# Machine Learning for Chemical and Materials Engineers/MTEN 633 Syllabus

**Spring 2025**

**Course Modality:**

This is an online course, which will be conducted fully online, asynchronously via Canvas. For more information on using Canvas and other supported learning tools, visit the IST Service Desk Knowledgebase.

**Instructor Information**

| Instructor | Email | Office Hours |
|---|---|---|
| **Rajesh N. Davé** | dave@njit.edu | Office hours will be by appointment via Zoom. Please email the instructor directly to set up a time. |

*I will respond to all emails/Inbox messages within 48 hours. Discussions and Reflections are not graded but submitting them is required for credit. Assignments will be graded weekly.

## General Information

### Course Description

This course builds upon foundational knowledge in mathematical and statistical tools or data science and data visualization to learn and apply machine learning approaches for solving problems in Engineering with emphasis on applications in Chemical and Materials. As an engineering course, the emphasis on case studies involving problem solving to augment mechanistic methods. It will include artificial intelligence (AI) topics such as machine learning and deep learning via methods such as classification, clustering, and neural networks (NNs), along with robust estimation including validation and error

quantification. Class projects will concern case studies from a broad range of disciplines and applications to chemical and materials engineering problems.

## Prerequisites/Co-requisites

Undergraduate degree in either Chemical, Mechanical, Electrical or Biomedical engineering, or in physics or chemistry. There are no co-requisites; students are encouraged to enroll in other courses in the data science certificate.

## Course Learning Outcomes

By the end of the course, students will be able to:

1. Construct, interpret, and manipulate datasets and databases relevant to chemical, materials or pharmaceutical engineering.
2. Utilize descriptive and inferential statistical methods.
3. Identify & extract the useful features from the dataset, perform dimensionality reduction, select training set and test data, apply regression methods to engineering problems.
4. Construct problem statements for labeled or unlabeled data and apply supervised or semi-supervised learning.
5. Apply classification and model identification methods including regression, as well as probabilistic and deterministic clustering in the context of engineering problems.
6. Apply deterministic (hard or fuzzy) memberships to cluster analysis and use such clustering techniques.
7. Identify the need for non-standard cluster prototypes including nonlinear manifolds. Apply cluster validity to assess goodness of fit.
8. Identify when real-life situations require treatment of noise and outliers in your data and select a relevant clustering method.
9. Identify situations where deep learning techniques such as neural networks are required.
10. Critically review the case studies from literature and as needed apply those learnings to applications in chemical, materials, or pharmaceutical engineering.

## Required Materials

- The instructor will provide class notes and research articles and online resources.
- Hands-on Machine Learning with Scikit-Learn, Keras, TensorFlow, Third Edition, Aurelien Geron,O'Reilly, 2022. ISBN: 9781098122478.

## Suggested Reading

(1) Machine Learning: An Algorithmic Perspective, Second Edition, Stephen Marsland, Routledge (Taylor and Francis Group) 2015. ISBN 9781466583283 (e-book) - Note this book may be somewhat outdated.

(2) A First Course in Machine Learning, Second Edition, Simon Rogers, Mark Girolami, Routledge (Taylor and Francis Group) 2017. ISBN 9780367574642 (e-book).

(3) Special Issue: Artificial Intelligence in Chemical Engineering, Volume 93, Issue 12, Chem. Ing. Tech., (Wiley Online) 12/2021.

Alternate or additional reference materials as appropriate will be specified at the time of offering.

## Grading Policy

NJIT Grading Legend

### Final Grade Calculation

Final grades for all assignments will be based on the following percentages:

| | |
|---|---|
| **Assignments** | **25%** |
| **Discussion Forums and Reflections** | **10%** |
| **Mid-semester literature report** | **25%** |
| **Capstone Project** | **40%** |

### Course Work

**Assignments: (25% of grade)** There will be frequent assignments meant to help you practice course concepts.

**Discussion Forums and Reflections: (10% of grade)** You are expected to participate in weekly discussion forums in Canvas. When all students participate in a discussion, it creates an active learning environment that will help you better understand the materials and be more successful in the class. Your participation also includes peer work assessment. You will post an initial response to the prompt by Fridays at 11:59pm and respond to two classmates by Sunday at 11:59pm of the week they are listed. For assigned

reflections, you will be submitting a very brief report each week. These are not graded but count for some credit based on your participation.

