NANYANG TECHNOLOGICAL UNIVERSITY SINGAPORE

COURSE CONTENT

Academic Year	2022/2023			**Semester**		2
Course Coordinator	Yong Keen Mun Kelvin					
Course Code	CB0494					
Course Title	Introduction to Data Science and Artificial Intelligence					
Pre-requisites	CE/CZ1003 Introduction to Computation Thinking BG2211 Introduction to Computational Thinking CH2107 Introduction to Computational Thinking CV1014 Introduction to Computational Thinking MS1008 Introduction to Computational Thinking MA1008 Introduction to Computational Thinking EE1005 From Computational Thinking to Programming RE1016 Engineering Computation					
No of AUs	3					
Contact Hours	LECTURES	0	LAMS/TEL (Online Videos and Resources)	16	EXAMPLE CLASSES (Hands-on Sessions and Seminars)	26
Proposal Date	20 September 2019					

Course Aims

In today's era of Information, 'Data' is the new driving force, provided we know how to extract relevant 'Intelligence'. This course will start with the core principles of Data Science, and will equip you with the basic tool and techniques of data handling, exploratory data analysis, data visualization, data-based inference, and data-focussed communication. The course will also introduce you to the fundamentals of Artificial Intelligence – state space representation, uninformed search, and reinforcement learning.

The course will motivate you to work closely with data and make data-driven decisions in your field of study. The course will also touch upon ethical issues in Data Science and Artificial Intelligence, and motivate you to explore the cutting-edge applications related to Big Data, Neural Networks and Deep Learning. Python will be the language of choice to introduce hands-on computational techniques.

Intended Learning Outcomes (ILO)

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

1. identify and define data-oriented problems and data-driven decisions in real life,
2. discuss and illustrate the problems in terms of data exploration and visualization,
3. apply basic machine learning tools to extract inferential information from the data,

4. compose an engaging “data-story” to communicate the problem and the inference,
5. outline the roles and requirements of artificial intelligence in practical applications,
6. discuss and explain fundamentals of state space search and reinforcement learning.

Course Content

	Topics	LAMS/TEL (Hours)	Example Classes (2-Hour Sessions)
1	**Data-Analytic Thinking** What is Data Science? – The core problems and solutions. Extracting Intelligence from Data – formulating problems. Introduction to Python.	1 3	Problem Formulation, Data Wrangling, Cleaning and Preparation (2 weeks)
2	**The Data Pipeline** Types of Data in various practical Data Science scenarios. Data Wrangling, Cleaning and Preparation using Python.	1	
3	**Data Presentation** Basic concepts in Statistics and Exploratory Data Analysis. Data Exploration and Data Visualization using Python. Case Studies involving Structured and Unstructured Data	2	Basic Statistics, Data Exploration and Visualization (2 weeks)
4	**Data-driven Inference** Basics of Machine Learning : Prediction and Classification. Prediction and Classification techniques using Scikit-Learn.	2	Prediction and Classification (2 weeks)
5	**Data-driven Identification** Basics of Machine Learning : Clustering and Anomalies. Clustering and Anomaly Detection using Scikit-Learn.	1	Clustering and Anomaly Detection (1 week)
6	**Digital Storytelling** Data-driven Dashboards, Websites and Presentations. Data Presentation using Python Notebooks and Plotly.	1	Data Presentation and Dashboards (1 week)

7	**Artificial Intelligence** What is Artificial Intelligence? – History and State-of-Art. Principles of problem solving and the State Space Search. Case Studies for State Space Search and Search Algorithms	2	Revision on Data Science (1 week)
8	**Reinforcement Learning and AI** Introduction to Reinforcement Learning in context of AI. Fundamentals of Markov Processes and Q-Learning.	2	Mini-Project Preparations/Discussions (2 weeks)
9	**Ethics in DS&AI** Ethical considerations and the idea of responsible DS&AI.	0.5	Mini-Project Presentations (2 weeks)
10	**State-of-the-Art in DS&AI** Progress in Big Data, Neural Networks and Deep Learning.	0.5	
	Check for Hours	= 16	= 26

Design Philosophy

The primary goal of this course is to enhance your "Digital Literacy" by introducing you to some real-life application of data-driven computational thinking and decision, so that you may observe the true power of your computing skills in handling practical problems. The course is planned in three parts – core data-science module, machine learning tool and techniques, and fundamentals of artificial intelligence.

