Data Science Life Cycle

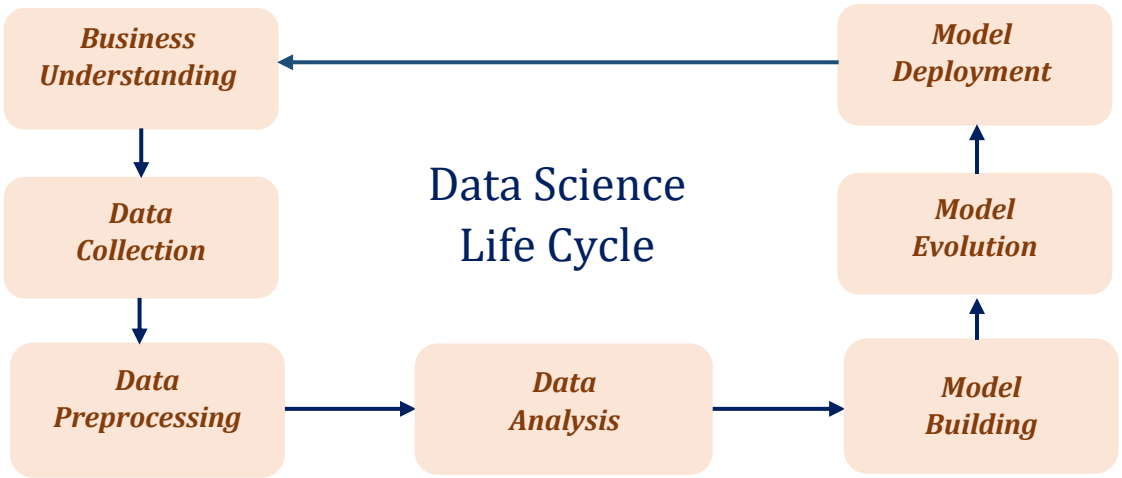
Learning Topics

- ✓ What is data science life cycle

1. What is Data Science Life Cycle:

Data science life cycle is a data science project flow planning that involves different types of steps in priority order.

Data science life cycle is the first step in every data science projects. Not only for data science, any type of software industry's projects starts with project life cycle building by technical experts. It explains the different types of stages involved in data science project. Let's start learning about all stages one by one.



1.1. Business Understanding:

Understanding the business is one of the important role in data science project. The business requirement and achieving goals are plays vital role in the project. In this stage, we collect all required information, data related to business and asking quarries about the data. Once all

business-related information received then we will move to next step called data collection.

1.2. Data Collection:

Data collection is the second step in data-science life cycles. In this stage we collect all required data related to business requirement. This collected data is called **raw data**. Because this data coming from different resources and in different formats. This all data we can't use directly in ML models, because it contains many unsupportable formats and data impurities.

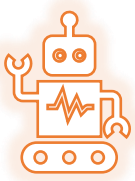
1.3. Data Preprocessing:

Next comes the data preprocessing stage. Right now, many preprocessing pipeline projects are going in the market. This consist of data cleaning, understand the outliers and unwanted data cleaning, removing useless columns, feature selection for correct training, feature encoding, feature scaling etc... The preprocessing is very important stage for all ML models training. A perfectly cleaned data make the perfect model. 90% of work in all ml projects is data preprocessing only remaining 10% is model building.

1.4. Data Analysis:

This concept includes the explanation of data, understanding the characteristics of data by presenting in graphical and numerical representation. The data analysis helps to select correct model. Analyzing information provides the hidden knowledge in the data and business-related information also.

1.5. Model Building:



Will select a suitable ML model based on the data analysis report. Once the model got selected, then will provide the training data to the model for training purpose.

1.6. Model Evolution:

Model evolution is nothing but testing the model. Once model got trained the next step is testing how accurate model got generated. If model gives expected accuracy, then model is ready for deployment.

1.7. Model Deployment:

Model deployment is nothing but Providing availability of the model for prediction future purpose.