ACM Data Science Task Force Course Example Template:

Intro to Machine Learning
Northwestern CT Community College
Winsted, CT
Crystal Wiggins
cwiggins@nwcc.edu

Link to complete course materials
https://drive.google.com/drive/folders/1o9elgfNjCxfiUjXwmXF6fCyRktP5K4F3?usp=sharing

Knowledge Areas that contain competencies (knowledge, skills, and dispositions) covered in the course

Knowledge Area	**Total Number of Contact Hours (credits)**
Machine Learning (ML)	*2.5*
Programming (PRG)	0.25

Where does the course fit in your undergraduate Data Science curriculum?
This course is required for the associate's degree and certificate. Students can take this course in their first or second year depending on where they start in math. The prerequisite for this course is the Data Science in R course.

Is this course from or used in other curricula/majors?
This course was designed for Data Science majors.

What is covered in the course?

This course focuses on machine learning as an integral tool for data science, including how to use data to automatically understand the world, make complex decisions, and even predict the future. Several algorithms will be introduced along with which language (Python or R) is better suited for which algorithm based on the particular goal in mind. Programming language(s) will be used.

Topics Covered:

1. Introduction to machine learning and the differences between ML and AI
2. Understanding the essentials of big data
3. Defining the role of algorithms in ML
4. How training works with algorithms

5. R & Python installations and basics
6. Matrices basics
7. Bayesian point of view on probability
8. Defining the most common error functions
9. How proper sampling methods are crucial in ML
10. Biased samples, overfitting, underfitting, and snooping
11. Decision tree basics
12. Cleaning data and the importance of outliers
13. Clustering data
14. Illustrating linear models for regression and classification
15. Neural network basics
16. Understanding what deep learning is
17. Basic formulation of linear support vector machines (SVM)
18. Intro to random forests
19. Classifying images
20. Intro to Natural Language Processing (NLP)
21. Intro to singular value decomposition (SVD)

What is the format of the course?
This course can be taught in all formats; online, hybrid, or face-to-face. There are lectures for each topic covered along with discussions for further student research. This course is 3 contact hours.

How are students assessed?
This course does not have exams (although it could be assessed via exams if preferred). Discussions are used to promote student research on several algorithms and their history in order to understand why and when each are used. Assignments are given when appropriate to combine the students' programming knowledge along with running these algorithms. Most of the algorithm's code is prewritten as this is an "intro" course. Students have not necessarily taken linear algebra prior to this course. Some math topics, such as vectors and matrices, are covered as needed.

Course tools and materials
- Machine Learning for Dummies by John Paul Mueller & Luca Massaron. ISBN: 978-1-119-24577-3 (paid)
- RStudio (free)
- Anaconda or Google Colabatory (free)

Why do you teach the course this way?

Goals:
Students will understand the basics of how machine learning is applied in real world problems and the algorithms used.

Outcomes:
Upon successful completion of this course, students will be able to:

1. Explain machine learning and its various tools

2. Describe theoretical foundations, algorithms, methodologies, and applications for machine learning
3. Evaluate a problem and decide which algorithm and language is best
4. Explain the difference between supervised, unsupervised and reinforcement learning

This course was created in Fall 2019 and piloted in Spring 2020. Students found this course easier than Data Science in R and overall loved the material. This course is taught with traditional lectures with assignments/discussions that are applied and geared toward how they will need to think when on the job.

Body of Knowledge coverage

KA	**Sub-domain**	**Competencies Covered**	**Hours (in a semester)**
ML	Intro to ML	● ML vs AI	3
ML	Big Data	● Essentials of Big Data	3
ML	Algorithms	**All algorithms are taught “in theory” as most students have not taken Linear Algebra. ● Naïve Bayes ● Decision Trees ● Regression and Classification ● Neural Networks ● SVM ● Random Forests ● Classifying Images ● NLP ● SV	27
PDA	Python	**R is covered in the prereq to this course ● Intro to Python ● Anaconda/Google Colab ● Basic syntax of Python	6
ML	Sampling	● Sampling Methods ● Biased Samples ● Overfitting/Underfitting ● Clustering Data	6