Lead Scoring Case Study Summary

Problem Statement

X Education is an online education company that sells courses to industry professionals. On any given day, many professionals who are interested in the courses land on their website and browse for courses.

The company has a high lead generation rate, but only 30% of leads convert into paying customers.

The CEO wants to develop Build a logistic regression model to assign a lead score between 0 and 100 to each of the leads which can be used by the company to target potential leads.

Dataset:

The company provide dataset from the past with around 9000 data points. This dataset consists of various attributes such as Lead Source, Total Time Spent on Website, Total Visits, Last Activity, etc. which may or may not be useful in ultimately deciding whether a lead will be converted or not

Steps taken to solve the problem and build the model is given as follows:

- Dataset is loaded Using Pandas libraries.
- Data Cleaning which involves handling null value treatment.
- Exploratory Data Analysis.
- Splitting of Data Set in to train and test data.
- Feature Scaling
- Feature Selection using RFE

- Model Training
- Model Evaluation

Conclusion:

Our Model has achieved the overall accuracy of 91% and the sensitivity of 86%, precision of around 89% and a recall of around 85% on the test data.

Top 3 feature which the model has predicted:

- Tags_Lost to EINS
- Tags_Closed by Horizzon
- Tags_Will revert after reading the email

We should focus on the following customer segments for lead conversion:

- Who revert after reading the email.
- Who spent time good amount of time on website.
- Working people