**Mid-Semester Literature Report: (25% of grade)** This is intended to be a literature survey report on your choice of various topics. You will prepare and submit an interim report (Milestone 1) followed by a written report and a presentation (Milestone 2). The criteria for earning a top grade in such an assignment is that your report could be used as a module on a specific topic during future course offerings.

**Capstone Machine Learning (ML) Project:** (**40% of grade**) There will be projects with regular milestones. You will have opportunities to iterate and revise your work based on peer and instructor feedback. The total grade is based on two intermediate progress reports/milestones, (Milestone 1) Selection and justification of the engineering topic and challenges in collecting and organizing data, (Milestone 2) Details of the proposed methodology and preliminary results. (Milestone 3) The final submission would require a written report, presentation slides and a recorded oral presentation. The criteria for earning a top grade for the final milestone is that your written report/presentation could be used as a major module on a specific topic during future course offerings.

## Feedback

I will deliver feedback on selected assignments using the comments feature in Canvas.

## Letter to Number Grade Conversions

Grades may be assigned with the following proposed rubric which is provided only as a guideline. The instructor will consider overall class performance to adjust the grading scale.

| | |
|---|---|
| A | 90-100 |
| B+ | 85-89 |
| B | 80-84 |
| C+ | 75-79 |

| | |
|---|---|
| C | 70-74 |
| F | 0-64 |

## Exam Information and Policies

This course does not have any exams. Per the NJIT Online Course Exam Proctoring Policy, this course will use authentic assessment, meaning you will be assessed and graded on your ability to deliver real-world outputs as well as your participation and feedback to other students.

## Policy for Late Work

Late homework and discussion assignments without prior written permission are usually not counted towards your final grade. Partial credit may be given depending on how late your submissions are. For the projects, extra time may be allowed but at a penalty of 10 % per day.

## Academic Integrity

*"Academic Integrity is the cornerstone of higher education and is central to the ideals of this course and the university. Cheating is strictly prohibited and devalues the degree that you are working on. As a member of the NJIT community, it is your responsibility to protect your educational investment by knowing and following the NJIT academic code of integrity policy.*

*Please note that it is my professional obligation and responsibility to report any academic misconduct to the Dean of Students Office. Any student found in violation of the code by cheating, plagiarizing or using any online software inappropriately will result in disciplinary action. This may include a failing grade of F, and/or suspension or dismissal from the university. If you have any questions about the code of Academic Integrity, please contact the Dean of Students Office at dos@njit.edu"*

## Netiquette

*Throughout this course, you are expected to be courteous and respectful to classmates by being polite, active participants. You should respond to discussion forum assignments in a timely manner so that your classmates have adequate time to respond to your posts.*

*Please respect opinions, even those that differ from your own, and avoid using profanity or offensive language.*

### Weekly Expectations

This course is organized by weekly modules. Each week, students must watch lecture/module videos, complete reading assignments, submit assignments and participate in a class discussion forum. Initial posts for the discussion forum are due by Friday at 11:59 p.m. and respond to two classmates discussion posts by Sunday at 11:59 p.m. Assignments, Reports and Reflections are due by Sunday at 11:59 p.m., unless otherwise indicated in the schedule below.

## Course Schedule

| Week | Topic | Reading/ Assignment | Due Dates |
|---|---|---|---|
| **1** | • What is Machine Learning (ML)<br>  ○ Artificial Intelligence (AI), ML, Deep Learning<br>  ○ What type of problems can be solved? A few examples are discussed.<br>• Read the special issue (Course reference materials #3) and pick one of those for a short report<br>• Discussion of class projects; details of the literature project | -Module 1: Peer Introductions<br>-Module 1: Discussion<br>-Module 1 Assignment<br>-Module 1: Reflection<br>-Preview Literature Report-Interim Report 1<br>-Read Ch. 2 of Grokking Book<br>-Read articles from Special Issue: Artificial Intelligence in Chemical Engineering | -Module 1: Peer Introductions and Discussion-Initial Post-**due Fri**<br>Peer Replies-**due Sun**<br>-Module 1 Assignment-**due Sun**<br>-Module 1: Reflection-**due Sun** |
| **2** | • Learning Problems<br>  ○ Supervised vs unsupervised<br>  ○ Classification and Regression<br>• Data driven models and mechanistic models<br>  • Role of ML<br>  • Hybrid models<br>  • Example: Powder flow data | -Module 2: Discussion<br>-Module 2: Assignment 1.1<br>-Module 2: Assignment 1.2<br>-Module 2: Reflection<br>-Preview Literature Report-Interim Report 1<br>-Preview Literature Report-Interim Report 2 | -Module 2: Discussion-Initial Post-**due Fri**<br>Peer Replies-**due Sun**<br>-Module 2: Assignment 1.1-**due Sun** |

