IMPERIAL
Department of Materials

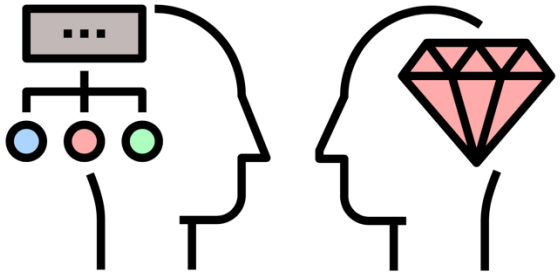
THE THOMAS YOUNG CENTRE TyC

Machine Learning for Materials

Research Challenge

Aron Walsh

Department of Materials
Centre for Processable Electronics



Module Assessment

Aim for working knowledge of ML with practical sessions and coursework

Computational exercises

Paired with each lecture

(Due at the end of each computer lab)

Research challenge

Assignment to complete

(details after Lecture 9)

Module Assessment

Aim for working knowledge of ML with practical sessions and coursework

Computational exercises

Completed - well done!

Research challenge

Individual assignment

(details today)

Research Challenge

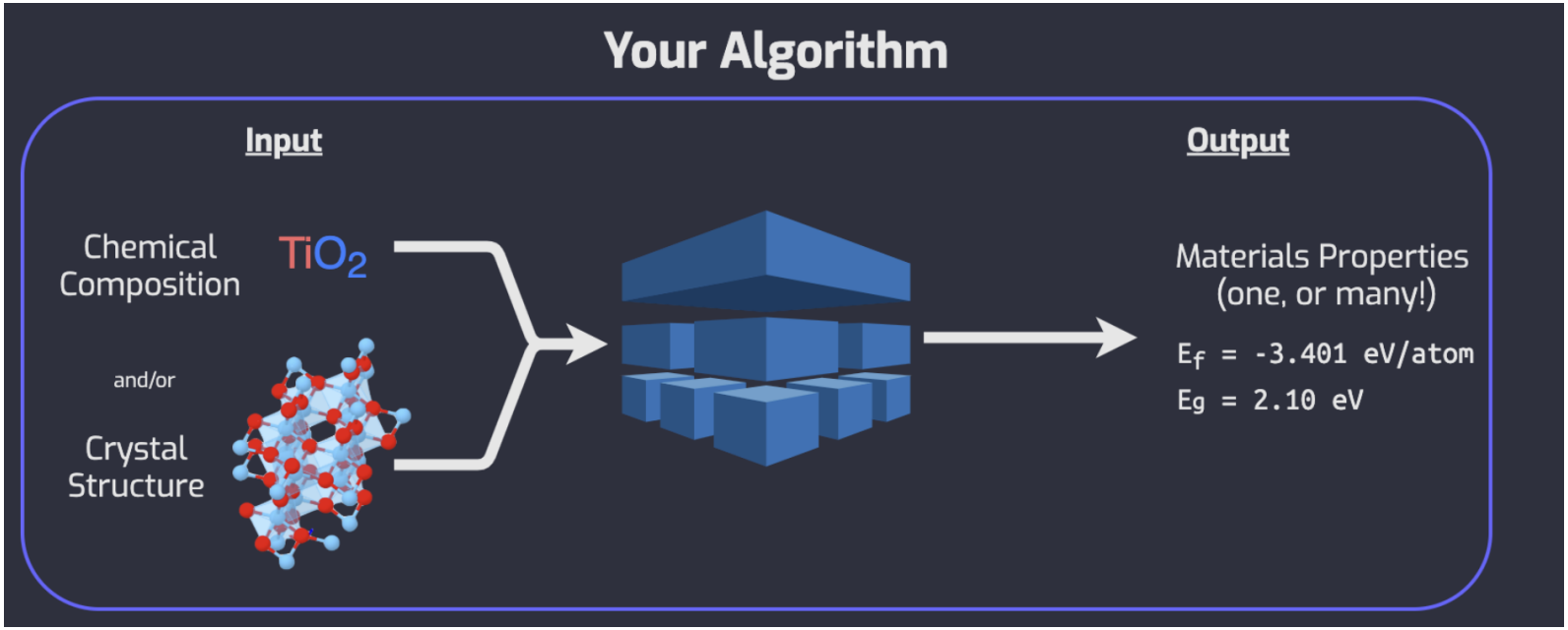
An opportunity to develop your practical skills. Goals:

- To apply the ML tools and data skills you have picked up so far
- To extend your knowledge through self-study, exploration, and cohort interactions
- To produce an annotated code with comparison to community benchmarks

Research Challenge

Each group is assigned a dataset from https://matbench.materialsproject.org

Your job is to produce an original model for the given classification or regression task



Some tasks use chemical composition only, while others use composition and structure

Research Challenge

The starting point is to check the literature.

Read the matbench paper and the models that have been tested

I. Data Preparation

II. Model Selection, Training & Testing

III. Discussion of Results

https://doi.org/10.1038/s41524-020-00406-3

Creative Solutions

There is great flexibility in programming with no unique solution for a given problem

You may be interested in speed or clarity, but ultimately want a robust code

- Check package manuals, e.g. https://matplotlib.org & https://scikit-learn.org

- Search https://stackexchange.com & https://github.com for ideas

Creative Solutions

Statement to be included in the submitted notebook

Large Language Model (LLM) Usage Declaration

- Did you use an LLM (e.g. GPT-4, Gemini, Co-Pilot)?
  - Specify tasks (e.g. code assistance)
  - Were any limitations/biases noted?
  - How did you ensure ethical use?

2025 Challenge Topics

One challenge per person has been randomly assigned

Challenge	Topic	Type	GTAs
A	Dielectric constant (4,764)	Regression (with structure)	Xia, Kinga
B	Experimental bandgap (4,604)	Regression (composition only)	Irea, Pan
C	Glass formation (5,680)	Classification (composition only)	Yifan, Fintan

Dataset details are provided in Notebook 9

GTA Assistance

Teaching assistants will be available
in the computer rooms:

Class 9 14:00-15:30

Class 10 14:00-15:30

The computer room is also booked on Feb 24th and 27th from 13:00-16:00 for self-study (no GTAs)

Submission deadline: 10th March 15:00

Challenge Submission

Two items submitted on Blackboard

1. Completed Jupyter notebook (.ipynb)

and

2. Recorded presentation* (max 5 min) where you introduce your code and your results on model training, selection, and performance

*Format is flexible. Could be recorded in PowerPoint, screenshare on Zoom, or plain video

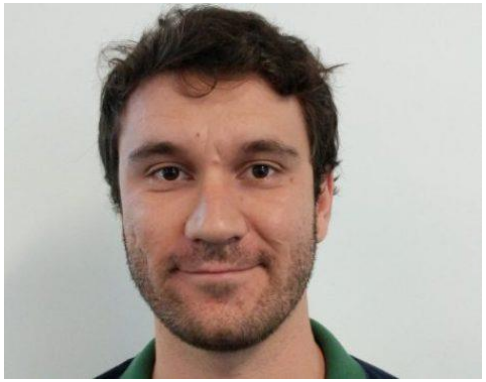
Challenge Assessment 2025

	Weight	Guidelines
Data Preparation	10 %	Apply appropriate pre-processing steps
Model Selection, Training and Testing	20 %	Justify model based on the problem, with appropriate validation and testing
Model Analysis and Discussion	20 %	Analysis of model performance, including high-quality plots
Python Code Quality	20 %	Clearly structured code with meaningful annotations
Recorded Presentation	30 %	Clarity and conciseness in model choices, results, limitations

Lecture 10

Final Class on Thursday at 1 pm

Guest lecture from Google Deepmind



Dr Ekin Dogus Cubuk

Senior Research Scientist

Module Feedback

Your feedback is valued & will help to shape the delivery for next year

Machine Learning for Materials

Go to

www.menti.com

Enter the code

3287 0914

Or use QR code

Appendix: Ethics of ML for Materials

How do these translate to the materials context?

Bias and Fairness

Influence on decision making processes

Transparency and Explainability

Interpretation of model predictions

Privacy and Data Protection

Collection, storage and using sensitive data

Social Impacts

From productivity increases to job displacements