



Course Syllabus – INST 414 Section 0101 – Spring 2021

Data Science Techniques

Instructor: Abdirisak Mohamed

Email: amoham70@umd.edu

Office: HBK 4120K

Office hours: Online (TBA)

TA: Anna Lenhart

Email: alenhart@terpmail.umd.edu

Office hours: Online (TBA)

UTA: Lana Dove

Email: lanavdove@gmail.com

Office hours: Online (TBA)

UTA: Joshua Iweala

Email: joshua.iweala@gmail.com

Office hours: Online (TBA)

Prerequisite: 1 course with a minimum grade of C- from (INST201, INST301); and minimum grade of C- in INST126, INST314, STAT100, MATH115, and PSYC100.

Restriction: Must be in Information Science program.

Learning Outcomes

This course will explore approaches to extract insights from large-scale datasets. The course will cover the complete analytical funnel from data extraction and cleaning to data analysis and insights interpretation and visualization. The data analysis component will focus on techniques in both supervised and unsupervised learning to extract information from datasets. Topics will include clustering, classification, and regression techniques. Through homework assignments, a project, exams and in-class activities, students will practice working with these techniques and tools to extract relevant information from structured and unstructured data.

Extended Course Description

This course explores the application of data science techniques to unstructured, real-world datasets including social media and open data sources. The course will focus on techniques and approaches that allow the extraction of information relevant for experts and non-experts in a wide range of areas including smart cities, transportation or public safety.

After successfully completing this course you will be able to:

- Collect and clean large-scale datasets
- Articulate the math behind supervised and unsupervised techniques
- Execute supervised and unsupervised machine learning techniques
- Select and evaluate various types of machine learning techniques
- Explain the results coming out of the models
- Critically evaluate the accuracy of different algorithms and the appropriateness of a given approach

Textbooks

(1) <https://github.com/jakevdp/PythonDataScienceHandbook>

(2) <http://greenteapress.com/wp/think-stats-2e/>

Technology

We will use the Python programming language with Jupyter notebooks. Here is a tutorial on how to use Jupyter : [Jupyter notebook / JupyterLab](#)

(1) Google Colab is a Jupyter notebook environment hosted by Google on the Cloud. It requires no setup to use and has the option to use GPUs as computing resources. Colab is free. You have to save your files

in GoogleDrive or in your local machine as the service is not always available. Uploaded files get deleted when the runtime is recycled: <https://research.google.com/colaboratory/faq.html>

A video on Google Colab: [Practical Introduction Colab](#)

(2) Anaconda

I will encourage you to download the free Anaconda environment on your local machine (Laptop), so that you can work with Python offline when you don't have access to internet:

<https://www.anaconda.com/products/individual#Downloads>

(3) JupyterHub

I requested from the IT department to offer us JupyterHub environment on a UMD server. If it works out, it would be another option like Colab.

Course Activities

(1) Homework Assignments

Every other week you will have an assignment that is designed to assess your mastery of the topics and techniques covered in the class and provide feedback to improve your understanding of the material.

(2) Quizzes

There will be a quiz every other week that is designed to test your knowledge from recent lectures (since the last quiz) and provide you rapid feedback to improve your understanding of the material. Quizzes will be taken on Elms/Canvas on the scheduled Day.

(3) Group Project

There will be a group project with teams of 3-5 people working on a data science project. This involves identifying a question, finding or developing a dataset, creating appropriate measures, conducting analyses, and preparing an appropriate information product based on the results.

The project will be graded on your ability to articulate an appropriate question, prepare the data, identify and perform reasonable methodology and study design, justify the appropriateness of certain machine learning approaches, articulate and conduct evaluations, analyze and interpret the results and create appropriate visualizations. You will be required to analyze your dataset using Python.

You must create your own GitHub account and push (upload) the project to a public repository. Projects will be submitted via ELMS for grading.

Exams

There will be one midterm and one final exam each worth 20% of your final grade. These exams provide an opportunity for you to test your understanding of the concepts, techniques, and problems associated with data science.

