Summary of Lead Scoring Case Study

We have been engaged by 'X Education' who wants us to use predictive modelling techniques to assign a lead score to each of the leads such that the customers with a higher lead score have higher conversion chance. CEO's target for lead conversion rate is around 80%. Their leads conversion rate is around 38% at present and they hope to increase the same.

Importing Libraries:

- The necessary libraries were imported to work with the data.
- NumPy, Panda, SKlearn, Matplotlib, Seaborn etc. were imported to help with data manipulation, model building and visualizations.

Data Cleaning:

- Columns with missing values of 45% or more were dropped,
- Missing values in numerical columns were imputed with mode values,
- Other activities like outliers' treatment, fixing invalid data, grouping low frequency values, mapping binary categorical values were carried out.

EDA:

- Performed univariate and bivariate analysis.
- Data imbalance checked and came to the conclusion that 'X Education' is currently able to convert only 38% of the Leads.

Data Preparation:

- Created dummies for categorical columns,
- Used feature scaling to standardize values, etc.

Model Building:

- Used RFE to select the top 12 features,
- Any features with P value grater than 0.05 was also dropped,

Model Evaluation:

- Accuracy was at 80% as per the requirement of the business,
- Confusion metrics were also prepared to check the specificity and sensitivity of the model.

Testing of Model:

- Evaluation metrics for train & test are very close to around 80%.
- Lead score was assigned.