

Fall 2021
BUS/CSE/IDS 2990: Data Science Literacy – Section 1
Tuesday/Thursday, 8:00 a.m. – 9:15 a.m.
Library Auditorium

Instructor

Dr. Jonathan Barlow
 Associate Director, National Strategic Planning and Analysis Research Center (NSPARC)
 jhb11@msstate.edu
 (662) 717-1059
 600 Russell St., NSPARC at the Mill

Office Hours: By Appointment

Course Description

Three hours lecture. An introduction to the subject of data science for students in all academic disciplines to gain literacy in the subject. Students will **not** be required to perform mathematics or computer programming.

Class Schedule and Due Dates

Date		Topic	Assignment
08-19-2021	TH	Introduction to the Field of Data Science, Part 1	First Day of Class
08-24-2021	TU	Introduction to the Field of Data Science, Part 2	Lecture, Readings
08-26-2021	TH	Sources and Types of Data, Part 1	Lecture, Readings
08-31-2021	TU	Sources and Types of Data, Part 2	Lecture, Readings
09-02-2021	TH	Data Science Tools and Resources	Lecture, Readings
09-07-2021	TU	Privacy, Confidentiality, and Ethics, Part 1	Lecture, Readings
09-09-2021	TH	Privacy, Confidentiality, and Ethics, Part 2	Lecture, Readings
09-14-2021	TU	Data Integrity, Management, Governance and Security, Part 1	Lecture, Readings
09-16-2021	TH	Data Integrity, Management, Governance and Security, Part 2	Lecture, Readings
09-21-2021	TU	Datafication: Representing the World with Data Objects, Part 1	Lecture, Readings
09-23-2021	TH	Datafication: Representing the World with Data Objects, Part 2	Lecture, Readings
09-28-2021	TU	Computational Aspects of Data Science (Barlow at SETA Conference)	Lecture, Readings
09-30-2021	TH	Statistical Inference / Learning, Part 1	Lecture, Readings, Video
10-05-2021	TU	Mid-Term Examination	Lecture, Readings
10-12-2021	TU	Statistical Inference / Learning, Part 2	Lecture, Readings, Video
10-14-2021	TH	Data Visualization	Lecture, Readings
10-19-2021	TU	Artificial Intelligence, Introduction Part 1	Lecture, Readings, Video
10-21-2021	TH	Artificial Intelligence, Introduction Part 2	Lecture, Readings, Video
10-26-2021	TU	Artificial Intelligence, Deep Learning Part 1	Lecture, Readings, Video
10-28-2021	TH	Artificial Intelligence, Deep Learning Part 2	Lecture, Readings, Video
11-02-2021	TU	The Practitioner's Perspective	Guest speaker who performs hands-on data science work. Lecture. Readings.
11-04-2021	TH	The Business Leader's Perspective	Guest speaker who depends upon data science to improve the performance of a business organization in terms of finance or marketing. Lecture. Readings.

11-09-2021	TU	Data Science in Policing, Security, and Corrections	Lecture, Readings. Sociology readings and recommended readings.
11-11-2021	TH	Data Science in Public Health and Education	Lecture, Readings
11-16-2021	TU	Data Science in the Smart City	Lecture, Readings
11-18-2021	TH	Data Science in Marketing and Politics	Lecture, Readings
11-23-2021	TU	Closing Argument: The Promise and Importance of Responsible Data Science	Lecture, Readings
11-30-2021	TU	Exam Review and Discussion	
12-08-2021	W	Final Exam, 8:00 - 11:00	

Learning Objectives

The goal of this course is to allow students from any academic discipline to become literate in the field of data science by understanding the role data scientists and data-science-literate professionals play in improving the performance of human institutions, organizations, businesses, and society.

Upon completion of this course, students will achieve literacy in:

- Key elements and concepts of data science as a field of inquiry.
- The data science lifecycle, including the source and types of data, datafication, visualization, methods of data analysis, and machine learning.
- The data science industry sector, including supporting technologies such as computing and storage; concepts of privacy, confidentiality, and security; and supporting disciplines such as cybersecurity and data governance.
- The ways that practitioners and data-science-literate professionals use data science to transform the modern world in many fields.

Course Materials

The course will make use of articles from popular and academic journals/books and publicly available materials. Readings will be available via links shared through email or on Canvas. The course will also make use of selected online, explanatory videos.

Class Assignments

An overview of the major course assignments and grading scale follows.

