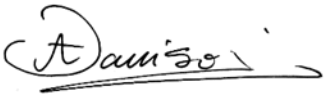
Approved by Chair:



Aug 20, 2020

Signature

COURSE SECTION INFORMATION

Applied Data Science

Teacher's Name Reza Dibaj

Email Reza.Dibaj@georgebrown.ca

Phone

Office

Out of Class Assistance

Course Code COMP 3122

Course Section

Academic Year 2020-2021

Term: Fall 2020

LIST OF TEXTBOOKS AND OTHER TEACHING AIDS:

Required:

1. Python Data Science Handbook: Essential Tools for Working with Data (by Jake VanderPlas)
   ISBN-10: 1491912057 **ISBN-13:** 978-1491912058

2. Free full text: under CC-BY-NC-ND license.
   GitHub repository: https://github.com/jakevdp/PythonDataScienceHandbook

Detailed Evaluation System

Assessment	Description	Outcome(s) assessed:	EES assessed:	Week	Weight
Lecture Quiz 5 x 2%	The best 5 out of 7 quizzes will count.	1,2,3,4,5,6	1,2,3,4,5	TBA	10%
Lab Test 2 x 6%	Hands-On test	1,2,3,4,5,6	1,2,3,4,5, 6,7,10	TBA	12%
Lab exercises	The best 8 out of 10	1,2,3,4, 5,6	1,2,3,4,5	TBA	8%
Assignment 5 x 5%	Five Individual assignment	1,2,3,4, 5,6	1,2,3,4,5,6, 7,8,9,10,11	TBA	25%
Mid-term Exam	Mixed format test on week 1 to week 7	1,2,3,4	1,2,4	9	20%
Final Exam	Mixed format test on week 1 to week 14	1,2,3,4,5,6	2,4,5	15	25%

Learning Schedule / Topical Outline (subject to change with notification)

TOPICAL OUTLINE:

WEEK	Topic	Outcome(s)	Content	Chapter/ Reference
1		1, 2	- Introduction to the field of Data Science and the relationship between AI, Machine Learning and Data Science - Historic overview - Overview of the book, tools and libraries used in the course - Introduction to IPython and Jupyter Notebooks - Administrative	1
2		1, 2	- Python recap: - Python data types - String parsing and formatting - Python loops, list comprehension, generators - Exceptions and error handling	
3	1	1	- Introduction to NumPy - Vectorized computation vs. Python loops - Two-dimensional arrays and NumPy broadcasting - Slicing NumPy arrays -	1,2
4	2	1	- Plotting data with Matplotlib - Scatter plots and correlation - Histograms and distributions	4
5	3	1, 2	- Introduction to Pandas - Working with public datasets, introduction to Kaggle	3,4
6		1, 2	- More on Pandas and visualization - Exploring data with descriptive statistics - Correlation and linear fitting -	3,4
7	4	1, 2	- Advanced DataFrame manipulations - Working with data from multiple sources - Handling Missing Data	3
8			**Intersession Week**	
9			**MID-TERM EXAM**	
10	5	3, 4, 5	- Patterns and structure in data - Recap of the concept of distance metric - Clustering - Dealing with more than 3 dimensions -	5
11	6	3, 4, 5	- Introduction to Statistical Machine Learning - Terminology: Supervised vs Unsupervised ML, Classification vs Regression - The concept of “model”	5

			- Linear regression as a simple example of supervised machine learning - K nearest neighbours (KNN) algorithm as both regression and classifier. - Introduction to scikit-learn API	
12	7	5, 6	- Dimensionality reduction for visualization - Recommender systems	5
13	8	4, 5, 6	- Model evaluation and comparison - Separation of test and training data - Overfitting and underfitting -	5
14	9	2, 5, 6	- Multidimensional regression - Regularization (Ridge and Lasso regression) - Feature engineering - Data pre-processing utilities in Sklearn - More on model evaluation - Overview of other related tools and resources with examples.	5
15			**FINAL EXAM**	

Please note: this schedule may change as resources and circumstances require.

For information on withdrawing from this course without academic penalty, please refer to the College Academic Calendar: http://www.georgebrown.ca/Admin/Registr/PSCal.aspx