

Summary

Following steps are used to find ways to get more industry professionals to join their courses

1. Data Cleaning and Preparation:

- a. As we can see that there are lot of columns where missing values are pretty high. These columns are not useful. Since there are 9240 datapoints in our dataframe, we can eliminate the columns having missing values greater than 3000 as they are of no use.
- b. Highest number of leads are from India. We can drop the column "Country" as well.

2. Preparing the data for modelling

3. Dummy variable creation:

- a. Our next step is to deal with the categorical variables present in the dataset. We will first take a look at the variables which are actually categorical.

4. Test-Train Split

- a. The next step is to split the dataset into training and testing sets. Splitting the dataset into 70% train and 30% test.

5. Model Building

- a. As we can see that there are a lot of variables present in the dataset which cannot be dealt with. So the best way is to select a small set of features from this pool of variables using RFE.

6. Calculating the Precision

- a. 0.782

7. Calculating Recall

- a. 0.775