

LEAD GENERATION – PREDICTION PROJECT

Context:

The EdTech industry has been surging in the past decade immensely, and according to a forecast, the Online Education market will be worth \$286.62bn by 2030 with a compound annual growth rate (CAGR) of 10.26% from 2018 to 2030. The modern era of online education has enforced a lot in its growth and expansion beyond any limit. Due to having many dominant features like ease of information sharing, personalized learning experience, transparency of assessment, etc, it is now preferable to traditional education.

In the present scenario due to Covid-19, the online education sector has witnessed rapid growth and is attracting a lot of new customers. Due to this rapid growth, many new companies have emerged in this industry. With the availability and ease of use of digital marketing resources, companies can reach out to a wider audience with their offerings. The customers who show interest in these offerings are termed as leads. There are various sources of obtaining leads for Edtech companies, like

- The customer interacts with the marketing front on social media or other online platforms.
- The customer browses the website/app and downloads the brochure
- The customer connects through emails for more information.

The company then nurtures these leads and tries to convert them to paid customers or sells these leads to education platforms for income generation. For this, the representative from the organization connects with the lead on call or through email to share further details.

Objective:

ExtraaLearn is an initial-stage startup that offers programs on cutting-edge technologies to students and professionals to help them upskill/reskill. With many leads being generated regularly, one of the issues faced by ExtraaLearn is to identify which of the leads are more likely to convert so that they can allocate resources accordingly. Now in this project, we use the data to:

- Analyse and build an ML model to help identify which leads are more likely to convert to paid customers,
- Find the factors driving the lead conversion process
- Create a profile of the leads who are likely to convert

Data Dictionary:

The data contains the different attributes of leads and their interaction details with ExtraaLearn. The detailed data dictionary is given below.

- ID: ID of the lead
- age: Age of the lead
- current_occupation: Current occupation of the lead. Values include 'Professional', 'Unemployed', and 'Student'
- first_interaction: How did the lead first interact with ExtraaLearn. Values include 'Website', 'Mobile App'
- profile_completed: What percentage of the profile has been filled by the lead on the website/mobile app = Values include Low - (0-50%), Medium - (50-75%), High (75-100%)
- website_visits: How many times has a lead visited the website

- `time_spent_on_website`: Total time spent on the website
- `page_views_per_visit`: Average number of pages on the website viewed during the visits.
- `last_activity`: Last interaction between the lead and ExtraaLearn.
 - Email Activity: Seeking details about the program through email, the Representative shared information with a lead like a brochure of the program, etc
 - Phone Activity: Had a Phone Conversation with a representative, Had a conversation over SMS with a representative, etc
 - Website Activity: Interacted on live chat with a representative, Updated profile on the website, etc
- `print_media_type1`: Flag indicating whether the lead had seen the ad of ExtraaLearn in the Newspaper.
- `print_media_type2`: Flag indicating whether the lead had seen the ad of ExtraaLearn in the Magazine.
- `digital_media`: Flag indicating whether the lead had seen the ad of ExtraaLearn on the digital platforms.
- `educational_channels`: Flag indicating whether the lead had heard about ExtraaLearn in the education channels like online forums, discussion threads, educational websites, etc.
- `referral`: Flag indicating whether the lead had heard about ExtraaLearn through reference.
- `status`: Flag indicating whether the lead was converted to a paid customer or not.

Observations:

- Once again the most important features are: `first_interaction_website`, `time_spent_on_website`, and `profile_completed_medium`.
- This confirms that the key features are related to the website.
- The best model is the tuned random forest model since it has balanced scores on both the train and test data sets.
- It can be arguable that the tuned decision model is also a very good option because the recall score on the train set is 0.93 and on the test set is 0.82. However, the difference between once set and the other is significantly big compared to the tuned random forest.

Conclusions:

- A thorough analysis on a data set provided by ExtraaLearn was carried out to identify the leads that might become paying customers.
- The EDA shows that 30% of the leads became paying customers. Also, most of the media channels do not have an impact on the conversion rate.
- Two modelling techniques are used: Decision Tree and Random Forest. In addition, models with default parameters and tuned models are considered.
- The best model is the Tuned Random Forest with generalized balanced results on both recall and precision.
- The analysis pointed out that the most important features are those related to the website such as the first interaction, the time spent of the website, profile completion, and number of page views per visit.

- Furthermore, the media channels do not impact the decision of a lead to become a paying customer.

Recommendations:

- The analysis shows that website-related variables are of great importance. ExtraaLearn should invest its resources in improving the user experience of the website.
- The first interaction with the website is the most decisive feature that will lead to becoming a paying customer. Research on making a lead that interacts first with the mobile app to visit the website as well might improve the conversion rate.
- The time spent on the website indicates the interest of the lead to buy the subscription by exploring the content and resources ExtraaLearn provides. Designing the website in such a way that these are easy to find and entertaining will increase the conversion rate.
- Finally, profile completion also plays a role in lead conversion rate. ExtraaLearn should research how to make the lead complete their profiles without blocking too much of the content on the website.
- The ad channels have a very low impact on the conversion rate. Two options arise: stop investing in ads in magazines or newspapers or give some special discount if the lead provides some code or something or the sort that is only available in these ad channels.