Course Outline

School:	Eng. Tech. & Applied Science
Department:	Information and Communication Engineering Technology (ICET)
Course Title:	Supervised Learning
Course Code:	COMP 247
Course Hours/Credits:	56
Prerequisites:	COMP 228, COMP 237, MATH 210
Co-requisites:	N/A
Eligible for Prior Learning, Assessment and Recognition:	Yes
Originated by:	mayy habayeb
Creation Date:	Fall 2020
Revision Date:	Summer 2021
Current Semester:	Winter 2024
Approved by:	ppesikan c/o

Chairperson/Dean

Students are expected to review and understand all areas of the course outline.

Retain this course outline for future transfer credit applications. A fee may be charged for additional copies.

This course outline is available in alternative formats upon request.

Acknowledgement of Traditional Lands

Centennial is proud to be a part of a rich history of education in this province and in this city. We acknowledge that we are on the treaty lands and territory of the Mississaugas of the Credit First Nation and pay tribute to their legacy and the legacy of all First Peoples of Canada, as we strengthen ties with the communities we serve and build the future through learning and through our graduates. Today the traditional meeting place of Toronto is still home to many Indigenous People from across Turtle Island and we are grateful to have the opportunity to work in the communities that have grown in the treaty lands of the Mississaugas. We acknowledge that we are all treaty people and accept our responsibility to honor all our relations.

Course Description

In this course, students will be introduced to supervised learning techniques and algorithms. Coursework cover the following algorithms: linear regression, logistic regression, decision trees, bayesian learning, support vector machines, sequence learning, k-nearest neighbors and ensemble techniques. The concepts of underfitting, overfitting, cross-validation, and kernel methods will be covered throughout the course. Students will practice building an end to end supervised learning project.

Program Outcomes

Successful completion of this and other courses in the program culminates in the achievement of the Vocational Learning Outcomes (program outcomes) set by the Ministry of Colleges and Universities in the Program Standard. The VLOs express the learning a student must reliably demonstrate before graduation. To ensure a meaningful learning experience and to better understand how this course and program prepare graduates for success, students are encouraged to review the Program Standard by visiting http://www.tcu.gov.on.ca/pepg/audiences/colleges/progstan/. For apprenticeship-based programs, visit http://www.collegeoftrades.ca/training-standards.

Course Learning Outcomes

The student will reliably demonstrate the ability to:

1. Explain the fundamental concepts and goals of supervised learning.
2. Explain and discuss the main challenges of machine learning.
3. Understand the life cycle of a machine learning project end-to end.
4. Practice and examine the elements of an analytic pre-processing pipeline: sampling, scaling, data transformation, sanity checking, data model development, evaluation, and scoring.
5. Design, code and test machine learning algorithms to solve problems and create predictive supervised learning models, using a variety of regression and classification algorithms (Linear regression, logistic regression, Naïve Bayes, Decision trees, support vector machines, ensembles of classifiers, k- nearest neighbors)
6. Discuss and explain the main concepts of Kernels.
7. Design, code, and test a “Web prediction service', that encapsulate and automate the core components of data transformation, model fitting, and scoring of new observations.
8. Evaluate various supervised learning algorithms to build software solutions for a variety of business problems.
9. Explore time series data and build forecasting models.

Essential Employability Skills (EES)

The student will reliably demonstrate the ability to*:

1. Communicate clearly, concisely and correctly in the written, spoken, and visual form that fulfills the purpose and meets the needs of the audience.
2. Respond to written, spoken, or visual messages in a manner that ensures effective communication.
3. Execute mathematical operations accurately.
4. Apply a systematic approach to solve problems.
11. Take responsibility for one's own actions, decisions, and consequences.

**There are 11 Essential Employability Skills outcomes as per the Ministry Program Standard. Of these 11 outcomes, the following will be assessed in this course.*

Global Citizenship and Equity (GC&E) Outcomes

N/A

Methods of Instruction

Engaging and interactive lecture content.

Lab demonstrations and tutorials.

Hands on practical lab exercises.

Interactive discussion forms and boards.

Team project.

