ASSIGNMENT

1. **Differentiate between gits and github.**

Ans: **Git:** Git is a version control system that lets you manage and keep track of your source code history. GitHub is a cloud-based hosting service that lets you manage Git repositories.

2. **What is version control system ?**

Ans: Data Version Control (DVC) is one of best open source tools available in the market for Machine Learning Models and Dataset Versioning. It is Git-compatible, Reproducible and there's a tons of supported protocols for store the datasets. You can find more information here: https://dvc.org/features.

3. **What is a repositary ?**

Ans: In Git, the repository is like a data structure used by VCS to store metadata for a set of files and directories. It contains the collection of the files as well as the history of changes made to those files. Repository in Git is considered as your project folder. A repository has all the project-related data. Distinct projects have distinct repositories.. A Git repository tracks and saves the history of all changes made to the files in a Git project. It saves this data in a directory called **.git**, also known as the repository folder.

4. **Difference between local and remote repositaries.**

Ans: A **local repository** is one that already exists or is already stored on your machine. It can be a repository that you already cloned before or was cloned outside of SourceTree (ex. using git commands

on the command line). You don't clone local repositories, you just need to add their folders to the SourceTree interface, as indicated in the UI to "*drag & drop repository folders*" into SourceTree, so that they will appear in the UI.

A **remote repository** is one that exists somewhere else, in Github, Gitlab, Bitbucket, or in any other server that hosts Git repositories. This is the one you need to **clone** into your own machine, where it can now become a **local** copy of the **remote repository**.

5. **What is a branch how do you create one in github ?**

Ans: Branches allows to develop features, fix bugs, or safely experiment with new ideas in a contained area of the repository. You always create a branch from an existing branch. Typically, you might create a new branch from the default branch of your repository.

Steps to Create the branch in github

1. At the top of the app, click Current Branch and then in the list of branches, click the branch that you want to base your new branch on.
2. Click New Branch.
3. Under Name, type the name of the new branch.
4. Use the drop-down to choose a base branch for your new branch.
5. Click Create Branch.

6. **What is github desktop ?**

Ans: GitHub Desktop is an application that enables you to interact with GitHub using a GUI instead of the command line or a web browser. GitHub Desktop is an open source project. You can see the roadmap, contribute to the project, or open an issue to provide feedback or feature requests.

7. Difference between classification and regression.

Ans:

Classification Algorithm	Regression Algorithm
The mapping function is used for assigning values to predefined groups.	The mapping function is used for the assignment of values to continuous output.
In Classification, the output element must be a discrete attribute.	In Regression, the output element must be of the constant type of real value.
The role of the classification algorithm is to map the input value(x) with the discrete output variable(y).	The role of the regression algorithm is to map the continuous output variable(y) with the input value (x).
Classification Algorithms are used for discrete data.	Regression Algorithms are used for continuous data.
In Classification, we strive to locate the judgment limit, which may split the dataset into different classes.	In Regression, we strive to find the best match rows, which can forecast the performance more accurately.
Classification Algorithms may be used to solve classification problems such as Voice Recognition, Identification of spam emails, Identification of cancer cells, etc.	Regression algorithms may be used to solve the regression problems such as House price prediction, Weather Prediction, etc.
The Classification algorithms can be classified into Multi-class Classifier and Binary Classifier.	The regression Algorithm can be further separated into Non-linear and Linear Regression.

7. Write a short note on Nominal data.

Ans:
Nominal data: Nominal data (also known as nominal scale) is a type of data that is used to label variables without providing any quantitative value. It is the simplest form of a scale of measure. Unlike ordinal data, nominal data cannot be ordered and cannot be measured.
Nominal data can be both qualitative and quantitative. However, the quantitative labels lack a numerical value or relationship (e.g.,

identification number). On the other hand, various types of qualitative data can be represented in nominal form. They may include words, letters, and symbols. Names of people, gender, and nationality are just a few of the most common examples of nominal data.

8. Write a short note on Categorial data.

Ans:

Categorial data: Categorical data is a collection of information that is divided into groups. I.e, if an organisation or agency is trying to get a biodata of its employees, the resulting data is referred to as categorical. This data is called categorical because it may be grouped according to the variables present in the biodata such as sex, state of residence, etc.

Categorical data can take on numerical values (such as "1" indicating Yes and "2" indicating No), but those numbers don't have mathematical meaning. One can neither add them together nor subtract them from each other.

9. What is encoding in machine learning.

Ans: Encoding is a technique of converting categorical variables into numerical values so that it could be easily fitted to a machine learning model. It is the process of converting the data or a given sequence of characters, symbols, alphabets etc., into a specified format, for the secured transmission of data. Decoding is the reverse process of encoding which is to extract the information from the converted format.

The purpose of encoding is to transform data so that it can be properly (and safely) consumed by a different type of system, e.g.

binary data being sent over email, or viewing special characters on a web page. The goal is not to keep information secret, but rather to ensure that it's able to be properly consumed.

10. Explain OHE (One Hot Encoding).

Ans: One hot encoding is the most widespread approach, and it works very well unless your categorical variable takes on a large number of values (i.e. you generally won't it for variables taking more than 15 different values. It'd be a poor choice in some cases with fewer values, though that varies.)

One hot encoding creates new (binary) columns, indicating the presence of each possible value from the original data.

One-hot encoding converts the categorical data into numeric data by splitting the column into multiple columns. The numbers are replaced by 1s and 0s, depending on which column has what value

One Hot encoder is used when:

- When the order does not matter in categorical features
- Categories in a feature are fewer.