Python for Materials Data Science
MSE 497
Instructors: Arun Mannodi Kanakkithodi, Michael Titus

Course Description
This course introduces Materials Science and Engineering students to Python programming for data analysis and visualization in materials research. Students will learn to install Python and essential scientific packages, import data from materials databases, and perform advanced data analysis. The curriculum covers creating publication-quality plots and interactive visualizations to effectively communicate research findings. Emphasis is placed on applying FAIR principles to data management and preparing data for export to standardized databases. By the end of the course, students will have developed practical skills in automating data workflows and implementing machine learning algorithms for materials property prediction. This hands-on course hosted on Github, with Jupyter notebooks accessed using Google Colab, prepares future materials scientists to leverage computational tools in their research and industry careers.

Learning Outcomes
1. Install and configure Python and relevant scientific packages (e.g., NumPy, Pandas, Matplotlib).

2. Demonstrate proficiency in basic Python programming concepts.

3. Access and import materials science databases using appropriate Python libraries.

4. Apply data cleaning and preprocessing techniques to materials datasets.

5. Perform statistical analysis on materials data using Python.

6. Implement machine learning algorithms for materials property prediction.

7. Create publication-quality visualizations of materials data using Python plotting libraries.

8. Develop interactive data visualization dashboards for materials science applications.

9. Apply FAIR (Findable, Accessible, Interoperable, Reusable) principles to data management in materials science.

10. Export processed data in standardized formats suitable for materials databases.

11. Automate data analysis workflows using Python scripts.

12. Implement version control for collaborative materials data projects.**Assignments**

We encourage you to use ChatGPT and other LLMs to generate code, even for your homework and final project. However, we also encourage you to write your own code, especially with the aim to learn Python. Therefore, we suggest that for every assignment, set a timer (15min – 1 h, depending on assignment length) and code on your own. After the time is up, utilize LLMs to help finish/correct the code. Alternatively, you can try a hybrid approach – write some code, ask LLMs to correct/modify it, or ask questions directly. Gemini is embedded in Google Colab and works seamlessly.

Grading

Attendance	20%
Assignments	60% (4x, 15% each)
Final Project	20%

Module 1 – Basics of Python

- Python Basics and Environment Setup
- Installing Python and IDEs**: Anaconda, Jupyter Notebook.
- Why Python for materials science? Compare briefly with Matlab.
- Python Syntax Basics: types (`int`, `float`, `string`, `list`, etc.), operators (`+`, `-`, `*`, `/`, `%`, `**`), input/output
- Using Jupyter Notebooks
- Conditional statements: `if`, `elif`, `else`.
- Loops: `for` and `while` (emphasize their use in data analysis and simulation).
- Simple functions with `def`.
- Function modularity with functions.

Module 2 - Python Libraries and Data Handling

- Introduction to Libraries, packages, and modules
- Introduction to `numpy`, `matplotlib`, and `pandas`.
- Introduction to `scipy`, `scikit-learn` / `sklearn`

Module 3 - Thermo-Calc / thermo-physical properties

- Outlier / error detection
- Cleaning up databases (incorrect data types, etc)
- Calculating new properties (thermal diffusivity)
- Histograms, Pearson correlations, PCA
- Dimensionality reduction: KMeans, MDS, tSNE, UMap
- **Assignment 1:** Create XX, Find Pareto front of thermal diffusivity vs. melting temperature

Module 4 – Semiconductor Properties

- Module 3, plus
- `Plotly` for interactive visualizations
- Machine learning: Gaussian Process, Random Forest, etc.
- `MatMiner` package for creating new ML descriptors
- Filtered dimensionality reduction
- **Assignment 2:** Discovery/Design of new materials

Module 5 – Images

- **Assignment 3**

Module 6 – Materials Project

- **Assignment 4**

Final Project