Western Engineering

DEPARTMENT OF ELECTRICAL AND COMPUTER ENGINEERING

AISE 3010 – Data Engineering and Machine Learning

COURSE DESCRIPTION: The aim of this introductory course is to provide a solid background in the fundamentals of data engineering and machine learning. The course covers 1) Introduction to data engineering methods include processing streaming data, data pipelines with CUDA C, big data capability on google cloud platform (GCP), etc. 2) involves recent state-of-the-art deep networks and its learning algorithms for applications related to pattern recognition, feature extraction, image processing, and data argumentation. Several recent classic deep learning models will also be included for further discussion, including AlexNet, VGG16/19, ResNet, Transformer, and Generative adversarial networks.

ACADEMIC CALENDAR:
https://www.westerncalendar.uwo.ca/Courses.cfm?CourseAcadCalendarID=MAIN_030385_1&SelectedCalendar=Live&ArchiveID=

The aim of this introductory course is to provide a solid background in the fundamentals of data engineering and machine learning. The course covers 1) Introduction to data pipelines, distributed data management, and streamline data processing; 2) Data manipulation and data structure for big data; and 3) Design and implementation of an engineering group project illustrating the machine learning and data engineering concepts being taught.

PRE OR COREQUISITES: Undergraduate courses in DS3000A/B, AISE3309A/B (or SE3309A/B), AISE2205A/B (or SE2205A/B), and AISE2251A/B (or SE2251A/B).

CEAB ACADEMIC UNITS: Engineering Science 100%

CONTACT HOURS: Timetable information is available at https://draftmyschedule.uwo.ca/.

LECTURE:	3 lecture hours/week
LAB:	2 lab hours/week

Required Textbook:

1. Adi Wijaya, Data Engineering with Google Cloud Platform, Packt Publishing,2022. Cost: $73.00
2. I. Goodfellow, Y. Bengio, A. Courvile, Deep Learning, MIT Press, 2016. Cost: $112.00

Other Required References:

1. Google Cloud Platform
2. Google Colab
3. Pytorch Tutorial: https://pytorch.org/tutorials/

Recommended References:

1. S. Haykin, Neural Networks and Learning Machine, Pearson, 2008.
   Cost: $220.00

GENERAL LEARNING OBJECTIVES (CEAB GRADUATE ATTRIBUTES)

Knowledge Base	A	Engineering Tools	A	Impact on Society	
Problem Analysis	A	Individual & Teamwork	D	Ethics and Equity	D
Investigation		Communication		Economics and Project Mgmt.	
Design		Professionalism		Life-Long Learning	

Notation: x represents the content level code as defined by the CEAB. blank = not applicable; I = introduced (introductory); D = developed (intermediate) and A = applied (advanced).

Rating: I – The instructor will introduce the topic at the level required. It is not necessary for the student to have seen the material before. D – There may be a reminder or review, but the student is expected to have seen and been tested on the material before taking the course. A – It is expected that the student can apply the knowledge without prompting (e. g. no review).

COURSE MATERIALS: Weekly content and guides for the laboratories will be available on the course OWL site. The material for this course will be taught in both lectures and labs; therefore, it is imperative that you attend each lecture and lab.

UNITS: SI

Course Topics and Specific Learning Outcomes	**CEAB Graduate Attributes Indicators**
1. Shadow networks and its learning methods	**ET2**
At the end of this section, students will be able to:	**KB1**
a. understand fundamental concepts of neural network	
b. describe in detail the specific shadow networks taught in class.	
c. understand automated and unsupervised labelling methods.	
d. understand regularization theory.	
e. write a program to implement algorithms taught in class for pattern recognition problems.	
2. Deep convolutional neural networks	**EE4**
At the end of this section, students will be able to:	**ET2**
a. identify broad categories of deep neural networks for variety tasks.	**KB3**

