

Summary Report

1. The Lead Score Case study assignment was prepared using logistic regression model. This model was the best fit because it contains various categorical variables.
2. At the initial stage when we looked at the data, there were around 9000 records out of which only 392 leads had a good score for getting converted.
3. We have observed that most of the leads are from INDIA and in terms of city, Mumbai has the highest number of leads.
4. There are a few variables which had a default level as 'Select' which basically means that the student had not selected any other option from that particular column which by default got recorded as 'Select' level. We have found that the column "How did you hear about X Education" and "Lead Profile" had high number of 'Select' level and we had to drop these two columns.
5. Then to draw some meaningful insights we selected few important categorical columns such as "Specialization", "What is your current occupation", etc.
6. Good numerical columns such as "TotalVisits", "Total Time Spent on Website" and "Page Views Per Visit" were helpful in determining the chances of the leads getting converted.
7. It is observed that leads who are ambitious in their career, who are looking for a better job opportunity, promotion or career switch have the high probability of getting converted.
8. If we observe the "Last Notable Activity" of the customers it is seen that those customers who are active in opening mails, reading SMS or attending phone calls have a high probability of getting converted.
9. In the dataset, we have observed that there are a high number of unemployed leads who had a low chance of getting converted, so the focus was maintained on working professional who had a high chance of getting converted.