| | | | |
|---|---|---|---|
| | | -Read pgs 1-33-Hands-on Machine Learning<br>-Any resource on inferential statistics | -Module 2: Assignment 1.2**-due Sun**<br>-Module 2: Reflection**-due Sun** |
| **3** | • Visualization of data<br>• Probability Theory – Introduction<br>• Inferential Statistics – Introduction<br>• Organizing and handling data – Linear models<br>• Training set and test data, feature extraction<br>• What about noisy data with outliers? Food for thought | -Module 3: Discussion<br>-Module 3: Assignment<br>-Literature Report-Interim Report 1<br>-Module 3: Reflection<br>-Preview Literature Report-Interim Report 2<br>-Read Ch. 2, 3, 4 and 8-Machine Learning<br>-Read Ch. 1 of A First Course in Machine Learning | -Module 3: Discussion-Initial Post**-due Fri**<br>Peer Replies**-due Sun**<br>-Module 3: Assignment**-due Sun**<br>-Literature Report-Interim Report 1**-due Sun**<br>-Module 3: Reflection**-due Sun** |
| **4** | • Introductory Linear algebra (Module 4.1 and 4.2)<br>• Recap: Organizing and handling data, training set and test data (Module 4.3)<br>• Regression (Module 4.4)<br>  ○ Linear<br>  ○ Multiple<br>  ○ Logistic | -Module 4: Discussion<br>-Module 4: Assignment 1<br>-Module 4: Assignment 2<br>-Literature Report-Interim Report 2<br>-Module 4: Reflection<br>-Preview-Literature Report-Final Report<br>-Read Ch 4-Hands-on Machine Learning<br>-Review basics of linear algebra | -Module 4: Discussion-Initial Post**-due Fri**<br>Peer Replies**-due Sun**<br>-Module 4: Assignment 1**-due Sun**<br>Module 4: Assignment 2**-due Sun**<br>-Literature Report-Interim Report 2**-due Sun**<br>-Module 4: Reflection**-due Sun** |
| **5** | • Applications of ML to chemical and materials engineering (Module 5.1)<br>• Multivariate dataset example (Module 5.2)<br>  ○ Individual powder properties | -Module 5: Discussion<br>-Module 5: Assignment<br>-Module 5: Reflection | -Module 5: Discussion-Initial Post**-due Fri**<br>Peer Replies**-due Sun** |

| | | | |
|---|---|---|---|
| | ○ Organizing the data (a) Apply normalization/standardization. (b) Getting ready for PCA<br>○ Moving up to the property of blends and their classification - How do we identify mixture rules?<br>● PCA for Dimensionality reduction (Module 5.3)<br>○ General introduction<br>■ Demos/tutorials in Modules 5.4 and 5.5<br>○ Limitations of PCA – cannot handle nonlinear relations between the features to be reduced<br>● PCA for Dimensionality reduction - demo for application to powder data example (Module 5.4)<br>● Demo/tutorial from Prof. Rich Cimino using the famous Fisher Iris data (Module 5.5) | -Preview-Literature Report-Final Report<br>-Read Class Notes from UVM<br>-Read 5.3.1 of A First Course in Machine Learning | -Module 5: Assignment-**due Sun**<br>-Module 5: Reflection-**due Sun** |
| **6** | ● Clustering in Machine Learning: Unsupervised learning methods<br>○ KNN clustering (Discussed this week)<br>○ K means Clustering (introduced this week)<br>○ Fuzzy C-Means Clustering<br>○ Methods that stretch and rotate individual clusters through eigenvalues and eigenvectors<br>○ Gustafson-Kessel Algorithm<br>○ Introduction to Nonlinear manifolds as cluster prototypes - circles/ellipsoids | -Module 6: Discussion<br>-Module 6: Assignment<br>-Literature Report-Final Report<br>-Module 6: Reflection<br>-Review linked web-pages within the slides of Modules 6.1 and 6.2 | -Module 6: Discussion-Initial Post-**due Fri**<br>Peer Replies-**due Sun**<br>-Module 6: Assignment-**due Sun**<br>-Literature Report-Final Report-**due Sun**<br>-Module 6: Reflection-**due Sun** |

