Customer Segmentation Classification

BY: Maram Alfaifi, Tahani Almutairi

Date: Dec 19, 2021

CONTENT

Introduction

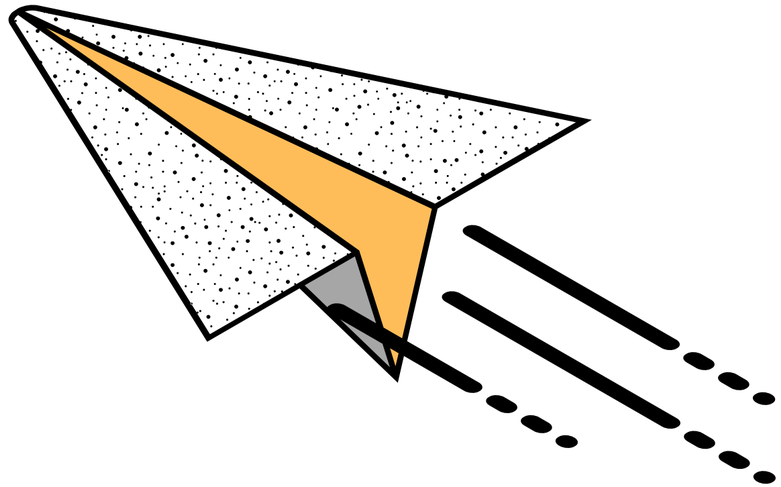
Data description

EDA

Model

Best Model

Conclusion



Introduction

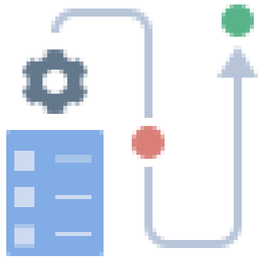
An automobile company is planning to enter new markets. The sales team categorized all customers into 4 segments, conducted segmented awareness and communication for different customer segment.

Business need

They plan to use the same strategy for new markets and have identified 2,627 new potential customers.
We have to help the manager to predict the right group of new clients.

Methodology

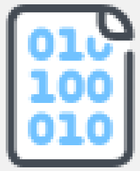
machine learning classification algorithms



