.*

## Conclusion

*In the end, Mobile Models perform better than laptop models because laptop dataset has very less data for training good models.*