- Readings and Videos. You will prepare for the concepts to be covered in class lectures by reading assigned articles and watching assigned explanatory videos.
- Mid-term Examination. A multiple-choice mid-term examination will assess student comprehension of course concepts over the first half of the semester.
- Final Examination. A multiple-choice final examination will assess student comprehension of course concepts over the second half of the semester.

Methods of Evaluation and Standards of Achievement

Assignment	Points
Attendance and Participation	25
Mid-term Examination	35
Final Examination	40

Letter Grade	Percentage
A	89.50—100% of points
B	79.50—89.49% of points
C	69.50—79.49% of points
D	59.50—69.49% of points
F	Less than 59.50% of points

Student Honor Code

Mississippi State has an approved Honor Code that applies to all students. The code is as follows: “As a Mississippi State University student, I will conduct myself with honor and integrity at all times. I will not lie, cheat, or steal, nor will I accept the actions of those who do.” Upon accepting admission to Mississippi State University, a student immediately assumes a commitment to uphold the Honor Code, to accept responsibility for learning, and to follow the philosophy and rules of the Honor Code. Student will be required to state their commitment on examinations, research papers, and other academic work. Ignorance of the rules does not exclude any member of the MSU community from the requirements or the processes of the Honor Code. For additional information, please visit: <http://honorcode.msstate.edu/policy>.

Title IX

MSU is committed to complying with Title IX, a federal law that prohibits discrimination, including violence and harassment, based on sex. This means that MSU’s educational programs and activities must be free from sex discrimination, sexual harassment, and other forms of sexual misconduct. If you or someone you know has experienced sex discrimination, sexual violence and/or harassment by any member of the University community, you are encouraged to report the conduct to MSU’s Director of Title IX/EEO Programs at 325-8124 or by e-mail to titleix@msstate.edu. Additional resources are available at Dean of Students Sexual Misconduct and Sexual Assault.

Disability Resource Center

Mississippi State University is committed to providing equitable access to learning opportunities for all students. The Disability Resource Center (01 Montgomery Hall) collaborates with students who have disabilities to arrange reasonable accommodations. If you have, or think you may have, a disability, please contact drc@saffairs.msstate.edu or 662-325-3335 to arrange a confidential discussion regarding equitable access and reasonable accommodations. Disabilities may include, but are not limited to, conditions related to mental health, chronic health, attention, learning, autism, brain injury, vision, hearing, mobility, speech, or intellectual disabilities. In the case of short-term disabilities (e.g., broken arm), students and instructors can often work to minimize barriers. If additional assistance is needed, please contact the Disability Resource Center.

Attendance Policy for Face-to-Face Instruction

This section is a face-to-face instructional class. Please refer to Academic Operating Policy 12.09, regarding attendance expectations and accommodations.

Continuity of Instruction

In the event that face-to-face classes are suspended due to extenuating circumstances, such as weather, the instructor will continue instruction in a manner that best supports the course content and student engagement. In this event, all instructors will notify students of the change via their university email address (the official vehicle for communication with students). At that time, they will provide details about how instruction and communication will continue, how academic integrity will be ensured, and what students may expect during the time that face-to-face classes are suspended. If a student becomes unable to continue class participation due to extenuating circumstances, (e.g., health and safety, loss of power, etc.) the student should contact their instructor and advisor for guidance. For additional guidance, please refer to Academic Operating Policy 12.09.

University Safety Statement

Mississippi State University values the safety of all campus community members. Students are encouraged to register for Maroon Alert texts and to download the Everbridge App. Visit the Personal Information section in Banner on your MyState portal to register. To report suspicious activity or to request a courtesy escort via Safe Walk, call University Police at 662-325-2121, or in case emergency, call 911. For more information regarding safety and to view available training including helpful videos, visit ready.msstate.edu.