| | | | |
|---|---|---|---|
| | ○ Robust Clustering (modification of FCM and others)<br>○ Introduction to Hierarchical clustering (Agglomerative and Divisive clustering)<br>● Validating classification – cluster validity<br>● Introduction to neural networks and deep learning | | |
| **7** | ● Recapping: Clustering techniques and K-means clustering<br>● Detailed discussion<br>○ The objective function of K-Means<br>○ What other forms of cluster center or cluster prototypes could be envisioned?<br>○ How to define distances from cluster prototypes?<br>○ Convergence of K-Means<br>■ Example: Image segmentation<br>○ K-means hyperparameters<br>○ Clustering of unlabelled data - big picture<br>○ Shortcomings of K-means | -Module 7: Discussion<br>-Module 7: Assignment<br>-Read Ch. 9 pgs. 235-255-Hands-on Machine Learning<br>-Module 7: Reflection<br>-Read section 9.1 from Bishop Book | -Module 7: Discussion-Initial Post-**due Fri**<br>Peer Replies-**due Sun**<br>-Module 7: Assignment-**due Sun**<br>-Module 7: Reflection-**due Sun** |
| **8** | ● Review of K-Means algorithm and introduction to Fuzzy C-Means<br>○ Hard versus fuzzy memberships – students are encouraged to review online Python tutorial<br>○ KMC and FCM have certain problems<br>● Recapping the issues and discussion of real-life data that is expected to have noise and outliers<br>● Demo of flow data to KMC<br>● Demo of FCM on a dataset and Python codes | -Module 8: Discussion<br>-Module 8: Assignment<br>-Preview Interim Capstone-Report 1<br>-Module 8: Reflection<br>-Chapter 9-Hands-on Machine Learning<br>-Read *FCM and Validity* | -Module 8: Discussion-Initial Post-**due Fri**<br>Peer Replies-**due Sun**<br>-Module 8: Assignment-**due Sun**<br>-Module 8: Reflection-**due Sun** |

| | | | |
|---|---|---|---|
| **9** | • ML Project and overview of two examples of using and applying machine learning tools in engineering<br>  ○ Reminder of the main steps in any ML project<br>  ○ Discussion of the ML project<br>• Performing Root Cause Analysis (RCA) - a tutorial with a hands-on example by Prof. Rich Cimino<br>• Modules on AI/ML applications to complex problems in Chemical Engineering – Hybrid Approach by Prof. Venkatsubramanian, Columbia University | -Module 9: Discussion<br>-Module 9: Assignment<br>-Preview Interim Capstone-Report 1<br>-Module 9: Reflection<br>-Read Brief Summary on Various Algorithms<br>-Read FCM on MATLAB | -Module 9: Discussion-Initial Post-**due Fri**<br>Peer Replies-**due Sun**<br>-Module 9: Assignment-**due Sun**<br>-Module 9: Reflection-**due Sun** |
| **10** | • Robust clustering via Noise Clustering Algorithm<br>  ○ Recapping K-Means, Fuzzy C-Means and bias due to noise, NC algorithm and examples<br>• Robust statistical concepts for outlier rejection<br>  ○ Initial discussion<br>• A demo with Python code to generate noisy data having multiple Gaussian clusters | -Module 10: Discussion<br>-Interim Capstone-Report 1<br>-Module 10: Reflection<br>-Preview Interim Capstone-Report 2<br>-Read Noise Clustering Paper | -Module 10: Discussion-Initial Post-**due Fri**<br>Peer Replies-**due Sun**<br>-Interim Capstone-Report 1-**due Sun**<br>-Module 10: Reflection-**due Sun** |
| **11** | • Further details of the NC algorithm and how it is implemented<br>  ○ This is a partitioning algorithm that also accomplishes detection of noise<br>  ○ Picking the “noise distance”<br>• The NC algorithm does outlier rejection like any M-estimator could but for C clusters simultaneously<br>  ○ A few advanced concepts<br>• A demo module for applying the FCM algorithm for the flow data | -Module 11: Discussion<br>-Module 11: Assignment<br>-Module 11: Reflection<br>-Preview Interim Capstone-Report 2<br>-Review of Robust Clustering | -Module 11: Discussion-Initial Post-**due Fri**<br>Peer Replies-**due Sun**<br>-Module 11: Assignment-**due Sun**<br>-Module 11: Reflection-**due Sun** |