A- 90.0-92.9% A 93.0-100%

B- 80.0-82.9% B 83.0-86.9% B+ 87.0-89.9%

C- 70.0-72.9% C 73.0-76.9% C+ 77.0-79.9%

D- 60.0-62.9% D 63.0-66.9% D+ 67.0-69.9%

F 0-59.9%

Grades

Homework	20
Quizzes	15
MidTerm Exam	20
Final Exam	20
Project	25

Excused Absences:

If an assignment due date or scheduled exam is a religious holiday for you or other university approved activity, please let me know at least one week in advance, so an alternate due date can be set. Missed exams with an excused absence must be made up within 2 weeks of the original deadline. Missed exams without a documented, excused absence cannot be made up.

Late Work:

Timely submission of the completed assignments is essential. Late assignments will be penalized. Late assignments will be penalized by 50% if they are turned in within one week of the due date and not accepted if they are more than one week late.

Other policies relevant to undergraduate courses are found here:

<http://ugst.umd.edu/courserelatedpolicies.html> Topics that are addressed in these various policies include academic integrity, student and instructor conduct, accessibility and accommodations, attendance and excused absences, grades and appeals, copyright and intellectual property.

OFFICE HOURS

Please visit the office hours (online) of the instructional team. This is an opportunity to ask questions about the material covered in the reading materials or in lecture. If you are having trouble in the course, please talk to us as soon as possible.

ACADEMIC DISHONESTY

Cheating in any form (copying, falsifying signatures, plagiarism, etc.) will not be tolerated. It will result in a referral to the Office of Student Conduct irrespective of scope and circumstances, as required by university rules and regulations. There are severe consequences of academic misconduct, some of which are permanent and reflected on the student's transcript. If you have any questions regarding the University's policies on scholastic dishonesty, please see <http://osc.umd.edu/OSC/Default.aspx> It is very important that you complete your own assignments, and do not share files (excluding raw data), partial work or final work.

University of Maryland Code of Academic Integrity

The University of Maryland, College Park has a nationally recognized Code of Academic Integrity, administered by the Student Honor Council. This Code sets standards for academic integrity at Maryland for all undergraduate and graduate students. As a student you are responsible for upholding these standards for this course. It is very important for you to be aware of the consequences of cheating, fabrication, facilitation, and plagiarism. For more information on the Code of Academic Integrity or the Student Honor Council, please visit <http://shc.umd.edu/SHC/Default.aspx>.

ACCOMMODATIONS

Please contact me as soon as possible if you think you might need any special accommodations for disabilities. In addition, please contact the Disability Support Services (301-314-7682 or <http://www.counseling.umd.edu/DSS/>). Disability Support Services will work with us to help create appropriate academic accommodations for any qualified students with disabilities. If you experience psychological distress during the course of the semester you can get professional help at the Counseling Center (301-314-7651 or <http://www.counseling.umd.edu/>)

Tentative Course Schedule

Week	Topic	Assignment	
January 25	Introduction Basic Python	HW 1 Due January 31	
February 1	Numpy, Pandas, Matplotlib	Quiz 1 Due February 7	
February 8	Probability Distributions	HW 2 Due February 14	
February 15	Linear Regression	Quiz 2 Due February 21	
February 22	Logistic Regression	HW 3 Due February 28	
March 1	Decision Trees	Quiz 3 Due March 7	
March 8	K-Means Clustering	HW 4 Due March 14	
March 15	Spring Break		
March 22	Midterm Exam	Due March 28	
March 29	Neural Networks TensorFlow / Keras PyTorch	Quiz 4 Due April 4	
April 5	Generative Models	HW 5 Due April 11	
April 12	Bayesian Inference	Quiz 5 Due April 18	
April 19	Causal Inference	HW 6	

		Due April 18	
April 26	Privacy / Adversarial Attacks	Quiz 6 Due May 2	
May 3	Group Project	Due May 9	
May 10	Final Exam	Due May 16	