

## Practice Assignment – Cognitive Application

### Theory Questions (Ayush Goyal, Manipal Institute of Technology, CSE, (9874787715))

#### 1. Differentiate between Rule-Based Approach and Machine Learning Approach?

Machine learning systems are probabilistic and rule-based AI models are deterministic. Machine learning systems constantly evolve, develop, and adapt its production in accordance with training information streams. Machine learning models utilize statistical rules rather than a deterministic approach.

The other major key difference between machine learning and rule-based systems is the project scale. Rule-based artificial intelligence developer models are not scalable. On the other hand, machine learning systems can be easily scaled.

Machine learning systems require more data as compared to rule-based models. Rule-based AI models can operate with simple basic information and data. However, machine learning systems require full demographic data details.

Rule-based artificial intelligence systems are immutable objects. On the other hand, machine learning models are mutable objects that enable enterprises to transform the data or value by utilizing mutable coding languages such as java.

#### 2. Explain the Lifecycle of an ML model?

The machine learning life cycle is the cyclical process that data science projects follow. It defines each step that an organization should follow to take advantage of machine learning and artificial intelligence (AI) to derive practical business value.

There are five major steps in the machine learning life cycle, all of which have equal importance and go in a specific order.

- a. **Define Project Objectives:** The first step of the life cycle is to identify an opportunity to tangibly improve operations, increase customer satisfaction, or otherwise create value.
- b. **Acquire and Explore Data:** The next step is to collect and prepare all of the relevant data for use in machine learning.
- c. **Model Data:** To gain insights from your data with machine learning, you must determine your target variable, the factor on which you wish to gain deeper understanding.
- d. **Interpret and Communicate:** One of the most difficult tasks of machine learning projects is explaining a model's outcomes to those without any data science background, particularly in highly regulated industries such as healthcare.
- e. **Implement, Document, and Maintain:** The final step is to implement, document, and maintain the data science project so that the hospital can continue to leverage and improve upon its models.

### **3. Differentiate between Supervised and Unsupervised learning? Mention real-time applications of each?**

In supervised learning, the algorithm “learns” from the training dataset by iteratively making predictions on the data and adjusting for the correct answer. While supervised learning models tend to be more accurate than unsupervised learning models, they require upfront human intervention to label the data appropriately. For example, a supervised learning model can predict how long your commute will be based on the time of day, weather conditions and so on. But first, you’ll have to train it to know that rainy weather extends the driving time.

Unsupervised learning models, in contrast, work on their own to discover the inherent structure of unlabelled data. Note that they still require some human intervention for validating output variables. For example, an unsupervised learning model can identify those online shoppers often purchase groups of products at the same time. However, a data analyst would need to validate that it makes sense for a recommendation engine to group baby clothes with an order of diapers, applesauce, and Sippy cups.

### **4. What do you mean by Clustering?**

Clustering is the task of dividing the population or data points into several groups such that data points in the same groups are more similar to other data points in the same group than those in other groups. In simple words, the aim is to segregate groups with similar traits and assign them into clusters.

Let’s understand this with an example. Suppose you are the head of a rental store and wish to understand preferences of your costumers to scale up your business. Is it possible for you to look at details of each costumer and devise a unique business strategy for each one of them? Definitely not. But what you can do is to cluster all your costumers into say 10 groups based on their purchasing habits and use a separate strategy for costumers in each of these 10 groups. And this is what we call clustering.

### **5. What is Regression? Explain with an example?**

Regression analysis is a statistical method that helps us to analyse and understand the relationship between two or more variables of interest. The process that is adapted to perform regression analysis helps to understand which factors are important, which factors can be ignored and how they are influencing each other. For the regression analysis to be a successful method, we understand the following terms:

**Dependent Variable:** This is the variable that we are trying to understand or forecast.

**Independent Variable:** These are factors that influence the analysis or target variable and provide us with information regarding the relationship of the variables with the target variable.

Let's understand the concept of regression with this example.

You are conducting a case-study on a set of college students to understand if students with high CGPA also get a high GRE score. Your first task would be to collect the details of all the students. We go ahead and collect the GRE scores and CGPAs of the students of this college. All the GRE scores are listed in one column and the CGPAs are listed in another column. Now, if we are supposed to understand the relationship between these two variables, we can draw a scatter plot. Here, we see that there's a linear relationship between CGPA and GRE score which means that as the CGPA increases, the GRE score also increases. This would also mean that a student who has a high CGPA, would also have a higher probability of getting a high GRE score. But what if I ask, "The CGPA of the student is 8.32, what will be the GRE score of the student?" This is where Regression comes in. If we are supposed to find the relationship between two variables, we can apply regression analysis.

In regression, we normally have one dependent variable and one or more independent variables. Here we try to "regress" the value of dependent variable "Y" with the help of the independent variables. In other words, we are trying to understand, how does the value of 'Y' change w.r.t change in 'X'.

## **6. How do we check the efficiency of an ML model? Mention the parameters.**

To evaluate the machine learning models, you will have to know the basic performance metrics of models. For example, accuracy, precision, recall, F1-score, or AUC values are important measures for classifiers. Following this, you are also to go through other supervised and unsupervised machine learning algorithms.

Evaluating your machine learning algorithm is an essential part of any project. Your model may give you satisfying results when evaluated using a metric say accuracy score but may give poor results when evaluated against other metrics such as logarithmic loss or any other such metric. Most of the times we use classification accuracy to measure the performance of our model, however it is not enough to truly judge our model.

Some parameters are: Classification Accuracy, Logarithmic Loss, Confusion Matrix, Area under Curve, F1 Score, Mean Absolute Error, Mean Squared Error.

**THE END**