	b. understand fundamental concepts of deep network layers. **c.** understand concepts of transfer learning, and training from scratch. **d.** understand how to apply them to process big datasets in pattern recognition problems in GPU based environment. **e.** develop algorithms based on deep convolutional neural networks for performing tasks such as image processing, object detection, and data augmentation. **f.** understand its limitation related to awareness of the principles of equity.	**PA2** **ITW2**
3.	**Data engineering with GCP & CUDA** At the end of this section, students will be able to: **a.** understand concepts of data life cycle **b.** describe how to develop a data warehouse with GCP tools. **c.** understand how to practice machine learning code using python in GCP. **d.** understand how to practice basic machine learning code using CUDA C for further GPU based optimization. **e.** understand how to leverage pre-built GCP AI models. **f.** understand how to process streaming data with dataflow. **g.** write a program to implement algorithms taught in class for pattern recognition problems.	**PA2** **KB3** **ET3**
4.	**Deep learning with large language models** At the end of this section, students will be able to: **a.** describe how to load and use pre-trained large language models. **b.** describe in detail the specific large language models taught in class. **c.** appropriately configure the algorithm according to the task including computer vision and language processing. **d.** understand the challenges associated with the ethical behavior of methods in applications.	**PA2** **ET1** **EE2** **KB1** **ITW2**

Evaluation

Course Component	Weight
Laboratory Assignments	4*5%
Final Project	10%
In-class random Quiz	10%
Midterm Test	20%
Final Examination	40%

Laboratory Assignments: There are 4 laboratory assignments. Some assignments will be individual while some will be assigned in group (up to two). One report can be submitted by each group. Most of the assignments will involve programming in Python or MATLAB. All assignments will be distributed

via OWL. All assignments are expected to be submitted via OWL by 11:55 pm on the due date. Each assignment is worth 5% of your overall grade.

In-class random Quiz: There will be 5–8 random quizzes conducted in class throughout the course, each worth approximately 1–2%. The primary purposes of these quizzes are to record attendance, review key concepts, and facilitate interactive discussion of challenging topics to enhance students' understanding of the course material. Students must attend all classes to earn these quiz marks. Quiz grading is based solely on completion, not on correctness.

If a student is absent during a quiz, no make-up opportunity will be provided, and the quiz score for that session will be recorded as 0.If a student needs to miss a class for a valid reason, they must notify the instructor by email at least 48 hours in advance.

Final Project: There is one final project which will combine two to three assignments to generate a project to deepen the understanding of our course knowledge. The project could be completed in group or individual. Several options will be provided. Student could pick one of them in order to consider the actual computation resource allocation.

Midterm Test: There will be one midterm exam that you are keeping up with the material being taught. They will take place during the normal lecture hour.

Final Examination: The final examination will be taking place during the regular examination period.

Missed Midterm Examinations: If a student misses a midterm examination, the exam will not be rescheduled. The student must follow the Instructions for Students Unable to Write Tests and provide documentation to their department within 24 hours of the missed test. The department will decide whether to allow the reweighting of the test, where reweighting means the marks normally allotted for the midterm will be added to the final exam. If no reasonable justification for missing the test can be found, then the student will receive a mark of zero for the test.

If a student is going to miss the midterm examination for religious reasons, they must inform the instructor in writing within 48 hours of the announcement of the exam date or they will be required to write the exam.

Late Submission Policy:

Please note that the assignment submission deadline includes flexibility in the form of a 48-hour submission window (grace period). As a result, the instructor reserves the right to deny any requests for academic consideration for assignments submitted after this grace period.

If students submit their assignments beyond the 48-hour grace period, a penalty of 10% per day will be applied for late submissions, up to a maximum of 3 days. After three days, late submissions will no longer be accepted.

Attendance: Any student who, in the opinion of the instructor, is absent too frequently from class, laboratory, or tutorial periods will be reported to the Dean (after due warning has been given). On the recommendation of the department, and with the permission of the Dean, the student will be debarred from taking the regular final examination in the course.