



ATENEO DE MANILA  
UNIVERSITY  
Loyola Schools

SYLLABUS

A. COURSE INFORMATION

COURSE NUMBER	CSCI 271			NO. OF UNITS	3
COURSE TITLE	DATA MINING				
PREREQUISITE/S	Basic programming experience				
DEPARTMENT/ PROGRAM	DISCS			SCHOOL	SOSE
SCHOOL YEAR	2024-2025			SEMESTER	2nd
INSTRUCTOR/S	John Paul C. Vergara				
VENUE	CTC 201B	SECTION	YW	SCHEDULE	W 1700-2000

B. COURSE DESCRIPTION

This course provides a comprehensive overview of approaches used to analyze structured and unstructured data. It includes the following: preliminary data visualization and pattern recognition, clustering, classification, regression, feature extraction analysis.

C. \*PROGRAM LEARNING OUTCOMES

*\*This section does not apply to a student who is not a major of the program under which this course is administered*

PROGRAM LEARNING OUTCOMES
PLO1: Perform independent, structured, multidisciplinary, and ethical research that is fact-based, verifiable, and publishable
PLO2: Recognize, interpret, and apply advanced concepts in Computer Science and ICT
PLO3: Initiate collaboration with domestic and/or international research teams
PLO4: Recognize socially-relevant problems that could be addressed via ICT, while keeping ethical implications in mind
PLO5: Design and devise effective solutions to unstructured, real-world problems via computational- and/or ICT-based approaches

### Alignment of the Course to the Program Learning Outcomes

	PLO1	PLO2	PLO3	PLO4	PLO5
CLO1		✓			
CLO2		✓			
CLO3		✓			
CLO4		✓			
CLO5		✓			
CLO6		✓			
CLO7	✓				✓

### D. COURSE LEARNING OUTCOMES

By the end of this course, students should be able to:

COURSE LEARNING OUTCOMES
CLO1: Understand data mining in the context of the data science umbrella.
CLO2: Compute for optimal clusters given an unlabeled dataset using various unsupervised learning approaches.
CLO3: Apply dimensionality reduction methods such as principal component analysis and <i>t</i> -SNE and provide interpretations.
CLO4: Label datapoints according to given categories using logistic regression.
CLO5: Elaborate on statistical classification using alternative classification algorithms.
CLO6: Create linear regression models and use the concept of regularization to improve performance.
CLO7: Apply the conceptual discussions through more general machine learning approaches, business applications and case studies.

**E. COURSE OUTLINE and LEARNING HOURS**

<b>Course Outline</b>	<b>CLOs</b>	<b>Estimated Contact or Learning Hours</b>
Introduction <ul style="list-style-type: none"><li>- Course Overview</li><li>- Data Mining and Machine Learning</li><li>- CRISP-DM</li></ul>	CLO 1,7	3
Classification <ul style="list-style-type: none"><li>- Nearest neighbors</li><li>- Decision trees</li><li>- Naïve-Bayes</li><li>- Selected Classification Methods (Reports)</li><li>- Validation</li><li>- Implementation</li></ul>	CLO 4,5,7	18
Regression <ul style="list-style-type: none"><li>- Linear Regression</li><li>- Decision Tree Regression</li><li>- Implementation</li></ul>	CLO 6,7	15
Midterm Reports and Midterm Exam	CLO 1,4-7	15
Clustering <ul style="list-style-type: none"><li>- K-means</li><li>- Hierarchical clustering</li><li>- Selected Clustering Methods (Reports)</li><li>- Scoring and validation</li><li>- Implementation</li></ul>	CLO 2,7	15
Principal Component Analysis <ul style="list-style-type: none"><li>- Concepts</li><li>- Implementation</li></ul>	CLO 3	6
Association Rules Mining <ul style="list-style-type: none"><li>- Concepts</li><li>- Implementation</li></ul>	CLO 7	9
Final Reports and Final Exam	CLO 2,3,7	9

## F. ASSESSMENTS AND RUBRICS

Assessment Tasks	Assessment Weight	CLOs
Homework and Quizzes	20%	CLO1 to CLO7
Projects/Reports	30%	CLO1 to CLO7
Midterm Exam	25%	CLO1, CLO2, CLO3, CLO4, CLO7
Final Exam	25%	CLO1, CLO5, CLO6, CLO7

