Summary

To summarize the industry problem and the solution of the problem, we have to use the logistic regression method.

We have to find out the Lead score of each candidate and then by analyzing it, we can now tackle the business problem.

The Company doesn't want to waste their time in convincing those candidates who are less interested in the program.

So the data science team comes to a solution and by using the machine learning technique and building model we can fix the problem.

So the total process for building the model and finding the Lead score are as follows.

1. First we need to read the data and remove the unnecessary columns that are not use of anything for the model building.

2. Then we have to create some dummy variables for the categorical columns which can make some effect on the Lead score.

3. As the data is so big and takes a lot of candidate inputs in it, we just split the data into train and test data.

4. Then we build a model by using machine learning and tested it over the train dataset.

5. After this we are using the RFE technique and manual approach in order to make the model more efficient.

6. In manual approach, we just dropped the columns which have high p Value and high VIF.

7. Then impute the confusion matrix and find out the different metrics of the model i.e. Specificity, sensitivity, Precision and Recall.

8. Then we calculate the Lead score by converting the value of Converted_prob column and Keep the value between 0 to 100.

9. So we concluded that, the company should focus on the clients

Who have scored more than 80 in Lead score and there the particular candidates which can be converted to pursue the course after the help of the marketing team of the company.