COS40007 Artificial Intelligence for Engineering

Week 1 Studio Activities

ILO	Understand the type of data used in the 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 the Exploratory Data Analysis (EDA) you conducted to your Tutor.

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

A range of engineering datasets require AI models to make intelligent decisions. AI models aim to make predictions, provide recommendations, identify anomalies, and more. Let us spend 15 minutes discussing your previous experience in Artificial Intelligence and Machine Learning. Your Tutor will organise to form a group of 3 and discuss the following topics:

- What engineering background do you have? Do you have any experience utilising artificial intelligence/machine learning solutions for engineering datasets? If yes, then what type of data you used? What have you explored? How did you do the data analysis? What tools did you use to develop your AI/ML models?
- Are you aware of any engineering datasets you want to use to develop your AI model? If yes, then what model do you plan to explore? What is the reason for selecting such a model?

- Have you used Python or any other tool to develop an AI model? What sort of challenges did you face in building such an AI model?

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

Overall, Machine Learning is a subdomain of artificial intelligence. It allows computers to learn and improve from experience without being explicitly programmed by programmers. It will enable 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 Python's simplicity, a vast collection of libraries, and many applications. It is an interpreted programming language known for its readability and straightforward 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. It supports object-oriented programming and is 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 is crucial in machine learning because It 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. Python is well-known for its readability and platform independence, making it the perfect language 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 must install the required library for machine learning in your created virtual environment. You can use pip or conda to install Numpy, pandas, Matplotlib, Scikit-learn, Tensorflow, Keras, Pytorch, OpenCV, and 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 the engineering process

Here is a sample dataset available online in various engineering disciplines from which you can start developing 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 successfully installed Python and the required libraries and 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)

Understanding data

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

Data Cleaning

Data cleaning involves systematically identifying and rectifying dataset errors, inconsistencies, and inaccuracies. It consists in removing duplicates, checking for outliers, and handling missing data. See section 4 of this link.

- Remove duplicate entries from the 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 the data

Studio Activity-3: Exploratory data analysis (EDA) (40 minutes)
Note: You can work on this later as homework if you cannot finish it in the studio.

Use section 5 of this link to complete 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 the 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.