



COURSE OUTLINE

Course identification

Name of programs – Codes:	COMPUTER SCIENCE TECHNOLOGY – ARTIFICIAL INTELLIGENCE AND MACHINE LEARNING – LEA.DQ COMPUTER SCIENCE TECHNOLOGY – MACHINE LEARNING – LEA.DR
Course title:	ADVANCED DATA MANAGEMENT
Course number:	420-A15-AS
Total number of course hours:	75 hours
Weighting:	2-3-3
Statement of the competency – Code:	Manage structured data – KP57 Manage unstructured data – KP58 Apply machine learning algorithms for supervised learning – KP59

Contribution of the course in the program

Course position

This course is located in the third semester of *Computer Science Technology – Artificial Intelligence and Machine Learning* (LEA.DQ) program and the second semester of *Computer Science Technology – Machine Learning* (LEA.DR) program. Its duration is 75 hours divided into 30 hours of theory and 45 hours of exercises plus approximately 45 hours of homework.

In LEA.DQ, it shares the development of KP57 and KP58 competencies with *Introduction to Artificial Intelligence* (420-A06-AS) offered in the first semester, *Applied Machine Learning I* (420-A14-AS) offered in the third semester, and *AI Applications in Industry* (420-A23-AS) and *Internship* (420-A24-AS) offered in the fifth semester. It also shares the development of KP59 competency with *Applied Machine Learning I* (420-A14-AS) and *Machine Learning and Neural Networks* (420-A16-AS) offered in the third semester, and *Applied Machine Learning II* (420-A18-AS) offered in the fourth semester.

In LEA.DR, it shares the development of KP57 and KP58 competencies with *Introduction to Artificial Intelligence* (420-A06-AS) offered in the first semester, *Applied Machine Learning I* (420-A14-AS) offered in the second semester, and *AI Applications in Industry* (420-A23-AS) offered in the third semester. It also shares the development of KP59 competency with *Applied Machine Learning I* (420-A14-AS) and *Machine Learning and Neural Networks* (420-A16-AS) offered in the second semester, and *Applied Machine Learning II* (420-A18-AS) offered in the third semester.

There are no prerequisites. Finally, even if it is not a prerequisite to any other course, the knowledge and skills developed in this course will be revisited in other courses of the two programs.

Scope of the course

During this course, the students will learn how to extract value from the datasets containing heterogeneous data from diverse sources (e.g., system logs, sensor networks, databases, medical records, tweets, etc.). They will first require data preparation to clean, transform, and enrich the raw data into a structure that is desirable for analysis. In this course, we will explore data-intensive applications, and the algorithms that are needed for different data management tasks. This course focuses on the fundamentals of data management for the purposes of producing information and supporting advanced techniques of business analytics

Course topics will include data preparation for analysis, data cleaning for modeling and data visualization for analytics and evaluation purposes. Using a combination of theory and practical exercises and case studies, the learner will develop the data acquisition, preparation and preliminary analysis skills that are a necessary pre-requisite to applying advanced business analytics to their data.

Upon completion of this course, the students will be able to demonstrate the ability to prepare, explore and validate sample data for business analysis and apply advanced data analysis tools and techniques towards the development of intelligent decision-making tools.

Course components (objective and standard of the competency)

Expected outcomes (achievement context of the competency)

The achievement context of this competency will reflect the conditions as they occur in the following settings: academic, professional, work, or life environment.

Achievement context appropriate to the competency – Manage structured data – KP57

- Based on the requirements.
- Based on structured data.
- Based on the appropriate algorithms.
- Using software, devices, and peripherals.
- In accordance with privacy policy.

Achievement context appropriate to the competency – Manage unstructured data – KP58

- Based on the requirements.
- Based on unstructured data.
- Based on the appropriate algorithms.
- Using software, devices, and peripherals.
- Using information sources.
- Using image processing libraries.
- Using image transformation techniques.
- Using convolutional neural networks for visual data.
- In accordance with the privacy policy.

Achievement context appropriate to the competency – Apply machine learning algorithms for supervised learning – KP59

- Using reference datasets
- Using machine learning libraries
- Using prototyping tools
- From experiments on GPUs or on the cloud
- In compliance with the code of ethics

Throughout the course, you will engage in various learning situations/activities so that by the end of the course, you will have met the expected outcomes.

Elements and performance criteria

The elements of an objective formulated in terms of the competency specify its essential components. They include only what is necessary in order to understand and master the competency. If the competency is described as a process, the elements are the steps for execution.

The performance criteria are the specific pre-established requirements upon which you and your teacher can objectively judge your development of the targeted competency. They are part of the description of this competency. They are prescriptive.

Sometimes an element appears in more than one course. If this is the case, a number indicates its complexity level: level one (1) being the simplest, level two (2), average, and level three (3), advanced, at the ministerial level.

