

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%.

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.

There are a lot of leads generated in the initial stage (top) but only a few of them come out as paying customers from the bottom. In the middle stage, you need to nurture the potential leads well (i.e. educating the leads about the product, constantly communicating, etc.) in order to get a higher lead conversion.

X Education wants to 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 higher lead score have a higher conversion chance and the customers with 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:

Through this project, the below mentioned processes were performed.

- Data Loading
- Data Exploration (EDA)
- Pre-processing
- Feature Engineering
- Outlier Analysis
- Model Building
- Model Performance Benchmarking
- Model Performance Evaluation
- Cross Validation and Hyperparameter Tuning
- Model Diagnosis - Probability Calibration, ROC AUC Curve, Precision Recall Curve

Insights from the Case study:

- More focus has to be on the Indian market, especially the Mumbai City which had the highest Conversion.
- Welinkak Website provides the best lead for conversion rate and also through reference and the focus can be on this category to increase the overall count of conversion.
- Google is helping for maximum lead generation and conversion, which has to be effectively utilized.
- SMS is the best option for Conversion and the website should be more engaging with relevant information. More time spent on the website the conversion chances are higher.
- The targeted person has to be the Unemployed people who are willing to join and engage in courses to upskill themselves and this has to be marketed and utilized effectively.
- Lead Quality mentioned as might also have better correlation of conversion and this too will have to be noted.