

Lead Scoring Case Study

BY: ANKIT CHAUHAN

AMIT SANTRA

ANJALI

Problem Statement

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.

The company markets its courses on several websites and search engines like Google. Once these people land on the website, they might browse the courses or fill up a form for the course or watch some videos. When these people fill up a form providing their email address or phone number, they are classified to be a lead.

The problem statement is to help X Education identify the most potential leads, or "Hot Leads," that are most likely to convert into paying customers. The company wants to build a logistic regression model to assign a lead score between 0 and 100 to each lead. A higher score indicates a higher conversion chance, while a lower score suggests a lower conversion chance. The goal is to increase the lead conversion rate, ideally reaching around 80%.

Analysis Approach

1. Importing libraries and inspecting the dataset.
2. Handling missing values and dropping unnecessary columns.
3. Data preparation, including converting binary variables and creating dummy features for categorical variables.
4. Test-train split and feature scaling.
5. Exploring correlation and dropping highly correlated variables.
6. Model building using logistic regression.
7. Performing feature selection using recursive feature elimination (RFE) and adjusting the model.
8. Evaluating the model's performance using accuracy, sensitivity, specificity, precision, and recall scores.
9. Plotting the ROC curve and determining the optimal cutoff point.
10. Making predictions on the test set and assigning lead scores.

Key Factors

The coefficients derived from the logistic regression model highlight the variables that significantly impact lead conversion. Some of the key factors identified include:

1. Total Time Spent on Website: Leads who spend more time on the website have a higher likelihood of conversion.
2. Lead Source_Olark Chat: Engaging with leads through the Olark Chat source increases conversion probabilities.
3. Last Activity_SMS Sent: Sending SMS messages to leads positively influences their conversion chances.
4. What is your current occupation_Working Professional: Working professionals exhibit a higher likelihood of conversion.
5. Lead Source_Reference: Leads generated through references have a significantly higher chance of conversion.

Communication Channels and Lead Nurturing

1. **Communication Channels:** The analysis underscores the importance of effective communication channels. Leads who engage with emails, open emails, visit specific website pages, or participate in Olark Chat conversations show a higher likelihood of conversion. These insights can guide X Education in tailoring their communication strategies to maximize engagement and conversion rates.
2. **Lead Nurturing:** The model emphasizes the significance of lead nurturing. Factors such as Last Activity_Showing interest and Last Notable Activity_Others demonstrate the importance of capturing and maintaining leads' interest throughout the customer journey. By nurturing leads and providing relevant information, X Education can increase conversion probabilities.

Results from 17th Model

In the 17th model, the logistic regression coefficients provide valuable insights into the factors that significantly influence lead conversion.

In the 17th model, the logistic regression approach yielded an overall accuracy of 0.7989, and the features that remain, are statistically significant and do not show any multi collinearity.

In the test run of our selected model, the accuracy was 0.7925. This indicates that the model correctly predicts lead conversion approximately 80% of the time. This accuracy is a strong indicator of the model's ability to identify potential conversions.

In the upcoming slides, we will interpret the results in business terms.

1. Do Not Email: This coefficient of -1.5563 suggests that leads who do not wish to receive emails have a significantly lower likelihood of conversion. X Education should respect the preferences of these leads and focus on alternative communication channels to engage with them effectively.
2. Total Time Spent on Website: The coefficient of 1.0050 indicates that the more time a lead spends on the website, the higher their conversion likelihood. This suggests that leads who show a greater interest and engagement with the website's content are more likely to convert. X Education should focus on nurturing these leads by providing relevant and valuable information to enhance their conversion chances further.
3. A free copy of Mastering The Interview: The coefficient of -0.3903 suggests that offering a free copy of the "Mastering The Interview" resource negatively impacts lead conversion. X Education might consider reevaluating the effectiveness of this particular offering or exploring alternative incentives to improve lead conversion.
4. Lead Source_Olark Chat: The coefficient of 0.7888 indicates that leads who come from the Olark Chat source have a higher likelihood of conversion. This implies that engaging with leads through the chat feature can be an effective strategy in driving conversions. X Education should prioritize engaging with leads from this source and provide timely and relevant information to increase their conversion chances.