Below are the elements of the competency and performance criteria for this course that are to be respected:

<p>Competency: Manage structured data – KP57</p> <p>General ministerial and institutional performance criteria:</p> <ul style="list-style-type: none"> – Autonomy – Initiative – Analytical mind and critical thinking 	
Elements of the competency	Performance criteria specific to each element
1. Analyze the data sources.	1.1 Exhaustive identification of data sources. 1.2 Accurate identification of data types. 1.3 Accurate analysis of validity and reliability of the data. 1.4 Effective analysis of the cost of obtaining data. 1.5 Proper determination of data confidentiality policies.
2. Collect the data.	2.1 Efficient identification of data acquisition methods. 2.2 Efficient construction of queries to extract data. 2.3 Correct execution of queries in big data frameworks. 2.4 Proper identification of MapReduce operations.
4. Finalize the preparation of the data.	4.1 Proper documentation of performed operations. 4.2 Proper identification of the data structure. 4.3 Accurate documentation of dataset characteristics. 4.4 Efficient implementation of data access object or API. 4.5 Proper visualization of data distribution.

Competency: Manage unstructured data – KP58

General ministerial and institutional performance criteria:

- Autonomy
- Initiative
- Analytical mind and critical thinking

Elements of the competency	Performance criteria specific to each element
1. Identify the source of the data.	1.1 Accurate analysis of the specifications of the data model. 1.2 Appropriate analysis of the documentation. 1.3 Efficient selection of the data. 1.4 Correct documentation of the findings.
2. Collect the data.	2.1 Proper identification of the data format. 2.2 Proper identification of the file system. 2.3 Efficient identification of data acquisition methods. 2.4 Efficient extraction of data samples. 2.5 Efficient storage of the data.
3. Clean the data.	3.1 Proper identification of corrupted data samples. 3.2 Proper identification of mislabelled data samples. 3.3 Efficient categorization of data samples. 3.4 Efficient implementation of automatic data correction procedures.
6. Finalize the preparation of the data.	6.1 Proper documentation of performed operations. 6.2 Proper identification of the data structure. 6.3 Accurate documentation of dataset characteristics. 6.4 Proper visualization of data distribution. 6.5 Proper application of parallel loaders for image data.

Competency: Apply machine learning algorithms for supervised learning – KP59

General ministerial and institutional performance criteria:

- Autonomy
- Initiative
- Analytical mind and critical thinking

Elements of the competency	Performance criteria specific to each element
9. Propose learning models.	9.1 Complete identification of effective learning algorithms with respect to the dataset. 9.2 Proper selection of performance evaluation measures. 9.3 Proper implementation of training and validation. 9.4 Proper visualization of loss and evaluation metrics for training and validation. 9.5 Efficient optimization of hyperparameters.

Course content/main themes

Listed below is the **essential** content to be covered in this course:

- Data retrieving from multiple sources (Web Services, CSV, Excel, Relational DB, etc.) in Python
- Descriptive data analysis in Python
- Data cleaning and preparation
- Data manipulation
- Data transformation
- Data quality checking
- Data sampling methods
- Feature engineering
- Multithreading and concurrent data processing
- Applying data for machine learning training, evaluating and testing
- Introduction to data visualization

Learning activities

Provided below are examples of learning activities that correspond to the competencies for this course. The learning activities are found in the course calendar that complements this course outline.

- Lab practice after teacher demonstration
- Problem solving
- Teamwork
- Case studies
- Project based learning
- Peer exchange

Terms for Evaluating Learning

The evaluation of your learning is based on two inseparable methods: formative evaluation and summative evaluation. These two evaluation types are formal. Detailed information on the evaluation schedule is found in the course calendar, under the “Formative and summative evaluation schedule” column.

Formative evaluation

Following a learning activity or learning period, time is set aside for introspection. You will determine what has been understood and achieved and seek to identify the nature and origin of weak areas. These designated periods consist of simple means: short tests, association games, logbooks, a portfolio, questions, creating of samples, etc.

Formative evaluation is frequent and covers as many aspects as possible. It takes place in class, individually or in groups, and leads to immediate decisions. **You are the one who assumes the bulk of the work during individual or group corrections, adjustments and other self-evaluation tasks. The purpose is not to determine grades.**

If you take the results of the formative evaluations seriously throughout the course, you will ensure preparedness for the summative evaluations. You will be able to make the necessary progress to acquire the targeted competency at the required level, according to the achievement context and pre-established performance criteria.

Below are some examples of formative evaluation methods that correspond to the targeted competencies for this course:

- Teacher Demonstration and feedback
- Homework and hands-on lab exercises
- Group work
- Problem solving
- Teacher’s feedback on every activity

Summative evaluation

Summative evaluations are less frequent. They take place later on, towards the middle and end of the semester. This gives you the time to integrate your learning and to learn how to apply it to situations related to the targeted competency. The summative evaluation material is prepared by your teacher according to the description of the course’s targeted competency: its elements, achievement context and performance criteria.

