Summary Report

An education company, X Education which sells online courses to industry professionals. The company advertises itself on many websites or search engines. Any professionals who find the course interesting lands on their websites and log in and shares their personal details such as email address or contact details and we have considered them as leads. Once their contact details have been saved then the sales department contact them and convince them to take up the course and we refer them as 'Hot leads'. The conversion to hot leads from leads is very low that is around 30 % .

The problem statements are :

- X Education gets a lot of leads but 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'

- If they successfully identify this set of leads, the lead conversion rate should go up as the sales team will now be focusing more on communicating with the potential leads rather than making calls to everyone.

The proposed solution is to classify the leads which is more likely to convert into a hot lead by building the model so that the sales team can focus more on these selected small group of leads and then the conversion rate increases from 30 percent to around 80 percent.

After cleaning the data and categorizing the less important data ion the other's column for each perspective field we can come to a conclusion on a few variables where we can focus more.

Based on the univariate analysis we have seen that many columns are not adding any information to the model, hence we can drop them for further analysis.

Upon creating a stable model we get our parameters as:

Overall accuracy of the model: 90.45

Sensitivity of our logistic regression model: 84.42

Specificity of our logistic regression model: 93.88