Core Data-Science Module

- Week 1 will teach you the premise of Data Science, and how to formulate data-oriented problems
- Week 2 will teach you how to wrangle acquired data to suit your needs, and how to get it cleaned
- Weeks 3 and 4 will introduce you to the art of presenting data, with basic exploratory data analysis

Machine Learning Tools

- Weeks 5 and 6 will dive into Machine Learning to explore the use of basic models in Data Science
- Week 7, right before the break, will introduce you to basic techniques of finding Patterns in Data
- Week 8 will tie together the ideas of Data Science and Machine Learning on a Digital Storyboard

Artificial Intelligence

- Weeks 9 and 10 will introduce you to the domain of Artificial Intelligence through Search Space
- Weeks 11 and 12 will extend the notion of AI to Reinforcement Learning and Markov

Processes

- Week 13 will end the course by exposing you to the ethical responsibilities of Data Scientists in using the tools and techniques of Artificial Intelligence, and will motivate you to probe deeper in the field

In due flow of the course, we will also refresh basic concepts in Statistics and Computing that you may have already seen in the previous semester. The new principles and techniques that you will learn in this course will be related to the practical tools of data analysis and state-space search, along with use and presentation of data in various forms and shape. You will also learn specific applications of DS&AI in your field of study, through real-life applications and case studies. We hope this will pique your interest!

Assessment (includes both continuous and summative assessment)

Component	Course LO Tested	Related Programme LO or Graduate Attributes	Weightage	Team/Individual	Assessment Rubrics
TEL participation and TEL MCQs	1,2,3,5,6	a,b,h,l	10%	Individual	Appendix 1
Online Quizzes based on MCQs	1,2,3,5,6	a,b,h	40%	Individual	Appendix 1
Exercises in Example Class	1,2,3,4,5,6	a,b,c,d,e,f,h,j	20%	Individual	Appendix 2
Mini Project in Example Class	1,2,3,4,5,6	a,b,c,d,e,f,i,j	30%	Team + Individual	Appendix 3
Total			100%		

Mapping of Course SLOs to EAB Graduate Attributes

Course Student Learning Outcomes	Cat	EAB's 12 Graduate Attributes*											
		(a)	(b)	(c)	(d)	(e)	(f)	(g)	(h)	(i)	(j)	(k)	(l)
EE0005 Introduction to Data Science and Artificial Intelligence	Core	●	●	◐	●	◐	○		○	◐	●		◐
Overall Statement	This course, as a part of the "Digital Literacy" program, aims to introduce you to the core techniques of data science, machine learning and artificial intelligence, including data manipulation, visualization, statistical modelling, inference, data presentation, state space search algorithms, and reinforcement learning, which constitute the toolbox for any Data Science & Artificial Intelligence practitioner.												
1. identify and define data-oriented problems and data-driven decisions in real life	(a), (b), (d), (f), (i), (j), (l)												
2. discuss and illustrate the problems in terms of data exploration and visualization	(a), (b), (c), (d), (e), (i), (j), (l)												
3. apply basic machine learning tools to extract inferential information from the data	(a), (b), (c), (d), (e), (i)												
4. compose an engaging "data-story" to communicate the problem and the inference	(a), (b), (e), (f), (h), (i), (j)												
5. outline the roles and requirements of artificial intelligence in practical applications	(a), (b), (d), (f), (h), (l)												
6. discuss and explain fundamentals of state space search and reinforcement learning	(a), (b), (c), (d), (e), (i)												

Legend:
- ● Fully consistent (contributes to more than 75% of Student Learning Outcomes)
- ◐ Partially consistent (contributes to about 50% of Student Learning Outcomes)
- ○ Weakly consistent (contributes to about 25% of Student Learning Outcomes)
- Blank Not related to Student Learning Outcomes

*The graduate attributes as stipulated by the EAB, are:

(a) **Engineering knowledge**: Apply the knowledge of mathematics, natural science, engineering fundamentals, and an engineering specialisation to the solution of complex engineering problems.

