University at Buffalo
The State University of New York

EAS 345: Introduction to Data Science

This course is partially support by NSF grant HDR-DSC-1924292

COURSE INFORMATION

3 Credits
Lecture – 3 credits
Lab – N/A

INSTRUCTOR INFORMATION

Dr. Bina Ramamurthy

- *Email:* bina@buffalo.edu
- ***All the class material including tests, assignments and grades will be posted on ublearns.***
- *Webpage:* http://www.cse.buffalo.edu/~bina/EAS345
- *Office:* Davis 345 (online this semester)
- *Office hours:* Mon and Wed: 10.00-11.30AM. (online: individual, private and/or group)

COURSE DESCRIPTION

This course is an introductory level course in data science. Students are introduced to fundamental concepts in data science. The course will focus on three major themes: Data characteristics, data science pipeline, and data-driven applications. Topics include: Data diversity, data products, data collection methods, data cleaning and formatting, storing and sharing data, privacy and confidentiality of data, data security, small data analysis, statistical analysis using R studio, big data analysis using tools such as Tableau, presentation of analytics using Jupyter notebook, data applications in healthcare and data applications for the society. Data for the discussion will be from sources such as data.gov, Pew research, Kaggle, and online crawl data. All concepts covered will be illustrated using hands-on experiments and problem-based learning activities, analyzing real-world data sets to develop data science skills.

Prerequisites: MTH 142 and (CSE 115, EAS 230, EAS 240, or EAS999TRCP), and approved engineering major or permission of the department.

TEXT BOOK

Programming skills for data science, Michael Freeman and Joel Ross, Addison Wesley Data and Analytics Series, ISBN: 978-0-13-513310-1, 2019.

COURSE ORGANIZATION / SCHEDULE

Week#	Description
1	Data diversity: Describe the diverse forms of data: structured, unstructured; numeric, text, audio, video, genetic, geo-location, IoT, devices, sensors
2	Data products: Explore data products or data-driven applications you use every day: (Google search), recommendation systems. social media applications
3	Data collection: Collect data from various sources using tools and surveys
4	Data cleaning & formatting: Clean, validate and transform data into format suitable for analysis
5	Storing and sharing data: Explore methods for storing data in single file, sheets, to cloud storage (AWS, Box)
6	Privacy and confidentiality of data: Discuss privacy and confidentially of data. Explore regulations such as HIPPA at a high level
7	Data Security: Discuss existing methods for encryption and digital signing
8	Small data analysis: Analyze data in sheets using built-in methods and plots.
9	Statistical analysis using R studio: Explore statistical analysis of data using R Studio integrated development environment

10	Big data analysis using tools such as Tableau: Explore visual analysis of using tools such as R Studio.
11	Presentation of analytics: Present the results using web-based executable notebooks such as R Studio.
12	Data applications in healthcare Apply data methods learned to healthcare data: application domain 1
13	Data applications for the society: Apply data methods learned to societal data: application domain 2
14	Putting it all together: from data to data pipeline and data products

STUDENT LEARNING OUTCOMES

Course Learning Outcome	Program Outcomes (CAC / EAC)*	Assessment Method(s)
An ability to design methods to collect data, clean and provision them for downstream analysis.	2 (CAC); 6(EAC)	Summative quiz, Exam
An ability to apply statistical methods to analyze data and discover knowledge.	1, 2 (CAC); 1, 6(EAC)	Summative quiz, Exam
An ability to derive conclusions based on data analysis and compare with real-word facts.	3(CAC); 3(EAC)	Summative quiz, Exam
An ability to recognize ethical challenges in data science problems.	4 (CAC); 4 (EAC)	Graded lab, Exam

* Student Outcomes from the Engineering Accreditation Commission (EAC) and Computing Accreditation Commission (CAC) of ABET have been adopted. See,

ABET EAC Criteria: https://www.abet.org/accreditation/accreditation-criteria/criteria-for-accrediting-engineering-programs-2019-2020/
ABET CAC Criteria: https://www.abet.org/accreditation/accreditation-criteria/criteria-for-accrediting-computing-programs-2019-2020/

Program Outcome Support
Blank: no coverage, 1: introduced 2: practices or reinforced 3: mastered

CAC:

PROGRAM OUTCOME	1	2	3	4	5	6
SUPPORT LEVEL	2	2	1	1	0	0

EAC:

PROGRAM OUTCOME	1	2	3	4	5	6	7
SUPPORT LEVEL	2	0	1	1	0	2	0

COURSE REQUIREMENTS

- **Quizzes** Students will be assessed through quizzes during lectures.
- Quizzes will be used to assess students on learning outcomes that are not amenable to be assessed through programming assignment.
- **Term project**: Students will work on a term project that spans the entire semester.

GRADING POLICY

Learning assessments will be graded based on rubric criteria and weighted according to the following breakdown.

Weight	Assessment / Assignment
10%	Class attendance
40%	Quizzes (online)
50%	Term project: incrementally developed over the semester with specific milestones to meet

Final Grades:

Grade	Quality Points	Percentage
A	4.0	93.0% -100.00%
A-	3.67	90.0% - 92.9%
B+	3.33	87.0% - 89.9%
B	3.00	83.0% - 86.9%
B-	2.67	80.0% - 82.9%
C+	2.33	77.0% - 79.9%
C	2.00	73.0% - 76.9%
C-	1.67	70.0% - 72.9%
D+	1.33	67.0% - 69.9%
D	1.00	60.0% - 66.9%
F	0	59.9 or below

Incompletes (I): A grade of incomplete ("I") indicates that additional course work is required to fulfill the requirements of a given course. Students may only be given an "I" grade if they have a passing average in coursework that has been completed and have well-defined parameters to complete the course requirements that could result in a grade better than the default grade. An "I" grade may not be assigned to a student who did not attend the course. Prior to the end of the semester, students must initiate the request for an "I" grade and receive the instructor's approval. Assignment of an "I" grade is at the discretion of the instructor. For more information on "I" grade see the catalog explanation here: https://catalog.buffalo.edu/policies/explanation.html

ACADEMIC INTEGRITY

Academic integrity is a fundamental university value. Through the honest completion of academic work, students sustain the integrity of the university while facilitating the university's imperative for the transmission of knowledge and culture based upon the generation of new and innovative ideas. See the link for more details:

https://www.buffalo.edu/academic-integrity.html

ACCESSIBILITY RESOURCES

If you have any disability which requires reasonable accommodations to enable you to participate in this course, please contact the Office of Accessibility Resources, 60 Capen Hall, 645-2608, and also the instructor of this course. The office will provide you with information and review appropriate arrangements for reasonable accommodations. See http://www.buffalo.edu/studentlife/who-we-are/departments/accessibility.html

COURSE MATERIALS

- Course Website (All relevant material will be provided through the course website)
- Examples of interesting real-world data sets that we might use include flu data from data.gov, FIFA soccer datafrom Kaggle, or mobile device usage data from Pew Research. Students will work with Jupyter notebooks (open source) and / or RStudio (academic package) to perform their work.