Data set

DATASET

- 8068 rows
- 11 columns



NUMERICAL FEATURES

2

CATEGORICAL FEATURES

6

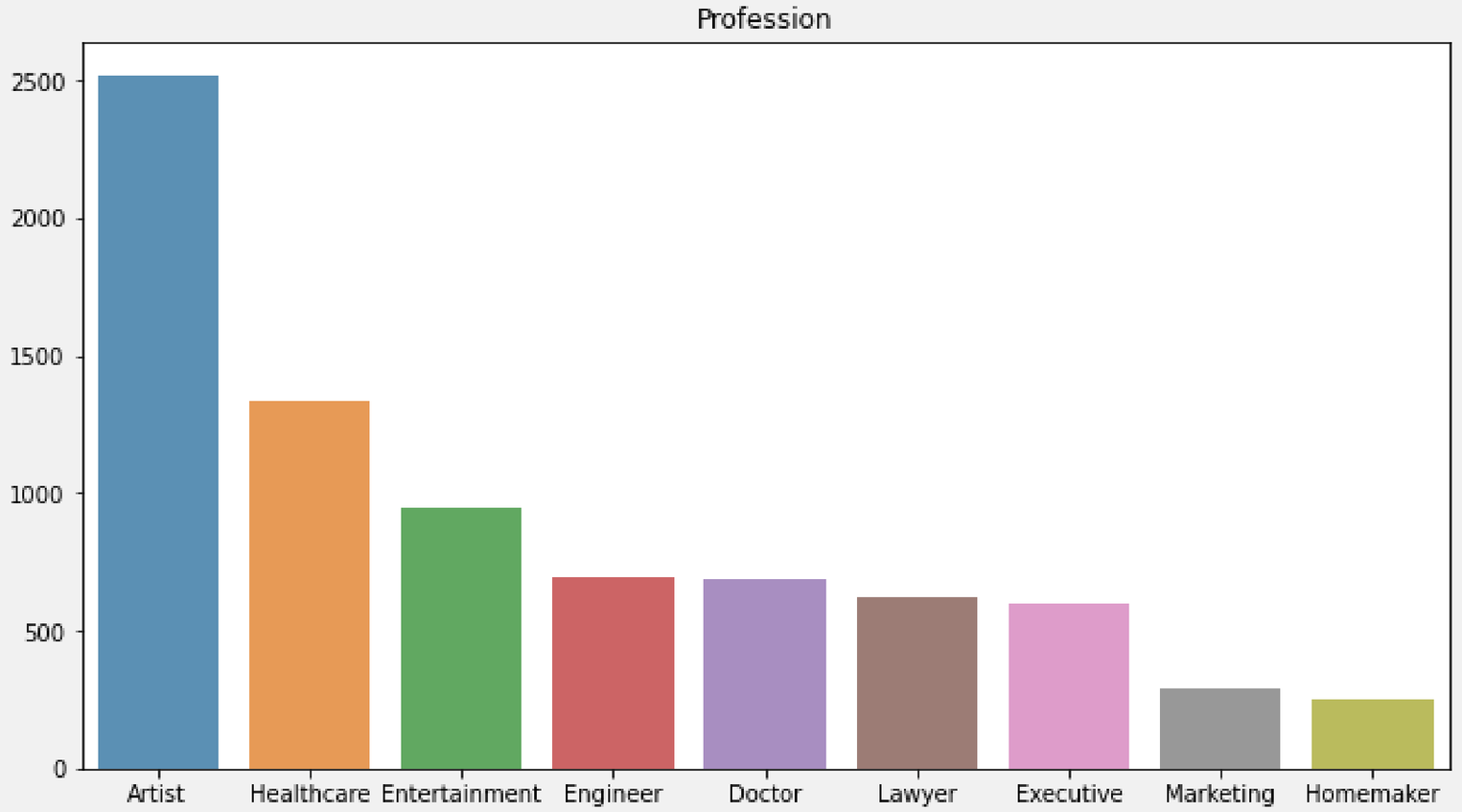
EDA

Step 1 Cleaning data

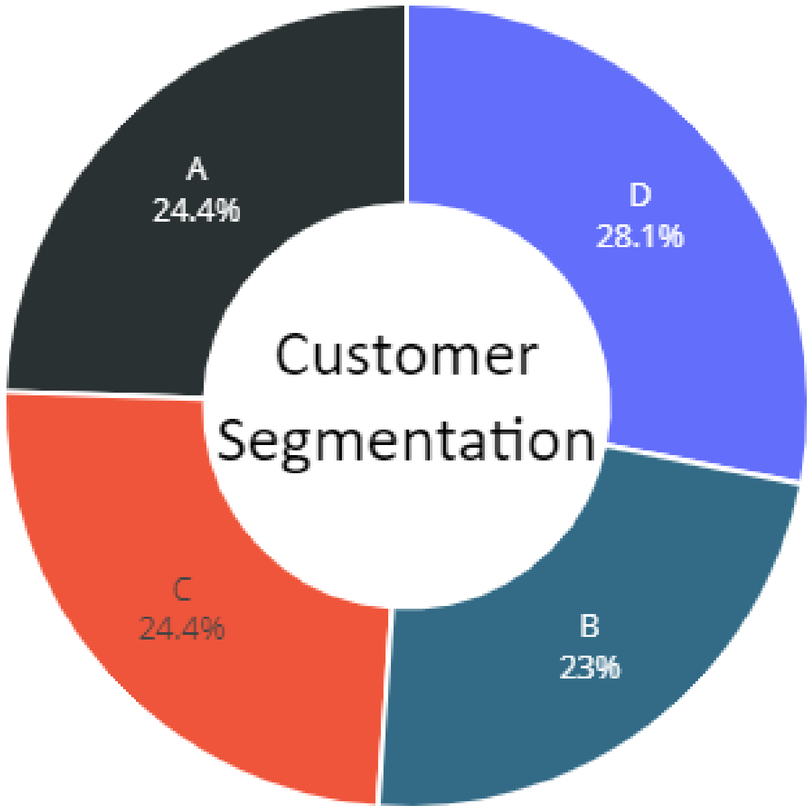
Step 2 Applying some feature improvements :