The work completed in summative evaluations is graded. The purpose is to determine what you have learned.

Below is the information on the summative evaluation schedule and details for this course, as well as the weighting of marks:

Evaluations	Weighting
Midterm Exam	30%
Project	30%
Final Exam	40%
Total	100%

Institutional requirements

Student's commitment

By registering for this course, you commit to:

- *obtain the necessary course materials at the start of the semester;*
- *respect the copyright;*
- *participate in the learning activities, formative and summative evaluation activities outlined in the course calendar;*
- *complete the work assigned to you;*
- *submit the work on time.*

Teacher's commitment

Your teacher commits to:

- *create varied learning situations that enable you to put into practice the knowledge, actions and professional behaviour of the targeted competency;*
- *plan sufficient and appropriate formative evaluation activities, involving correction and improvement, that provide frequent feedback, allowing you to be well informed of your progress;*
- *provide summative evaluations that correspond to the course's targeted competency;*
- *evaluate work according to the applicable criteria, in a fair and equitable manner within a reasonable time.*

The Institutional Policy on Evaluating Learning (IPEL) is applied to all institutional programs. Listed below are a few of its clauses:

Written language (article 5.7)

The teacher is responsible for identifying spelling and grammar errors and for allocating the corresponding number of marks for any given summative evaluation.

Below is the % – based on language requirements – that can be attributed to each summative evaluation:

- *up to 5 %*

Class attendance (article 5.12)

Attendance and participation in classes and evaluations are mandatory for all students.

The teacher has the responsibility of monitoring attendance and of evaluating the reasons justifying student absences from classes.

A student whose absences exceed the allowable number for the course could be denied access to the final exam for that course.

Plagiarism and fraud (article 5.16)

Plagiarism, attempted plagiarism or complicity in plagiarism during an assignment or any evaluated task contravenes the rules. This includes (but is not limited to):

- *the whole or partial presentation (reference, paraphrase, summary, translation, insertion) of the work of another (text, illustration, film, music, etc. on paper or online) as one's own, or failing to cite a source;*
- *the use of another student's exam during an exam;*
- *the use of an assignment done for another course or a project already submitted in the past, which is passed off as an original work.*

Fraud, attempted fraud or complicity in fraud constitutes an infraction.

This includes (but is not limited to):

- *the possession or use of any unauthorized document, material or equipment during an exam, including the use of technological tools;*
- *the execution of an evaluated task by another person;*
- *the substitution for another person during an exam, assignment or any evaluated task;*
- *the possession of the questions or answers of the exam;*
- *the obtainment of any aid not authorized in advance by the teacher.*

Plagiarism, attempts at plagiarism or fraud, or collaboration in plagiarism or fraud are prohibited and considered serious offences. Thus, any instances of plagiarism or fraud will lead to a grade of '0' for the assignment in question. In addition, a note will be made in the student's file and the student will receive a written notice from his or her Program Directorate to that effect.

In the case of recidivism, in the same course or in another course, the student will be given a grade of '0' for the course in question. A second note is made in the student's file and the student will receive a summons from his or her Program Directorate. For a third offence, he or she may be expelled from the College.

Submission of work and tests (article 5.8)

All assignments must be submitted in class at the time designated by the teacher. Any late submissions result in a grade of zero (0).

Upon presentation of an official supporting document or valid reason for the absence, the student may request an extension from the teacher, who may accept or refuse the student's work and apply a penalty for the lateness.

Program Directorates do not accept student work. Assignments must be submitted directly to the teacher.

Rules and regulations to follow

Late arrivals

The teacher may refuse to admit to the classroom any student arriving late. A late arrival is considered an absence for that period.

Note: Students arriving late must recognize that the information they missed will not be repeated. Late students are therefore responsible for asking their peers about the material they missed. Arriving after the break, as well as leaving before the end of the class, may result in one or more hours of absence.

Eating and drinking in class

Eating and drinking are prohibited in the classrooms, locker rooms and Documentation Centre. Food may only be eaten in the cafeteria, vending machine areas and student lounges.

Mandatory course material

- Laptop with specifications mentioned on the college's website.
- LaSalle College. Bring Your Own Device. 2017. < <http://www.lasallecollege.com/future-students/bring-your-own-device> >

Bibliography for this course

- Material will be provided by the teacher via Omnivox
- MCKINNEY, W. *Python for Data Analysis: Data Wrangling with Pandas, NumPy, and IPython*. United States: O'Reilly Media, 2017.
- HEYDT, M. *Learning Pandas*. Second Edition. India: Packt Publishing, 2017.

Academic Studies Directorate approval:
