

ORGANIC ANALYSIS

BUSINESS OBJECTIVE

The business goal is to decide whether a bank customer will subscribe to a bank product or not.

This will help to target our marketing efforts for this product.

DATA SCIENCE TASK

We will perform a cluster analysis by using the naïve bayes classification method to model the probability of subscribing to the product.

SELECT DATA

We will use 222223 records of 10 input variables (features) and one binary target variable (whether the customer subscribed to the product or not)

DATA EXPLORATION

DATA ANALYSIS

We applied the following steps in our analysis. The details of the analysis may be seen in the appendix.

We modeled the data using a Nive Bayes Classification method which yielded the prior and conditional probabilities for all model variables. The model is own in the appendix. These variables are interesting on their own as a look at how each feature variable impacts the target variable.

1. DemAffl
2. DemAge
3. DemCluster
4. DemClusterGroup
5. DemGender
6. Der
7. DemTVReg
8. PromClass
9. PromSpend
10. PromTime
11. TargetBuy
12. TargetAmt

APPLY ANALYSIS

We can use the Naïve Bayes Classification model to classify new customers as to their likelihood to subscribe or not. Customers likely to subscribe could be given a marketing message to entice to subscribe.

DEPLOY MODEL

We would create a random Control group and two experimental groups of the same number as the control one which the NB Model predicts will not subscribe. We would expect the subscribe experimental group to have a higher subscription rate compared to the control group and the non-subscribe experimental group to have a lower subscription rate compared to control group.

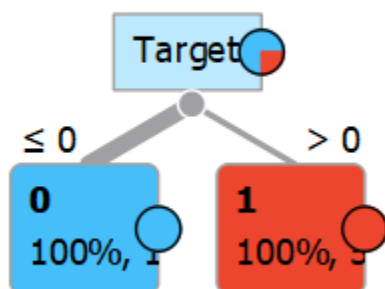
ASSESS RESULTS

We will evaluate the subscription rates of the control and two experimental groups to see if there was a statically significant and business value difference. If not, we will re-analyze the problem.

response rate for the targeted customers compared to the response rate for the random sample to see if there was a greater response rate for the xyz.

STRENGTHS OF XYZ ANALYSIS

CLASSIFICATION TREE



LOGISTIC REGRESSION ANALYSIS

Preprocessing

Rank - Orange

Scoring Methods

☒ Information Gain

☒ Information Gain Ratio

☐ Gini Decrease

☐ ANOVA

☐ χ^2

☒ ReliefF

☐ FCBF

Select Attributes

☐ None

☐ All

☒ Manual

☐ Best ranked:

		#	Info. gain	Gain ratio	ReliefF
1	N TargetAmt		0.808	0.814	0.346
2	N DemAge		0.093	0.047	0.051
3	N DemAffl		0.070	0.035	0.018
4	C DemGender	3	0.044	0.035	0.023
5	N PromSpend		0.010	0.005	0.002
6	C PromClass	4	0.009	0.005	0.050
7	C DemClusterGroup	7	0.003	0.001	0.038
8	N DemCluster		0.003	0.001	0.011
9	N PromTime		0.003	0.001	0.013
10	C DemTVReg	13	0.001	0.000	0.026
11	C DemReg	5	0.000	0.000	0.009
12	N ID		0.000	0.000	0.019
13	N Feature 1		0.000	0.000	0.020

Predictions - Orange

Show probabilities for:

