Nanyang Technological University
Division of Chemistry and Biological Chemistry

Academic Year	AY19/20	**Semester**	1
Course Coordinator	Lu Yunpeng		
Course Code	CM4044		
Course Title	Artificial Intelligence in Chemistry		
Pre-requisites	PS0001 and PS0002 OR Approval by the School		
No of AUs	3		
Contact Hours	Lecture (2 hours) and Laboratory (3 hours)		
Proposal Date	20 February 2019		

Course Aims

This course aims to equip you with working knowledge of how artificial intelligence (AI) can be applied to chemistry. As a start, you will learn the rigorous way to construct Python projects for AI. Next, you will learn how to apply AI to analyze chemical data, either for visualization, classification, quantitative determination, or for pursuing insights into chemical process. We will introduce multiple case studies and analyze how AI has impacted the various sub-disciplines of chemistry. The course will improve your communication skills by requiring you to deliver technical presentations to a broad audience. As a result of this course, you will develop the necessary skills to prepare yourself for a career that involves harnessing AI for research and industry applications.

Intended Learning Outcomes (ILO)

By the end of this course, you (as a student) would be able to:

Python Packages for Data Science and AI

1. Use Numpy or other similar packages for scientific computation
2. Use Matplotlib or other similar packages for scientific data visualization
3. Use Regular Expression or other similar packages for pattern searching in text information for data extraction
4. Use Pandas or other equivalent packages to load different format of data for analysis
5. Identify the functionalities in scikit-learn and other equivalent packages
6. Apply scikit-learn or other package to build machine learning models
7. Identify the relevant functionalities in TensorFlow or other equivalent packages
8. Build deep learning models by using TensorFlow or other equivalent/variant packages

AI applications in Chemistry

9. Identify interesting data-driven questions in chemistry.
10. Formulate meaningful study problems that you want to explore in chemistry

11. Collect/extract relevant data, visualize and perform exploratory analysis on data.
12. Perform machine learning models to extract meaningful insights from data.
13. Implement above techniques with Python
14. Present your analysis results and problem solution via an engaging written communication.

Course Content

Python Packages for Data Science and AI

1. Introduction to numpy package
2. Introduction to matplotlib package
3. Introduction to regular expression package
4. Introduction to Pandas package
5. Introduction to scikit-learn package
6. Introduction to TensorFlow package

AI applications in Chemistry

7. Data in chemistry: Data source, format and extraction
8. Introduction to data-driven scientific questions in chemistry
9. Visualization of chemical data and simple analysis based on primitive methods
10. Machine learning applications to data analysis in chemistry
11. Deep learning in data analysis in chemistry

Assessment (includes both continuous and summative assessment)

Component	Course ILO Tested	Related Programme LO or Graduate Attributes	Weighting	Team/Individual	Assessment rubrics
1. Group Projects	All	Competence, Communication, Civic-mindedness, Character, Creativity.	40%	Team (20%) and Individual (20%)	Appendix 2, 3
2. Lab assignments	All	Competence	20%	Individual	Appendix 1
3. Examination	1-13	Competence, Creativity, Communication.	40%	Individual	Point-based marking (not rubrics based)
Total			*100%*		

Formative feedback

You will receive written and verbal feedback from the lecturer for Components 2 & 3.

You will receive summative group feedback on the group project in component 1.

Learning and Teaching approach

Approach	**How does this approach support students in achieving the learning outcomes?**
Lectures	Present the key ideas and important steps used to solve different types of problems.
Lab Tutorials	Develop proficiency in problem solving skills. Reinforce concepts already covered in the lectures. Give an opportunity for weaker or more reserved students to clarify doubts.
Group projects	Train the class on teamwork and cohesion, as well as to boost confidence for weaker students. Develop communications skills. Students will be able to learn the importance of teamwork.

Reading and References

1. Stuart Russell and Peter Norvig (2016) *Artificial Intelligence: A Modern Approach, 3rd edition.* Pearson. ISBN-13: 978-0136042594
2. Hugh Cartwright (2008) *Using artificial intelligence in chemistry and biology: a practical guide, 1st edition.* CRC Press. ISBN: 9780849384141, 0849384141
3. Aurélien Géron (2017) *Hands-On Machine Learning with Scikit-Learn and TensorFlow, 1st edition.* O'Reilly. ISBN-13: 978-1-4919-6229-9
4. Hugh Cartwright (2003) Applications of artificial intelligence in chemistry, 1st edition, Oxford University Press. ISBN 019855737X, 0198557361
5. Rajarshi Guha and Andreas Bender (2012) Computational approaches in cheminformatics and bioinformatics, Wiley InterScience. ISBN: 9781118131411, 111813141x

Course Policies and Student Responsibilities

(1) General

You are expected to complete all assigned pre-class readings and activities, attend all tutorial classes punctually and take all scheduled assignments and tests by due dates. You are expected to participate in all tutorial discussions and activities.

