Customer Retention

The first step I have taken to find null values and find out that there are no null values have been in the dataset.

Again, I have checked by using the .info method and found out that it has no-null values.

Then I used the .describe method and find out that the mean and std difference is correct. The other difference like min, 25%, 50%, 75%, and max ok.

Now, we have found out that most of the data is in string and the model can only be run in the numeric numbers. So now we are going to convert the string data into numeric data and proceed with our model building.

So, firstly I have checked what type of data our columns hold. If it is categorical or regression only, we can proceed with the model. If it is categorical then we can easily convert the data and proceed further. But if it is regression, we have to do something else to proceed further.

So we checked and find out that our data is categorical and we can easily convert them into numeric form and used them. Then we used the Label encoder and convert our all data into numeric form.

After that, we checked the correlation between our columns and find out that most of the columns are 60 to 90% the same as the others columns. So we drop those columns which have this type of accuracy.

After using the drop method we have found out that our columns are now just 56 previously it is 71 columns and now this.

We can now build our model and proceed with the train test split method and proceed further.