Lead scoring case study-summary

Problem Statement:

➔ An education company named X Education sells online courses to industry professionals.
➔ The company markets its courses on several websites and search engines like Google. Once these people land on the website, they might browse the courses or fill up a form for the course or watch some videos. When these people fill up a form providing their email address or phone number, they are classified to be a lead.
➔ X Education gets a lot of leads, its lead conversion rate is very poor.
➔ To make this process more efficient, the company wishes to identify the most potential leads, also known as Hot leads

Objective:

➔ Build a logistic regression model to assign a lead score between 0 and 100 to each of the leads which can be used by the company to target potential leads. A higher score would mean that the lead is hot, i.e. is most likely to convert whereas a lower score would mean that the lead is cold and will mostly not get converted.
➔ There are some more problems presented by the company which your model should be able to adjust to if the company's requirement changes in the future so you will need to handle these as well. These problems are provided in a separate doc file. Please fill it based on the logistic regression model you got in the first step. Also, make sure you include this in your final PPT where you'll make recommendations.

Important steps in case study:
➔ Data understanding
➔ Data cleaning
➔ Exploratory data analytics
➔ Data preparation(dummy variables creation, scaling)
➔ Test-Train split
➔ Model building
   ◆ RFE variable selection
   ◆ Fitting the model and check P value and VIF
➔ Model evaluation
   ◆ Accuracy
   ◆ Sensitivity
   ◆ Specificity
   ◆ Precision
   ◆ Recall
➔ Predicting for Test set
➔ Model evaluation for test set

From the model we can conclude that these are the important parameters as per our model.
- ➔ TotalVisits
- ➔ Total Time Spent on Website
- ➔ Lead Origin_Lead Add Form
- ➔ Lead Origin_Lead Import
- ➔ Lead Source_Welingak Website
- ➔ Do Not Email_Yes
- ➔ Last Activity_Olark Chat Conversation
- ➔ Last Activity_SMS Sent
- ➔ Country_Not provided
- ➔ What is your current occupation_Working Professional
- ➔ What matters most to you in choosing a course_Not provided
- ➔ Last Notable Activity_Had a Phone Conversation
- ➔ Last Notable Activity_Unreachable

From the model evaluation below are the results we got:

- ➔ Accuracy score 80%
- ➔ Sensitivity as 78%
- ➔ Specificity as 81%
- ➔ Precision as 73%
- ➔ Recall as 78%

We can improve the process of conversion by making some changes in our method, like we can use more email communication, can increase more advertisement in Welingak website, target more working professionals, and give importance to people who try to reach by phone, we can also increase our call support to have a smooth call support.

Track total visits, time spent on the website for all people and try to contact them regularly to convert them, also have some more approaches like offers, coupons, spot discount, combo packages, to attract more customers to convert.