(2) Absenteeism

Absence from the midterm without a valid reason will affect your overall course grade. Valid reasons include falling sick supported by a medical certificate and participation in NTU's approved activities supported by an excuse letter from the relevant bodies. There will be no make-up opportunities for CA components.

All project assignments must be submitted on time. Failure to do so will affect your score.

Academic Integrity

Good academic work depends on honesty and ethical behaviour. The quality of your work as a student relies on adhering to the principles of academic integrity and to the NTU Honour Code, a set of values shared by the whole university community. Truth, Trust and Justice are at the core of NTU's shared values.

As a student, it is important that you recognize your responsibilities in understanding and applying the principles of academic integrity in all the work you do at NTU. Not knowing what is involved in maintaining academic integrity does not excuse academic dishonesty. You need to actively equip yourself with strategies to avoid all forms of academic dishonesty, including plagiarism, academic fraud, collusion and cheating. If you are uncertain of the definitions of any of these terms, you should go to the academic integrity website for more information. Consult your instructor(s) if you need any clarification about the requirements of academic integrity in the course.

Course Instructors

Instructor	Office Location	Phone	Email
Lu Yunpeng	CBC-06-23	65132747	yplu@ntu.edu.sg

Planned Weekly Schedule

Week	Topic	Course ILO	Readings/ Activities
1	Introduction to Numpy in scientific computing	1	Readings: Study several program using numpy package Activities: Students will be asked to develop their simple

			program in lab tutorials
2	Introduction Matplotlib in data visualization	2	Readings: Study several program using Matplotlib package Activities: Students will be asked to develop their simple program in lab tutorials.
3	Introduction to Regular Expression in pattern searching in text for data	3	Readings: Study several program using Regular Expression package Activities: Students will be asked to develop their simple program in lab tutorials.
4	Introduction to Pandas package	4	Readings: Study several program using Pandas package Activities: Students will be asked to develop their simple program in lab tutorials.
5	Introduction to scikit-learn package in machine learning	5,6	Readings: Study several program using scikit-learn package in machine learning Activities: Students will be asked to develop their simple program in lab tutorials.
6	Introduction to tensorflow package in deep learning	7,8	Readings: Study several program using tensorflow package in machine learning Activities: Students will be asked to develop their simple program in lab tutorials.
7	Data in chemistry	9	Readings: Lecture notes and reference

			books of Rajarshi Guha and Andreas Bender (2012) and Hugh Cartwright (2008). Activities: Students will discuss about general concept of data and the specialty of chemical data. They will also learn to how to retrieve data from different sources.
8	Introduction to data-driven scientific questions in chemistry	9, 10	Readings: Lecture notes, reference books of Rajarshi Guha and Andreas Bender (2012) and Hugh Cartwright (2008). Activities: Students will discuss the concept of cheminformatics, particularly, in quantitative structure-activity relation and molecular descriptor in lab tutorial.
9	Machine learning examples in chemistry	1-6, 9-13	Readings: Lecture notes, reference books of Rajarshi Guha and Andreas Bender (2012) and Hugh Cartwright (2008). Group project will be assigned. Lab practice on several machine learning models on chemical data

10	deep learning in chemistry	1-4, 7-13	Readings: Lecture notes, reference books of Rajarshi Guha and Andreas Bender (2012) and Hugh Cartwright (2008). More group project will be assigned for deep learning Lab practice on several deep learning models on chemical data
11	Group Project	1-13	Group discussion on the project and collaborative work.
12	Group Project	1-13	Group discussion on the project and collaborative work.
13	Group Project Presentations	14	Both lecture and lab tutorial time will be used for project presentation.

Appendix 1: Assessment Criteria for Lab Assignments
Standards Criteria

Levels of Performance	Criteria Description
A+ (Exceptional) A (Excellent)	Provides clear, efficient, working and well-documented code; evidence of programing understanding and concern for code efficiency beyond getting correct solution. Demonstrated ability to develop multiple approaches to programming task, and understanding of their respective advantages.
A- (Very good) B+ (Good)	Provides clear, efficient, working and well-documented code; evidence of programing understanding.
B (Average) B- (Satisfactory) C+ (Marginally satisfactory)	Working but limited documentation of code.
C (Bordering unsatisfactory)	Write the code with lots of help from TA and instructor.

