MS0003 – INTRODUCTION TO DATA SCIENCE AND ARTIFICIAL INTELIGENCE

Academic Year	2023	Semester	2
Course Coordinator	Kedar Hippalgaonkar		
Course Type	Foundational Core (FC)		
Pre-requisites	SC/CE/CZ1003 Introduction to Computational Thinking and Programming		
	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 Introduction to Computational Thinking		
	RE1016 Introduction to Computational Thinking		
AU	3		
Grading	Letter Grading		
Contact Hours	Lectures 13 hours Tutorials/Labs 26 hours		
Proposal Date			

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 in Materials Science 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

Teaching Week	Topics	LECTURES (Hours)	Example Classes (2-Hour Sessions)
1	Introduction - Data-Analytic Thinking What is Data Science? – The core problems and solutions. Extracting Intelligence from Data – formulating problems.	1	Introduction - Data-Analytic Thinking What is Data Science? – The core problems and solutions. Extracting Intelligence from Data – formulating problems.
2	Basic Data Acquisition and Handling	1	Basic Data Acquisition and Handling
3	Basic Statistics and Exploratory Data Analysis	1	Basic Statistics and Exploratory Data Analysis
4	Linear Regression	1	Linear Regression
5	Classification	1	Classification
6	Clustering and Anomalies	1	CA 1
7	Visualization	1	Clustering and Anomalies
Recess Week			
8	Neural Networks	1	Visualization
9	Large Language Models	1	Neural Networks & Large Language Models
10	Time Series Modeling	1	CA2
11	Introduction to Real-Life Datasets	1	Project Lab I
12	Strategies for improved performance on project datasets I	1	Project Lab II
13	Strategies for improved performance on project datasets II	1	Project Lab III
Check for Hours		13	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.

Week 1: Introduction to Data-Analytic Thinking

- Understand the essence of Data Science and its real-world applications.
- Explore the fundamental problems in Data Science and learn potential solutions.
- Develop the skill of formulating data-related problems effectively.

Week 2: Basic Data Acquisition and Handling

- Dive into the world of data acquisition, including sources, types, and collection techniques.
- Master the art of data cleaning and preprocessing to ensure data quality.
- Learn about data integration, transformation, storage, and retrieval methods.

Week 3: Basic Statistics and Exploratory Data Analysis

- Explore descriptive statistics to understand data distribution and central tendencies.
- Get hands-on experience with inferential statistics, hypothesis testing, and confidence intervals.
- Delve into the power of Exploratory Data Analysis (EDA) and data visualization.

Week 4: Linear Regression

- Understand the concept of regression analysis and its practical applications.
- Learn how to perform simple and multiple linear regression.
- Evaluate and interpret regression models for making data-driven decisions.

Week 5: Classification

- Explore classification problems and their significance in data science.
- Dive into logistic regression for binary classification.
- Get acquainted with decision trees, random forests, and model evaluation metrics.

Week 6: Clustering and Anomalies

- Discover clustering algorithms like K-Means and DBSCAN for grouping data points.
- Learn techniques for identifying and handling anomalies in datasets.
- Apply clustering and anomaly detection methods to real-world scenarios.

Week 7: Visualization

- Master advanced data visualization techniques to communicate insights effectively.
- Explore interactive data visualization tools and platforms.
- Create informative data visualizations as part of a project.

Week 8: Neural Networks

- Dive into the world of artificial neural networks (ANNs) and deep learning.
- Understand the architecture of feedforward neural networks.

Week 9: Large Language Models

- Explore Natural Language Processing (NLP) and its applications.
- Study language modeling and the use of pretrained models like GPT-3.

Week 10: Time Series Modeling

- Gain expertise in handling time series data.
- Perform time series analysis, including decomposition and forecasting.
- Apply ARIMA models and address seasonality and trends in time series data.

Week 11: Introduction to Real-Life Datasets

- Discover real-world data sources and the ethical considerations of using them.
- Learn data wrangling techniques for preparing real-life datasets.
- Understand the unique challenges and opportunities posed by real-life data.

Week 12: Strategies for Improved Performance on Project Datasets I

- Focus on feature engineering to create meaningful predictors.
- Explore model tuning, hyperparameter optimization, and cross-validation.
- Address data imbalance and bias issues in project datasets.

Week 13: Strategies for Improved Performance on Project Datasets II

- Dive deeper into advanced techniques like ensemble learning (stacking, bagging, boosting).
- Emphasize model interpretability and explainability.
- Learn deployment strategies for data science projects and present final projects for peer review.

