



Syllabus CS 212 *Introduction to Data Science and Machine Learning*

COURSE CODE / COURSE LEVEL: CS 212

Decided and approved by the Department in line with existing course coding; this is done prior to submission to Academic Council

COURSE NAME: *Introduction to Data Science and Machine Learning*

TOTAL NO. OF CONTACT HOURS: 45

CREDITS: 3

PREREQUISITES: CS 160 and MA 100/101

COURSE DESCRIPTION

This course introduces students to the main concepts of data science, focusing on the practical application of machine learning and deep learning models for classification and prediction. The course explores the statistical foundations, computational techniques, and ethical considerations essential for building and deploying effective AI solutions. Through a project-based learning approach, students will gain hands-on experience applying data science methodologies to real-world problems. Using open, pre-existing datasets, they will learn to formulate, execute, and evaluate data science projects. Topics covered include descriptive statistics, elementary probability theory, basics of linear algebra; supervised and unsupervised learning, parametric and nonparametric decision-making; AI bias, fairness and accountability.

SUMMARY OF COURSE CONTENT

- Python libraries for data science
- Introduction to methodology for data science projects
- Data cleaning and preparation
- Exploratory data analysis and visualizations
- Modeling: unsupervised, supervised, and deep learning
- Performance assessment and model selection
- Bias in ML

LEARNING OUTCOMES

Students completing this course will be able to:

- Analyze data using statistical methods and programming tools
- Identify the appropriate machine learning task for a given problem
- Apply machine learning techniques to solve problems
- Communicate data-driven insights effectively

TEXTBOOK

- *An Introduction to Statistical Learning*, by G. James, D. Witten, T. Hastie, R. Tibshirani, J. Taylor. Springer - available for download at <https://www.statlearning.com>. 978-3-031-38747-0
- Data mining with Python : theory, application, and case studies, by Di Wu. CRC Press. 9781040010402

GRADING POLICY

Assessment methods:

Assignment	Guidelines	Weight
Home assignments	Four assignments based on reporting coding debugging and interpretation of results	20
Quizzes	Three in-class quizzes, each on the most recent topics	30
Midterm exam	One in-class assessment at the end of the first half of the course	20
Final project	Implementation of a data science project, to verify the knowledge acquired by the students in the course. Students will be asked to present their project in class.	20
Attendance and participation	Attendance and participation are fundamental, due to the creative and theoretical nature of the topics.	10

Assessment criteria:

Grade A characteristics:

Work of this quality directly addresses the question or problem raised and provides a coherent argument displaying an extensive knowledge of relevant information or content. This type of work demonstrates the ability to critically evaluate concepts and theory and has an element of novelty and originality. There is clear evidence of a significant amount of reading beyond that required for the course

Grade B characteristics:

This is highly competent level of performance and directly addresses the question or problem raised. There is a demonstration of some ability to critically evaluate theory and concepts and relate them to practice. Discussions reflect the student's own arguments and are not simply a repetition of standard lecture and reference material. The work does not suffer from any major errors or omissions and provides evidence of reading beyond the required assignments

Grade C characteristics:

This is an acceptable level of performance and provides answers that are clear but limited, reflecting the information offered in the lectures and reference readings.

Grade D characteristics:

This level of performances demonstrates that the student lacks a coherent grasp of the material. Important information is omitted and irrelevant points included. In effect, the student has barely done enough to persuade the instructor that s/he should not fail.

Grade F characteristics:

This work fails to show any knowledge or understanding of the issues raised in the question. Most of the material in the answer is irrelevant.

Grade scale

A	=	94- 100%	B	=	84-86%	C-	=	70-73%
A-	=	90-93%	B-	=	80-83%	D+	=	67-69%
B+	=	87-89%	C+	=	77-79%	D	=	60-66%
			C	=	74-76%	F	=	0-59%

ATTENDANCE REQUIREMENTS:

Attendance is to be considered mandatory and will be part of the final grade. Students will be granted 2 absences without penalty. Any other absences will only be excused with medical certificates or permission from the Dean's Office.

Examination policy

A major exam (midterm or final) cannot be made up without the permission of the Dean's Office. The Dean's Office will grant such permission only when the absence was caused by a serious impediment, such as a documented illness, hospitalization or death in the immediate family (in which you must attend the funeral) or other situations of similar gravity. Absences due to other meaningful conflicts, such as job interviews, family celebrations, travel difficulties, student misunderstandings or personal convenience, will not be excused. Students who will be absent from a major exam must notify the Dean's Office prior to that exam. Absences from class due to the observance of a religious holiday will normally be excused. Individual students who will have to miss class to observe a religious holiday should notify the instructor by the end of the Add/Drop period to make prior arrangements for making up any work that will be missed.

ACADEMIC HONESTY

As stated in the university catalog, any student who commits an act of academic dishonesty will receive a failing grade on the work in which the dishonesty occurred. In addition, acts of academic dishonesty, irrespective of the weight of the assignment, may result in the student receiving a failing grade in the course. Instances of academic dishonesty will be reported to the Dean of Academic Affairs. A student who is reported twice for academic dishonesty is subject to summary dismissal from the University. In such a case, the Academic Council will then make a recommendation to the President, who will make the final decision

STUDENTS WITH LEARNING OR OTHER DISABILITIES

John Cabot University does not discriminate on the basis of disability or handicap. Students with approved accommodations must inform their professors at the beginning of the term. Please see the website for the complete policy

SCHEDULE

Week 1	Introduction to data science projects Python libraries for data science Exploratory data analysis.	
Week 2	Data cleaning and preprocessing of data. Supervised learning- Regression: linear regression	
Week 3	Supervised learning - Classification: logistic regression, decision trees, k-nearest neighbors	
Week 4	Deep Learning - Artificial Neural networks with various architectures	
Week 5	Unsupervised learning: clustering ML bias	

OVERVIEW OF KEY BIBLIOGRAPHIC WORKS FOR THE COURSE

R for Data Science, by H. Wickham, M. Çetinkaya-Rundel, and G. Grolemund. O'Reilly Media - available for download at <https://r4ds.hadley.nz/>. 978-1492097402

Python Data Science Handbook, by Jake VanderPlas. O'Reilly (available online at <https://jakevdp.github.io/PythonDataScienceHandbook/>). 978-1491912058