

In21-S4-CS3121 - Introduction to Data Science

Activity 2 - Data Science Process In Class Activity

Group K

Data Science Process

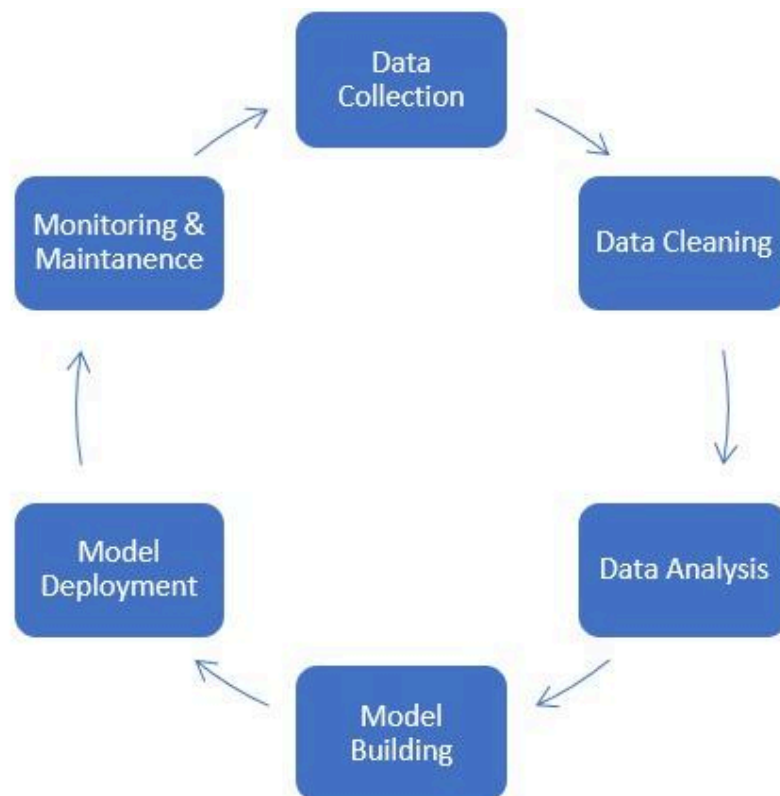
What is Data Science?

The modern world is data-driven. Every major decision in almost every field is highly dependent on data and information. But raw data doesn't tell anything. This is where Data Science comes in and analyses this raw data to gain insights that would assist in decision making process.

What is Data Science Process?

To gain insightful information from raw data, some key steps have to be performed. These steps are called the Data Science process.

We can show this process in this way,



Data Collection

This step includes obtaining data from various sources, including databases, surveys, Web scraping etc. But getting data is not sufficient because data is not in the right form. Any useful insights cannot be obtained from this raw data.

Data Cleaning

Here, the collected data is cleaned. It includes deciding what to do with missing values in some data, how to structure data for the required task.

Data Analysis

We try to visualise data with different methods and try to find any patterns in the data. Identifying these patterns assists in developing a model to extract useful information from our data.

Model Building

This is a sector that involves statistics and many mathematical theories to determine how we can simulate patterns that we observed in the obtained data. Then we implement and perform tests upon the model built and fine-tune it to increase accuracy of the model.

Model Deployment

Here we deploy our application to be used in real world data. Also We have to keep Monitoring and maintain the system for upcoming new data.

Monitoring and maintenance

This involves continuously tracking the performance of deployed models and making necessary updates or improvements based on new data and feedback. This step ensures that the model remains effective and relevant over time, reflecting changes in the underlying data and business requirements.

Conclusion

An approach for resolving complicated issues and deriving value from data is the data science process. In order to comprehend the issue, analyse data, create models, and provide stakeholders with useful insights, each step is essential. Organisations may successfully use data to innovate and accomplish their goals by adhering to this method.

Group Members

210518H Ranasinghe K.S
210450P Pathirana L.P.T.R
210483T Prabashwara D.G.H
210588U Senarathna L.P.S.U.K
210460V Perera I.T.M.