

SUMMARIZED REPORT

Case Study: Lead Scoring

Objective:

We aim to improve the conversion rate of X Education and meeting the CEO's expectation of having a conversion rate of 80 per cent by creating a machine learning model that would predict if a particular user can be converted. We also aim to recognise key factors that effect the conversion rate of X Education, improving which can help X Education improve their conversion rate.

Process:

- **Data Collection:** X Education has provided the data.
- **EDA - Data Cleaning:** We then handle missing values by dropping columns and by imputing values, either by mode for categorical data or by median for numerical data. We then explore the data provided to us using various plots for a better understanding of the data provided to us by X. We also have handled outliers by capping the values. The data provided had both raw as well as derived data, that has been created by the sales team of X Education. We remove all the derived data as when the model is deployed this data will not be available to the model. We also remove some columns that are skewed as it may have a negative impact in the model.
- **Data Pre-processing:** We have used standard scaler to scale the numeric variables also created dummy variables for the categorical data. We have cleaned raw data in 7:3 ratio for training the model and testing the model.
- **Feature Scaling :** We have scaled the numeric variables using the Standard Scaling.
- **Feature Selection & Model Building:** Using RFE we select the top 18 features that have the most effect on the conversion. of a particular lead. We than manually drop features based on their P-values and VIF values.
- **Model Evaluation:** We have calculated Metrics like Sensitivity, Specificity, Accuracy Precision and recall to evaluate the model. Confusion metrics was also evaluated.
- **Model Test :** We than tested our model on the test data to check if the results are as expected.

Recommendations:

X Education needs to improve their engagement on Welingak website, they also must target the working professionals as they are the users that are more attracted by the services X Education provides. X Education can have a huge boost by starting a referral program with incentives to both the person giving reference and the person joining X Education because of the referral. They can also improve the lead conversion score by increasing the time spent on their website and by improving the Olark chat service.

Conclusion:

We can see that all the objectives stated above are satisfied and have achieved a sensitivity of above 80%, which indicates that we have met the CEO's expectation. We have also found the key factors that indicate if a particular user can be converted, this would help the sales team of X Education use their limited human and financial effectively and efficiently.

