How Data Science is Used in Healthcare Field

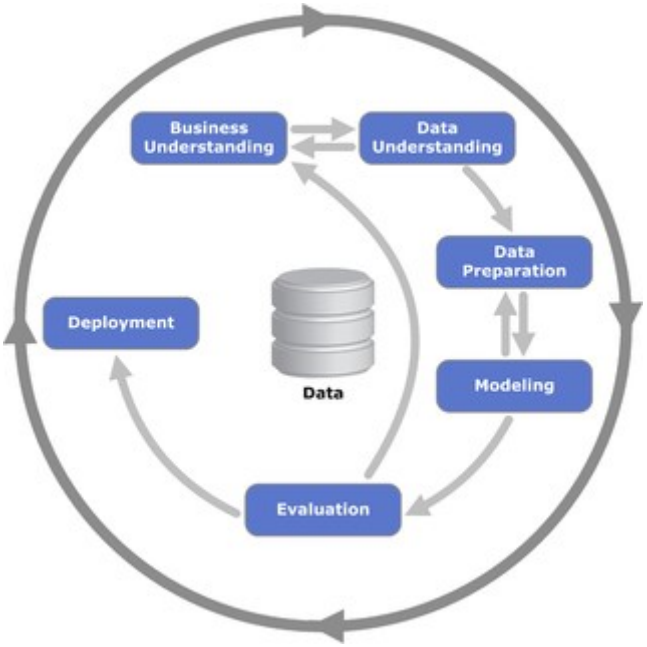
Healthcare is one of the most vital parts of human society. It involves medical practitioners, like doctors and/or specialised doctors, nurses, etc, that treat the patient with the help of various medical instruments/tools. These medical tools are usually tools that are used to help medical practitioners to measure something from the patients. From the measurement, the doctor then can decide what are the next procedures to treat the patients. Obviously, the results from the medical tools are data in various forms, whether in image, text, etc. This is where data science comes in handy.

Data science is basically an interdisciplinary study that combines maths and statistics, computer science, and domain experts to extract information from data in various forms that can be used as an insight into making a decision. Maths, statistics, and computer science are used to gather, clean, analyse, and modelling of the data that with the help of domain expertise, which is the knowledge of a particular field from where the data from/will be applied to, will create an insight that can help making any decision in that particular field.

One of the fields where data science can be used in healthcare is in the processing of the data gained from the medical tools that measure something, such as MRI, X-Ray, etc. We can use machine learning models to detect any anomalies from the data that are gained from the tools or to predict the disease based on the data.

CRoss Industry Standard Process for Data Mining (CRISP-DM) is a standard data science methodology that is widely used by data scientists everywhere. It has 6 sequential phases, which are:

1. Business understanding
2. Data understanding
3. Data preparation
4. Modelling
5. Evaluation
6. Deployment



The CRISP-DM cycle for data processing in healthcare system are as follow:

1. Business Understanding

   In this phase, the goal is to have a business understanding of where the data will be used. This is the so-called domain expertise aspect of data science. What we have to do here is to:

   a. Determine business objectives. Here is the part where we have to determine what the objective of the system is in this business. The objective of the system is to make an analysis and prediction from any of the data from the medical instruments.
   b. Situation assessment. This part is to assess the situation, for example how will the data be gained, how the transfer of the data, etc. We will take the data from each medical tool and make a unified system to transfer the data.
   c. Producing project plan. This part is to create the project, like what model will be used for the predictor of each data from each instrument.

2. Data Understanding.

   This part is the part to understand the data. This part is also the part to collect/gather the data. The way to understand the data is by doing an Exploratory Data Analysis (EDA). By performing EDA, we can familiarise ourselves with the data, from what types of the data, the visualisation of the data, etc.

3. Data Preparation

This part is usually the longest/most time-consuming phase of the process as it involves data pre-processing, data cleaning, and also the handling of the missing data. Because we will be dealing with medical patients data, which is protected by the law, we have to do some approval procedure from the patients to get their approval of the data to be used in this process. After that, we can use the data to prepare it, pre-process, clean, and handle the missing value in the data.

4. Modelling

   After we have the data ready, the next step is to create a model to do the various objectives defined in the first phase. The modelling process will involve a lot of trial-and-error. Because we will be making models for tasks, detection and prediction, we need to use a suitable algorithm for each task and also for each tool that is used.

5. Evaluation

   Basically evaluation and modelling is an inseparable part because they work together. After each model is made, we have to evaluate whether or not the model performed the way we wanted to or not.

6. Deployment

   The last part is to deploy the system into the healthcare system.

Github: https://github.com/OkaDarmayasa/ZEN_AML_Topic-1-2

References

Data Science. (n.d.). Exasol. Retrieved September 19, 2022, from https://www.exasol.com/glossary-term/data-science-definition/

Hotz, N. (2022, August 8). *What is CRISP DM?* Data Science Process Alliance. Retrieved September 20, 2022, from https://www.datascience-pm.com/crisp-dm-2/