



Lead Scoring

OAISWARI SHAW

10/18/2022

Background

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%. Now, although X Education gets a lot of leads, its lead conversion rate is very poor.

For example, if, say, they acquire 100 leads in a day, only about 30 of them are converted. To make this process more efficient, the company wishes to identify the most potential leads, also known as 'Hot Leads'. If they successfully identify this set of leads, the lead conversion rate should go up as the sales team will now be focusing more on communicating with the potential leads rather than making calls to everyone.

Objective

Build 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 customers. 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.

Data Understanding

- ❑ A total of 9240 samples were given in input file with 37 columns.
- ❑ Most of the features are of categorical types and are of Yes/No types.
- ❑ From the count plot of city, we can see that maximum leads are from Mumbai.
- ❑ For Most of the leads, having better career prospects, matters the most while choosing a course.
- ❑ The current occupation of maximum number of leads is Unemployed.
- ❑ Majority of the leads are specialised in Finance Management, followed by Human Resource & Manufacturing Management.
- ❑ Majority of the leads are from India.
- ❑ Majority of the leads were the customers tagged as -will revert after reading the email.
- ❑ Total Time Spent on Website is highly correlated with Converted.