Topical Outline of Contact Hours

1. **Introduction to the Field of Data Science (3 Hours)**
 - a. Course Overview and Orientation
 - b. Defining the nascent field of Data Science
 - c. Data Science literacy and why it is important
 - d. What knowledge, skills, and abilities makes someone a data scientist?
 - e. Functional differentiation of different kinds of Data Scientists (Management, Technical, Hands-On, etc.)
 - f. Data Science in:
 - i. Academia and Research
 - ii. Business and Industry (Digital Transformation)
 - iii. Society
 - g. What is/are Data? (Definition, Distinctions Based on Use, etc.)
 - h. Datafication
 - i. Understanding the Practical High-Level Flow or Process
 - i. Identify or create data source / design an experiment
 - ii. Data collection, cleaning, munging, wrangling (“data are never clean”), storage
 - iii. Mapping / understanding the data
 - iv. Exploratory analysis (visualization) to find patterns, build models
 - v. Routinization or Expansion of Analysis to Entire Dataset.
 - vi. Data as integral part of operational processes
 - vii. Communication of insight gained, application of insight to new problems
2. **The Sources and Types of Data, Big Data (3 Hours)**
 - a. What to notice in each of the following categories:
 - i. Format and content of the data source
 - ii. Purpose / Intention for which it is created or collected
 - iii. Alternative Purposes to which it could be put
 - iv. Structure / Lack of Structure
 - v. Social Meaning of Data Source
 - b. Category 1: Data produced by Human Actions (Manual)
 - i. Information / Measurements Gathered as part of an Experimental Design
 - ii. Case Management
 - iii. Digital Communications (Email, SMS)
 - iv. Social Media (Twitter, Facebook)
 - v. Mass Communication
 - vi. Opinion Polling / Surveys / Political Polling
 - c. Category 2: Data produced by Sensors (Data Acquisition Systems)
 - i. Cameras
 - ii. Internet of Things
 - iii. Personal / Private - Smart thermostat, health-monitoring bracelet, wearable
 - iv. Public - Smart City Sensors – cameras, smart transportation, trains
 - v. Modern manufacturing and industrial production
 - vi. Farming
 - vii. Chemical and Other Manufacturing (for Automation of Maintenance)
 - d. Category 3: Data from Computer Systems
 - i. Application Logs
 - ii. Operating System Logs
 - iii. Access Logs – (example: Google Analytics)
 - e. Discussion: Public vs. Alternative Data and Competitive Advantage
3. **Data Science Tools and Resources (3 Hour)**
 - a. Goal: Awareness
 - b. Programming Languages: R, Python, Scala, etc.
 - c. Tools: SAS
 - d. Frameworks: TensorFlow, PyTorch, SciKit, etc.
 - e. Notebooks: Jupyter, etc.

- f. Competition Sites for Learning: Kaggle
- g. Online Datasets for Social Science, Health, Humanities, the Arts, Physics, etc.
- 4. **Privacy, Confidentiality, and Ethics (3 Hours)**
 - a. Three Important Concepts important for all data
 - i. Privacy
 - ii. Confidentiality
 - iii. Security
 - b. Relevant Laws
 - i. 1914 Federal Trade Commission Act
 - ii. 1970 Fair Credit Reporting Act
 - iii. 1974 Privacy Act
 - iv. 1986 Electronic Communications Privacy Act
 - v. 1986 Computer Fraud and Abuse Act
 - vi. 1999 Gramm-Leach-Bliley Act
 - vii. 1996 HIPAA
 - viii. 2003 CAN-SPAM
 - ix. 2009 Health Information Technology for Economic and Clinical Health Act
 - x. 2012 – FERPA
 - xi. Other Federal Examples
 - xii. State Level Laws (e.g., 2018 California Consumer Privacy Act)
 - xiii. International Laws - 2016 GDPR
 - c. Ethics of Data Use
 - i. General Principles
 - ii. Industry- or Job-Specific Principles
 - iii. Preservation, Destruction, Non-Use
 - iv. Attempts to Detect and Address Bias in Analytics
 - d. Using Data in Academia
 - i. Research Exceptions
 - ii. IRB: Social Science, Psychology, Health, and other Disciplines
- 5. **Data Integrity, Management, Governance, and Security (3 Hours)**
 - a. Definitions
 - b. Data Governance + Examples
 - c. Case Studies of Data Breaches (Yahoo, Equifax, SolarWinds)
 - d. Organizations must have a Data Storage and Management Strategy to ensure Privacy, Confidentiality, and Security
 - e. Industry Standard Strategy - NIST Framework
 - i. NIST: Identify
 - 1. Definition
 - 2. Inventory
 - 3. Governance
 - 4. Risk Assessment
 - 5. Risk Management
 - ii. NIST: Protect
 - 1. Definition
 - 2. Access Controls
 - a. Access vs. Entitlements
 - b. Hierarchical User Roles
 - 3. Awareness and Training
 - 4. Data Security Safeguards
 - 5. Processes and Procedures to Protect Information
 - 6. Maintenance
 - 7. Protective Technologies
 - 8. Throughout Data Lifecycle
 - a. Acquisition / Transfer
 - b. Storage
 - c. Cleansing