Text and other Instructional/Learning Materials

Text Book(s):

1- Hands-On Machine Learning with Scikit-Learn and TensorFlow by Aurélien Géron

Publisher: O'Reilly Media, Inc. 3rd edition

Release Date: September 2022

Online: https://learning.oreilly.com/library/view/hands-on-machine-learning/9781098125967/?sso_link=yes&sso_link_from=centennial-college

2- Artificial intelligence a modern approach by Stuart J. Russell and Peter Norvig, forth edition Published by Pearson Education, Inc. ISBN-13: 978-0-13-461099-3 ISBN-10: 0-13-461099-7

3- Machine Learning for Time-Series with Python by Ben Auffarth Published by Packt Publishing 2021.

Available online: https://learning.oreilly.com/library/view/machine-learning-for/9781801819626/?sso_link=yes&sso_link_from=centennial-college

4- The Hundred-Page Machine Learning Book by Andriy Burkov

Online Resource(s):

Safari IT Books online

Online documentation of libraries: scklearn, anaconda, pandas, numpy, matplotlib

Material(s) required for completing this course:

Lecture and lab course html notes/python scripts and videos

Evaluation Scheme

- Lab assignments: Four Individual lab assignments, cover the following areas:

Lab assignment # 1: Classification
Lab assignment # 2: Support Vector Machines
Lab assignment # 3: Decision Tress
Lab assignment # 4: Ensemble learning and random forests

- Lab exercises: These are notebooks/scripts that contain the main tools/ libraries/ functions/classes required for data pre-processing and representation students are required to replicate the functionalities with minor amendments and submit as proof of work:
  Lab exercise #1 : Numpy
  Lab exercise #2 : Pandas
  Lab exercise # 3 : Matplotlib
  Lab exercise # 4 : Linear algebra (optional)
- Quizzes: Three short individual quizzes
- Online participation (discussion boards): Active significant participation on the discussion boards
- Test #1: Hands-On & theory covering Week 1 - 6 materials
- Group project assignment two parts: Students will work in groups to develop a complete end to end supervised learning project on a real dataset to solve a business problem for the public. Final deliverable is a predictive supervised learning model (classifier) and deployed as an analytics service API. Submission is in two parts. Part #1 will cover deliverables 1&2 and is worth 13%, due week #10. Part #2 builds on part #1 and is the full project i.e. all deliverables and worth 13% due end of course.

Evaluation Name	CLO(s)	EES Outcome(s)	GCE Outcome(s)	Weight/100
Lab assignments	1, 4, 5, 8	1, 3, 4, 11		20
Lab exercises	3, 4, 5	4, 11		6
Quizzes	1, 2, 3, 6, 8	1, 2		15
Online participation (discussion boards)	1, 2, 8	1, 2		8
Test #1	1, 2, 4, 5, 6	1		25
Group project assignment two parts	3, 5, 7, 8, 9	1, 2, 11		26
Total				100%

If students are unable to write a test they should immediately contact their professor or program Chair for advice. In exceptional and well documented circumstances (e.g. unforeseen family problems, serious illness, or death of a close family member), students may be able to write a make-up test.

All submitted work may be reviewed for authenticity and originality utilizing Turnitin®. Students who do not wish to have their work submitted to Turnitin® must, by the end of the second week of class, communicate this in writing to the instructor and make mutually agreeable alternate arrangements.

When writing tests, students must be able to produce official Centennial College photo identification or they may be refused the right to take the test or test results will be void.

Tests or assignments conducted remotely may require the use of online proctoring technology where the student's identification is verified and their activity is monitored and/or recorded, both audibly and visually through remote access to the student's computer and web camera. Students must communicate in writing to the instructor as soon as possible and prior to the test or assignment due date if they require an alternate assessment format to explore mutually agreeable alternatives.

Student Accommodation

The Centre for Accessible Learning and Counselling Services (CALCS) (http://centennialcollege.ca/calcs) provides programs and services which empower students in meeting their wellness goals, accommodation and disability-related needs. Our team of professional psychotherapists, social workers, educators, and staff offer brief, solution-focused psychotherapy, accommodation planning, health and wellness education, group counselling, psycho-educational workshops, adaptive technology, and peer support. Walk in for your first intake session at one of our service locations (Ashtonbee Room L1-04, Morningside Room 190, Progress Room C1-03, The Story Arts Centre Room 285, Downsview Room 105) or contact us at calcs@centennialcollege.ca, 416-289-5000 ext. 3850 to learn more about accessing CALCS services.

