Lead Scoring

Case Study

PROBLEM STATEMENT

► Industry professionals can purchase online courses from X Education.

► Although X Education receives a lot of leads, it has an extremely low lead conversion rate. For instance, only approximately 30 of 100 leads they could in a day might actually be converted.

► The goal of the business is to find the most promising leads, commonly referred to as "Hot Leads," in order to increase the efficiency of this process.

► The lead conversion rate should increase if they are successful in identifying this group of leads because the sales staff will now be concentrating more on connecting with the potential leads rather than calling everyone.

Objective

- They aim to create a model that detects hot leads for that purpose.
- The setting up of the model for future use.
- What are the most promising leads ? Asks X education

SOLUTION

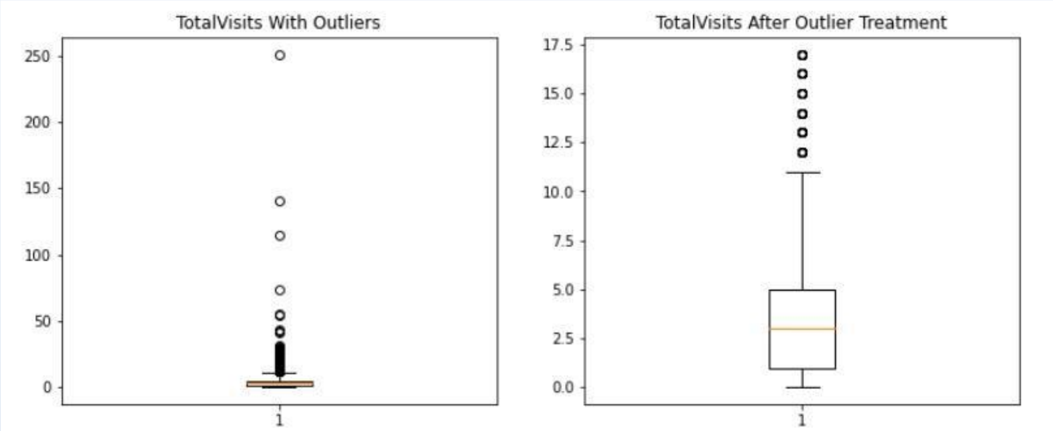
DATA MANIPULATION

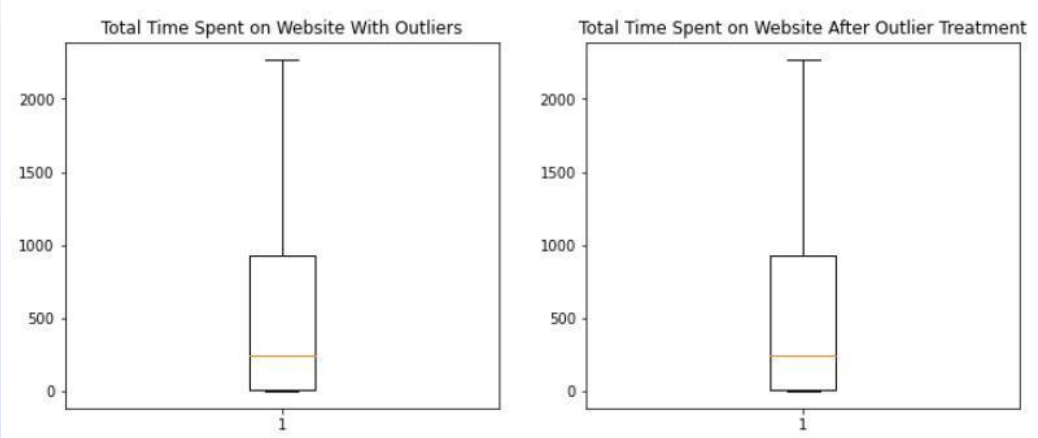
- Verify and deal with duplicate data.
- Verify and handle missing and NA values.
- Remove columns from the analysis if they have a significant number of missing values.
- If necessary, value impugnation.
- Examine and deal with data outliers.

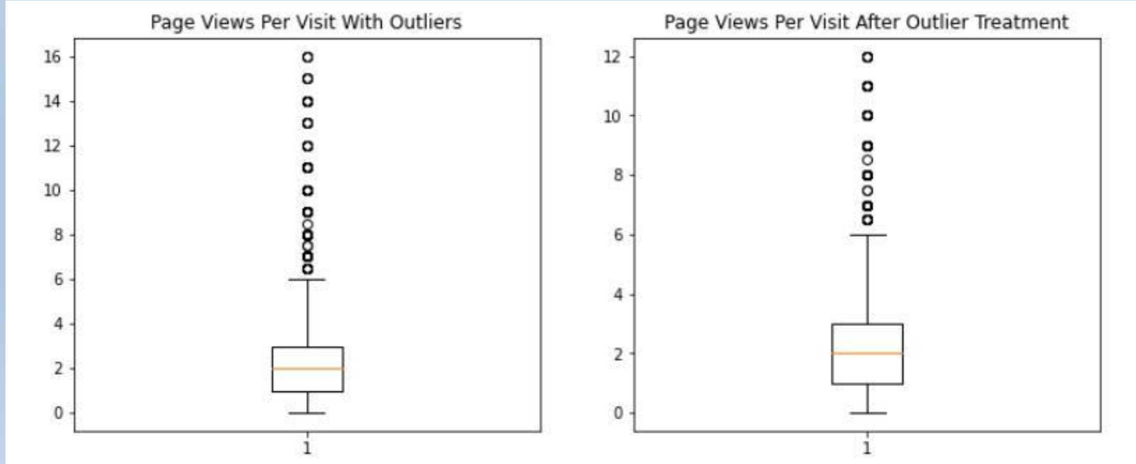
Exploratory Data Analysis

- Analysis of univariate data: value count, variable distribution etc
- Bivariate data analysis : patterns between the variables and correlation coefficients etc
- Data encoding, feature scaling, and dummy variables
- Using logistic regression to create models and make predictions is a classification strategy
- Model verification
- Presentation of model
- Recommendations and conclusions.