(b) **Problem Analysis**: Identify, formulate, research literature, and analyze complex engineering problems reaching substantiated conclusions using first principles of mathematics, natural sciences, and engineering sciences.

(c) **Design/development of Solutions**: Design solutions for complex engineering problems and design system components or processes that meet the specified needs with appropriate consideration for public health and safety, cultural, societal, and environmental considerations.

(d) **Investigation**: Conduct investigations of complex problems using research-based knowledge and research methods including design of experiments, analysis and interpretation of data, and synthesis of the information to provide valid conclusions.

(e) **Modern Tool Usage**: Create, select, and apply appropriate techniques, resources, and modern engineering and IT tools including prediction and modelling to complex engineering activities with an understanding of the limitations.

(f) **The engineer and Society**: Apply reasoning informed by the contextual knowledge to assess societal, health, safety, legal, and cultural issues and the consequent responsibilities relevant to the professional engineering practice.

(g) **Environment and Sustainability**: Understand the impact of the professional engineering solutions in societal and environmental contexts, and demonstrate the knowledge of, and need for the sustainable development.

(h) **Ethics**: Apply ethical principles and commit to professional ethics and responsibilities and norms of the engineering practice.

(i) **Individual and Team Work**: Function effectively as an individual, and as a member or leader in diverse teams and in multidisciplinary settings.

(j) **Communication**: Communicate effectively on complex engineering activities with the engineering community and with society at large, such as, being able to comprehend and write effective reports and design documentation, make effective presentations, and give and receive clear instructions.

(k) **Project Management and Finance**: Demonstrate knowledge and understanding of the engineering and management principles and economic decision-making, and apply these to one's own work, as a member and leader in a team, to manage projects and in multidisciplinary environments.

(l) **Life-long Learning**: Recognize the need for, and have the preparation and ability to engage in independent and life-long learning in the broadest context of technological change.

Formative feedback

TEL participation and TEL MCQs : This is an online exercise. You will see you scores, your answers, the correct answers, feedback on your incorrect answers, and explanations for the correct answers, immediately after you have submitted your answers online.

Online Quizzes based on MCQs : These are online exercises too. You will see you scores, your answers, the correct answers, feedback on your incorrect answers, and explanations for the correct answers, immediately after you have submitted your answers online.

Exercises in Example Class : Individual feedback will be provided to you after proper evaluation of your submissions. The answers will be discussed in the class, and you will also get to know the basic score statistics of the other students in the same cohort.

Mini Project in Example Class : You will be guided in choosing the topic, and the instructor will also help you during the course of the project, as and when required. Regular interactions with the instructor will be arranged to monitor your progress, and to provide you with constructive criticism.

Learning and Teaching approach

Approach	**How does this approach support students in achieving the learning outcomes?**
LAMS/TEL (Online Video)	Topics will be delivered as a series of online videos lectures, and you will also be provided reference materials for self-study to achieve the ILOs.
Example Class (Face-to-Face)	Example Classes will be used for seminar sessions for students to discuss, debate and clarify the contents of the online LAMS/TEL contents, as well as hands-on sessions to equip students with practical knowledge on data science, machine learning and artificial intelligence, and to guide in terms of the design and implementation of a mini project, to achieve the ILOs.

Reading and References

There is no single textbook for the course. The following books and resources will be used as references.