Use of Dictionaries

- Any dictionary (hard copy or electronic) may be used in regular class work.

Program or School Policies

N/A

Course Policies

Late lab exercises and assignments will be accepted up to three calendar days beyond due date, and will bear a 10% late penalty reduction, per day.

College Policies

Students should familiarize themselves with all College Policies that cover academic matters and student conduct.

All students and employees have the right to study and work in an environment that is free from discrimination and harassment and promotes respect and equity. Centennial policies ensure all incidents of harassment, discrimination, bullying and violence will be addressed and responded to accordingly.

Academic Honesty

Academic honesty is integral to the learning process and a necessary ingredient of academic integrity. Forms of academic dishonesty include cheating, plagiarism, and impersonation, among others. Breaches of academic honesty may result in a failing grade on the assignment or course, suspension, or expulsion from the college. Students are bound to the College’s AC100-11 Academic Honesty and Plagiarism policy.

To learn more, please visit the Libraries information page about Academic Integrity https://libraryguides.centennialcollege.ca/academicintegrity and review Centennial College's Academic Honesty Module: https://myappform.centennialcollege.ca/ecentennial/articulate/Centennial_College_Academic_Integrity_Module_%202/story.html

Use of Lecture/Course Materials

Materials used in Centennial College courses are subject to Intellectual Property and Copyright protection, and as such cannot be used and posted for public dissemination without prior permission from the original creator or copyright holder (e.g., student/professor/the College/or third-party source). This includes class/lecture recordings, course materials, and third-party copyright-protected materials (such as images, book chapters and articles). Copyright protections are automatic once an original work is created, and applies whether or not a copyright statement appears on the material. Students and employees are bound by College policies, including AC100-22 Intellectual Property, and SL100-02 Student Code of Conduct, and any student or employee found to be using or posting course materials or recordings for public dissemination without permission and/or inappropriately is in breach of these policies and may be sanctioned.

For more information on these and other policies, please visit www.centennialcollege.ca/about-centennial/college-overview/college-policies.

Students enrolled in a joint or collaborative program are subject to the partner institution's academic policies.

PLAR Process

This course is eligible for Prior Learning Assessment and Recognition (PLAR). PLAR is a process by which course credit may be granted for past learning acquired through work or other life experiences. The PLAR process involves completing an assessment (portfolio, test, assignment, etc.) that reliably demonstrates achievement of the course learning outcomes. Contact the academic school to obtain information on the PLAR process and the required assessment.

Topical Outline (subject to change):

Week	Topics	Readings/Materials	Weekly Learning Outcome(s)	Instructional Strategies	Evaluation Name and Weight	Evaluation Date
1	Course Overview. Machine learning fundamentals and applications(review). Batch and Online Learning. Main Challenges of Machine Learning. Linear regression (review). Polynomial Linear Regression. Gradient Descent (review).	Chapters 1 & 4 Hands on machine learning with scikit-learn (Aurélien) Chapter 19 “Artificial intelligence a modern approach” (Russell & Norvig)	Explain and discuss the main challenges of machine learning. List various applications of machine learning. Explain and understand the math/logic behind linear regression models and model evaluation. Explain and understand the logic behind polynomial linear regression. Explain and discuss the concept of gradient descent algorithm (Batch, Stochastic & mini-batch). Design code and test and evaluate linear and polynomial linear regression models.	Lecture online content. Lab online tutorial. Videos. Discussion boards.	Lab exercise #1 due "Numpy" Lab exercise #2 Pandas made available.	end of week #1
2	Regularization. Classification (review). Logistic regression (review). Softmax regression. Multi-class classification strategies. Multi label classification. Confusion matrix. Receiver operating characteristic (ROC) (review).	Chapter 3 & 4 Hands on machine learning with scikit-learn (Aurélien). Chapter 19 “Artificial intelligence a modern approach” (Russell & Norvig).	Explain the concept of regularization to address overfitting and discuss the types of regularization Ridge Regression, Lasso Regression and Earlynet. Explain and understand the math/logic behind Logistic regression. Explain the concept of multiclass-classification in supervised learning. Explain and discuss the strategies to address multiclass classification. Explain the concept of Softmax regression (multinomial regression). Explain the concept of multi-labels in supervised learning. Explain and discuss the confusion matrix metrics and ROC ,AUC. Design, code, test and evaluate and multi-class models.	Lecture online content. Lab online tutorial. Videos. Discussion boards.	Lab exercise #2 "Pandas" due. Lab exercise #3 "Matplotlib" made available. Lab assignment #1 made available. Discussion board #1 made available.	end of week #2.
3	Support vector machines. Linear SVM Classification. Linear kernel.	Chapter 5 Hands on machine learning with scikit-learn (Aurélien). Chapter 19 “Artificial intelligence a modern approach”	Explain the concept of support vector machine(SVM). Explain and discuss the concept of Linear SVM classification. Explain and compare the difference between hard margin classification and soft margin classifications.	Lecture online content. Lab online tutorial. Videos. Discussion boards.	Lab exercise # 3 "MatplotLib" due. discussion board #1 due. Discussion board #2	end of week #3.

