Problem Statement : X Education sells online courses to industry professionals.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%.

Business Goal : X Education needs help in selecting the most promising leads, i.e. the leads that are most likely to convert into paying customers.

The company needs a model wherein you a lead score is assigned 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:

* The overall conversion rate of around 39%

* The maximum conversion happened from Landing Page Submission Also there was only one request from quick add form which got converted.

* It can be seen that major conversion in the lead source is from google

* Major conversion has happend from the emails that have been sent

* It can be noticed that major conversions happened when calls were made. However, it can also be seen that 2 leads opted for "Do Not Call", but they still got converted.

* Last activity value of 'SMS Sent' had more conversion

* More conversion happend with people who are unemployed. It can also be noticed from the above data that - Out of 7 business men, 4 got converted - Out 10 housewives, all 10 leads got converted.

* Conversion rate is high on leads who are not through search

* Since "Newspaper Article" column now has only one value for all rows - "No" , it is safe to drop this column

* Since "X Education Forums" column now has only one value for all rows - "No" , it is safe to drop this column

* Since Newspaper column has only one row with "Yes" as the value and further since this lead did not get converted and rest of all the values are "No", we can safely drop the column

* It can be noticed above that there were 2 leads that came from digital advertisement of which one lead got converted

* It can be seen that a total of 6 leads came through recommendations of which 5 leads got converted

* Conversion rate is high on leads who do not want a free copy of Mastering Interviews

* It can be noticed that the conversion rate is high for "SMS Sent"

* Hence we can see that the final prediction of conversions have a target of 80% (79.8%) conversion as per the X Educations CEO's requirement . Hence this is a good model.* Overall Metrics - Accuracy, Confusion Metrics, Sensitivity, Specificity, False Postive Rate, Positive Predictive Value, Negative Predicitive Value on final prediction on train set

* Hence we can see that the final prediction of conversions have a target rate of 79% (78.5%) (Around 1 % short of the predictions made on training data set)

* Overall Metrics - Accuracy, Confusion Metrics, Sensitivity, Specificity on test set

summary:

- While we have checked both Sensitivity-Specificity as well as Precision and Recall Metrics, we have considered the optimal cut off based on Sensitivity and Specificity for calculating the final prediction.

- Accuracy, Sensitivity and Specificity values of test set are around 81%, 79% and 82% which are approximately closer to the respective values calculated using trained set.

- Also the lead score calculated in the trained set of data shows the conversion rate on the final predicted model is around 80%

- Hence overall this model seems to be good.