Lead Scoring Case study

- **Members :**
  - **Member 1:**
    - **Student Name:** **Hafizullah Mahmudi**
    - **Student Id:** **APFE19M00734**
    - **IIIT Role Number:** DDS1950112
  - **Member 1:**
    - **Student Name:** **Kiran Sunilkumar**
    - **Student Id:** **APFE19804043**
- **2nd March, 2020**

X Education Online Services

Lead Scoring Case Study

Problem statement

To build a Logistic Regression Model to predict whether a lead for online courses for an education company named X Education would be successfully converted or not.

Business Objective and Problem

X Education is a company that sells online course to industry professional. The company markets its products on several website. Most of the company's leads are not converted but the company wants to increase the conversion rate to a round 80%.

In this case study, we will use the existing data and recommend ways how the company can increase its leads to around 80% by finding the most promising leads (hot leads) and label each lead with score between 0 and 100.

Analysis Approach

At first the data was reviewed, null values checked and then outliers were identified and treated.

Columns with single value and high missing values were excluded from the analysis since they do not playing role in the models.

After outlier treatment, data were normalized to unify the scale of the numbers and dummy variables were created as the preparation step for analysis.

Logistic Regression which is one of the recommended methods of machine learning was used to identify leads and calculated the score for each leads to find how likely they are to be converted.

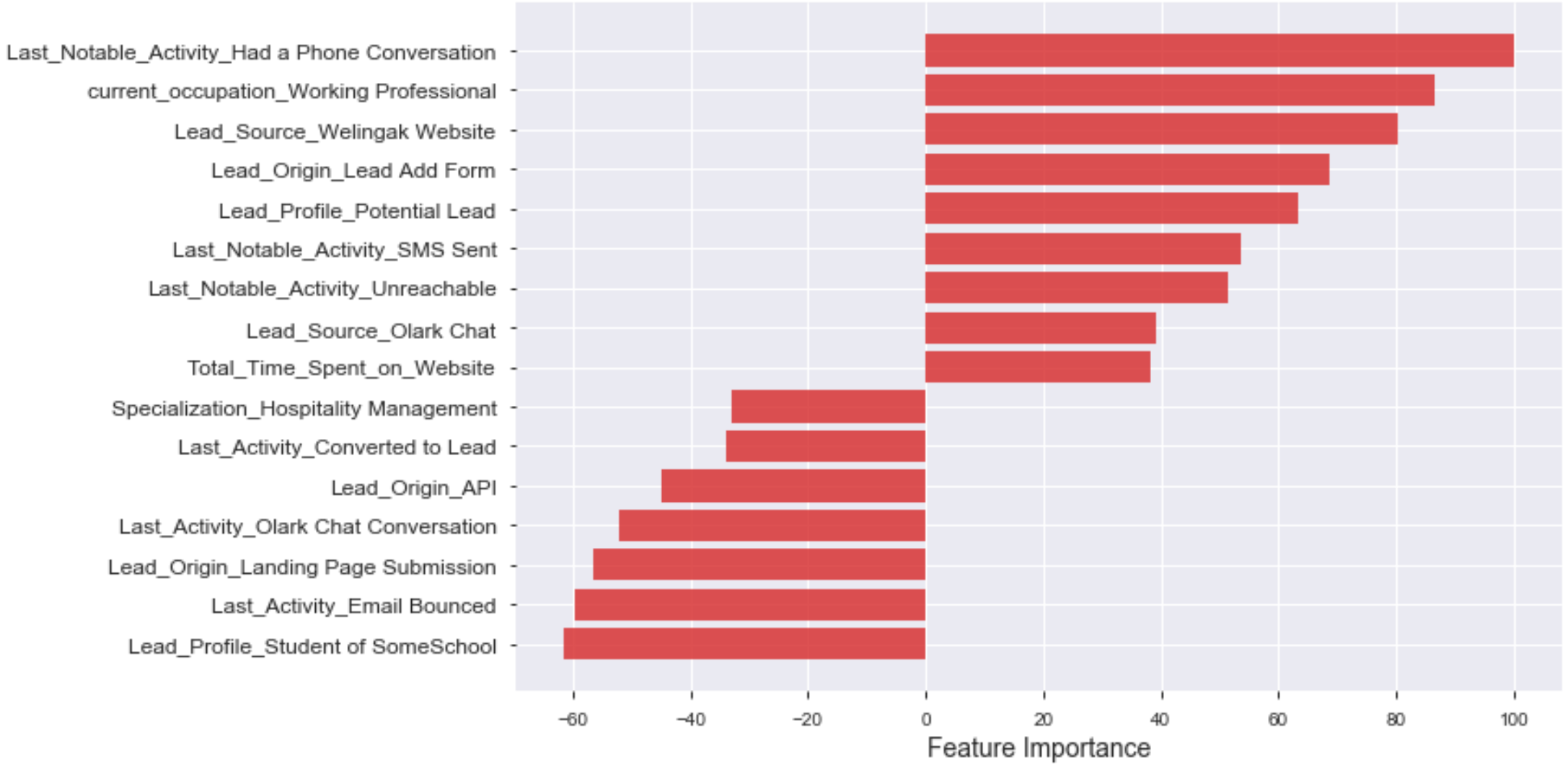
Key variables/drivers for probability of lead conversion

After applying different models and study of significance of variables p-values and VIF, the following variables were found to be most important that is recommend X Education to put more focused on them:

'Total_Time_Spent_on_Website',
'Lead_Origin _API',
'Lead_Origin_Landing Page Submission',
'Lead_Origin_Lead Add Form',
'Lead_Source_Olark Chat',
'Lead_Source_Welingak Website',
'Last_Activity_Converted to Lead',
'Last_Activity_Email Bounced',
'Last_Activity_Olark Chat Conversation',
'Specialization_Hospitality Management',
'current_occupation_Working Professional',
'Lead_Profile_Potential Lead',
'Lead_Profile_Student of SomeSchool',
'Last_Notable_Activity_Had a Phone Conversation',
'Last_Notable_Activity_SMS Sent',
'Last_Notable_Activity_Unreachable']

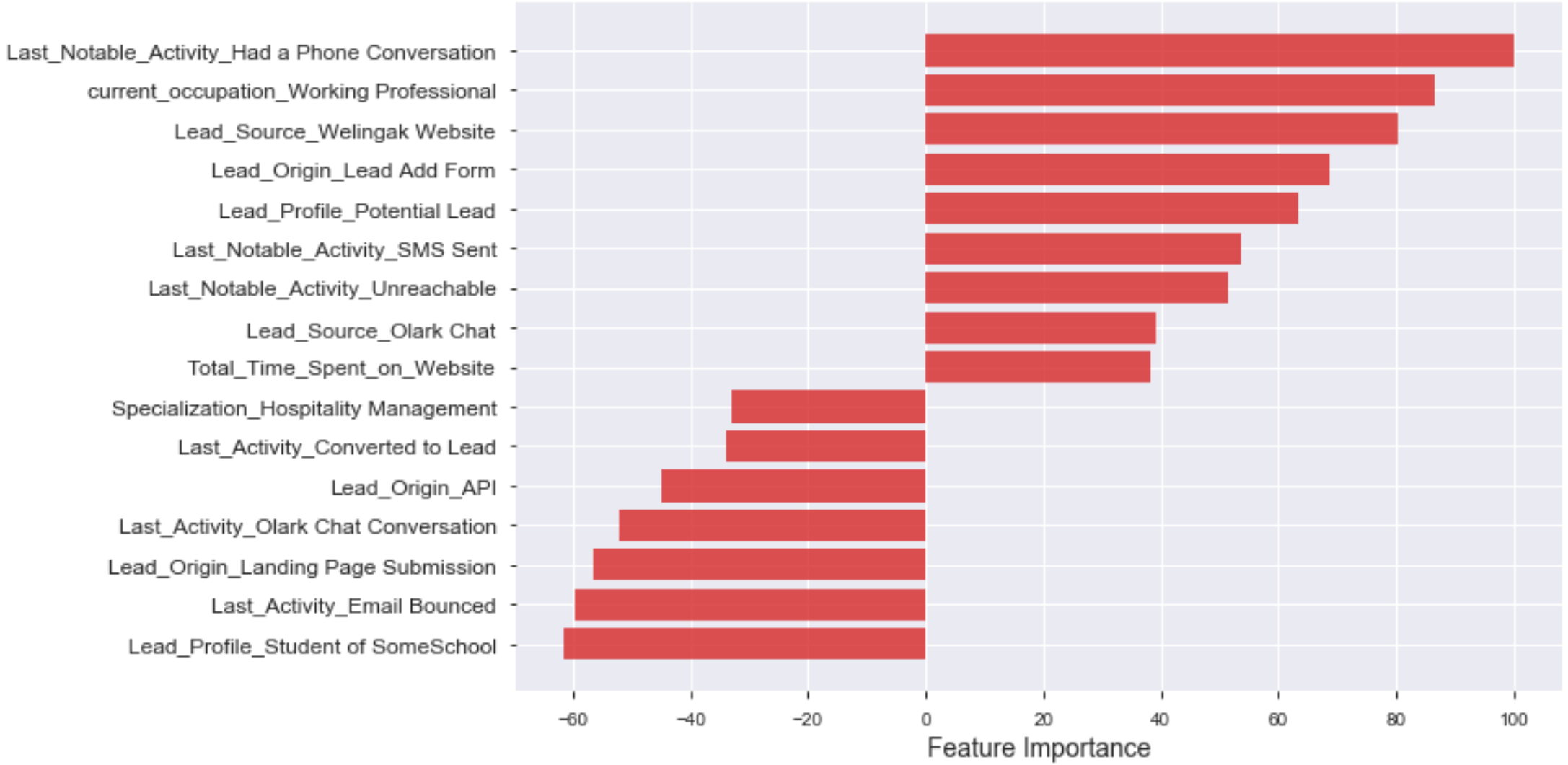
Conversion Probability Increase - I

According to the finding of the models, increasing the value of the following features/variables will increase the probability of conversion:



Conversion Probability Increase - II

According to the finding of the models, decreasing the value of the following features/variables will increase the probability of conversion:



Most Important Variables/Features

Following are top 3 three features are the most important ones that have high impact on the conversion and lead scoring.

	Importance
Last_Notable_Activity_Had a Phone Conversation	100.000000
current_occupation_Working Professional	86.693937
Lead_Source_Welingak Website	80.407611

Recommendations

- Following variables contribute most to the probability of leads to converted so it is a good investment to put most

	Importance
Last_Notable_Activity_Had a Phone Conversation	100.000000
current_occupation_Working Professional	86.693937
Lead_Source_Welingak Website	80.407611

- X Education may target a lower value of threshold if the company adds extra staff to the pool.
- The company may choose a higher threshold value if the company want to be strict and tends to target those leads who are very likely to be converted.

THE END