

Lead Score Case Study

Anuja Phadtare

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. Moreover, the company also gets leads through past referrals. Once these leads are acquired, employees from the sales team start making calls, writing emails, etc. Through this process, some of the leads get converted while most do not. The typical lead conversion rate at X education is around 30%.

X Education has appointed you to help them select the most promising leads, i.e. the leads that are most likely to convert into paying customers. The company requires you to build a model wherein you need to assign a lead score to each of the leads such that the customers with a higher lead score have a higher conversion chance and the customers with a lower lead score have a lower conversion chance. The CEO, in particular, has given a ballpark of the target lead conversion rate to be around 80%.

Solution Methodology

Data Preprocessing:

- Identify and handle duplicate records.
- Address and manage NA and missing values.
- Eliminate columns with a significant number of missing values that are not beneficial for analysis.
- Impute missing values as required.
- Detect and manage outliers in the dataset.

Exploratory Data Analysis (EDA):

- Univariate analysis: Assess value counts and distribution of variables.
- Bivariate analysis: Examine correlation coefficients and relationships between variables.

Data Transformation:

- Perform feature scaling.
- Generate dummy variables and encode the data for analysis.

Solution Methodology (Contd.)

Model Development:

- Utilize logistic regression for model creation and predictions.

Model Validation:

- Determine the optimal cut-off using the ROC curve.
- Evaluate accuracy, precision, recall, and specificity of the model.

Model Presentation:

- Present the model's performance metrics and visualizations.

Conclusion and Recommendations:

- Summarize the findings and provide actionable recommendations based on the analysis.