This weekly lesson plan provides a structured and comprehensive curriculum for mastering various aspects of data science. Students will gradually build their knowledge and skills in data analysis, machine learning, and real-world data applications throughout the semester.

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 materials science, through real-life applications and case studies. We hope this will pique your interest!

Assessment (Includes both continuous and summative assessment)

Component	ILO Tested	EAB Graduate Attributes	Weightage	Team / Individual	Rubrics
1. CA1 (Hand-on Lab Quiz)	1,2,3,4,5,6	a,b,c,d,e,f,g	30%	Individual	Refer to the description below
2. CA2 (Hands-on Lab Quiz)	1,2,3,4,5,6	a,b,c,d,e,f,g	30%	Individual	
3. Mini Project in Example Class	1,2,3,4,5,6	a,b,c,d,e,f,g	40% (Team 28% Individual 12%)	Team & Individual	Appendix 1
Total			**100%**		

Description of Assessment Components

CA1 (Hands-on Lab Quiz)
You will take this first hands-on Lab Quiz during the semester, based on the material covered during the Labs or the Example Classes. You will need to code for this quiz (at least major part of it), and the maximum score for the Lab Quiz is 30% of your total marks.

CA2 (Hands-on Lab Quiz)
You will take this second hands-on Lab Quiz during the semester, based on the material covered during the Labs or the Example Classes. You will need to code for this quiz (at least major part of it), and the maximum score for the Lab Quiz is 30% of your total marks.

Mini Project in Example Class
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. Mini-Project will be graded out of 100 points, with 70 points for the Team Exercise (code, presentation, report) and 30 points for Individual contribution. The Individual contribution will be judged based on an Oral Evaluation after the project presentation. The score for the Mini-Project, graded out of 100, will then be scaled down to 40% of your total marks.

EAB Graduate Attributes[1]	
a)	**Engineering Knowledge** Apply the knowledge of mathematics, natural science, engineering fundamentals, and an engineering specialisation as specified in WK1 to WK4 respectively to the solution of complex engineering problems.
b)	**Problem Analysis** Identify, formulate, research literature, and analyse 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 systems, 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 (WK8) 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 problems, with an understanding of the limitations.
f)	**The Engineer and Society** Apply reasoning informed by contextual knowledge to assess societal, health, safety, legal and cultural issues and the consequent responsibilities relevant to professional engineering practice and solutions to complex engineering problems.
g)	**Environment and Sustainability** Understand and evaluate the sustainability and impact of professional engineering work in the solution of complex engineering problems in societal and environmental contexts.
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**

[1] *Reference: EAB Accreditation Manual*

	Demonstrate knowledge and understanding of engineering 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** Recognise 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

Students will receive feedback on their projects from their project supervisors and Teaching Assistants (TAs) during their weekly meetings and project preparation phase. They will also receive formal feedback on their individual performance after their CAs.

Learning & Teaching Approach

Approach	**How does this approach support students in achieving the learning outcomes?**
Lectures (Face-to-Face)	Topics will be delivered in class, 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 students to practice coding with hands-on sessions to equip students with practical knowledge of data science, machine learning and artificial intelligence, and to guide them in terms of the design and implementation of a mini project, to achieve the ILOs.

Readings & References

There is no single textbook for the course. The following books and resources will be used as references and if necessary, notes will be provided.

1. Python Data Science Handbook : Jake VanderPlas : O'Reilly (2016)
2. An Introduction to Statistical Learning : James, Witten, Hastie, Tibshirani (2021)

Additional resources, if required, will be shared with you in the Lectures and Example Classes.

Course Policy & Student Responsibility

(1) General
Students are expected to attend the lecture briefings, complete the in-class learning assignment and tutorials, and contribute to group projects.

As this module involves group project, students are expected to take responsibility in planning the group project in advance and execute.

(2) Absenteeism

This module requires you to contribute to teamwork. Therefore, absence from group discussions and experiment planning is unacceptable without a valid reason/supporting documents. Otherwise, this will affect your overall course grade.

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 recognise 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 about the definitions of any of these terms, you should refer to the Academic Integrity Handbook for more information. Consult your instructor(s) if you need any clarification about the requirements of academic integrity in the course.

On the use of technological tools (such as Generative AI tools), different courses/assignments have different intended learning outcomes. Students should refer to the specific assignment instructions on their use and requirements and/or consult your instructors on how you can use these tools to help your learning.

Instructor	Office Location	Phone	Email
Dr Kedar Hippalgaonkar	N4.1	NIL	kedar@ntu.edu.sg
Dr Leonard Ng Wai Tet	N4.1	NIL	leonard.ngwt@ntu.edu.sg

Planned Weekly Schedule

Please refer to Course Content