- Feature Seelction :
  - select columns
  - drop unneceerary columns (e.g.ID,Work_Experience).
- Feature Engineering :
  -Encoding categorical features (6features).

What are the professions of customers?



What are the Customer Segmentation ?



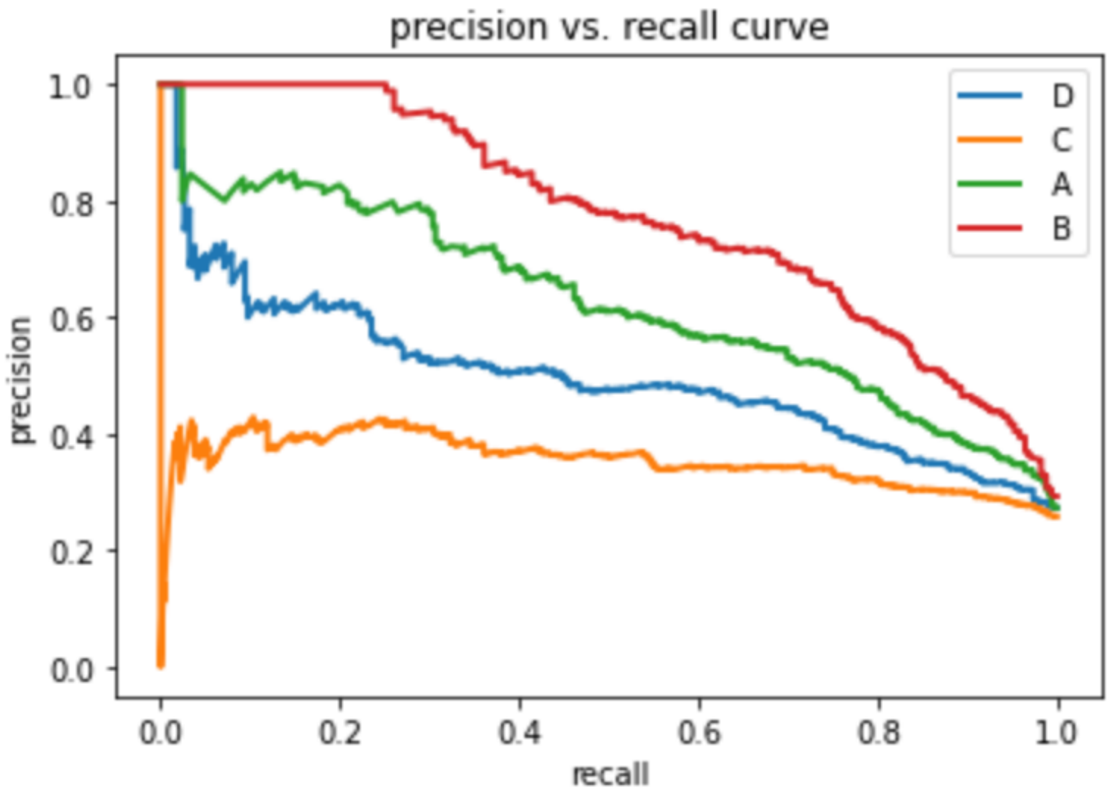
Baseline Model

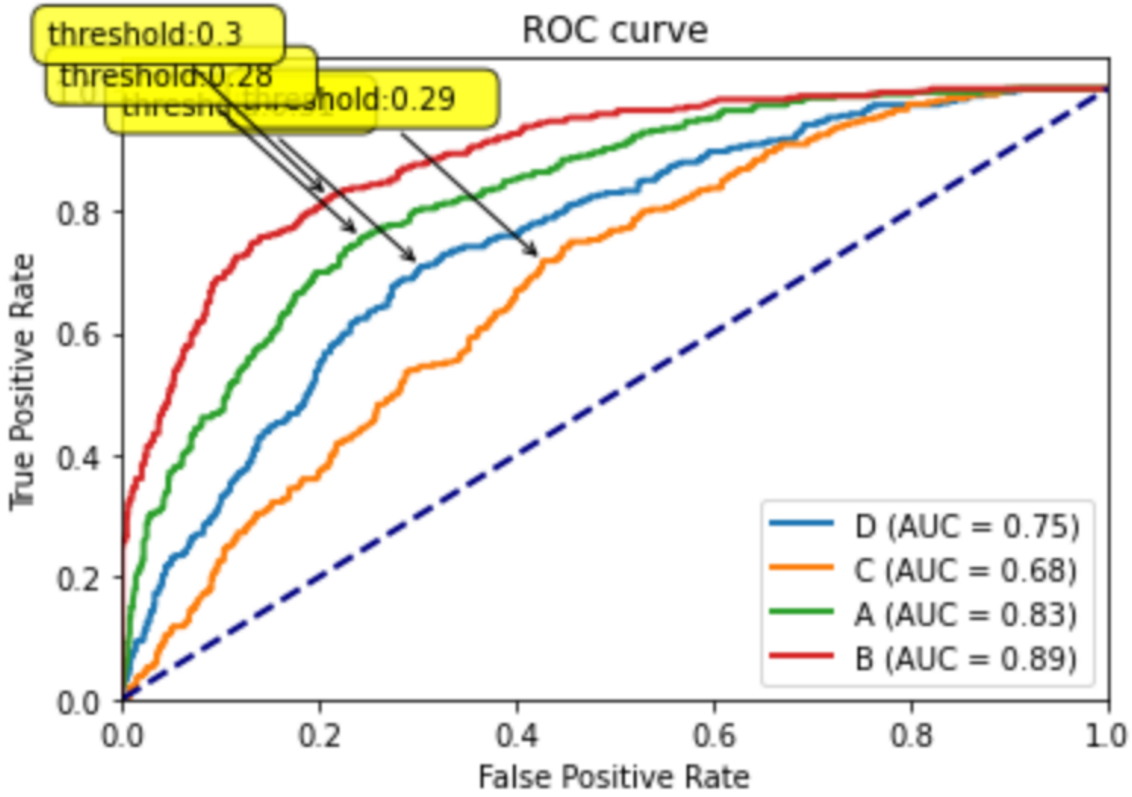
MODEL	precision	recall	f1-score	accuracy
Baseline	0.056	0.237	0.091	0.237

