LEAD SCORING CASE STUDY

**Expectations**

The following case study is primarily intended to facilitate a discussion on modeling choices and to have a tangible problem to discuss. As a guideline, you should expect to spend around 4 hours to complete this exercise. The assignment does not have to be completed all at once.

**Once completed, submit your deliverables to HR team coordinator**

**Problem Description**

An education company named X Education sells online courses to industry professionals. On any given day, many professionals who are interested in the courses land on their website and browse for courses.

The company markets its courses on several websites and search engines like Google. Once 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. Additionally, the company also gets leads through referrals. Once these leads are acquired, employees from the sales team start making calls, writing emails, etc. Through this process, some of the leads get converted while most do not. The typical lead conversion rate at X Education is around 30%.

There are a lot of leads generated in the initial stage, but only a few of them come out as paying customers. In the middle stage, the company nurtures the potential leads (i.e. educating the leads about the product, constantly communicating, etc. ) in order to get a higher lead conversion.

X Education has appointed you to help them select the most promising leads for follow-up, i.e. the leads that are most likely to convert into paying customers.

**The Task**

Build a model to assign a score between 0 and 100 to leads which can be used by the company to target potential leads. A higher score would indicate 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.

**Data Supplied:**

Leads Data Dictionary.xlsx:  
Leads.csv

**Deliverables:**

• The code that you wrote to solve the problem. (R or Python is preferred)

• Anything that helps a reader understand how well the model will work (if needed)

• Answers to the questions below.

Questions: Please answer the following questions.

1. How long did it take you to solve the problem?

2. What software language and libraries did you use to solve the problem?

3. What steps did you take to prepare the data for the project? Was any cleaning necessary?

4. What algorithmic method did you apply? Why? What other methods did you consider?

5. What features did you use? Why?

6. How did you train your model? During training, what issues concerned you?

7. How did you assess the accuracy of your predictions? Why did you choose that method? Would you consider any alternative approaches for assessing accuracy?

8. Which features had the greatest impact? How did you identify these to be most significant?