**AIOPS Assignment 3**

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

*It is an open source version control system for machine learning projects.*

*In ML Projects, we may have multiple versions of data for different model iterations. We have to track all the model and data changes. DVC helps in governing data. DVC runs alongside Git.*

1. How is DVC different from git and GitHub?

*Git and Github store different versions of the code, DVC stores different versions of data and model files.*

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

*dvc init*

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

*DVC can be used when we have to experiment the performance of model with different versions of data. Eg: When you have an one hot encoded data for categorical variables as one version and frequency encoded variables as another version, we will be able to compare the model performances with the two datasets and the different model versions and pick the best performing model for production.*

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

*dvc repro --pipeline*

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

*dvc metrics show*

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

*Github blocks files larger than 100 MB. This is where DVC can be used when we have large datasets, DVC uses a remote repository to store all the data and models. Using a pointer to the remote location, the actual data can be accessed.*

