ENPM808W Syllabus

Course: ENPM808W – Data Science
Semester: Spring 2017
Instructor: S. Farshad Bahari, Ph.D.
Day; Time: Thursdays, 4:00 PM - 6:40 PM
Location: JMP2222
Email: fbahari at umd dot edu
Office Hour: Thursdays, 6:40 PM - 6:50 PM at TBD, and by appointment

Required Textbook
No required textbook

Recommended Resources

- G. James, D. Witten, T. Hastie and R. Tibshirani. Introduction to Statistical Learning with Applications in R. Springer 2013. A free PDF of this book can be downloaded at their site.

Course Objective

Data science encapsulates the interdisciplinary fields for creating data-centric products and applications towards specific needs. This course’s goal is to provide an introduction to data science from the perspective of security expert. Due to increased sophistication of targeted attacks and insufficiencies of traditional passive security measures in mitigating persistent targeted hackers, data science with the use of Big Data techniques is considered as a promising alternative for identifying threats and attacks. This specialized field includes (a) systems for storing and processing massive data sets; (b) data analytics for extracting information from such data sets; (c) security for assessing the cyber attacks.

This course focuses mainly on the following topics via case studies and projects from applications: (a) data managements and systems, (b) exploratory and statistical data analysis, (c) data and information visualization, (d) the presentation and communication of analysis results.
Grading System

Students are responsible for all material discussed in class, covered in lecture notes, and posted on course page on Canvas. Active participation in lectures, locally or remotely, is highly encouraged although is not mandatory. All the lecture material will be recorded and will be available on the course page especially for online students.

The grade for the course will be based on homework/mini-project/reading assignments, and a final project, which their respective contributions to the overall grade are given below. Homework/mini-project/reading assignments are due weekly and need to be submitted electronically to the course page no later than the beginning of the following lecture, i.e. by 4 pm on Thursdays. For the final project, each team needs to pre-record their presentation and submit it by the final exam date to the course page*.

<table>
<thead>
<tr>
<th>Mini-projects/homework assignments</th>
<th>60%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Final Project</td>
<td>40%</td>
</tr>
<tr>
<td>To be submitted to the course page by</td>
<td></td>
</tr>
<tr>
<td>the final exam day (TBD)</td>
<td></td>
</tr>
</tbody>
</table>

It is the student’s responsibility to inform the instructor of any intended absences for religious observations in advance. Notice should be provided as soon as possible but no later than the end of the adjustment period.

Students with disabilities who need any special accommodations should contact me as soon as possible with their approved DDS letter.

Course Syllabus

The following list of lecture topics will vary in terms of pace and emphasis:

Course Logistics, Introduction, What is Data Science

Introduction of tools for Data Science – R, RStudio, git, GitHub

Data sets, measurements and data types, best practices for managing collaborations and organizing projects
Data Management and Manipulation, Munging, Scraping, Wrangling, Preparation for Tidy Data

Exploratory Data Analysis (EDA data visualization, data summarization, data transformation, handling missing data,)

Statistical and probabilistic models (exploration to statistical inference, testing), Databases

Machine Learning models for classifications

Linear modeling for classifications (regression and prediction), Bias, Linear regression

Decision Tree and Random Forest

Nearest Neighbor Classifier k-NN, Validation, Cross-validation

Principal Components Analysis (PCA)

Nonlinear methods

Support Vector Machine and Performance Evaluation

Best Practices in Machine Learning

Experimental Design in Data Science

Interactive Data Visualization

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 of the Student Honor Council, please visit http://shc.umd.edu/SHC/HonorPledgeInformation.aspx.

Course Evaluations
If you have any suggestions for improving this class, please do not hesitate to tell me or the TA during the semester! At the end of the semester, please also do not forget to provide your feedback using the campus-wide CourseEvalUM system. Your comments will help make this course better.

Right to Change Information

Although every effort has been made to be complete and accurate, unforeseen circumstances arising during the semester could require the adjustment of any material given here. Consequently given due notice to students, the instructor reserves the right to change any information on this syllabus or in other course materials.