

# Summary

The Analysis of X education was done to find more ways to get industry professionals to join their courses. Data provided has information regarding the time spent, how the lead reached the site and conversion rate.

Following Steps are used :

1. **Cleaning data:**

- a. Columns having 40% or more null values were dropped.
- b. Instead of value "select" Null was replaced
- c. In categorical features null were replaced by mode and in numerical null were replaced by median
- d. Other activity like outlier treatment, grouping low frequency data were done

2. **EDA**

- a. Checking the data imbalance (only 38% were converted)
- b. Removing all the irrelevant columns as they will just increase the complexity

3. **Data Preparation:**

- a. Creating dummy variables for categorical features.
- b. Splitting data into train and test sets in the 70 30 ratio.
- c. Scaling the data using fit transform

4. **Model Building**

- a. Using RFE to select top 15 features which will highly impact the decision
- b. After building a model, manual feature reduction is performed based on p value (p value > 0.05) and VIF > 5 was also rejected

5. **Model Evaluation**

- a. **With the help of confusion matrix cut-off value was selected to be 0.35**
- b. **at cut-off 0.35 approx 80% accuracy was obtained.**
- c. **for cut-off at 0.38 similar accuracy was found but other than that for all other value it was less than 80%**

6. **Prediction**

- a. **Scaling on test set was performed**
- b. **Prediction was done on test dataset at optimum cut-off 0.35 and confusion matrix was created.**
- c. **Accuracy was close to 80%, which is also close to train accuracy.**

## **Recommendations.**

- **Focus should be more on working professionals.**
- **Some time and interest can be given to curating more personalized Emails and SMS, as there are leads which will revert back after reading it.**
- **Giving right Tags is imported, as it will be very helpful to target the right people.**