Teaching the Essentials of Machine Learning in the Context of Quantitative Information Literacy

Anonymous Authors[1]

Abstract

X College, is a small liberal arts postsecondary institution in the United States. An information literacy course, Calling Bull, serves as an introductory data science class as well as a prerequisite-free quantitative literacy class. In this context, we spend a week discussing machine learning, with an emphasis on facial recognition algorithms. The emphasis is on the general algorithmic approach, critical inquiry of the process and careful interpretation of results presented in research or other decision-making. This module relies on the use of Open Educational materials, discussion, and careful attention to issues of marginalization and algorithmic justice.

1. Introduction

Calling Bull is an open educational course developed by quantitative biologist Carl Bergstrom and information scientist Jevin West. Its goal is to prepare students for a world of big data by introducing students to tools and techniques for sorting through information in the data economy and understanding what drives information in the digital and scientific ecosystems.

Bergstrom and West published a book by the same name in 2020, but their website

`http://callingbullshit.org/`

(and the k-12 friendly companion website callingbull.org), has a suggested syllabus, case studies and links to lectures on YouTube (Bergstrom & West, 2021).

In 2019, I adopted and adapted Calling Bull for a specific context at X College in the Digital and Computational Studies program as part of a broader data science set of courses. X College is a small liberal arts school in the United States, and the Digital and Computational Studies program is a new interdisciplinary department which is designed to bridge multiple disciplines with computation and digital.

The goals of the Calling Bull course from an institutional perspective were 1) collaborate with a theme of interest to social science majors, such as economics and politics, 2) provide a gentle introduction to programming with R, 3) to reinforce key skills such as interpreting graphs and understanding uncertainty in science as well as introduce data science, and 4) meet the quantitative literacy standards. All students regardless of major take one such quantitative literacy course. I found Calling Bull a particularly compelling theme for a quantitative literacy course because it captures key quantitative reasoning and critical thinking skills that I want all graduates of our college to have.

The course learning objectives for students are[1]:

> **This course is designed as a community learning journey**
> Together, we will:
>
> - Metacognitively engage in contemporary issues in equity and social justice related to their digital world, community, and identity. (Think about this class every time you hear the news, make daily choices, or even put your shoes on.)
> - Play with computational ideas creatively, using a growth mindset which values revision and experimentation and demonstrate community leadership skills as a collaborator that shares strengths, builds weaknesses, and contributes to a broader shared understanding. (Participate in teamwork in respectful ways that allow people to relax and play with ideas).
> - Recognize and translate between algebraic, numeric, visual, and verbal representations

[1]Anonymous Institution, Anonymous City, Anonymous Region, Anonymous Country. Correspondence to: Anonymous Author <anon.email@domain.com>.

Preliminary work. Under review by the Teaching Machine Learning Workshop at ECML 2021. Do not distribute.

[1]Most of the parentheticals are borrowed with permission directly from the syllabus on the Calling Bull website (Bergstrom & West, 2021)

of data. (Remain vigilant for bull contaminating your information diet and recognize said bull whenever and wherever you encounter it.)

- Design models of and computationally investigate ideas in practical and professional spaces through and communicate the process and meaning to others. (Figure out for yourself precisely why a particular bit of bull is bull, provide a statistician or fellow scientist with a technical explanation of why a claim is bull - using R and employing proper data visualization techniques where necessary, and provide your "casually racist uncle" with an accessible and persuasive explanation of why a claim is bull.)

To support these course learning objectives, I blended the Calling Bull curriculum from Bergstrom and West with programming instruction and projects in R, and supplemented with daily data visualization activities, metacognitive reflections, and other activities. A full set of curriculum from can be found on QUBESHub.org. An aggregated list of this and other related community contributed curriculum can be found at the Calling Bull Instructors group at `callingbull.qubeshub.org`.

An emphasis of all courses taught by Digital and Computational Studies program faculty is the explicit attention to human rights and social justice, interrogating racism, sexism, bigotry, and other forms of exclusion in the design of digital spaces and algorithms. In this context, students have multiple opportunities to ask questions about how our assumptions and biases affect our data collection, models, and visualizations, as well as affect the interpretation and communication of our data analysis and model results.

