
LEAD SCORE CASESTUDY SUMMARY

X Education sells online courses to industry professionals. It needs a model to differentiate between leads and hence assign a lead score.

The CEO, in particular, has given a ballpark of the target lead conversion rate to be around 80%.1.

Data Cleaning:

- a) Checking the null values in columns and dropping them when they have more than 40% values of missing.
- b) Dropping the columns with very less variance to offer such as the column that has only 'No'.
- c) Replacing Select with Nan and then imputing them according to the columns – categorical/ Numerical, mean/median etc.

EDA:

- a) Checking for outliers
- b) Univariate and bivariate analysis of the data – categorical and numerical
- c) Correlation between the numerical columns

Creating Dummy Variables and Scaling:

- a) Dummy variables were created for categorical variables
- b) Scaling was done using StandardScaler

Train and Test Split:

- a) The data was split into train and test set with the ratio of 70-30

Model Building:

- a) Obtained around 20 variables using RFE and manual feature elimination was done on these variables to obtain a stable model with p-value less than 0.05 and VIF less than 5 for all the columns.

Model Evaluation:

- a) A confusion matrix was made. Later an optimum cut off value was used to find accuracy, sensitivity and specificity which was around 80%.

Prediction:

- a) Prediction on test data with optimum cut off of 0.34 resulted in accuracy of ~81%, specificity of ~80% and sensitivity of ~82%.

Precision – Recall:

- a) A cut off of 0.41 was found to with precision of around 86% and recall of around 85% on test data.

Conclusion:

Major indicators that a lead will get converted to a hot lead:

- Lead Origin_Lead Add Form: A lead sourced from Lead Origin_Lead Add Form is more likely to get converted
- Occupation_Working Professional: Working professionals are more likely to get converted.
- Lead_Source_Welingak website: A lead sourced from Welingak Website is more likely to get converted.
- Last Activity_SMS Sent: A lead having SMS sent previously are more likely to get converted.
- Lead Source_Olark Chat: A lead sourced from Olark Chat is more likely to get converted

Major indicators that a lead will NOT get converted to a hot lead:

- Last_Activity_Olark chat conversation: Customers under olark chat, are less likely to get converted into hot leads.
- Lead Ongin_Landmg Page Submission: Customers under Lead Ongin_Landmg Page Submission, are less likely to get converted into hot leads.
- Do Not Email: Customer who choose Do Not Email are less likely to get converted into hot leads.

Recommendations:

- The company should use a leads score threshold of 41 to identify "Hot Leads" as at this threshold, Sensitivity Score of the model is around 81% which is as good as CEO's target of 80%.