

### **DSM2022 Booth Camp Project Guidelines**

This is an open-ended project which tests the student's ability to solve the industrial-level problem from the ground up. It will be done in a group of five (5) students. Each student will show his/her contribution to the overall project. Upon successful completion of the project and other grade components, you will be awarded your certificate for participation in this Booth Camp.

### **Expectations**

At the end of the project, the student should be able to

- Understand the meaning and application of data science concepts
- Effectively apply data analysis tools
- Author a proficient industrial-level report
- Demonstrate the ability to tell insightful stories from data
- Horn-in on team-building skills
- Demonstrate and improve your presentation skills
- Proficiently use all the tools during the Booth Camp

Your group can choose from any of these two major data headings.

1. Structured Data- This type of data involves a table of rows and columns. You can develop Regression or Classification models for the following areas:
  - Finance: examples stock data
  - COVID-19
  - Health Data
  - Oil and Petroleum Data
  - Other structured data sources
  
2. Unstructured Data- This involves data that is not a table of rows and columns. Examples include images, text, speech, video etc. You can develop models in:
  - Natural Language Processing
  - Image detection
  - Object detection

Please note that the project is open-ended and so you own the direction of the project. We will give some data sources for these two data headings.

### **Examples of Projects:**

- Breast cancer dataset for logistic regression
- MNIST handwritten data set for image recognition
- Boston House Pricing Prediction
- Stock Price Prediction
- Predicting the dynamics of COVID-19 Infection in Ghana or Nigeria

### **Sources of Data**

1. [https://www.who.int/emergencies/diseases/novel-coronavirus-2019?adgroupsurvey={adgroupsurvey}&gclid=CjwKCAjws--ZBhAXEiwAv-RNL0fVOVA5dkbt\\_DevcE5v-yW2bQynwc4vP3Po9xc\\_I5riv9MQ0EtOJR0CHtAQAvD\\_BwE](https://www.who.int/emergencies/diseases/novel-coronavirus-2019?adgroupsurvey={adgroupsurvey}&gclid=CjwKCAjws--ZBhAXEiwAv-RNL0fVOVA5dkbt_DevcE5v-yW2bQynwc4vP3Po9xc_I5riv9MQ0EtOJR0CHtAQAvD_BwE)
2. <https://archive.ics.uci.edu/ml/datasets.php>
3. <https://finance.yahoo.com/quotes/OCR,dataset/view/v1/>

