

Assignment 3

1. What is DVC, and why is DVC used?

DVC stands for data version control. dvc file is a small text file that points to your actual data files in remote storage. The .dvc file is lightweight and meant to be stored with your code in GitHub. When you download a Git repository, you also get the.

DVC is built to make ML models shareable and reproducible. It is designed to handle large files, data sets, machine learning models, and metrics as well as code.

2. How is DVC different from git and GitHub?

While Git is used to store and version code, DVC does the same for data and model files. Git can store code locally and also on a hosting service like GitHub, Bitbucket, or GitLab.

DVC: Open-source Version Control System for Machine Learning Projects. It is an open-source Version Control System for data science and machine learning projects. It is designed to handle large files, data sets, machine learning models, and metrics as well as code.

3. Which command can be used to initialise a DVC project?

To initialize DVC command is => **dvc init**

4. In What all use cases DVC can be used?

If you store and process data files or datasets to produce other data or machine learning models, and you want to

- track and save data and machine learning models the same way you capture code;
- create and switch between versions of data and ML models easily;
- understand how datasets and ML artifacts were built in the first place;
- compare model metrics among experiments;
- adopt engineering tools and best practices in data science projects.

5. Which command can be used to reproduce the entire pipeline?

Command to reproduce pipeline => **dvc repro**

6. Which DVC command can be used to check metrics?

Below are the 2 commands used for dvc metrics :

dvc metrics show

dvc metrics diff

7. Can we store a large amount of Data on GitHub? Justify.

No, we cannot added large data files in github. Files that you add to a repository via a browser are limited to 25 MB per file. You can add larger files, up to 100 MB each, via the command line.