Below, I present a module constructed for our version of the Calling Bull class, which uses the Jevin and West lecture videos and suggested readings as a foundation. These are supplemented with with other freely available resources - videos by Joy Buolamwini and a research paper by Garg et al. utilizing word embeddings (Garg et al., 2018) to develop a broader view of computing ethics and human rights beyond concerns driven by phrenology pseudoscience and to broaden the voices of those represented in the conversations about the ethics, process, uses and abuses of machine learning models. This module occurs about halfway through the semester, just before their mid-term independent project.

2. Machine Learning Module

The course is designed as a series of one week modules for two 80 minute classes. However, I am presenting the most recent version which was conducted online in a compressed 7 week module. All one-week modules were reallocated to a single hour and 45 minutes synchronous session with additional asynchronous time. In the two classes per week format, I would assign reading for the first day to introduce the topic, and then a R lab experience on the second day that uses that theme, typically based on a case study, that would be due at the end of the week along with a written reflection. The big data module is an exception to this general rule - the case studies are part of the thematic introduction, and the "lab" is in essence a racial equity intervention.

2.1. Pre-Module reading

Students are asked to watch or read the following items before class. With the exception of the research paper, all are from the Calling Bull website.:

- Videos: Lecture 5 Big Data on YouTube
- Case Study: Criminal Machine Learning
- Case Study: Machine Learning about Sexual Orientation?
- Research Paper: Word embeddings quantify 100 years of gender and ethnic stereotypes (Garg et al., 2018).

Each YouTube "lecture" is broken into a handful of shorter videos, and in total are typically less than one hour. During the Big Data Lecture 5 videos, Bergstrom and West introduce the basic idea of machine learning and how training models based on biased data and assumptions can lead to faulty results. They then verbally debunk the validity of two papers which use machine learning algorithms on photographs to determine criminality and sexuality. These arguments are also presented as case studies on the Calling Bull website - and students are asked to read and comment on these and the Garg et. al research paper through the use of Perusall[2] (Perusall, 2021). The use of Perusall makes it possible for students to interact social around the text before class. I am able to read these responses before class and make adaptive changes to the in-class lesson plan described below based on any misconceptions or student questions.

The research paper uses word embeddings, which combine natural language processing and machine learning methods, to explore and expose stereotype bias as expressed through language over time (Garg et al., 2018). Unlike the questionable papers highlighted in the case studies, Garg et al.

[2]Perusall is a collaborative annotation software. The Perusall grading algorithm itself is presented in the first day of the semester as an example of a black-box proprietary algorithm driven by machine learning, and so is brought up as a reprise in this class as well. Given the critique we make in class, I relax many of conditions of the Perusall automated grading algorithm and all scores less than 100% are manually re-graded by a teaching assistant.

showcases cutting edge interdisciplinary techniques, utilizes critical quantitative inquiry, and exposes bias of machine learning models, all while exploring an interesting research question. This paper provides a disciplinary bridge between computer science and quantitative critical social science in which the use of machine learning as a research tool acts against racism and sexism instead of propagating it. The use of word embeddings to measure stereotype bias leverages the fact that machine learning is biased by the data (here historical documents) with which is it represented. This expands the scope of the discussion from image classification for criminality and sexuality to using machine learning to find language associations that reveal historical perspectives on gender, race and ethnicity. Finally, the visualizations presented in the results section are accessible to the audience which at this point has completed a regression project with Pearson correlation coefficient interpretation.

2.2. In-class discussion of readings

On the whole, the class uses critical pedagogy which seeks reframe the students as knowledge holders and constructors (Freire & , Firm). After a data visualization activity and sharing reflections from students on their learning from the previous week, we open with a discussion of the videos and readings for the day. I begin with crowdsourcing definitions for big data and data science, after narrowing in on a co-constructed definition, I then ask "What does it mean that data science is changing our methodologies?" We then discuss various approaches to doing science, such as hypothesis driven, data-driven, and/or mathematical model-driven using the Rule-of-Five framework (Diaz Eaton et al., 2019). This helps us derive a common understanding of language and approaches to research questions. The goals is to help students see an inclusive framework for modeling by which data science is one approach, decentering a particular idealization of any one approach in an effort to decolonize our discussion of methods.

