Lead Scoring Case Study

Bussiness Problem

- An education company named X Education sells online courses to industry professionals. On any given day, many professionals who are interested in the courses land on their website and browse for courses. When these people fill up a form providing their email address or phone number, they are classified to be a lead. Moreover, the company also gets leads through past referrals. X Education has appointed us to help them select the most promising leads, i.e. the leads that are most likely to convert into paying customers.

Solution

We have been provided with a leads dataset from the past with around 9000 data points. This dataset consists of various attributes.

- The target variable, in this case, is the column 'Converted' which tells whether a past lead was converted or not wherein 1 means it was converted and 0 means it wasn't converted.

- We built a logistic regression model to assign a lead score between 0 and 100 to each of the leads which can be used by the company to target potential leads. A higher score would mean that the lead is hot, i.e. is most likely to convert whereas a lower score would mean that the lead is cold and will mostly not get converted.

Summary Report

Thia Model contains,6351 training records From total now of records.This Is Generalised Linear Model,Binomial,Containing best 15 features .

Dep. Variable:	Converted	No. Observations:	6351
Model:	GLM	Df Residuals:	6335
Model Family:	Binomial	Df Model:	15
Link Function:	Logit	Scale:	1.0000
Method:	IRLS	Log-Likelihood:	-3278.6
Date:	Tue, 21 Jan 2025	Deviance:	6557.2
Time:	15:59:30	Pearson chi2:	6.63e+03
No. Iterations:	5	Pseudo R-squ. (CS):	0.2585
Covariance Type:	nonrobust		

Summary Report

This report tells about coefficient, standard error, P value.

	coef	std err	z	P>|z|	[0.025	0.975]
const	-3.2303	0.485	-6.659	0.000	-4.181	-2.280
Lead Origin	0.4386	0.063	6.912	0.000	0.314	0.563
Lead Source	0.1023	0.013	8.085	0.000	0.078	0.127
Do Not Email	-1.5613	0.150	-10.388	0.000	-1.856	-1.267
TotalVisits	0.0243	0.009	2.818	0.005	0.007	0.041
Page Views Per Visit	-0.0618	0.021	-2.948	0.003	-0.103	-0.021
Last Activity	0.1167	0.014	8.337	0.000	0.089	0.144
Country	-0.0023	0.004	-0.558	0.577	-0.010	0.006
Specialization	-0.0245	0.007	-3.578	0.000	-0.038	-0.011
What is your current occupation	1.0714	0.104	10.319	0.000	0.868	1.275
What matters most to you in choosing a course	-0.6544	0.084	-7.762	0.000	-0.820	-0.489
Search	-0.6869	0.895	-0.768	0.443	-2.441	1.067
Lead Profile	-0.7089	0.052	-13.617	0.000	-0.811	-0.607
City	-0.0173	0.017	-0.996	0.319	-0.051	0.017
A free copy of Mastering The Interview	-0.3451	0.075	-4.613	0.000	-0.492	-0.198
Last Notable Activity	0.0410	0.015	2.651	0.008	0.011	0.071

Summary report

This reports about VIF.it stataes that low VIF value is better for feature.

	Features	VIF
16	X Education Forums	inf
15	Newspaper Article	inf
11	What is your current occupation	61.83
23	Lead Profile	43.62
10	How did you hear about X Education	19.43
27	Last Notable Activity	16.31
7	Last Activity	13.04
8	Country	9.19
9	Specialization	8.26
24	City	4.85
1	Lead Source	4.78
12	What matters most to you in choosing a course	4.34
6	Page Views Per Visit	4.26
0	Lead Origin	3.61
5	Total Time Spent on Website	2.37
26	A free copy of Mastering The Interview	2.29
18	Digital Advertisement	2.01
4	TotalVisits	1.91
19	Through Recommendations	1.40
13	Search	1.21
2	Do Not Email	1.15
3	Do Not Call	1.00
14	Magazine	NaN
17	Newspaper	NaN
20	Receive More Updates About Our Courses	NaN
21	Update me on Supply Chain Content	NaN
22	Get updates on DM Content	NaN
25	I agree to pay the amount through cheque	NaN

Summary Report

This is confusion matrix.

It gives True positive,True negative,False positive and False negative values to calculate :

Sensitivity,specificity,precision,recall .

- ([[1433, 291], [399, 600]],

Thank U