- d. Access Controls and Roles
 - i. Physical Security
 - ii. Hardware Security
 - iii. Software Security
 - iv. Network Security
 - e. Auditing and Monitoring
 - f. Destruction
 - 9. Encryption, Decryption, Authentication
 - a. Storage (at rest) and transfer (in motion)
 - b. Ciphers (symmetric / asymmetric)
 - c. Example: Passwords
 - d. Example: Browser Security
 - e. Example: Corruption of Navigation Info
 - 10. Physical Security
 - a. Locks
 - b. Keys (Physical, codes, biometrics, Multi-Factor)
 - c. Separation
 - d. Access Records
 - e. Prevention of Line-of-Sight / Earshot Exploits
 - f. Shredding / Destruction
 - g. Social Hacking
 - 11. Maintenance of Protection
 - a. Changing Keys / Passwords
 - b. Certificate Renewal
 - c. Review
 - d. Bias towards minimal entitlements until escalation is needed.
 - e. Etc.
 - iii. NIST: Detect
 - 1. Definition
 - 2. Continuous Monitoring
 - 3. Logs
 - 4. Intrusion Testing (automated, white-hat)
 - 5. Awareness of Public Notices
 - 6. Zero-Day Exploits
 - 7. Physical Monitoring (heat, motion, sounds)
 - 8. Reputation Monitoring
 - 9. Artificial Intelligence Approaches
 - iv. NIST: Respond
 - 1. Definition
 - 2. Planning for Response
 - 3. Communications
 - 4. Mitigation
 - 5. Improvements
 - v. NIST: Recover
 - 1. Definition
 - 2. Advance Recovery Planning
 - 3. Improvements
 - 4. Communication
 - 5. Reputation Restoration
 - f. Additional Considerations for the Cloud(FISMA)
6. **Datafication, Representing the World with Data Structures (3 Hours)**
 - a. Data Objects and Properties
 - b. Object-Oriented Thinking
 - c. Representation of Things, Processes, Human Systems as Objects
 - d. Property Types (Strings, Dates, Times, Numbers, etc.)
 - e. Data Models – Conceptual, Logical, Physical

- f. Concepts: Flat vs. Normalized
 - g. Database Types
 - i. Relational (Also introduce SQL)
 - ii. Unstructured (Key Value) + Indexing
 - iii. Graphs
 - iv. Blockchain
 - h. Difference Between *Data Model*, *Statistical Model*, *Machine Learning Model*
 - i. Proxies for Unavailable Attributes
 - j. Key Takeaway: The properties of an object form the basis for data analysis / operations / strategy
7. **Computational Aspects of Data Science (3 Hours)**
- a. Balancing Costs: Time, Electricity, Hardware
 - b. Storage and Retrieval of Data (Storage Strategies)
 - i. Persistent Storage
 - ii. System Memory / Short-Term
 - iii. Storing big data cheaply – large redundant volumes, Google FS, Big Table, Hadoop
 - iv. Retrieving Data, Pagination, Micro-Batching, “Real Time”
 - v. Storing big data expensively (common enterprise options)
 - vi. How Databases relate to Storage
 - 1. In-Memory
 - 2. Persistence
 - vii. Digitizing Old Paper Data
 - c. Networking Aspects of Data
 - i. Transmission of Objects across a Network
 - ii. Transport Protocols (REST, SOAP)
 - iii. Encryption
 - d. Processing Data
 - i. Traditional supercomputing
 - ii. Large, networked commodity computing
 - iii. GPUs and TPUs
 - iv. Specialized Tasks (e.g., Bitcoin) - TPUs
 - v. Cloud versus On-Premises
 - vi. What about Quantum Computers?
 - e. Reducing the Size of a Computing Task
 - i. Sampling
 - ii. Reduction of Precision
 - iii. Batch Processing
 - iv. Parallel Processing
 - v. Extended MapReduce Example
 - f. Data Science and Data Engineering
 - i. Building Analysis Pipelines
 - g. Spark + Hadoop
8. **Statistical Inference / Learning (3 Hours)**
- a. Probabilistic Systems
 - b. Populations and samples
 - c. Probability distributions used as foundations for statistical modeling
 - d. Prediction
 - e. In-class discussion of Video example: <https://www.youtube.com/watch?v=1UV1Q14oiL8>
 - f. Regression (Linear, Logistic), Decision Trees, Support Vector Machines, k-Nearest
 - g. Models and Predictions
 - i. What is a “model” (e.g., climate model, model for baseball success, etc.)
 - ii. Fitting a model to data
 - iii. Creating a model
 - iv. Using a model
 - v. Famous models in different academic areas
 - h. Other Examples – Polling, Prevention of Fraud, Opioids, Workforce

