CH-457

AI for chemistry

Schwaller Philippe

Cursus	Sem.	Type
Chimiste	MA2, MA4	Opt.

Language of teaching	English
Credits	2
Session	Summer
Semester	Spring
Exam	During the semester
Workload	60h
Weeks	14
Hours	**3 weekly**
Courses	1 weekly
Exercises	2 weekly
Number of positions	

Summary

The AI for Chemistry course will focus on teaching students how to use machine learning algorithms and techniques to analyze and make predictions about chemical data. The course will cover topics such as the basics of machine learning, common algorithms and their applications in chemistry.

Content

The AI for Chemistry course will focus on teaching students how to use machine learning algorithms and techniques to analyze and make predictions about chemical data. The course will cover topics such as the basics of machine learning, common algorithms and their applications in chemistry, and how to implement these algorithms using programming languages such as Python. Students will also learn how to visualize and interpret the results of their analyses and how to evaluate the accuracy and reliability of their predictions. The course will include hands-on projects and exercises where students can apply the concepts and techniques they have learned to real-world chemical data.

In a machine learning course in chemistry, students will also learn about the challenges and limitations of using machine learning in the field of chemistry. This will include discussions about the need for high-quality and diverse training data, the potential for overfitting and bias in the algorithms, and the importance of domain expertise in the interpretation of the results. Students will also learn about different approaches to machine learning in chemistry, such as supervised and unsupervised learning. They will explore the use of different types of machine learning models, such as decision trees and neural networks. Overall, a machine-learning course in chemistry will provide students with the skills and knowledge they need to apply machine-learning techniques to a wide range of problems and challenges in the field of chemistry.

The outline for the course:

- Python in Chemistry: This module will get the students up to speed with Python and commonly used modules in chemistry.

- Introduction to machine learning: This module will provide an overview of the basics of machine learning, including common definitions and concepts, the types of problems that can be addressed with machine learning, and the steps involved in a typical machine learning project. It will introduce students to the most used machine learning algorithms and techniques, including decision trees, k-nearest neighbours, and neural networks. Students will learn how to implement these algorithms in a programming language such as Python and will practice using them to analyze and make predictions on chemical data.

- Machine learning in chemistry: This module will focus on the application of machine learning in the field of chemistry. It would cover common tasks and challenges in chemistry that can be addressed with machine learning, such as predicting chemical properties and reactions, analyzing large datasets, and identifying patterns and trends in chemical data.

- Evaluating and interpreting machine learning models: This module will cover key concepts and techniques for evaluating and interpreting the results of machine learning models, such as accuracy, precision, recall, and confusion matrices. Students would also learn how to visualize and present their results and how to evaluate the reliability and

generalizability of their models.

- Advanced topics in machine learning for chemistry: This module will explore more advanced topics in machine learning for chemistry, such as unsupervised learning, deep learning, and transfer learning. Students will have the opportunity to apply these techniques to more complex chemical data and problems and would learn about the challenges and limitations of using machine learning in chemistry.

Keywords

articifial intelligence, machine learning, molecular design cycle, chemical reactions, data in chemistry, Python

Learning Prerequisites

Required courses

The course assumes basic programming knowledge, such as:

- computer programming in Python (see Information, Computation, Communication CS-119(k))
- https://www.kaggle.com/learn/python (another way to get familiar with basic Python programming)

Recommended courses

- Machine learning for physicists (PHYS-467)

Important concepts to start the course

- A knowledge of Python programming and machine learning will be helpful, but the course is open to all.

Here some excellent resources:
- [Andrew White's deep learning for molecules & materials book](https://dmol.pub
- [MolSSI Education Resources](http://education.molssi.org/resources.html#programming)
- [Greg Landrum's RDKit blog](https://greglandrum.github.io/rdkit-blog/)
- [Esben Bjerrum's Cheminformania](https://www.cheminformania.com)
- [iwatobipens' blog](https://iwatobipen.wordpress.com)
- [Rocío Mercado's dl-chem-101](https://github.com/rociomer/dl-chem-101)
- [Jan H. Jensen's Machine Learning Basics](https://sites.google.com/view/ml-basics/home)
- [Pat Walter's Practical Cheminformatics With Open Source Software](https://github.com/PatWalters/practical_cheminformatics_tutorials)

Learning Outcomes

By the end of the course, the student must be able to:

- Produce chemical data in a machine-readable format
- Represent molecules and reactions in different chemical representations
- Apply machine learning models to chemical tasks
- Plan a machine learning project
- Assess / Evaluate if machine learning models are suited for a given task

Transversal skills

- Assess progress against the plan, and adapt the plan as appropriate.
- Plan and carry out activities in a way which makes optimal use of available time and other resources.
- Communicate effectively with professionals from other disciplines.
- Demonstrate the capacity for critical thinking

- Demonstrate a capacity for creativity.
- Make an oral presentation.
- Write a scientific or technical report.

Teaching methods

1h lecture
2h hands-on exercises (bring your own laptop)

Assessment methods

Final project : For the assessment, students will have the opportunity to apply the concepts and techniques they have learned throughout the course to a real-world problem in chemistry. This could be a research project, a case study, or a practical application of machine learning in chemistry. Students will work individually or in small groups to complete their projects, and will present their findings to the class in the last lecture and write a 4-page report.

Resources

Websites

- https://schwallergroup.github.io/ai4chem_course/

Moodle Link

- https://go.epfl.ch/CH-457