Model table

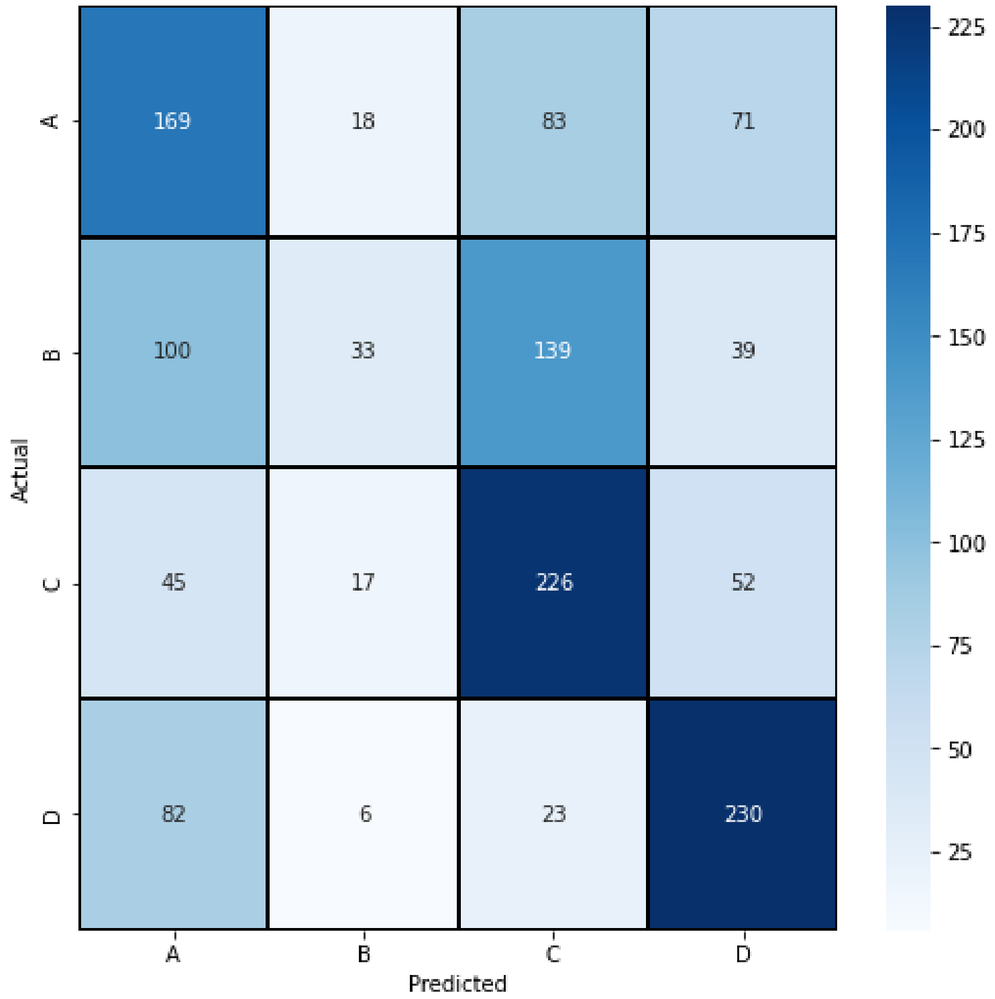
MODEL	precision	recall	f1-score	accuracy
KNN	0.508	0.513	0.506	0.513
logistic Regression	0.503	0.511	0.516	0.511
Descision Tree	0.508	0.495	0.497	0.495
Random Forest	0.487	0.501	0.481	0.501
XGBClassifier	0.522	0.525	0.520	0.525
Naive Bayes	0.460	0.474	0.444	0.474

Curve


precision vs. recall curve
precision
recall
1.0
0.8
0.6
0.4
0.2
0.0
0.0
0.2
0.4
0.6
0.8
1.0
D
C
A
B


ROC curve
threshold:0.3
threshold:0.28
threshold:0.29
True Positive Rate
False Positive Rate
0.0
0.2
0.4
0.6
0.8
1.0
D (AUC = 0.75)
C (AUC = 0.68)
A (AUC = 0.83)
B (AUC = 0.89)

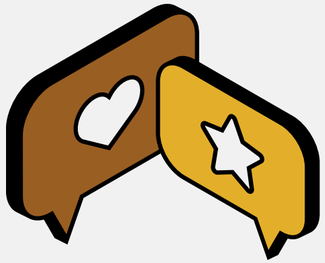
Confusion Matrix



Conclusion

Summary of data modeling

It can be found that all the performance are generally not very good,However Refer to the accuracy list above, XGBClassifier seems to be the best approach. However, would suggest to have a better sampling again for better data modeling.



THANK YOU

Maram alfaifi
Tahani almutairi