### RUBRICS:

Excellent:	Demonstrates thorough understanding of all concepts 95-100% of questions answered correctly Programs fully compliant with specifications, programs working on all provided test cases
Very Good:	Demonstrates thorough understanding of most concepts 85-94% of questions answered correctly Projects compliant with specifications, programs working on most test cases
Good:	Demonstrates reasonable understanding of critical concepts 75-84% of questions answered correctly, with some questions partially correct Projects generally compliant with specifications, programs working on important test cases
Satisfactory:	Demonstrates some understanding of critical concepts 65-74% of questions answered correctly, with some questions partially correct Projects compliant with essential specifications, program working on some test cases
Fail:	Demonstrates little or no understanding of critical concepts Less than 65% of questions answered correctly Projects not compliant with essential specifications

## G. TEACHING and LEARNING METHODS

TEACHING & LEARNING METHODS & ACTIVITIES	CLOs
Lectures	CLO1 to CLO7
Quizzes, Seatwork, Homework	CLO1 to CLO7
Projects, Reports	CLO1 to CLO7
Exams	CLO1 to CLO7

## H. REQUIRED READINGS

- Lecture slides and any reading material posted in our Canvas modules
- Han, Kamber, Pei, Data Mining: Concepts and Techniques 3<sup>rd</sup> Ed, 2012
- Russell, Stuart J., Norvig, Peter, Artificial Intelligence: A Modern Approach, 4<sup>th</sup> Ed, 2020

## I. SUGGESTED READINGS

- Hastie, Trevor, et. al . An Introduction to Statistical Learning. Springer, 2009.
- Kevin. Machine Learning: A Probabilistic Perspective. The MIT Press, 2011.
- Deisenroth, A Aldo Faisal, and Cheng Soon Ong. Mathematics of Machine Learning. Cambridge University Press, TBP.

## J. GRADING SYSTEM

93-100	A	Excellent
87-92	A-	Very Good
81-86	B+	Good
75-80	B	Fair
69-74	C	Unsatisfactory (no graduate credit)
<69	F	Failure (no graduate credit)

## K. CLASS POLICIES

- This course shall have synchronous sessions once a week (approximately one hour per week), but students should plan to spend at least 3 hours a week on the material posted. I will ask students to report on specific topics.
- Projects, reports, and other work shall be submitted through the Canvas online system, unless otherwise indicated. Deadlines will be indicated, although some leniency will be applied due to difficulties stemming from the pandemic.
- There will be around 6 quizzes given. The quiz with the lowest score will be dropped.
- Cheating will not be tolerated. Cheating in any requirement will result in a *minimum* penalty of having a grade of 0 for that requirement, and will be reported to the appropriate authorities, as provided for by the Student Handbook. Duplicate work will merit penalties for *both* the student who copied and the student from whom the work was copied.
- Students are expected to comply with the DISCS Academic Integrity Policy. With each submission, students must include a certification that their work is substantially their own and not copied from others. In addition, students must clearly acknowledge and specify any help from outside sources such as other classmates, the Web, books, etc., that they received while doing their projects. Failure to acknowledge such may be interpreted as intellectual dishonesty. Refer to the following references for DISCS-related policies:  
[DISCS Policies 2024-2](#)  
[DISCS Academic Integrity Policy](#).
- AI Policy: Unless otherwise specified, students may use artificial intelligence platforms such as ChatGPT. Even if permissible, students are required to disclose the use of such tools and cite them appropriately. Disclosure should come in the form of a formal declaration within the work submitted. Any use of AI without full disclosure falls under plagiarism, cheating, and dishonesty and will be penalized as such. This is to encourage the responsible and efficient use of AI while upholding integrity, transparency, and excellence.
- Participants (students and instructor) are expected to conduct themselves and treat others with respect, ensuring an inclusive and safe learning environment. Participants of this class are expected to abide by the LS Gender Policy and the Code of Decorum and Administrative Rules on Sexual Harassment, Other Forms of Sexual Misconduct, and Inappropriate Behavior. Any form of discrimination (based on sex, gender, marital status, parental status, sexual orientation, gender identity and expression) has **no place** in this class. See the following links for more information:
  - <https://www.ateneo.edu/central/policies/he-gender>
  - <https://www.ateneo.edu/college/current-students/handbook>
- Additional policies, with due consultation with the students, may be implemented by the teacher to adapt to the class environment, and will be posted on the Canvas site. Students are advised to be aware of such updates.

## H. CONSULTATION HOURS

NAME OF FACULTY	EMAIL	DAY/S	TIME
John Paul Vergara	jpvergara@ateneo.edu	MTh W	2-3pm 4-5pm