Job –Hackthon

1) For predicting the customer is lead or not I used following approach.

2) First analyse the given data from this I came to know that Customer Product column is having missing values, and this column is very high importance. So to fill this nan values I used following approach

3) First I used label encoding technique to convert strings into numbers then from this column I created another supplementary column which takes Boolean masking of the values present in the data.

4) By this feature engineering technique model learn the Customer Product data very neatly

5) As problem is of classification so I used LightGbm classifier algorithm for predicting the probabilities of is lead or not columns

6) Finally model performance is evaluated by Area under the curve method.