| | | | |
|---|---|---|---|
| | and compare the performance of KMC and FCM<br>○ Examine classification from each method<br>● Demo module comparing the performance of KNN, KMC, and FCM | | |
| **12** | ● About linear and nonlinear cluster prototypes<br>○ Introduction to prototypes other than centroids and available algorithms by Prof. Davé<br>○ Clustering based on distance metric for Shells and Manifolds<br>○ Extensions by others as rectangular shells and super-quadrics<br>● NC application examples for selected datasets<br>○ Applying NC algorithm in Python<br>● NC application for the flow dataset and analysis | -Module 12: Discussion<br>-Interim Capstone-Report 2<br>-Module 12: Reflection<br>-Preview ML Capstone - First Draft<br>-Read Adaptive Fuzzy Clustering<br>-Read Fuzzy C-Shells<br>-Read Adaptive Fuzzy C-Shells | -Module 12: Discussion-Initial Post-**due Fri**<br>Peer Replies-**due Sun**<br>-Interim Capstone-Report 2 **-due Sun**<br>-Module 12: Reflection-**due Sun** |
| **13** | ● Cluster validity – Basic Concepts<br>● Module on applying ML for Exploratory Data Analysis (EDA) by Mr. B. Kurra: A tutorial on EDA that forms an important step in any ML project<br>○ It is applied to parts remaining useful life data<br>● Example of a popular robust clustering technique - DBSCAN<br>○ DBSCAN implementation on flow data and test its robustness | -Module 13: Discussion<br>-Module 13: Assignment<br>-Module 13: Reflection<br>-Preview ML Capstone - First Draft<br>-Preview ML Capstone - Peer Review<br>-Preview ML Capstone - Final Submission<br>-Read Cluster Validity for C-Shells | -Module 13: Discussion-Initial Post-**due Fri**<br>Peer Replies-**due Sun**<br>-Module 13: Assignment-**due Sun**<br>-Module 13: Reflection-**due Sun** |
| **14** | ● Deep learning – introduction to Neural Networks (NNs) - Prof. Joshua Young, CME, NJIT | -Module 14: Discussion<br>-ML Capstone -First Draft<br>-ML Capstone -Peer Review | -Module 14: Discussion-Initial Post-**due Fri** |

| | | | |
|---|---|---|---|
| | • Application of NN to a research problem - Prof. Joshua Young, CME, NJIT<br>• Demo of NN to parts remaining useful life data<br>• Discussion and Critique of final projects | | Peer Replies-**due Sun**<br>-ML Capstone -First Draft-**due Mon**<br>-ML Capstone -Peer Review-**due Mon**<br>Peer Reviews-**due Wed** |
| **15** | • Final project submission/discussion addressing the issues raised by the instructor and peers<br>• Complex research problem requiring ML tools and skills by Prof. Kerri-Lee Chintersingh<br>○ Quantitative Characterization of Single Particle Combustion using X-ray Phase Contrast Imaging and Machine Learning | -Module 15: Discussion<br>-ML Capstone -Final Submission | -Module 15: Discussion-Initial Post-**due Tues**<br>Peer Replies-**due Wed**<br>-ML Capstone -Final Submission-**due Thurs** |

## Additional Information and Resources

**Accessibility:**
This course is offered through an accessible learning management system. For more information, please refer to Canvas's Accessibility Statement.

**Requesting Accommodations:**

The Office of Accessibility Resources and Services works in partnership with administrators, faculty, and staff to provide reasonable accommodations and support services for students with disabilities who have provided their office with medical documentation to receive services.

If you are in need of accommodations due to a disability, please contact the Office of Accessibility Resources and Services to discuss your specific needs.

## Resources for NJIT Online Students

NJIT is committed to student excellence. To ensure your success in this course and your program, the university offers a range of academic support centers and services. To learn more, please review these Resources for NJIT Online Students, which include information related to technical support.