

Lead Score Assignment Summary:

Process Followed:

- Data understanding - Understood the data and the target variable (converted) to analyse.
- Missing Values Evaluation - Checked the file for missing data. The columns that had more than 40% of the data were removed. For the remaining columns which were mainly categorical, mode of the column was imputed. The remaining columns where there were less than 2% missing values, the missing value rows were removed. 98% of the total data was retained after the missing value evaluation.
- EDA - Analysed all the columns that were retained.
- Data cleaning and Data Pre processing - Removed the columns that did not add in any context to the analysis. Created dummy variables for the categorical columns and scaled the numerical columns using Standard Scaling.
- ML Algorithm used - Logistic Regression.
- Used RFE technique for feature selection. Started with 20 features.
- Model Creation - Created and revised models with the optimal p-value and VIF values were achieved. Below are the final features with VIF values.

	Features	VIF
2	Lead Origin_Landing Page Submission	3.41
15	Last Notable Activity_Modified	2.71
4	Lead Source_Direct Traffic	2.14
11	Specialization_Others	2.14
10	Last Activity_Olark Chat Conversation	1.94
0	Do Not Email	1.86
14	Last Notable Activity_Email Opened	1.85
9	Last Activity_Email Bounced	1.76
3	Lead Origin_Lead Add Form	1.54
7	Lead Source_Welingak Website	1.37
16	Last Notable Activity_Olark Chat Conversation	1.37
5	Lead Source_Organic Search	1.29
8	Last Activity_Converted to Lead	1.24
1	Total Time Spent on Website	1.19
12	What is your current occupation_Working Profes...	1.16
17	Last Notable Activity_Page Visited on Website	1.10
13	Last Notable Activity_Email Link Clicked	1.06
6	Lead Source_Referral Sites	1.03

-

- Model Evaluation: ROC Curve Area = 0.88. The ROC achieved was 88%, indicating that the model is good.
- The metric evaluation on Train and Test data are as below:

Train Data:

- Accuracy : 80.7 %
- Sensitivity : 83 %
- Specificity : 79.3 %

Test Data:

- Accuracy : 80.2 %
- Sensitivity : 80.9 %
- Specificity : 79.8 %

•

- Final Features with parameters

```
: Lead Origin_Lead Add Form                2.688760
What is your current occupation_Working Professional  2.606756
Lead Source_Welingak Website                2.592299
const                                         1.760204
Total Time Spent on Website                 0.991559
Lead Source_Direct Traffic                 -0.364692
Lead Source_Organic Search                 -0.384839
Lead Source_Referral Sites                 -0.698451
Specialization_Others                      -1.043635
Last Activity_Converted to Lead            -1.095930
Last Activity_Olark Chat Conversation       -1.164906
Last Activity_Email Bounced               -1.182122
Lead Origin_Landing Page Submission        -1.315186
Last Notable Activity_Email Opened         -1.444833
Last Notable Activity_Olark Chat Conversation -1.453328
Do Not Email                              -1.555558
Last Notable Activity_Page Visited on Website -1.743767
Last Notable Activity_Email Link Clicked   -1.773033
Last Notable Activity_Modified             -1.781831
dtype: float64
```

•

Recommendations:

- The company should make calls to Working professionals
- The company should make calls where the lead source is Welingak Website
- The company should make calls where the lead origin is Lead Add Form
- The company should make calls to leads that have spent time on the website

The company should not make calls to:

- Leads that have lead source as direct traffic, organic search or referral sites
- Leads whose Last activity is converted to lead, Olark Chat Conversation, Email Bounced
- Leads whose lead origin is Landing page submission
- Leads whose last notable activity is Email opened, Olark Chat Conversation, Page Visited on Website, Email link clicked or modified
- Leads with Do Not Email