Week	Topics	Readings/Materials	Weekly Learning Outcome(s)	Instructional Strategies	Evaluation Name and Weight	Evaluation Date
		(Russell & Norvig).	Explain and discuss the concept of linear kernels. Explain and discuss the concept of Hinge loss. Design, code, test and evaluate a linear SVM classifier.		made available.	
4	The kernel trick. Nonlinear SVM. Classification. Polynomial Kernel. Gaussian RBF Kernel. SVM Regression.	Chapter 5 Hands on machine learning with scikit-learn (Aurélien). Chapter 19 “Artificial intelligence a modern approach” (Russell & Norvig).	Explain and discuss the concept of the kernel trick. Explain and discuss the concept of polynomial kernels. Explain and discuss the concept of the Gaussian RBF Kernel. Explain and discuss the concept of the SVM Regression. Design, code, test and evaluate a nonlinear SVM classifier.	Lecture online content. Lab online tutorial. Videos. Discussion boards.	Lab assignment #1 "Classification " due. Lab assignment #2 "Support vector machines" made available. Quiz #1	end of week #4
5	End to end machine learning project phases. Scikit-Learn’s API. Handling Text and Categorical Attributes. Stratified sampling. Transformation Pipelines. Feature scaling (review). Fine tuning models – Grid Search. Deployment and monitoring.	Chapter 2 Hands on machine learning with scikit-learn (Aurélien).	List and explain all phases of a machine learning project. Explain the importance of Use the pipeline class and apply it. Use various transformers. Discuss the need for stratified sampling and apply it. Apply transformation on categorical attributes using various transformer. Apply feature scaling to data sets. Fine tune a model using Grid search. Discuss deployment options. Build an end to end small machine learning project.	Lecture online content. Lab online tutorial. Videos. Discussion boards.	Discussion board #2 due. Discussion board #3 made available.	
6	Analytics as a web service. Lambda architecture. Sharing Models with Prediction Services. The architecture of a prediction service. Clients and Servers.	Lecture content	Explain the differences between “batch-oriented” and “stream processing”. Explain the lambda architecture. Explain and discuss how to build 'prediction services', web applications that encapsulate and automate the core components of data transformation, model fitting, and scoring of new observations.	Lecture online content. Lab online tutorial. Videos. Discussion boards.	Lab assignment #2 "Support vector machines" due.	