Classes in data

☒ Show classification errors

Restore Original Order

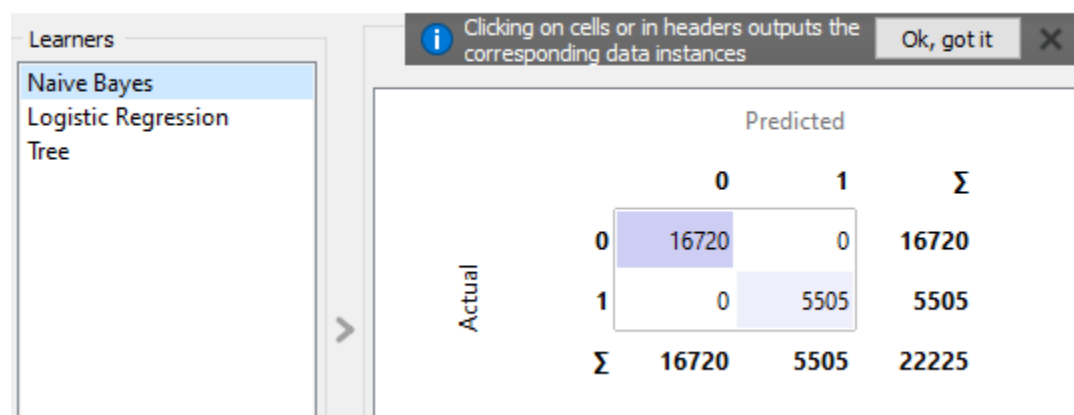
	Naive Bayes (0)	Naive Bayes (1)	gistic Regression	gistic Regression	Tree (0)	Tree (1)	Fold	Feature 1	ID	DemAffl	DemAge	DemCluster	DemClusterGroup	DemGender	De
0.999981	1.94327e-05	0.60097	0.39903	1	0	1	3525	8014957	5	76	52	F	F	Midian	
0.999918	8.18735e-05	0.552213	0.447787	1	0	1	658	2752220	9	55	28	D	F	Scottish	
0.999923	7.67946e-05	0.506782	0.493218	1	0	1	363	1117629	7	?	35	D	F	Midian	
0.999993	6.57249e-06	0.783778	0.216222	1	0	1	19077	44148196	7	49	51	F	U	Midian	
0.999995	5.38083e-06	0.763918	0.236082	1	0	1	10639	27748928	5	37	4	A	?	South E	
6.30217e-06	0.999994	0.56087	0.43913	0	1	1	4727	10053654	?	30	?	?	F	Midian	
0.00315369	0.999646	0.703264	0.296736	0	1	1	8020	17674107	23	34	50	F	M	?	
0.000407959	0.999592	0.761169	0.238831	0	1	1	13618	32987230	10	61	14	B	F	South E	
0.000292768	0.999707	0.760813	0.239187	0	1	1	11997	29974682	6	36	7	B	F	South E	
0.999995	4.73788e-06	0.690821	0.309179	1	0	1	9574	24270121	7	61	51	F	U	Scottish	
0.999906	9.35281e-05	0.603525	0.316475	1	0	1	11139	28769619	11	52	21	C	M	South E	
0.999981	1.85922e-05	0.776229	0.223771	1	0	1	21564	51180801	5	52	22	C	M	North	
0.999908	9.24447e-05	0.721667	0.278333	1	0	1	13840	33322666	12	57	25	C	M	Midian	
0.999975	2.45891e-05	0.7837	0.2163	1	0	1	15536	35294465	7	61	37	E	F	North	
0.999938	6.19968e-05	0.679663	0.320337	1	0	1	6105	12965111	12	52	7	B	F	South E	
0.999906	9.37681e-05	0.587735	0.412265	1	0	1	6831	14570335	10	41	47	F	?	North	
0.000291179	0.999709	0.696702	0.303298	0	1	1	7474	16546857	11	62	18	C	F	North	
0.999734	0.000266202	0.573236	0.426764	1	0	1	5613	11827831	7	42	19	C	F	Midian	
0.999812	0.000187811	0.699618	0.300382	1	0	1	14991	34779570	12	47	28	D	M	Midian	
0.999991	9.05475e-06	0.690816	0.309184	1	0	1	13695	33086750	6	50	42	E	U	North	
0.000003	17.04579e-05	0.748945	0.259155	1	0	1	1479	9278123	?	79	77	D	M	South E	

☒ Show performance scores

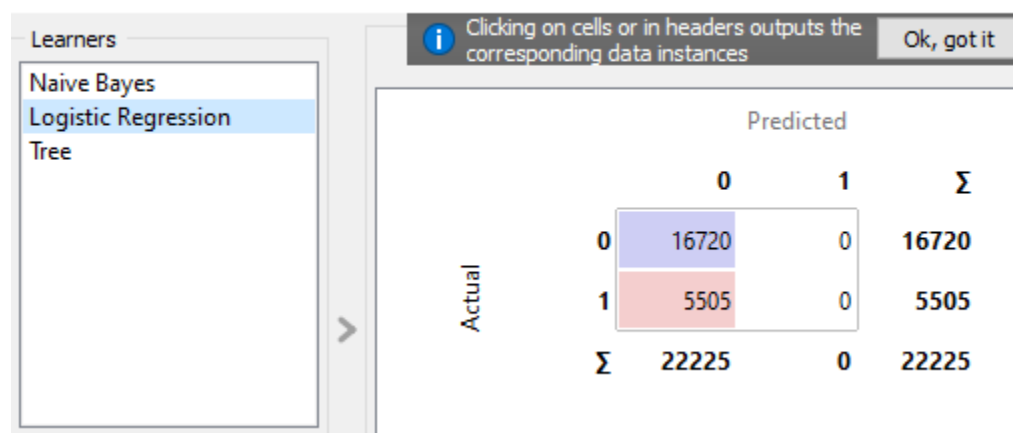
Target class: (Average over classes)

