**Step by Step Approach to solution & Methodology**

- Following data is about an ABC product that is concerned about its sales. Predicting its type of customer and finding out the pattern of the customers gives more reach to same type of customers. Using given data, we can do prescriptive analysis for finding out behaviour of sales of ABC product. Some of factors that’s shows us customers are from different groups or from different category.

- Very first step is that I need to check null values, as there are no NAN values are available. Continuing with exploratory data analysis to know about each feature and relation between them. Further I check with the outliers, outliers may affect on final result or it ay increase error. So, to avoid error I drop outliers using IQE method. After done with outliers I convert categorical features into numerical variables using label encoding.

- Finally, data have all numerical variable however some of features like Var1, Var2 having different units as compared to other features. For make all units same, scale the data using standardization. Standardization makes all features having mean of 0 and variance is 1.

- After, standardization to train the model I split the original data into train and test format. So that I will train the model on training data and test on test data. While splitting the data I used hold out method as data is not overfitting.

- Finally, I used lazy predicter to find out accuracy of all classifiers. Using lazy predict I get top 4 models that has maximum highest score amongst others.

- From 4 models I build all 4 model and see the difference between them. LGBM model gives better count of true positive and true negative values as compared to XGB model.

- I also used some other performance matrices like precision and recall. Precision shows me exactness of model and recall shows me completeness of model.

- To increase accuracy and avoiding overfitting hyperparameter tuning has done and achieve perfect parameter to model.

**Inferences**

After all the data pre-processing steps, many classification models have been tried and tested after splitting the data into train and test.

Here, we chosen LGBM classifier is final model as we are getting high accuracy, precision and recall.

One of the advantages of using LGBM is handled large size of data in less time and also It focuses on increasing accuracy.

In the final prediction we check confusion matrix as performance matrix and we get true positive rate is 2795 and true negative is 2172. Based on this result company can reach out to same customers that are interested to buy ABC product.