CASE STUDY SUMMARY

The requirement is to find the right features that can be targeted to gain more students to take up a course in the X Education portal. The provided data contains features that are inputted by the user/potential students and the sales team. The following steps are taken to create a logistic model for this business need.

1. EDA:

EDA process included observing, cleaning, and manipulating the data. Most of the features had a single value for more than 90% of the rows, such features are dropped. Some features had considerable missing values, and the rows pertaining to those are dropped keeping in mind to keep at least 60% of the data retained.

2. Visualization:

Pair plots are used to see the relations between the features and the target variable. It can be seen that as Total time spent on website increase, the conversion rate increases. The trend goes same with the Total visits.

3. Dummy Variables and Scaler:

To have an efficient ML model, it's important to differentiate between the values in the categorical feature. Creation of dummy variable helps in diving deep into each field of a feature and converting them as feature to best explain the trend. For numeric variables, to normalize the values MinMax scaler.

4. Model Building:

The first step that comes in Model building is the Train-Test split(i.e 70-30 split respectively), the model is trained on train set and predicted with the unseen, unbiased Test set.

The model is built using RFE selection method to choose 15 preferred features for the further analysis. The important features are selected and the features that show high VIF and p-value are dropped. The VIF for the variables are maintained <5 and p-value <0.05

5. Model Evaluation:

The metrics such as Accuracy, Specificity and Sensitivity need to be determined, for this a confusion matrix was created, an optimum cut-off of 0.43 was derived from this matrix, which was used on the conversion probability to determine the final prediction

6. Model Prediction:

The columns chosen for Train data are used for Test Set also, and the scaler is just to transform the numerical. The model well with all the metrics maintained over 75%. Metrics also included Precision and Recall.

Recommendations:

- The total time spent and more number of visits to website play a crucial role in getting a conversion

CASE STUDY SUMMARY

- Those who have opted for Do No Email on the portal have less interested in getting into the course
- Interestingly, phone conversation shows a higher conversion rate.
- Wellingak website shows higher coefficient for getting enrolled to the course
- If the lead origin is **Lead Add Form,** the chances for buyers purchasing this course is very high compared to other means.