

APPLIED MACHINE LEARNING

CSCI 490-0J13/680-0K13

Spring 2022

Instructor:	Pratool Bharti	Time:	Tue/Thu 3:30 - 4:45p
Email:	pbharti@niu.edu	Place:	Psyc-Computer Science 156 (Online)

Office Hours: Tue/Thu (2:00 - 3:30p)

Descriptions/ Objectives: This course offers a hands-on approach for implementing state-of-the-art machine learning algorithms in Python programming language. It focuses more on the practical aspects of these algorithms and less on the theory behind them. Topics include machine learning techniques in data preparation, data visualization, feature selection, model interpretation, and evaluation. Tools include Scikit-learn and Keras/Tensorflow for training, evaluation, interpretation and debugging of machine learning algorithms.

Prerequisites: A good understanding of Probability, Calculus and Linear Algebra would be very helpful for this course. Familiarity with Python programming and basic use of NumPy, Pandas and Matplotlib is expected. Students enrolled for Python course in Spring 2022 may enroll for this course simultaneously. A good reference is the [Python Data Science Handbook](#) by Jake VanderPlas. It's online for free and available as a notebook at the link above. I highly recommend going through it.

Programming Platform: Students are required to create a free account in Google Colab (<https://colab.research.google.com>). All programming assignments are required to be done using Jupyter Notebook on Google Colab server. Details will be discussed in the first class.

Recommended References: Students need to consult following recommended book regularly. Online version is available free of cost. Other required materials will be posted on Blackboard.

- Muller C. A and Guido S, *Introduction to Machine Learning with Python: A Guide for Data Scientists*, An O'Reilly Media publication; 1st edition (October 21, 2016)
- Kuhn M and Johnson K, *Applied Predictive Modeling*, Springer; 1st ed. 2013, Corr. 2nd printing 2018 edition (March 30, 2018)
- Goodfellow I, Bengio Y and Courville A, *Deep Learning*, An MIT Press book, 2016,
Online: - <http://www.deeplearningbook.org>.

Grading Policy: 5/6 Assignments (60%), Midterm 1 (20%), Midterm 2 (20%), Quizzes (Extra credits).

Grade	Average
A	≥ 90
B	≥ 80
C	≥ 70
D	≥ 60
F	< 60

Important Dates:

Midterm #1 March 22, 2022
 Midterm #2 May 05, 2022

Course related Announcements: All course related materials and announcements will be posted on Blackboard. Students should login at least once everyday in the Blackboard.

Collaboration: You are welcome and encouraged to discuss course materials and reading assignments with other students. Please limit discussion of homework to general approaches. You are not allowed to share code between submissions or submission groups. For homeworks submitted individually, each individual is required to write their own solution. Collaboration is not permitted for any of the exams.

Assignments Policy: There will be approximately 5-6 individual assignments. Most or all of the assignments will be well-structured Python programs and written reports. Graduate students are required to do an additional assignment/ questions. There will be likely no extensions. The maximum possible points that will be awarded will decrease by 10% of the original maximum point value per day that they are late. Maximum late submission allowed is 48 hours. Assignments submitted later than 48 hours will be assigned 0 score.

Attendance Policy: Regular attendance is essential and expected.

Academic Integrity: Academic dishonesty is strictly prohibited. You are expected to do your own work on the assignments, programs and exams. Cheating includes, but is not limited to, copying work from other students, from other textbooks, from the Internet, or allowing others to do the same, without proper citation of source material. Please follow the link for more information <https://www.niu.edu/academic-integrity/students/plagiarism/index.shtml>. Lack of knowledge of the academic honesty policy is not a reasonable explanation for a violation.

Special circumstances: Students with special needs (disability accommodation, religious observances, required military service, major illness or other unexpected events) are encouraged to contact the instructor as soon as possible. Having a lot of work for your other classes, being busy at your job, and network problems are not special circumstances; they are normal circumstances that everyone has.

Disability accommodations: The instructor will provide all of the accommodations to which you are entitled by law. If you need an accommodation for this class, you must provide a notification letter from the Disability Resource Center. If you wish to take your exams at the DRC office, you must also follow DRC regulations with regard to exam scheduling. Therefore you should contact the DRC as soon as possible. They are located on the Campus Life Building, Suite 180, and can be reached at 815-753-1303 or drc@niu.edu. Also, please contact me privately as soon as possible to discuss possible accommodations. Please note that you will not be required to disclose your disability, only your accommodations. The sooner you let me know your needs, the sooner I can assist you in achieving your learning goals in this course.

Tentative Course Outline:

Week	Date	Class Topics
1	01/18	Introduction, Course Overview
	01/20	Math Essentials for Machine Learning
2	01/25	Data Visualization (Matplotlib)
	01/27	Introduction to Supervised Learning

3	02/01	Introduction to Supervised Learning (continue)
	02/03	Data processing for Missing values and Imbalanced dataset
4	02/08	Linear Models for Regression (Lasso, Ridge and ElasticNet)
	02/10	Linear models for Classification (Regularized Logistic Regression)
5	02/15	Trees, Forests, Gradient Boosting & XGBoost
	02/17	Trees, Forests, Gradient Boosting & XGBoost (continue)
6	02/22	Model Evaluation (Bias and Variance)
	02/24	Model Evaluation (AUC RUC, Average Precision)
7	03/01	Model Interpretation and Feature Selection
	03/03	Model Interpretation (LIME and SHAP)
8	03/08	Learning with Imbalanced Data
	03/10	Revision Day
9	03/15	University Closed
	03/17	University Closed
10	03/22	First Mid-term Exam
	03/24	Dimensionality Reduction (PCA, LDA and TSNE)
11	03/29	Clustering and mixture models (K-Means, Hierarchical Linkage)
	03/31	Clustering and mixture models (DBSCAN, Guassian Mixture Models)
12	04/05	Working with text data (Tokenization, Normalization and N-grams)
	04/07	Working with text data (Data cleaning)
13	04/12	Topic models for text data
	04/14	Word and document embeddings
14	04/19	Word and document embeddings (Continue)
	04/21	Neural Networks
15	04/26	Keras and Convolutional Neural Nets
	04/28	Keras and Convolutional Neural Nets (Continue)
16	05/03	Revision Day
	05/05	Second Mid-term Exam