

Summary

We looked at information about people visiting X Education to see how we can get more industry professionals to take their courses. We checked things like how people found the website, how long they stayed, and if they signed up for courses. Here's what we did:

1. Cleaning the data: We fixed some missing information and made sure everything was tidy. We changed some empty parts to "not provided" so we didn't lose too much info. We grouped people from India and other places.
2. Checking the data: We quickly looked at the data to see if everything looked okay. We found that some categories didn't matter much, but the numbers looked good and there were no strange values.
3. Making pretend categories: We made up categories for things like where people came from. We got rid of the made-up ones we didn't need. For numbers, we scaled them down to make them easier to work with.
4. Splitting the data: We divided the data into two parts, with 70% for training and 30% for testing.
5. Model building: Firstly, we used RFE to select the top 15 important variables. Then, we manually removed the remaining variables based on their VIF values and p-values. We kept the variables with VIF values less than 10 and p-values less than 0.05.

	Features	VIF
16	What is your current occupation_Unemployed	9.87
5	Lead Origin_Landing Page Submission	6.82
15	Specialization_not provided	6.31
18	What is your current occupation_not provided	5.75
12	Specialization_Management_Related	5.33
3	Page Views Per Visit	2.83
7	Lead Source_Olark Chat	2.80
1	TotalVisits	2.27
4	A free copy of Mastering The Interview	2.25
6	Lead Origin_Lead Add Form	2.15
17	What is your current occupation_Working Profes...	2.05
11	Specialization_Business Administration	1.39
10	Lead Source_Welingak Website	1.36
8	Lead Source_Organic Search	1.36
2	Total Time Spent on Website	1.33
13	Specialization_Media and Advertising	1.22
14	Specialization_Travel and Tourism	1.22
0	Do Not Email	1.11
9	Lead Source_Social Media	1.03

6. Model Evaluation:

- a. ROC - 3.5
 - b. Training set
 - i. Accuracy - 80.19 %
 - ii. Sensitivity is 65.87 %
 - iii. Specificity is 89.0 %
 - c. Test set
 - i. Accuracy - 80.11%
 - ii. Sensitivity is 74.93 %
 - iii. Specificity is 83.31 %
7. Prediction: Prediction was done on the test data frame and with an optimum cut off as 0.35

It was found out that the following matter a lot for customers to convert

- 1. Lead Origin in Lead Add Form
- 2. Working professionals
- 3. Lead source
 - a. Lead Source_Welingak Website
 - b. Lead Source_Olark Chat
 - c. Lead Source_Social Media
- 4. Total Time Spent on Website