CONFUSION MATRICES

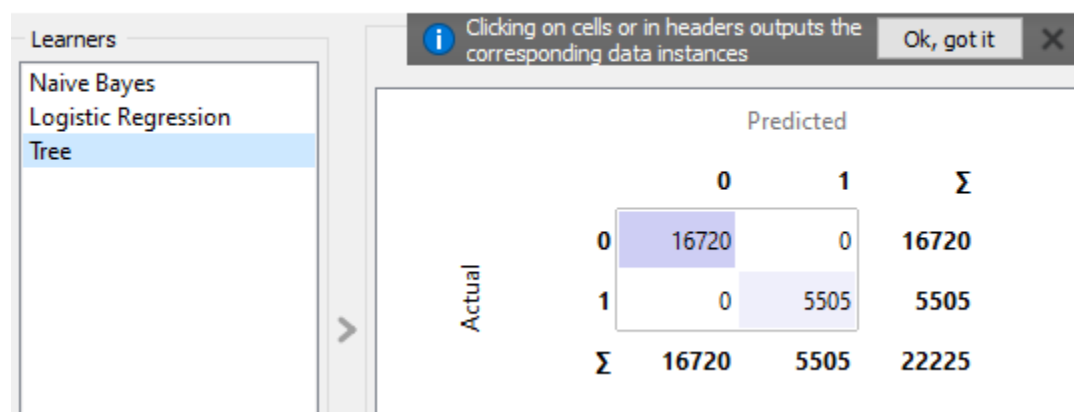
Confusion Matrix - Orange



Confusion Matrix - Orange



Confusion Matrix - Orange



APPENDIX

X	ID	DemAffl	DemAge	DemCluster
Min. : 1	Min. : 140	Min. : 0.000	Min. :18.0	Min. : 1.00
1st Qu.: 5556	1st Qu.:11694023	1st Qu.: 6.000	1st Qu.:44.0	1st Qu.:14.00
Median :11112	Median :28748786	Median : 8.000	Median :54.0	Median :27.00
Mean :11112	Mean :26055403	Mean : 8.712	Mean :53.8	Mean :27.19
3rd Qu.:16668	3rd Qu.:37454020	3rd Qu.:11.000	3rd Qu.:64.0	3rd Qu.:38.00
Max. :22223	Max. :52856469	Max. :34.000	Max. :79.0	Max. :55.00
		NA's :1085	NA's :1508	NA's :674
DemClusterGroup	DemGender	DemReg	DemTVReg	PromClass
Length:22223	Length:22223	Length:22223	Length:22223	Length:22223
Class :character	Class :character	Class :character	Class :character	Class :character
Mode :character	Mode :character	Mode :character	Mode :character	Mode :character

PromSpend	PromTime	TargetBuy	TargetAmt
Min. : 0.01	Min. : 0.000	Min. :0.0000	Min. :0.0000
1st Qu.: 0.01	1st Qu.: 4.000	1st Qu.:0.0000	1st Qu.:0.0000
Median : 2000.00	Median : 5.000	Median :0.0000	Median :0.0000
Mean : 4420.59	Mean : 6.565	Mean :0.2477	Mean :0.2947
3rd Qu.: 6000.00	3rd Qu.: 8.000	3rd Qu.:0.0000	3rd Qu.:0.0000
Max. :296313.85	Max. :39.000	Max. :1.0000	Max. :3.0000
	NA's :281		

	education			
Y	primary	secondary	tertiary	unknown
no	0.17216881	0.50978759	0.27613751	0.04192845
yes	0.12800948	0.46450237	0.33180095	0.07587678

	default	
Y	no	yes
no	0.979323644	0.020687535
yes	0.990568720	0.009526066

	housing	
Y	no	yes
no	0.4348854	0.5651258
yes	0.6540758	0.3460190

	loan	
Y	no	yes
no	0.8541140	0.1458971
yes	0.8957820	0.1043128

	contact		
Y	cellular	telephone	unknown
no	0.61878703	0.06987703	0.31135271
yes	0.85312796	0.05218009	0.09483412

	duration	
Y	[,1]	[,2]
no	222.3756	209.9930
yes	539.3555	437.6109

	poutcome			
Y	failure	other	success	unknown
no	0.10565120	0.03633874	0.01453885	0.84349357
yes	0.09957346	0.06639810	0.15170616	0.68251185