C- (Unsatisfactory)	Limited code documentation or demonstration of conceptual understand.
D (Deeply unsatisfactory) F (0-44)	Lack of demonstrated conceptual understanding. Non-functional code.

Appendix 2: Assessment Criteria for Group Project (20%)
Standards Criteria

Levels of Performance	Criteria Description
A+ (Exceptional) A (Excellent)	Provides clear and meaningful study questions; appropriate methods for data presentation, manipulating and exploration; efficient, working and well-documented code; evidence of programing understanding and concern for code efficiency beyond getting correct solution. Takes an original approach to the questions; very well structured reports with good interpretations of results; evidence of excellent ability to apply knowledge taught in the course while thinking outside the box; provides clear, efficient, working and well-documented code Clearly identifies, illustrates and critically examines implications of the project in wider context of society. Provide source acknowledgement in standard citation format. All references and citations are present and correctly written.
A- (Very good) B+ (Good)	Takes a conventional approach to the question; good interpretation of results; evidence of ability to apply knowledge taught in the course; provides clear, efficient, working and well-documented code. Describes conventional links between project and wider context of society with clear illustrations, or identifies and examines implications of the project in the wider context of society. Provides source acknowledgement in standard citation format. One or two references or citations missing or incorrectly written.
B (Average) B- (Satisfactory) C+ (Marginally satisfactory)	Takes a conventional approach to the question; limited interpretation of results; evidence of some (but not significant) ability to apply knowledge taught in the course; working but limited documentation of code. States conventional links between project and wider context of society without clear illustrations, or acknowledges obvious implications of the project on the wider context of the society. Provides minimal source acknowledgement. Some information does not contain a citation.

C (Bordering unsatisfactory) C- (Unsatisfactory)	Limited understanding of process; incorrect or miss-interpreted results; limited evidence of ability to apply knowledge taught in the course. Non-functional or limited code documentation. Makes some weak connections or missed some obvious implications of the project and the wider context of society. Many references and citations are missing. Format has technical errors or is presented in inconsistent styles.
D, F (Deeply unsatisfactory)	Inadequate in addressing the question; incorrect and/or miss-interpretation of results; lacks structure and focus, and is mostly or wholly off topic; inadequate capacity to apply knowledge taught in the course; non-functional code. OR failure to submit the report. Makes little to no connection between the project and the wider context of society, or missed some obvious negative implications pf the project on the wider context of society. References and citation errors detract significantly from paper. Little or no acknowledgment of sources.

Appendix 3: Assessment Criteria for Individual Contribution in Group Project (20%)

	Fail standard (0-39%)	Pass standard (40-75 %)	High standard (76-100 %)
Individual Contribution	Little contribution to the project Silent on the ideas of others Little or no interaction with group members Absent or was often late and leaving early Clueless when question on basic material/concepts	Participate meaningfully in the project Show a willingness to discuss the ideas of others Cooperate with other group members Was present for most meetings, seldom late or leaving early Lead or Facilitate discussions Demonstrate familiarity with most materials/concepts when question	Contribute significantly in the development of the project Constructively critique and build on the ideas of others Play an instrumental role in getting group members to cooperate Was present and punctual for all meetings Lead and Facilitate discussions Demonstrate a high degree of familiarity with materials/concepts when question, often

			with detailed elaboration.

CBC Programme Learning Outcome

The Division of Chemistry and Biological Chemistry (CBC) offers an undergraduate degree major in Chemistry that satisfies the American Chemical Society (ACS) curricular guidelines and equips students with knowledge relevant to the industry. Graduates of the Division of Chemistry and Biological Chemistry should have the following key attributes:

1. Competence

Graduates should be well-versed in the foundational and advanced concepts of chemical science, be able to evaluate chemistry-related information critically and independently, and be able to use complex reasoning to solve emergent chemical problems.

2. Creativity

Graduates should be able to synthesize and integrate multiple ideas across the curriculum, and propose innovative solutions to emergent chemistry-related problems based on their training in chemistry.

3. Communication

Graduates should be able to demonstrate clarity of thought, independent thinking, and sound scientific analysis and reasoning through written and oral reports to audiences with varying technical backgrounds. They should also be able to effectively engage other professional chemists in collaborative endeavours.

4. Character

Graduates should be able to act in responsible ways and uphold the high ethical standards that the society expects of professional chemists.

5. Civic-mindedness

Graduates should be aware of the impact of chemistry on society, and how chemistry can be applied to benefit mankind. They should also be aware of and uphold the best chemical safety practices.