1. Python Data Science Handbook : Jake VanderPlas : O'Reilly (1st edition)
2. An Introduction to Statistical Learning : James, Witten, Hastie, Tibshirani
3. Artificial Intelligence: A Modern Approach : Russell and Norvig (3rd edition)

Additional resources, if required, will be shared with you in the LAMS/TEL videos and Example Classes.

Course Policies and Student Responsibilities

As a student of the course, you are required to abide by both the University Code of Conduct and the Student Code of Conduct. The Codes provide information on the responsibilities of all NTU students, as well as examples of misconduct and details about how students can report suspected misconduct. The University also has the Student Mental Health Policy. The Policy states the University's commitment to providing a supportive environment for the holistic development of students, including the improvement of mental health and wellbeing. These policies and codes concerning students can be found in the following link: http://www.ntu.edu.sg/SAO/Pages/Policies-concerning-students.aspx

Academic Integrity

Good academic work depends on honesty and ethical behavior. Quality of your work as a student relies on adhering to the principles of academic integrity and to the NTU Honor 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 of NTU, it is important that you recognize your responsibilities in understanding and applying the principles of academic integrity in all the work you do at the University. 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, and 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 Instructor(s)

Instructor	Office Location	Phone	Email
Yong Keen Mun Kelvin	N1.2-B2-26B	69081989	kmyong@ntu.edu.sg

Planned Weekly Schedule

Week	Topic	Course LO	Readings	Example Class Activities
1	**Data-Analytic Thinking** What is Data Science? – The core problems and solutions. Extracting Intelligence – formulating problems. Introduction to the Python programming language.	1,2	Online Video (LAMS/TEL)	Defining a Data Science Problem in real-life. Familiarization with Python tools for DS.
2	**The Data Pipeline** Types of Data in various practical Data Science scenarios. Data	1,2	Online Video (LAMS/TEL)	Extraction, Wrangling, Cleaning, Preparation of Data using Pandas.

	Wrangling, Cleaning, Preparation.			
3	**Data Exploration** Basic concepts in Statistics and Exploratory Data Analysis.	1,2	Online Video (LAMS/TEL)	EDA using Case Studies involving Structured and Unstructured Data
4	**Data Presentation** Data Exploration and Data Visualization using Python.	2,4	Online Video (LAMS/TEL)	Visualization tools in Python and the basics of Data Visualization
5	**Data-driven Predictions** Prediction using techniques of Regression and Time Series	2,3	Online Video (LAMS/TEL)	Using Prediction tools from Scikit-Learn.
6	**Data-driven Classification** Classification using techniques of Decision Trees and Support Vectors	2,3	Online Video (LAMS/TEL)	Using Classification tools from Scikit-Learn.
7	**Data-driven Identification** Clustering and Anomaly Detection.	2,3	Online Video (LAMS/TEL)	Using Clustering tools from Scikit-Learn.
8	**Digital Storytelling** Data-driven Dashboards, Websites and Presentations.	2,4	Online Video (LAMS/TEL)	Data Presentation using Notebooks and Plotly.
9	**Artificial Intelligence** What is Artificial Intelligence? – History and State-of-Art. Principles of problem solving and State Space.	5,6	Online Video (LAMS/TEL)	Case Studies for State Space Search and Search Algorithms
10	**Uninformed Search** Search Algorithms : breadth-first, depth-first, IDA, uniform-cost.	5,6	Online Video (LAMS/TEL)	Case Studies for State Space Search and Search Algorithms
11	**Reinforcement Learning** Introduction to Reinforcement Learning in context of AI. Basics of Markov Processes and Q-Learning.	5,6	Online Video (LAMS/TEL)	Case Studies for Reinforcement Learning
12	**Reinforcement Learning** Introduction to Reinforcement Learning in context of AI. Basics of Markov Processes and Q-Learning.	5,6	Online Video (LAMS/TEL)	Case Studies for Reinforcement Learning
13	**Ethics and State-of-the-Art** Ethical considerations and the idea of responsible DS&AI. Progress in Big Data, Neural Net,	1,5	Online Video (LAMS/TEL)	Ethical considerations and the idea of responsible DS&AI.