Week	Topics	Readings/Materials	Weekly Learning Outcome(s)	Instructional Strategies	Evaluation Name and Weight	Evaluation Date
	Serialization and deserialization of a predictive model. Flask.		Design, code and test machine learning web service.			
7	Review & Test (weeks 1-6) and introduction of group project	Chapters 1 - 5 Hands on machine learning with scikit-learn (Aurélien) Chapter 19 “Artificial intelligence a modern approach” (Russell & Norvig)	N/A	Live interaction session	Test #1. Discussion board #3 due.	
8	Decision Trees. The CART Training Algorithm. Entropy and information gain. Imbalanced datasets.	Chapter 6 Hands on machine learning with scikit-learn (Aurélien). Chapter 19 “Artificial intelligence a modern approach” (Russell & Norvig).	Explain and discuss the concept of Decision trees as classifiers. Explain the concept of Entropy and information gain. Explain the computational complexity of decision trees. Design, code, test and evaluate a decision tree classifier. Discuss and explain the problem of imbalanced datasets.	Lecture online content. Lab online tutorial. Videos. Discussion boards.	Quiz #2	
9	Gini impurity. Pruning decision trees. Regularization of hyper-parameters.	Chapter 6 Hands on machine learning with scikit-learn (Aurélien). Chapter 19 “Artificial intelligence a modern approach” (Russell & Norvig).	Explain and discuss the concept of Decision trees as a regression predictor. Explain and discuss the concept of Gini impurity. List all parameters of a decision tree and discuss how to fine tune these parameters to avoid over fitting. Explain the concept of pruning decision trees. Design, code, test and evaluate a decision tree for regression predictions.	Lecture online content. Lab online tutorial. Videos. Discussion boards.	Lab assignment #3 "Decision trees" made available.	
10	Ensemble Learning. Voting classifiers. Stacking. Boosting.	Chapter 7 Hands on machine learning with scikit-learn (Aurélien). Chapter 19 “Artificial intelligence a modern approach” (Russell & Norvig).	Explain and discuss the concept of ensemble learning across multiple hypothesises. List the benefits of using ensemble learning. Explain and discuss the concept of stacking in ensemble learning. Explain and discuss the concept of boosting in ensemble learning. Design, code, test and evaluate an ensemble voting classifier using different classification	Lecture online content. Lab online tutorial. Videos. Discussion boards.	Submit code/ report and present/demo part #1 of project assignment. Lab assignment #4 " Ensemble	During class for onsite. During virtual session for online.

Week	Topics	Readings/Materials	Weekly Learning Outcome(s)	Instructional Strategies	Evaluation Name and Weight	Evaluation Date
			algorithms.		learning and random forests" made available. Discussion board #4 made available.	
11	Bagging & Pasting. Random forests.	Chapter 7 Hands on machine learning with scikit-learn (Aurélien) Chapter 19 “Artificial intelligence a modern approach” (Russell & Norvig).	Explain and discuss the concept of bagging in ensemble learning. Explain and discuss the concept of random forest algorithms. Design, code, test and evaluate an ensemble bagging classifier using different classification algorithms.	Lecture online content. Lab online tutorial. Videos. Discussion boards.	Discussion board #4 due.	
12	Quantifying uncertainty. Joint probabilities. Naïve Bayes models(review). k-nearest neighbors.	Chapter 12 “Artificial intelligence a modern approach” (Russell & Norvig).	Explain and discuss the concept of generative models in machine learning. Explain and discuss the probability model Explain and discuss the concept of Naïve Bayes models. Explain and discuss the math logic behind k-nearest neighbors’ algorithm. Apply Naïve Bayes and k-nearest neighbors to a variety of business problems.	Lecture online content. Lab online tutorial. Videos. Discussion boards.	Quiz #3 Lab assignment #4: " Ensemble learning and random forests" due.	assignment end of week #12
13	Time series data Machine learning for time series Forecasting Course review.	Chapters 1,2,3,4,5 & 7 Machine Learning for Time-Series with Python (Ben auffarth)	Explain the concept of time series data, and use cases. Explore the most common statistical and machine learning approaches for time series data. Explore forecasting with Moving Averages and Autoregressive Models. Build forecasting models using time series data. Build machine learning models using time series data.	Lecture online content. Lab online tutorial. Videos. Discussion boards.	N/A	
14	Project Presentation	Lecture & Lab material	Develop a web analytics service.	Project presentation and code submission	Submit final code/report and present/demo	During class for onsite. During

Week	Topics	Readings/Materials	Weekly Learning Outcome(s)	Instructional Strategies	Evaluation Name and Weight	Evaluation Date
					part #2 of project assignment.	virtual session for online.