PR W11 SK5004 Machine Learning and Artificial Intelligence

Naura Kasihna Risat (10322037)

> **Problem 1:**
>
> Explain in brief about data science.

Answer: Data science is a scientific process aimed at transforming raw data into meaningful information that can support smarter decision-making. It involves collecting data, analysing patterns, building models, and communicating results. Data science is not a stand-alone field, it draws on other areas such as mathematics, programming, and domain expertise to be truly effective.

> **Problem 2:**
>
> What are the differences between data, data science, and data scientist?

Answer:

- **Data** refers to raw facts (numbers, text, images, or any information) that has not yet been processed or interpreted.
- **Data science** is the structured process of using tools, techniques, and theory to extract insight and value from that raw data. It includes stages like data collection, cleaning, analysis, modeling, and communication.
- **Data scientist** is the person who performs this work. Their role includes gathering relevant data, preparing it for analysis, applying statistical and machine learning methods, interpreting the results, and clearly communicating their findings to decision-makers.

> **Problem 3:**
>
> Explain about the four foundational aspects of data science?

Answer:

- Mathematics : Provides the logical structure behind data analysis, including concepts like functions, relationships, and abstract reasoning. These help build and understand models used in data science.
- Technology : Includes tools and programming languages such as Python, Jupyter Notebook, and GitHub. These allow data scientists to process and manage data efficiently.
- Visualisation : Focuses on representing data in visual formats like plots or dashboards, making it easier to understand and communicate
- Communication : Essential for presenting findings to others. This includes writing code with good documentation, explaining results in notebooks, and creating concise and effective reports tailored to the target audience.

Problem 4:

List link on PyPI for installing Jupyter Notebook, Matplotlib, NumPy.

Answer:

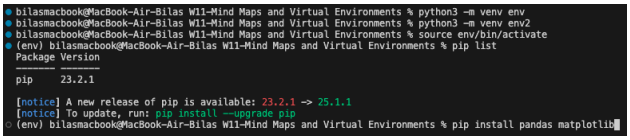
- **Jupyter Notebook**: https://pypi.org/project/notebook/
- **Matplotlib**: https://pypi.org/project/matplotlib/
- **NumPy**: https://pypi.org/project/numpy/

Problem 5:

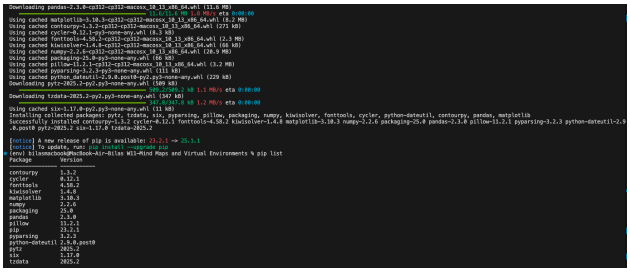
Create a virtual environment, install some packages, and save information to requirements.txt, create other virtual environment and use requirement.txt. Show the screenshots for all processes.

Answer:

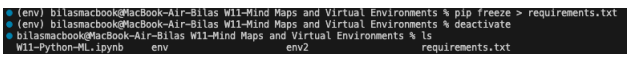
- Create two virtual environments and install packages in the first one.



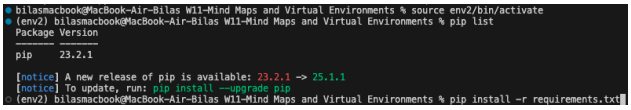
- Packages are successfully downloaded.



- Save the installed packages to a `requirements.txt` file.



- Exit the first environment and activate the second one.



- Install the packages from `requirements.txt` into the second environment successfully.