	Deep Learning.			

Appendix 1 : Assessment Criteria for TEL MCQs

You will complete 13 online LAMS/TEL sessions, including embedded MCQs (or similar). The maximum score is 10% of your total marks. You will take 1 online theory quiz (MCQs) based on LAMS lectures during the semester. The maximum score is 40% of your total marks.

Appendix 2 : Assessment Criteria for Exercises in Example Class

You will take 1 online Lab Quiz during the semester, based on the material covered during the Labs or the Example Classes. The maximum score for the Lab Quiz is 20% of your total marks.

Appendix 3 : Assessment Criteria for Mini-Project

You will submit the code(s) for data analysis, the visualization dashboard, and a final report to illustrate the Mini-Project – both the problem and the solution. You will need to do a presentation also for the Mini-Project. Mini-Project will be graded out of 100 points. The Mini-Project accounts for 30% of your total marks.

Criteria	Standards		
	Fail standard (0-40 %)	Pass standard (41-74 %)	High standard (75-100 %)
Identify the core definition of the problem, and plan the data-driven solution. (LO 1, 3, 5)	Identifying completely wrong definitions of the problems, and planning solutions that are somewhat related but are not the actual solutions expected for the problems.	Identifying the correct and relevant definitions of the problems in line with the course materials, planning solutions reasonably in line with solutions expected for the problems, and trying to relate the course materials to the planned solutions. Accuracy and clarity can be further improved.	Identifying the correct and relevant definitions of the problems in line with the course materials, planning technically accurate steps for the solutions that are expected for the problems, and clearly connecting the course materials to the planned solutions.
Explore the data effectively and devise required models to solve the problems. (LO 2, 3, 6)	Ad hoc analysis of the data and arbitrary steps in building the model without properly connecting the concepts with relevant concepts from the course. No or little evidence of critical evaluation of the proposed solution.	Logical exploration of the data that demonstrates a good understanding of the concepts from the course, and building models with reasonable accuracy to solve the problems. Reasonable evidence of critical thinking related to the proposed solution, and producing solutions with some degree of intuition and justification (rigorous	Clear logical flow of data exploration of that demonstrates a good understanding of the concepts from the course (and beyond), and building models with high accuracy to solve the problems. Extensive evidence of critical thinking related to the proposed solution, and producing solutions with clear intuition and proper

		steps for model-building or validation of models and results may be missing).	justification, including rigorous steps for model-building and validation of the models and results.
Overall Editorial Standard of the Solution and the Final Report. (LO 4)	Disorganised format and arrangement of the code and report, without any comment or little/no mention of references/resources.	Clear logical flow and well-formatted arrangement of the code and report, with all essential components. Reasonable comments and reasonable documentation of references /resources.	Clear logical flow and well-formatted arrangement of the code and report, with all essential components. Detailed set of technical comments to illustrate the choices made towards the solution, and to highlight the inferences. Proper documentation of references /resources.

Your Individual contribution (20 points out of 100) towards the Mini-Project will be judged based on an Oral Evaluation, as per the following rubrics.

Criteria	**Standards**		
	Fail standard (0-40 %)	**Pass standard (41-74 %)**	**High standard (75-100 %)**
Understanding of the Project and Individual Contribution. (LO 1, 2, 3)	Little understanding of problem definition, solution techniques, data exploration and machine learning tools used in the project. Individual contribution is too low compared to the team-mates.	Decent understanding of problem definition, solution techniques, data exploration and machine learning tools used in the project. Individual contribution to the project is proportional to the team size and project difficulty.	Clear understanding of problem definition, solution techniques, data exploration and machine learning tools used in the project. Individual contribution to the project is significantly high compared to team-mates.