Accusaga Assignment

Name: Sumit Santra

Ph No: 7276592088

So the Problem statement was to do a predictive modeling which can predict the persons who will be Buying a ABC Product from a Store and the Target variable is a Categorical Variable in which there are two Classes 0 – No and 1 – yes We need to Predict the Classes based on the Independent Columns in the Dataset.

1. Firstly I Started the Project By importing all the python libraries which are used.
2. After Importing all the Libraries I imported the data and started understand the dataset.
3. While Understanding the dataset I came to know that there are 7 columns in which our last column is out target column and rest of them are independent columns.
4. After understanding the dataset I started looking for null values in the dataset and there where no null values.
5. As Customer ID column was not giving much information about the dataset so I dropped the column and use other columns for model building
6. I visualize the numerical column to know if there are any outliers in the data set and found out some of them and removed them using Inter Qunatile Range Technique using Python.
7. After this i looked at the skewness of the dataset and found that the data is right skewed and fixed them using Square root and Cube root.
8. Then I Converted all the categorical Variable to numeric variable so that our model can understand it I done it using Label Encoder.
9. After Doing this step I did Splitting of the dataset into two parts Training and testing set so that we can train our model on training set and test our model on test set.
10. I used Classification algorithms like Logistic regression, Decision Tree Classifier, Random Forest, Adaboost, SVC, XGboost for this Dataset and got Adaboost as the best model by looking at the metrics such as accuracy score, Classification Report , Confusion Matrix and AUC ROC Curve.
11. As GridSearchCV was taking to much time so I didn’t performed Hyperparameter Tuning otherwise our model would have perform better.
12. I’ve created a final dataframe and added some styling to get the probability of person those are going to buy the ABC Product.