COMP 4151/6151: Introduction to Data Science. Fall 2021

Time: TR 1pm – 2:25pm. **Location**: DH 119
Instructor: Vinhthuy Phan (vphan@memphis.edu)
Office: Dunn Hall 309. **Office Hours**: by appointment

Course Description
COMP 4151/6151. A hands-on and programming-intensive introduction to data science and applications of data mining and machine learning techniques to analyze real data sets. Specific topics include data collection, cleaning, manipulation, and visualization, clustering and developing models to make predictions, and ethical aspects of data science.

Prerequisites: COMP 2150, MATH 4614 or MATH 4635, or permission of instructor.

In this course, students will learn how to visualize and analyze real-word data. Examples include building models and tools to predict home prices based on available information (location, size, neighborhood, etc.), recommending movies based on a person's preference, predicting admissions into graduate schools, and many more.

Course Objectives:

- Students will understand fundamental data structures such as series and dataframes.
- Students will be able to manipulate dataframes with groupby.
- Students will be able to manipulate dataframes with pivoting.
- Students will be able to select data from dataframes.
- Students will be able to create charts to visualize numerical and categorical data.
- Students will be able to build models to cluster data using existing machine learning libraries.
- Students will be able to build regression models and make predictions using existing machine learning libraries.
- Students will be able to build classification models and make predictions using existing machine learning libraries.

Recommended Textbooks:

- Python for Data Analysis, 2nd Edition, 2017. by Wes McKinney, O'Reilly.
- The Data Science Desgin Manual, by Steven Skiena, Springer. https://link.springer.com/book/10.1007%2F978-3-319-55444-0

Grading:

Class attendance	5%
Homework assignments	30%
Exams	40%
Project	30%

Grading scale:
A ≥ 94 A- ≥ 90 B+ ≥ 86 B ≥ 83 B- ≥ 80 C+ ≥ 76 C ≥ 73 C- ≥ 70 D+ ≥ 60 D ≥ 50 F < 50

Tentative agenda

1. Python, iterative patterns

2. Dictionaries, Series and Data frames (Pandas)
3. Visualization of numerical data: iris dataset
4. Visualization of categorical data
5. Correlation, linear regression
6. Time series data
7. K-means clustering, clustering evaluation (Rand, silhouette)
8. Hierarchical clustering, connectivity, DBScan, density
9. Classification: K-nearest neighbor, K-nearest neighbor classifier
10. Decision trees.
11. Cross validation, precision, recall.
12. Support vector machine
13. Ensemble methods: random forest, AdaBoost.
14. Feature selection: decision tree, random forest, chi-squared test
15. Dimensionality reduction
16. Parameter optimizations

Plagiarism or cheating behavior in any form is unethical and detrimental to proper education and ***will not be tolerated*.** All work submitted by a student (projects, programming assignments, lab assignments, quizzes, tests, etc.) is expected to be a student's own work. The plagiarism is incurred when any part of anybody else's work is passed as your own (no proper credit is listed to the sources in your own work) so the reader is led to believe it is therefore your own effort. Students are allowed and encouraged to discuss with each other and look up resources in the literature (including the internet) on their assignments, but ***appropriate references must be included for the materials consulted****,* and appropriate citations made when the material is taken verbatim.

If plagiarism or cheating occurs, the student will receive a failing grade on the assignment and (at the instructor’s discretion) a failing grade in the course. The course instructor may also decide to forward the incident to the University Judicial Affairs Office for further disciplinary action. For further information on U of M code of student conduct and academic discipline procedures, please refer to: **http://www.people.memphis.edu/~jaffairs/**

Special accommodation:

If you need special accommodation, please let the instructor know immediately.

Latest information related to COVID-19:

https://www.memphis.edu/coronavirusupdates/communications/index.php