Learning Outcomes

This course explores the application of data science techniques to unstructured, real-world datasets including social media and geo-referenced sources. The course will focus on techniques and approaches to extract information relevant for experts and non-experts in areas that include smart cities, public health and disaster management.

After successfully completing this course you will be able to:

- Understand and discuss state-of-the-art solutions in the field data science
- Envision, design and evaluate data science solutions for real-world problems
- Write data science results in a scientific manner

Recommended Resources

This book is not required but presents an overview of the themes we will be discussing in class. All the reading materials – mostly papers- will be provided in class.

Social Physics: How Good Ideas Spread – The lessons from a new science
Alex Pentland
Scribe Publications (2014)
ISBN-10: 1922247553

Course website: https://piazza.com/umd/fall2018/inst750/home

Activities, Learning Assessments, & Expectations for Students

This is a seminar-type class that requires active student participation. Students will be required to actively participate and will be evaluated with three types of activities:

1. Paper presentation and Discussion
Students will be required to present several research papers during the semester. They will be required to identify the most salient contributions and to lead a discussion with the rest of the class. When not presenting, students will be required to read the paper that is being discussed and to actively participate in the discussion. Part of the grade is based on that participation.
2. Semester-long Research Project
Students will be required to pick one of the proposed projects or to propose a project of their own. The project will be a semester-long, multi-student effort. Students will be required to design and develop a full fledged data science project from data collection, to data cleaning, to applying different applied machine learning techniques and to discuss insights that can be extracted from the methods. Students are required to give a presentation of intermediate results every two weeks, in class. Students will receive feedback from the professor.

3. Paper writing
Students are required to write a technical paper with the main contributions of their project. The paper is required to be in conference format. Students are encouraged, although not required to submit their paper to an ACM workshop or conference.

Campus Policies
It is our shared responsibility to know and abide by the University of Maryland’s policies that relate to all courses, which include topics like:

- Academic integrity
- Student and instructor conduct
- Accessibility and accommodations
- Attendance and excused absences
- Grades and appeals
- Copyright and intellectual property

Please visit www.ugst.umd.edu/courserelatedpolicies.html for the Office of Undergraduate Studies’ full list of campus-wide policies and follow up with me if you have questions.

Course-Specific Policies

No computers, phones or tablet devices are permitted during our class meeting unless used for class purposes.
I expect you to make the responsible and respectful decision to refrain from using your cellphone in class. If you have critical communication to attend to, please excuse yourself and return when you are ready. For more information about the science behind the policy watch: http://youtu.be/WwPaw3Fx5Hk

Late Work
Any assignment submitted up to three days after the deadline will get half credit. If you do this, please send me an email so that I know you have submitted it late. Assignments submitted after that will *not* be graded unless the student provides a formal letter (from the dean, doctor,...) to justify the special circumstances. If you envision not being able to meet a deadline, not being able to lead a paper presentation or discussion, not being able to participate in a discussion or not being able to attend your bi-weekly project presentation, please let me know one week in advance.

Get Some Help!

Taking personal responsibility for you own learning means acknowledging when your performance does not match your goals and doing something about it. I hope you will come talk to me so that I can help you find the right approach to success in this course, and I encourage you to visit http://tutoring.umd.edu to learn more about the wide range of campus
resources available to you. In particular, everyone can use some help sharpen their communication skills (and improving their grade) by visiting [ter.ps/writing](ter.ps/writing) and schedule an appointment with the campus Writing Center. Finally, if you just need someone to talk to, visit [counseling.umd.edu](counseling.umd.edu).

**Names/Pronouns and Self Identities**

The University of Maryland recognizes the importance of a diverse student body, and we are committed to fostering equitable classroom environments. I invite you, if you wish, to tell us how you want to be referred to both in terms of your name and your pronouns (he/him, she/her, they/them, etc.). The pronouns someone indicates are not necessarily indicative of their gender identity. Visit [trans.umd.edu](trans.umd.edu) to learn more.

Additionally, how you identify in terms of your gender, race, class, sexuality, religion, and dis/ability, among all aspects of your identity, is your choice whether to disclose (e.g., should it come up in classroom conversation about our experiences and perspectives) and should be self-identified, not presumed or imposed. I will do my best to address and refer to all students accordingly, and I ask you to do the same for all of your fellow Terps.

**Grades**

Grades are not given, but earned. Your grade is determined by your performance on the learning assessments in the course and is assigned individually (not curved). If earning a particular grade is important to you, please speak with me at the beginning of the semester so that I can offer some helpful suggestions for achieving your goal.

All assessment scores will be posted on the course ELMS page. If you would like to review any of your grades (including the exams), or have questions about how something was scored, please email me to schedule a time for us to meet in my office.

Late work will not be accepted for course credit so please plan to have it submitted well before the scheduled deadline. I am happy to discuss any of your grades with you, and if I have made a mistake I will immediately correct it. Any formal grade disputes must be submitted in writing and within one week of receiving the grade.

<table>
<thead>
<tr>
<th>Learning Assessments</th>
<th>Category Weight</th>
</tr>
</thead>
<tbody>
<tr>
<td>Paper Presentation and Discussion (up to 5 papers)</td>
<td>25%</td>
</tr>
<tr>
<td>Project (one in group)</td>
<td>50%</td>
</tr>
<tr>
<td>Paper writing (one in group)</td>
<td>25%</td>
</tr>
</tbody>
</table>

Final letter grades are assigned based on the percentage of total assessment points earned. To be fair to everyone I have to establish clear standards and apply them consistently, so please understand that being close to a cutoff is not the same as making the cut (89.99 ≠ 90.00). It would be unethical to make exceptions for some and not others.

<table>
<thead>
<tr>
<th>Final Grade Cutoffs</th>
</tr>
</thead>
<tbody>
<tr>
<td>+ 97.00%  + 87.00%  + 77.00%  + 67.00%</td>
</tr>
<tr>
<td>A  94.00%  B  84.00%  C  74.00%  D  64.00%  F  &lt;60.0%</td>
</tr>
<tr>
<td>- 90.00%  - 80.00%  - 70.00%  - 60.00%</td>
</tr>
</tbody>
</table>
Course Schedule

Class 1: Introduction to Advanced Data Science

Class 2: Project Discussion

Class 3: Understanding Human Mobility


3. E. Zagheni and I. Weber. You are where you e-mail: using e-mail data to estimate international migration rates. WebSci 2012, pp. 348-351

Project Presentations:

Class 4: Localizing Social Media


Project Presentations:
Project Proposals due next week: October 3rd, 2018, by midnight.

Class 5: Content Pollution in Social Media


Project Presentations:

Class 6: Social Networks: Structural Properties


Project Presentations:

Class 7: Social Networks: Diffusion


Project Presentations:

Class 8: GeoSocial Media


Class 9: Applications I - Health


3. Adam Sadilek and Henry Kautz. Modeling the Impact of Lifestyle on Health at Scale. *Sixth ACM International Conference on Web Search and Data Mining (WSDM)*, 2013

Project Presentations:

Paper updates due November 7th before midnight.

Class 10: Applications II – Computational Sociology


Project Presentations:

Class 11: Applications III - Smart Cities


Project Presentations:

Class 12: Privacy of Data, Biases and Fairness


Project Presentations:

Class 13: Final Project Presentations
Class 14: Final Project Presentations

Final papers due December 12th before midnight.