Assignment: Data Science Scenarios

In this first assignment, we ask you to think about some real-life process or problem in different problem domains, and how you can improve it using the Data Science process. Think about the following:

1. Which data can you collect?
2. How would you collect it?
3. How would you store the data? How large the data is likely to be?
4. Which insights you might be able to get from this data? Which decisions we would be able to take based on the data?

Try to think about 3 different problems/processes and describe each of the points above for each problem domain.

Solution

Problem1: Highway Engineering

I would like to think about a problem I described on my Relational Databases course which is Highway Traffic Management System.

1. Which data can you collect?

   Answer : some of the features that we need to collect are:

   - Vehicle Speed - Lane Width -Number of Lanes
   - Number of Vehicles per Segment - Weather Conditions
   - Time Series for Vehicles in previous Data
2. How would you collect it?

   Answer: I believe we need IoT system to collect live data from the highway

3. How would you store the data? How large the data is likely to be?

   The Data is considered to be Big Data, because it has many different features that to be collected in 24/7, and the type of the data can defer from structured like the data being collected from the sensors up to unstructured ones like the images collected by the cameras are trained to captured desired pretrained classes.

About storing the data , I have no idea such that I need to search exactly .

Click here for some useful external resources

4. Which insights you might be able to get from this data? Which decisions we would be able to take based on the data?

   Many things :

   1- Monitor and Control the current conditions of the highway elements (vehicles, accidents, required maintenance , and more)
   2- Ensure safety in terms of the driving roles , such as putting the seat belt, handling phone, drink or eat during driving.
   3- Warning the far drivers if a vehicle being shut down for any reason in the mid of the highway, so the screens that are hanged above the highway show warning messages to slow down for a specific speed that is recommended by the calculations of the sensors.
   4- The project is related to computer vision, highways engineer, and weather conditions fields. Also, it is indirectly related to social analysis from the side of the people ethics on driving, I believe that driving is a one reflection to the community.

Problem 2: E-Learning

Assume that I am thinking for an entrepreneurial project about E-Learning for different disciplines in the undergraduate level course.

1. Which data can you collect?

   - Text data about academia resources and perform some text mining
   - Number of students in each university in the level of country
   - Surveys for the idea acceptance among a sample study between students. Also, we include on it some information about the internet availability, speed, subscription info, dedicated hours

2. How would you collect it?

   - By conducting the surveys

3. How would you store the data? How large the data is likely to be?

   I believe that the data is not too much big, most likely would be structured

4. Which insights you might be able to get from this data? Which decisions we would be able to take based on the data?

   - Sentiment analysis
   - Educational interests among the students
   - Supporting decision making on the proper time to prepare the materials (Time management and cost control)

Problem 2: Vaccination Control During Pandemic

We all know how the world handled the COVID-19 pandemic during2019-2020

1. Which data can you collect?

   - GPS Location for each registered device
   - Demographics information
   - The Governmental daily reports and reflecting it on demographics

2. How would you collect it?

   - Creating a formal application monitored and updated by the government for each citizen health status

3. How would you store the data? How large the data is likely to be?

   Using the governmental servers, I think we need to ensure that is ready and the infrastructure required is ready

4. Which insights you might be able to get from this data? Which decisions we would be able to take based on the data?

   - The public safety through reflecting the data results I application
   - The personal safety by recommending the citizen to the allowed areas can be visited if they are having negative result
   - If the citizen has positive result, they would have steps to deal with that based on the updated data on the application