

LEAD SCORING CASE STUDY

Objective:

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. 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. 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 leads. A higher score i.e. is most likely to convert whereas a lower score would mean that the lead is cold and will mostly not get converted. There are some more problems presented by the company which your model should be able to adjust to if the company's requirement changes in the future so you will need to handle these as well.

Steps Followed

- Reading Data
- Cleaning Data
- EDA
- Creating Dummy
- Splitting data into train and test set
- Building Model
- Making Predictions
- Model Evaluation
- ROC Curve
- Prediction on test set

- Precision- Recall

Learning

- working on the "Lead Scoring" dataset provided practical experience in handling real-world data analysis tasks, including data cleaning, exploratory analysis, modeling, and evaluation. The assignment enhanced our understanding of logistic regression.

Details of files given

- The python file showing coding and data analysis
- Subjective questions answers in pdf format
- Final Presentation
- Summary on what's done in the entire python file in pdf format