5. **Lead Source_Reference:** With a coefficient of 3.5838, leads who come through references have a significantly higher likelihood of conversion. This underscores the importance of word-of-mouth and referrals in generating high-quality leads. X Education should focus on building and nurturing relationships with current customers to encourage more referrals and leverage the power of positive recommendations.
6. **Last Activity_Email Opened:** The coefficient of 0.8404 suggests that leads who open emails are more likely to convert. X Education should focus on creating compelling email content to capture the attention of leads and drive them towards conversion.
7. **Last Activity_Page Visited on Website:** The coefficient of 0.6151 indicates that leads who visit specific pages on the website have a higher conversion likelihood. X Education should analyze the pages that lead to higher conversions and tailor their website content and user experience to encourage leads to visit these pages.
8. **Last Activity_SMS Sent:** With a coefficient of 2.0734, sending SMS messages to leads significantly increases their likelihood of conversion. X Education should leverage SMS communication to engage with leads effectively and provide personalized information that encourages conversion.

9. Last Activity_Showing interest: The coefficient of 1.1094 suggests that leads who demonstrate interest in X Education's offerings have a higher likelihood of conversion. X Education should identify and prioritize leads who have shown interest and tailor their communication to address their specific needs and concerns.
10. What is your current occupation_Working Professional: The coefficient of 2.6325 indicates that working professionals have a significantly higher likelihood of conversion. X Education should focus on targeting and engaging with this segment, providing specialized content and offerings that cater to their professional development needs.
11. Last Notable Activity_Modified: The coefficient of -0.7767 suggests that leads with the last notable activity as "Modified" have a lower likelihood of conversion. X Education should analyze the reasons behind this and consider strategies to engage with these leads more effectively to increase their conversion chances.
12. Last Notable Activity_Others: The coefficient of 1.8021 indicates that leads with other notable activities have a higher likelihood of conversion. X Education should identify the specific activities falling under this category and leverage them to drive conversions effectively.

By considering these coefficients and their corresponding variables, X Education can prioritize their sales and marketing efforts, focusing on the factors that significantly impact lead conversion and tailoring their strategies to maximize conversion rates.

Lead Scores

A higher score indicates a higher conversion chance, while a lower score suggests a lower conversion chance.

We can see in the image that a lead has been assigned a score of 100 which means that it has highest chance of getting converted.

Leads are assigned as low as 7 which means the chances of conversion is very low.

X Education can prioritize their efforts and focus on leads with a higher conversion potential.

	Converted	Lead Number	Converted_Prob	final_predicted	Lead Score
0	0	3271	0.193890	0	19
1	1	1490	0.965466	1	97
2	0	7936	0.170969	0	17
3	1	4216	0.748603	1	75
4	0	3830	0.141577	0	14
5	1	1800	0.742194	1	74
6	0	6507	0.439351	1	44
7	0	4821	0.301091	0	30
8	1	4223	0.956930	1	96
9	0	4714	0.384427	1	38
10	1	3977	0.490952	1	49
11	0	4209	0.072814	0	7
12	0	2320	0.091052	0	9
13	0	3744	0.072814	0	7
14	1	2570	0.384427	1	38
15	1	985	0.682251	1	68
16	1	1946	0.724436	1	72
17	0	7813	0.149096	0	15
18	0	7031	0.145849	0	15
19	0	2044	0.066275	0	7
20	0	7906	0.159328	0	16
21	1	2461	0.381566	1	38
22	0	8154	0.204663	0	20
23	1	5822	0.996774	1	100
24	0	4513	0.072814	0	7

Conclusion

Based on the analysis, the presentation provides actionable recommendations to improve lead conversion:

1. Prioritize leads who spend more time on the website and engage with Olark Chat.
2. Utilize SMS communication to enhance engagement and conversion.
3. Focus on working professionals and leverage their specific needs and interests.
4. Foster referrals and word-of-mouth by providing an exceptional customer experience.
5. Tailor communication strategies to open and click rates for emails.
6. Provide personalized content based on lead interests and preferences

Thank You