COS40007 Artificial Intelligence for Engineering

Week 1 Studio Activities

ILO	Understand type of data in Engineering process.
Aim	Introduction to python for ML and DL Installation and setup Use some real-world engineering process data for preliminary analysis and exploration
Resources	Books: 1. Prosise, Jeff. Applied machine learning and AI for engineers. " O'Reilly Media, Inc.", 2022. 2. Raschka, Sebastian, Yuxi Hayden Liu, and Vahid Mirjalili. Machine Learning with PyTorch and Scikit-Learn: Develop machine learning and deep learning models with Python. Packt Publishing Ltd, 2022. Web Resources: 1. https://www.geeksforgeeks.org/machine-learning-with-python/
Requirements for submission to be marked as complete	Demonstrate and explain the outcome of your Exploratory data analysis (EDA) to your Tutor

Previous experience in Machine Learning (ML) and Artificial Intelligence (AI)

There are range of engineering datasets that require AI models to make intelligent decisions. The purpose of AI models is to make predictions, provide recommendations, identify anomalies and many more. Let's spend 15 minutes to discuss your previous experience in Artificial Intelligence and Machine Learning. Your tutor will organise to form a group of 3 and discuss among yourself about the following topics:

- What engineer background you have? Did you have any previous experience of utilising Artificial Intelligence/Machine Learning solution for any engineering datasets? If yes, then what type of data your used? What have you explored? How you did the data analysis? What tools you used to develop your AI/ML models?
- Are you aware of any engineering datasets that you want to use to develop your AI model? If yes, then what model you plan to explore? What is the reason of selecting such a model?

- Have you used python or any other tool to develop AI model? What sort of challenges you faced to build such AI model

Introduction to Machine Learning and Artificial Intelligence

- Machine Learning is making the computer learn from studying data and statistics
- Machine Learning is a step into the direction of Artificial Intelligence (AI)
- Machine Learning is a program that analyses data and learns to predict the outcome.

Overall, Machine Learning is a subdomain of artificial intelligence. It allows computers to learn and improve from experience without being explicitly programmed by programmers, and It is designed in such a way that allows systems to identify patterns, make predictions, and make decisions based on data.

Machine Learning with Python

Python, a versatile programming language, has become a good-to-go choice for all to start with, and it helps many machine learning enthusiasts due to Pythons simplicity, a vast collection of libraries, and a large number of applications. It is interpreted programming language known for its readability and clear syntax. It provides various libraries and frameworks that simplify machine learning development. Python's versatility and active community make it an ideal language for machine-learning projects and supports object-oriented programming, most commonly used to perform general-purpose programming. Python is used in several domains like Data Science, Machine Learning, Deep Learning, Artificial Intelligence, Networking, Game Development, Web Development, Web Scraping, and various other domains.

Python has a crucial role in machine learning because Python provides libraries like NumPy, Pandas, Scikit-learn, TensorFlow, and Keras. These libraries offer tools and functions essential for data manipulation, analysis, and building machine learning models. It is well-known for its readability and offers platform independence. These all things make it the perfect language of choice for Machine Learning.

Studio Activity 1: Setting Up Python for Machine Learning (30 minutes)

Step 1: Install Python and Required Libraries

- Begin by installing Python on your system.
- You can download the latest version from the official Python website.
- It is good to create a virtual environment for your python version. This allows you to manage multiple python versions in your system. You can use this tutorial to create a virtual environment for your python version

- Now you'll need to install the required libraries for machine learning in the virtual environment you created. You can you pip or conda to install Numpy , pandas, Matplotlib, Scikit-learn, Tensorflow, Keras, Pytorch, OpenCV, Yolo V5, labelme

Step 2: Choose an Integrated Development Environment (IDE)

Select an IDE for writing and executing your Python code. Install the IDE on your machine. Some popular options include
Jupyter Notebook, PyCharm, and Visual Studio Code.

Data in engineering process

Here is sample dataset available online in various engineering disciplines that you can start to develop your AI model.

1. Combined cycle power plant in Electrical Engineering
2. Fault diagnosis problem of electromechanical device in Mechanical Engineering
3. Faulty Steel Plates in civil engineering
4. Water Quality in Water Engineering
5. Liquid Battery Electrolyte Formulation in Electronics
6. Battery and Heating data in real driving cycle in Energy
7. Breast cancer diagnosis in Biomedical engineering

Assumptions: In Activity-1 you already successfully installed python and required libraries and you already know how to write code in python.

Studio Activity-2: Pick a dataset closely relevant to your engineering discipline and conduct some initial data exploration using python pandas (30 minutes)

Understand your data

- Download your selected dataset
- Convert your data in .CSV format if data is not that format already
- Convert any categorical values to numerical values (e.g., 0,1,2,…)
- Data collection: Use pandas to read CSV data
- Understand your data (using pandas).
    - Check how many rows and columns are there
    - What are the columns names and their types

Data Cleaning
Performing data cleaning involves a systematic process to identify and rectify errors, inconsistencies, and inaccuracies in a dataset. Data cleaning process involves removing duplicates, checking for outliers, handling missing data. See section 4 of this link

- Remove duplicate entries from your data (use pandas drop_duplicates)
- Check if there are any outliers in the data and remove outliers. (use boxplots as provided an example here)
- Check outliers' correction after removing them
- Check if there are any missing values in your data

Studio Activity-3: Exploratory data analysis (EDA) (40 minutes)
Note: You can work on this later as homework if you unable to finish this in the studio

Use section 5 of this link as completing this task. This task provides an example of EDA for the cement manufacturing dataset. Perform a similar EDA on your selected dataset.

- Identify your target variable and Predicators.
- Conduct and univariate analysis for Predicators
- View a summary statistic of all the variables
- Conduct a Multivariate analysis
- Identify pairwise correlations among the variables and plot them in a heatmap

At this end of this activity Create a summary report of your EDA similar to what is described in Section 5 of this link. Prepare this report of your EDA before your next Studio session.