OUTLIERS CHECK







MODEL BUILDING

- dividing the data into sets for training and testing

- Performing a train-test split is the first fundamental step in regression; we have selected a 70:30 split ratio.

- For Feature Selection, Use RFE

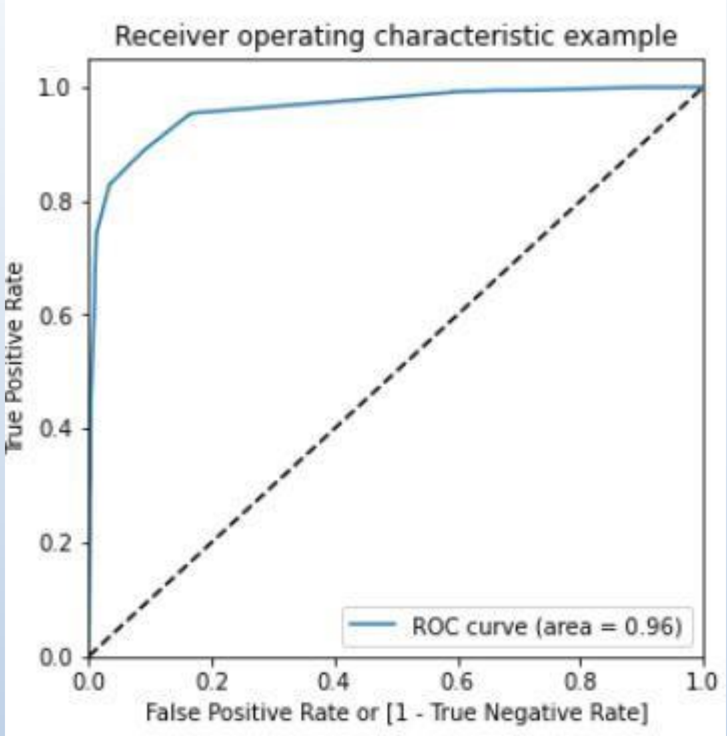
- Running RFE with 15 output variables

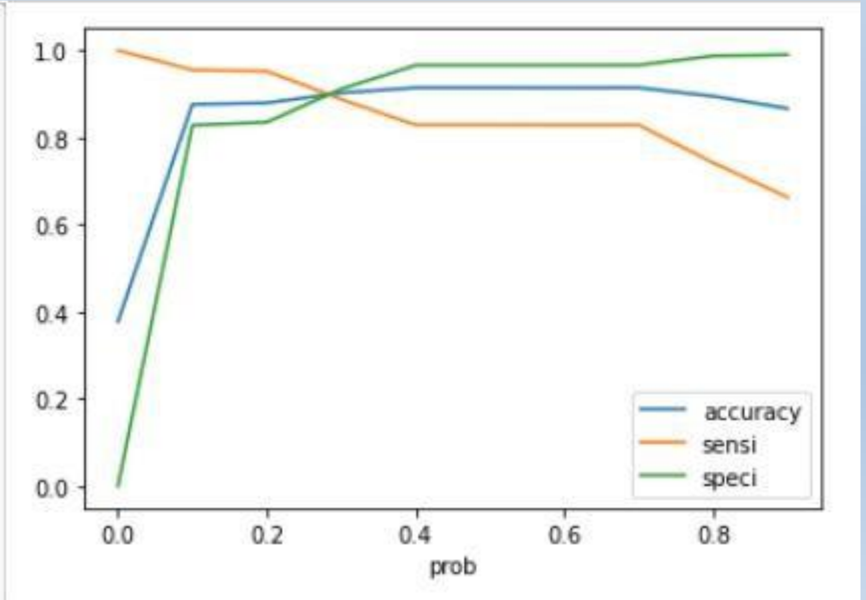
- Removing variables from the model whose p-value is higher than 0.05 and vi value is higher than 5

- Forecasts based on the test data set

- 94.26% overall accuracy

ROC Curve





0.35 is the optimal cutoff in the 2nd graph

PREDICTION ON TEST SET

- We must standardise the test set and ensure that the exact identical columns are present in our final train dataset before making predictions on the test set.

- We then began predicting the test set after completing the previous stage, and the updated prediction values were saved in a new data frame.

- Following this, we evaluated the model by determining its accuracy, precision, and recall.

- The accuracy score was 94.26%, the precision 81.65%, and the approximate recall 91.4%.

- This demonstrates that the accuracy, precision, and recall scores for our test prediction are within acceptable bounds.

- This demonstrates the stability, accuracy, and recall/sensitivity of our model.

- In order to detect hot leads, a lead score is built on a test dataset; the greater the lead score, the better the possibility.

THANK YOU

BY –

Mounika Prudhvi