9. **Data Visualization (1 Hour)**
 - a. Definition and Purpose
 - b. Communication, Persuasion, and Understanding
 - c. Tufte and “The Visual Display of Quantitative Information”
 - d. History
 - i. Examples from Humanities, Cartography, History
 - ii. William Playfair – Bar Chart, Line Chart, Area Chart, Pie Chart, Circle Graph
 - e. Principles for Excellence in Statistical Graphics with Examples
 - f. Exploratory Visualizations (e.g., Histograms, Box Plot, Heat Map, Pareto, Scatter)
 - g. Explanatory Visualizations (e.g., Infographics)
 - h. Discussion of Business Intelligence Systems and Dashboards
10. **Artificial Intelligence – Intro and Machine Learning (3 Hours)**
 - a. Definition of AI (Moving Goalposts)
 - b. Bacon, Turing, Reductionism, Importance of Subject-Matter Expertise
 - c. Precursors: Rule-Based Systems, Expert Systems, Algorithms
 - d. Comparison of rule-based vs. machine learning language translation example from Google Translate
 - e. A Machine Learning approach to learning tasks
 - f. Types of Learning: Supervised, Unsupervised, Semi-Supervised, Reinforcement
 - g. AI Now
 - i. Search / Optimization
 - ii. Probabilistic Methods
 - iii. Classifiers
 - iv. Artificial Neural Networks
 - h. Positive Predictions about AI / Fear and Skepticism / Impacts on Labor Market and Society, Concerns of Bias
11. **Artificial Intelligence – Artificial Neural Networks and Deep Learning (3 Hours)**
 - a. Neural network concepts – node, weight, layers, activation
 - b. Walkthrough of Neural Network Training based on Homework Video
 - i. Input layers
 - ii. Hidden layers
 - iii. Back-propagation
 - iv. Activation
 - c. A Few Types of ANNs and their Applications
 - i. Feed-forward
 - ii. Radial Basis
 - iii. Kohonen Self-Organizing
 - iv. Recurrent
 - v. Convolutional
 - vi. Modular
 - vii. Generative Adversarial + Examples
 - d. The impact of model size: GPT-3 and OpenAI
 - e. Term: “Deep Learning”
 - f. Learning to Rank, Learning to Recommend
 - g. Applications (Dating Apps, Advertising, Psychological Persuasion, Sentiment Analysis, Netflix movie recommender, Facebook, Amazon, Linked-In)
12. **Specific Applications of Data Science (12 Hours)**
 - a. The Practitioner’s Perspective in Data Science and Data Engineering
 - i. Jobs for practitioners in Data Science in many fields
 - ii. Pure Science
 - iii. Data Engineering
 - b. The Business Leader’s Perspective
 - i. Data Science Literacy and Competitive Advantage
 - ii. Business Organizations
 - iii. Financial and Investing Advantages
 - iv. Digital Transformation in Industrial Processes: Digital Transformation Social Mindset

- c. Data Science in Policing, Security, and Corrections
 - i. Recent Data-Driven Policing Strategies
 - ii. Crime Prevention
 - iii. Predicting Recidivism
 - iv. Recommended: “Big Data Sociology” from *The Sociological Review*
 - v. Advanced Recommendation: C. Wright Mills, *Sociological Imagination*
- d. Data Science in Public Health and Education
 - i. Health and Health Informatics
 - ii. Personalized Health and Education
 - iii. Public Health
- e. Data Science in the Smart City
 - i. Defining “Smart City”
 - ii. Infrastructure and Sensors
 - iii. Government, Policy, and Program Performance
 - iv. Examples and Experiments
- f. Data Science in Marketing and Politics
 - i. Public Opinion, Representation, and Gerrymandering
 - ii. Micro-Targeting
 - iii. Examples: Cambridge Analytica, “Fake News”
- g. Closing Argument: The Promise and Importance of Responsible Data Science
 - i. Ten Commandments for Responsible Data Science
 - ii. Next Steps at MSU – Research, Academics, and Entrepreneurship

Total: 45 Hours