

## Summary

X-Education is an Education company that sells online courses to industry professionals. We were tasked with analysing the people, also known as leads, who are interested in enrolling to one of the courses. The objective is to identify “Hot Leads”, i.e. people with the highest potential of enrolling.

We set about this task by first understanding the data available to us. We performed exploratory data analysis and first discarded the data columns which did not provide any statistically significant data. We then performed data cleansing to remove rows with unwanted values, and, also imputed data where necessary. Finally, we scaled the data so that the machine learning model is and its results are more readable.

Once we completed prepping the data using steps above along with a few additional steps to make the data ingestible by our machine learning model, we starting by build out a model which can accurately predict whether a lead was “hot” or “not”. The entire process of the data analysis, cleansing and model building provided us with two huge benefits:

1. It gave us a good machine learning model which can predict with over 90% accuracy whether a lead can be converted or not.
2. It gave us a deeper understanding of all the features/data points available and how each of the features is affecting the conversion.

The top 5 features/variables that has the highest importance in predicting whether a lead is hot or not is listed below in no particular order:

- Lead Origin
- Specialization
- Lead Source
- Last Activity
- Last Notable Activity
- Tags

A few other key points to keep in mind:

1. People who have a tag of “Will revert after reading email” have a very high chance of being converted and these people will need to be treated as priority 1 hot leads.
2. People whose phones are switched off or leave their phone ringing are extremely unlikely to enrol to any of the courses. So we can reduce the energy and effort spent to try and convert these leads.
3. Even though direct traffic and google sources have extremely high number of leads, the conversion rate is not satisfactory. Efforts should be made to improve the conversion rate for these sources which are extremely effective in producing leads.
4. People who have a tag of “Already a student” also have a high chance of not enrolling for another course. Actions need to be taken to better understand why this is so that the underlying issues can be addressed.

Keeping the above factors in mind and with the help of our prediction algorithm, X-Education can definitely work on not only creating more leads, but also on increasing the conversion rate of these leads there by achieving their target.