

Data Science

Data Science:

- field of study that uses mathematics, statistics, programming and domain knowledge to extract meaningful insights from data.
- Data Science is the process of collecting, cleaning, analyzing, and interpreting data to make informed decisions.

Key Steps in Data Science:

1. Data Collection
2. Data Cleaning
3. Data Analysis
4. Model Building
5. Interpretation and Communication

Data Science Lifecycle:

1. Problem Definition
2. Data Collection
3. Data Cleaning (Data Preprocessing)
4. Data Exploration (EDA - Exploratory Data Analysis)
5. Model Building
6. Model Evaluation
7. Deployment
8. Communication & Reporting
9. Maintenance & Iteration

Data Science tools -

1. Jupyter Notebook (with Anaconda Distribution)

- jupyter notebook

2. Google Colab

3. VS Code

4. Pycharm

5. Cursor AI

Intro to Miniconda & Anaconda -

- Miniconda - Lightweight version of Anaconda
- Anaconda - full distribution with hundreds of packages for data science and machine learning

{ Conda list
Conda install Numpy

Conda workflow -

- Check available Environment - `conda env list`
- Create new environment - `conda create -n gemhavenv`
- `conda create -n envname python=3.11`
- Activate environment - `conda activate envname`
- Install Package - `conda install -C conda-forge numpy`
channel maintain channel with cutting-edge packages
(a source for package)
- `conda install jupyter`
- Deactivate environment - `conda deactivate -`

Anaconda Navigator -

- Graphical user interface (GUI) that comes with Anaconda distribution

Jupyterlab & Jupyter notebook -

• Jupyter Notebook -

- Simple interactive environment to write and run code
- Each document is saved as a .ipynb file.
- Cell in Jupyter Notebook is a container where you can write and run code, markdown, raw text.
- install : `conda install jupyter`
- open : `jupyter notebook`

• Jupyter lab -

- Multi-Panel Interface
- Better Navigation
- Extensions
- Integrated tools
- install : `conda install jupyterlab`
- open : `jupyter lab`