

Lead Scoring

In nearly any client-focused business, the acquisition of new customers is a high priority requiring a large investment of time, money, and labor. In order to have the best outcomes, new leads generated (say, by a marketing team) must be appropriately nurtured (by sales staff) until the sale is made, and we can say the lead has been converted. Naturally, however, not all leads are of the same quality, and spending valuable time on hopeless leads can be an inefficient use of energy and resources.

In my personal sales experience at a private adult education provider, I've wasted plenty of time writing multiple emails and calls to potential students that would go unanswered. As the only sales advisor in our small branch, that was time better spent drumming up new leads or ensuring a strong relationship with current students that I had already converted.

In larger companies, with teams of sales personnel and more time to follow up on every lead, another conflict can arise if individual employees perceive that leads are unfairly distributed among team members, whether by the luck of the draw or by biased actors.

One possible solution to both of these problems is a reliable model of lead scoring. Incoming leads can be assessed based on multiple features, including data explicitly collected from the potential client (such as personal identifiers like job or location, or answers to questions about their specific motivation in choosing a product) and data implicitly gathered from them (such as time they've spent on the website, or whether they have responded to communication). Each lead is then given a score indicating how likely they are to convert to a sale based on the available information.

With robust lead scoring, a company can simultaneously pursue two pathways to better sales outcomes:

- 1. Better Lead Conversion Rate**

Attention to lead score will make sure sales staff are prioritizing the leads most likely to convert, and conversely deprioritizing unlikely sales. This should lead to a greater overall conversion rate, as good leads aren't likely to be accidentally ignored.

- 2. Better Lead Acquisition**

By identifying the features most important to determining lead score, the model gives evidence for prioritizing certain markets over others. For example, if school applicants from major cities have a higher conversion rate than rural ones, marketing efforts should be focused on those areas.

Additionally, lead scoring can help ensure fair assignment of leads among sales staff, because the model impartially provides a numerical assessment of lead quality that can then be distributed evenly.

X Education Case Study

As real data on leads is not generally available for free use (largely to protect privacy), I will be using data simulated for educational purposes by UpGrad IIIT-B, accessed via [kaggle.com](https://www.kaggle.com). The data represents data on over 9,000 unique leads for X Education, a fictionalized online adult education company.

Typically, a potential student will consider, and even submit an application to, multiple schools before deciding which option is the best fit for their academic goals, finances, and schedule. It is essential, therefore, for schools to make sure they provide lots of support for students throughout the application process, to make sure any students who are undecided follow through and enroll. According to the problem statement provided on Kaggle, “The typical lead conversion rate at X Education is around 30%.” By focusing efforts on students more likely to convert, X Education should be able to increase their sales greatly.

The goal of this project is to build a model to predict the likelihood of conversion for each lead based on our gathered features, and then assign a numerical score reflecting that probability. Armed with this score, company sales management can then make an informed decision of how to assign and treat incoming leads. Meanwhile, by looking at the percentage of high- and low-scoring leads, the company can adjust marketing techniques to try to maximize the number of strong leads generated.