Syllabus Project ENGL 380
Group: Steven Taruc, Jake Esprabens, Montserrat Dabkowski

An Ethical Dive into Data

Meeting Times:
MW 2-4pm

Course Description: The application of mathematics and statistics in the development of technology has progressed the digital world to great lengths with extremely accurate algorithms, big data processing, and artificial intelligence. Throughout our daily lives, we encounter these analytical methods when we use our smartphones, watch sports, or even vote in an election. These applications create automation in complex tasks efficiently, or have the ability to optimize problems that were once unsolvable. Although, this constantly evolving field is controversial in our current climate. Social and moral issues arise within data science through ill-intent or violations in data collection and storage. In essence, Data Science is a double-edged sword in affecting the technical and ethical outcomes of technology. To gain a deeper understanding of data science in the real world, we must consider the following questions: What are the correct methods to collect data? What are the benefits of using machine learning and artificial intelligence and how do they work? What must we be wary about when applying these algorithms? Lastly, what are its consequences and how can we prevent them?

Course Objectives: In this course, we will investigate the technical and ethical sides of mathematics and statistics, and the role they play in the real world. Since this is an upper-level curriculum, our main goal is to gain knowledge and critical thinking on algorithms, statistical methods, proofs, and advanced techniques used in industry. We will also cover the importance of collecting and analyzing data honestly and justly. In addition, we will study the effects of mathematics, statistics, and data science in real world examples. To accomplish these goals, we will assign practice sets throughout the quarter and journal discussions on current events. The beginning of the quarter will cover basic definitions and mathematical proofs pertaining to theory. Then, we will dive into research and overarching benefits of data science. The middle of the quarter focuses on technical and ethical applications of data science and how we can evaluate their results. It is also important to consider entry into the data science field because it

continues to be a significant field in our advancing world. Therefore, we will cover access to technology near the end of the quarter. The final project presentation pertaining to your specific major will allow you to work in an industry-like work environment and progress your career interests. Overall, this course is intended to set you up in the right direction for your career in STEM.

Summary of Learning Objectives:
- Ability to apply correct statistical and mathematical methods in the data science field
- Understand mathematical theory behind algorithms and utilize proofs to gain a deeper reasoning to specific methods
- Learn how to honestly and justly collect and analyze data and understand how this affects society in the real world.
- Become aware of ethical concerns in AI/ML implementation to prevent consequences or to efficiently remediate them.
- Strengthen student involvement with current events by analyzing recent articles and groundbreaking discoveries in tech.
- Allow students to be wary and lenient with the access to technology in education to enforce inclusivity.
- Work on team-based projects in an industry-level setting to prepare for future careers and to practice public-speaking skills.

Required Readings:
- Ambalina, Limarc. "Global Coronavirus Dataset Library." *Lionbridge AI*, 26 May 2020, lionbridge.ai/datasets/coronavirus-datasets-from-every-country.
- Bracken, Becky. "Trump Site Alleging AZ Election Fraud Exposes Voter Data." *Threatpost*, 12 Nov. 2020, threatpost.com/trump-site-alleging-az-election-fraud-exposes-voter-data/161068.
- Buranyi, Stephen. "Rise of the Racist Robots - How AI is Leaning All Our Worst Impulses." *The Guardian*, www.theguardian.com/inequality/2017/aug/08/rise-of-the-racist-robots-how-ai-is-learning-all-our-worst-impulses. Accessed 7 November, 2020.
- Faghmous, James H., and Vipin Kumar. "A Big Data Guide to Understanding Climate Change: The Case for Theory-Guided Data Science." *Big Data*, 2014, www.ncbi.nlm.nih.gov/pmc/articles/PMC4174912.
- Gil Press. "A Very Short History Of Data Science." *Forbes*, 15 Oct. 2014, www.forbes.com/sites/gilpress/2013/05/28/a-very-short-history-of-data-science/?sh=32df21a755cf.
- Grus, Joel. *Data Science from Scratch: First Principles with Python*. 2nd ed., O'Reilly Media, 2019.
- Jackson, Debra L. and Paul A. Newberry, *Critical Thinking: A User's Manual, ed 2.* Cengage Learning, 2015.
- Narang, Nirbhay, and Mehul Jangir. "Analyzing Data on the Spread of COVID-19 Using Statistical Tools to Predict the Inflexion Point of the Virus in Italy." *Indian Journal of*

Computer Science, vol. 5, no. 2 & 3, 2020, p. 7. *Crossref*,
doi:10.17010/ijcs/2020/v5/i2-3/152206.
- O'Neil, Cathy. *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy*. Reprint, Crown, 2017.
- Ratschat, Jonathan. "How to Collect Data for Your Analysis - Towards Data Science." *Medium*, 6 June 2020,
  towardsdatascience.com/how-to-collect-data-for-your-analysis-a8bc58043e64.
- Rosen, Kenneth H. *Discrete Mathematics and Its Applications, ed 7.* McGraw Hill, 2012.
- Schneier, Bruce. *Data and Goliath: The Hidden Battles to Collect Your Data and Control Your World*. Reprint, W. W. Norton & Company, 2016.
- Wexler, Natalie. "How Classroom Technology Is Holding Students Back." *MIT Technology Review,*
  www.technologyreview.com/2019/12/19/131155/classroom-technology-holding-students-back-edtech-kids-education/. Accessed 8 November, 2020.
- Wong, Alia. "Why Millions of Teens Can't Finish Their Homework." *The Atlantic*,
  www.theatlantic.com/education/archive/2018/10/lacking-internet-millions-teens-cant-do-homework/574402/. Accessed 5 November, 2020.
- Zawadzki, Jan. "Storytelling for Data Scientists - Towards Data Science." *Medium*, 6 Jan. 2020, towardsdatascience.com/storytelling-for-data-scientists-317c2723aa31.
- Zheng, Alice. *Evaluating Machine Learning Models*. O'Reilly, 2015.

Required Films / Videos:
- "Humans Need Not Apply." *YouTube*, uploaded by CGP Grey, 13 Aug. 2014, www.youtube.com/watch?v=7Pq-S557XQU.
- *Moneyball*. Directed by Bennett Miller, Columbia Pictures, *2011*
- Tenner, Ted. "Unintended Consequences." *TED,*
  https://www.ted.com/talks/edward_tenner_unintended_consequences/transcript?language=en#t-393318. Accessed 7 November, 2020.
- *The Imitation Game*. Directed by Morten Tyldum, Black Bear Pictures*, 2014*
- *The Social Network.* Directed by David Fincher, Columbia Pictures, *2010*

Grading Breakdown:

Attendance	10%
Problem Sets	25%
Journal Entries	25%
Final Presentation	40%

Problem Sets Grading Criteria: Problem sets will be assigned and collected according to the schedule. Each problem set is worth 10 points. Three computational problems and/or proofs will be graded closely, each out of 3 points, for a total of 9 possible points. The remaining 1 point will be awarded for completion. (This grading criterion is in keeping with how math/statistics problem sets are graded within the departments). This is to assess your ability to apply the computational and proof methods presented in this course.

Journal Entries Grading Criteria: Journal entries will be assigned and collected according to the schedule. Throughout the quarter, you will be asked to explore and reflect on the ethical implications (see the course objectives for key ethical concepts) of technology as it applies to your major, concentration, and even a minor if applicable (pure math, applied math, teaching math, statistics, and/or data science) in 500-700 words. Each journal entry is worth 10 points. 5 points will be awarded for reaching the word minimum, the remaining 5 points will be awarded for addressing the journaling prompt as it relates to the key ethical concepts.

Final Project/Presentation: Since this is a capstone course for those in the math and statistics majors (which span multiple concentrations and applicable minors), your future career paths look vastly different (e.g. those of you in the pure concentration will likely attend graduate programs in math, those of you in the teaching concentration will likely move to attain your single subject teaching credential, and those of you with the data science minor will likely work in computer science adjacent fields). Thus, to make this course most meaningful to you as you move on from Cal Poly, you will be completing a final project (in groups of 2-3) and giving ~15 minute presentations during week 10, instead of a written final exam. The goal is for you to research one area of technology that is most closely related to your major, concentration, and minor if applicable, and consider its ethical implications. For example, if you are in the pure concentration you might research topological data science methods and examine how these methods can be used to serve highly targeted ads. You will need at least three credible resources (this can include research papers, textbooks, and Wolfram's MathWorld). See the final project and presentation grading rubric.

Final Project/Presentation Grading Rubric:

Criteria	Excellent: 8	Good: 5	Lacking: 3	Severely Lacking: 0
Background and introduction	Presenter effectively introduces major, concentration, and/or applicable minor, and how the covered topic relates.	Presenter somewhat effectively introduces major, concentration, and/or applicable minor, and how the covered topic relates.	Presenter does not introduce major, concentration, and/or applicable minor, or how the covered topic relates.	Presenter does not introduce major, concentration, and/or applicable minor, and how the covered topic relates.

Methods and Research	Presenter effectively introduces the methods and/or theory used in applicable proofs or computations. Presenter has at least 3 credible resources (linked at the end of presentation or turned in separately).	Presenter somewhat effectively introduces the methods and/or theory used in applicable proofs or computations (this includes too much or too little detail). Presenter has at least 2 credible resources.	Presenter does not introduce methods and/or theory that is used in applicable proofs or computations and would be helpful to the audience's understanding. Presenter has at least 1 credible resource.	Presenter fails to cover any methods and/or theory that is used in the applicable proofs or computations. Presenter has no credible resources.
Ethical considerations	Presenter effectively introduces the ethical considerations of their chosen topic	Presenter introduces the ethical considerations of their chosen topic, but may be difficult to follow	Presenter introduces ethical considerations of their chosen topic, but does not allocate sufficient thought or time to it	Presenter fails to introduce any ethical considerations of their chosen topic
Visual aids and creativity	Visual aids are aesthetically pleasing and aid the audience in understanding methods or ethical considerations	Visual aids help the audience in understanding methods or ethical considerations but is somewhat lacking aesthetically	Visual aids while aesthetically pleasing do not help the audience in understanding methods or ethical considerations (the aids may be too technical to be helpful)	Visual aids do not help the audience in understanding methods or ethical considerations and are not aesthetically pleasing
Presentation and speaking skills	Presenter consistently speaks clearly and maintains good eye contact	Presenter speaks clearly and mostly maintains good eye contact	Presenter mostly speaks clearly and attempts eye contact	Presenter is consistently difficult to understand and avoids eye contact

Schedule:

Week	Monday Meeting (2-4pm)	Wednesday Meeting (2-4pm)	Assignments (due Friday 11:59pm on Canvas)
1: What is Data Science/Statistics/Math?	**Topics:** ● What is Math and Statistics? ● What is Data Science? **Readings/Films:** ● *The Imitation Game* (2014)	**Topics:** ● The History of Data Science ● How Data Science has Affected our Current Lives **Readings/Films:** ● *A Very Short History of Data Science* (Gil) ● *Humans Need Not Apply* (CGP Grey)	**Journal Entry:** Discuss the readings or a current event that relates to these fields.
2: Methods of proof + philosophy basics	**Topics:** ● Methods of Proof Basics: Propositional logic, truth tables, Venn diagrams ● Intro to Philosophy Basics: Normative ethics (deontology, consequentialism, eudaimonism) and applied ethics	**Topics:** ● Proof Methods: direct, cases, contradiction, contrapositive, induction ● Philosophy: logical fallacies	**Problem Set:** Selected exercises from Rosen's Discrete Math and Jackson's Manual
3: Benefits of Data Science	**Topics:** ● How has Data Science Benefited us in the Past? **Readings/Films:** ● *Moneyball* (2011)	**Topics:** ● Recent Examples of How Data Science has Benefited Us **Readings/Films:** ● *Analyzing Data on the Spread of COVID-19 Using Statistical Tools to PRedict the INflexion Point of the Virus in Italy* (Narang, Nirbhay, and Mehul) ● *A Big Data Guide to Understanding Climate Change* (Faghmous and Vipin)	**Journal Entry:** Discuss current beneficial uses of data science in our world
4: Ethics of Data Science	**Topics:** ● What does it mean to be ethical when using Data Science? **Readings/Films:** ● *Data and Goliath: The Hidden Battles to Collect Your Data and Control Your World.* (Part II: What's at Stake) (Schneier)	**Topics:** ● Examples of Ethical Uses of Data Science ● Examples of Unethical Uses of Data Science **Readings/Films:** ● *The Social Network* ● *Weapons of Math Destruction* (Chapter 5. Civilian Casualties: Justice in the Age of Big Data) (O'neil)	**Journal Entry:** Discuss the readings or other current examples of the role ethics plays in the field of data science.
5: Basic research and unintended	**Topics:** ● What are unintended	**Topics:** ● Algebraic homology review	**Journal Entry:** Discuss an area of

consequences	consequences? ● Famous examples of unintended consequences **Readings/Films:** ● Ted Tenner's Ted Talk *Unintended Consequences* (Tenner)	● Algebraic homology's unintended application to data science and its implications **Readings/Films:** ● *Rise of the Racist Robot* (Buranyi)	research that applies to your concentration/ major and what some of the unintended consequences of the research findings may be
6: Collecting data and sampling methodology	**Topics:** ● Types of data and where to extract them from. ● Data organization **Readings/Films:**: ● *How to Collect Data for Your Analysis* (Ratschat)	**Topics:** ● Correct and ethical sampling techniques ● Random sampling, bootstrap sampling, distributions. **Readings/Films:** ● *Election Fraud exposes Voter Data*	**Journal Entry:** Subset a dataset from the Global Coronavirus Dataset Library and discuss observations from this dataset.
7: Machine learning and Artificial Intelligence methods	**Topics:** ● Quantitative algorithms. (Linear Regression, K-Nearest Neighbors Regression) ● Bayes Theorem	**Topics:** ● Classifiers (Logistic Regression, K-Means, Decision Trees, LDA, SVM) ● Deep Learning (Neural Networks) **Readings:** ● *Data Science from Scratch: First Principles with Python* (Chapter 12: Machine Learning) (Grus)	**Journal Entry:** From the Global Coronavirus Dataset Library, develop a machine learning algorithm of your choice and discuss your findings.
8: Evaluating models and interpretations	**Topics**: ● Metrics of evaluation (MSE, precision and recall, accuracy) ● Cross Validation (training and testing data) **Readings/Films:** ● *Evaluating Machine Learning Models* (pg. 7-18) (Zheng)	**Topics:** ● Interpretability versus predictability trade-off **Readings/Films:** ● Storytelling for Data Scientists (Zawadzki)	**Problem Set:** Using the algorithm you developed from the Coronavirus dataset: train and test your model, evaluate its performance, and interpret its results.
9: Access to technology	**Topics:** ● Inclusivity of technology access in teaching ● Big tech's interest in education **Readings/Films:** ● *Why millions of teens can't finish their homework* (Wong)	**Topics:** ● How the technology you use in the classroom affects student learning and students morale **Readings/Films:** ● *How classroom technology is holding students back* (Wexler)	**Journal Entry:** Address one specific barrier that technology represents and how you would address this issue as an educator (focus on middle/high school math/stats)
10: Final Presentations	Presentations	Presentations	

Attendance:
Attendance is mandatory for each class period and is part of your grade. Regular attendance has been proven in many studies to be one of the greatest factors influencing academic success. Moreover, it is easy to get lost in these concepts if you are not regularly attending the class. Regularly attending class, asking questions, and participating in discussions are all strongly recommended because they will help you understand the difficult statistical and mathematical concepts, theories behind the machine learning algorithms, the ethics behind the field of data science, and will keep you updated with the current uses of these implementations.

Academic Honesty:
While this class does promote and encourage discussions with other students, there will be no toleration for plagiarising other works. This includes copying other students' discussions, using quotes from resources as your own, and working together on exams. This is a reminder that the ultimate goal of this class is to learn about the applications, ethics, and effects of data science in our world. Cheating will not help you with these goals. A first instance of cheating will result in a zero on the assignment and a second instance of cheating will result in an automatic failure in this class. Please take a moment to review [Cal Poly's Academic Integrity Policy](#).

Accessibility and Diversity:
- Disability Resource Center:
    - If you wish to request an accommodation due to a suspected or documented disability, please inform me and contact the Disability Resource Center as soon as possible.
- Student Writing Center:
    - The Student Writing Center offers one-on-one sessions staffed by peer learning facilitators trained to help at any stage in the writing process and with writing assignments.
- Diversity
    - This class is created for people of all different races, genders and backgrounds. The primary goal of this class is for you to learn about the backbones of data science and how it is applied in the real world. All students are encouraged to participate in the discussions and class periods because it will ultimately help the rest of the students understand the material. If you ever feel uncomfortable in class because of your race, gender, or background, please contact me and I will do my best to make a more inclusive environment.

Email Policies:
If there are any questions about assignments that you are comfortable posting to the whole class, please post your question in the discussion forum titled, Questions. If you would rather not post your questions publicly or if you have personal questions, please feel free to email me at _____@calpoly.edu.