

## Assignment 1:

1. Explain the use case and type of analytics and statistics involved in data analysis.

⇒ The types of analytics are:

### a. Descriptive analytics:

The descriptive analytics helps to analyze the past data by the means of various statistical methods. In this process the answer of the question "What happened?" is understood clearly.

Use case:

The hospital uses descriptive analytics to find the total patients who suffered from common cold or fever.

### b. Diagnostic analytics

Diagnostic analytics is used to analyze the cause of past event in more detail. It answers the question "Why did it happen?"

Use case:

Descriptive analytics helps the hospital or doctor to find the core reason behind the people suffering from common cold or fever. Eg: High consumption of ice-cream, cold-drinks ... etc.



### Bc. Predictive Analytics

Predictive analytics uses historical data and machine learning to forecast the future events or outcomes. It answers the question "What is likely to happen in future?"

use case

By analyzing the historical data the doctors estimates the possible number of patients who might come to visit them after suffering from common cold.

### d. Prescriptive analytics

The predictive analytics combines all descriptive, diagnostic and predictive analytics to recommend the best possible action or decision. It answers the question "What should we do to get the best result?"

use case:

The hospital might run an awareness program or health events to aware the people about the various diseases.



2. Compare and contrast the trade-off between Exploration and Exploitation in Reinforcement Learning.

- In Reinforcement Learning (RL) the trade off between exploration and exploitation is a fundamental concept that significantly affects the learning process and performance of an agent.

a. Exploration

Exploration is the process of performing new actions to discover their effect and gather more information about the environment. This helps to discover new strategies and improve future decision.

Key characteristics:

1. Riskier :- Involve taking action that may lead to unknown result.
2. Long-term - Learning :- Helps the user discover better actions.

3. Short-term - sacrifice :- It might lead to lower immediate reward.

Example:

The person might explore different path to reach to point "B" starting from point "A".



b. Exploitation agent

Exploitation involves using the ~~user's~~ current knowledge to choose actions that maximize the outcome based on the past experience. It is about using the best-known strategy at the current moment.

Key characteristics:

1. Less risky :- The agent selects action it already knows are rewarding

2. Efficiency : It is more efficient for short terms.

3. Miss-better-strategy :- The agent might miss the best strategy and may become less productive.

Q.No.3. What are the basic workflows / process in a machine learning project.

The process of ML project is given below:

1. Problem definition :

- clearly define the problem you want to solve.



2. Data Collection :

Collect the data required to train the model. The data can be collected from database or public dataset.

3. Data preprocessing:

Clean and prepare the data for analysis. This step contains handling the missing values, converting data types, encoding and many more.

4. Exploratory Data Analysis (EDA)

EDA is done to understand the data patterns and get more insights from into the dataset.

5. Feature engineering

Create or Identify the features that may increase the model performance.

6. Model selection

Select the best algorithm / model according to the problem type and data characteristics.

7. Model Training

Train the selected model using the training data from the dataset. perform hyperparameter tuning to increase model performance.



### 8. Model Evaluation

Evaluate the performance of the trained model by using the metrics like  $R^2$ , RMSE, F1 Score and others.

### 9. Model Deployment

Deploy the trained model to a production environment.

### 10. Monitoring and maintenance

Monitor the model performance overtime, update the model with new data.