We then break into groups[3] with the following prompt:

1. List 5 major insights from videos
2. How did these insights about big data manifest in the readings?
3. Pick at least one of the 3 readings and post on the Jamboard (Google, 2021): Which reading your group picked and two ways the big data concerns or potential bullshit issues manifest in that reading.

[3] I take care to construct the groups for this modules because of the sensitive nature of the content and discussions. When we are face-to-face, I let them choose groups, and online, I tend to group students based on what I can discern using weekly reflections about their personal challenges, struggles, and understanding about social justice in computer science.

In a face-to-face context, we groups might write this on the whiteboard or group-by-group report outs. This is followed by a co-construction of a list of issues that emerge and how they manifest in the various readings. I then tie up any specific points that have not yet been mentioned in any of the case studies.

Most students understand that the phrenology studies presented in the Calling Bull Case Studies are "bullshit," because there is no reasonable biological basis for criminality or sexuality. They can also point to the bias in the sourcing of the images in the criminality case study as a source of bias. Students also typically arrive at the idea that sexuality is not a binary, which presents an issue for classification model output. The same can be asked about whether criminality she be subject to binary classification as well. Students also clearly can understand the impact and potential legal issues surrounding the use of machine learning models and image classification to detecting criminality. However, they are often unprepared when I ask "Why would someone want to classify someone's sexuality based on their face?"

To help prompt this discussion - because, often, uncomfortable silence follows - I next ask about places in the country or world in which having a sexual identity other than heterosexual or acting on ones sexual identity is considered illegal. As of May 2021, there are 69 countries in which homosexuality has been criminalized (BBC News Reality Check Team, 2021). Fifteen states in the United States do not offer full protections against discrimination based on sexuality (Wikipedia, 2021). This allows students to consider the implications of such technology, which re-opens the door to the criminality discussion, as we ask if there are avenues of scientific inquiry that should not be pursued for the sake of scientific curiosity, particularly when the evidence for such pursuits is non-existent.

In the case of the research paper, students may also readily make the connection between what they know about the history of gender and racial discrimination and bias in the United States and the findings in the paper. Depending on the time allotment and students questions in the Perusall assignment, we might review one or two of the findings and associated visualizations together. Discussion of this paper sets us up quite well for the next module, which also focuses on gender and racial bias in training sets and models trained on those sets.

2.3. Through the lens of Joy Buolamwini

To launch the second phase of the class, I ask the following two questions:

- Where do you see examples of big data being used to benefit your life?

- Where do you see examples of data or algorithms that are biased?

This conversations adds to an overall uses and abuses conversation, but with an emphasis on personal impact. Depending on the student's axes of privilege or marginalization, they may have more positive or negative personal experiences with some of the automated decision-making models that run their life, from Netflix algorithms and Google ads to credit approval and policing. This portion of the class is intended to provide insight as to the continued invisibility and disregard for Blackness in technology as manifested by underrepresentation among faces in image training sets. My goal is for white students to develop both empathy and understanding related to this form of oppression and to identify ways to work towards greater justice in computing.

As we enter this difficult conversation space, I conduct a short grounding in and give permission for folx to move and breathe as needed, particularly folx of color. I introduce Joy Buolamwini as a graduate student and researcher at the Massachusetts Institute of Technology who founded the Algorithmic Justice League, and preface the series of three videos as an evolution of her scholarship with respect to algorithmic justice. In successive order and with small individual reflection breaks, I show the following three video clips:

1. Ted Talk: The Coded Gaze (Boulamwini, 2021b)
2. Gendershades (Boulamwini, 2021c)
3. Ain't I a woman? (Boulamwini, 2021a)

I also emphasize that I appreciate the rich infusion of visual poetry aside the computer science justice issue of the last piece, and mention that it is why Joy Buolamwini is called the "Poet of Code." After another short individual reflective break, I again ask students to discuss in groups the "take-home messages" for these videos.

Students never fail to mention the quote from the first video "Who codes matters, how we code matters and why we code matters(Boulamwini, 2021b)." Many semesters, I have had someone ask about fixing the particular algorithm or the particular training data set. In addition to acknowledging the fix for the specific problem at hand, I steer them to recognizing a systemic problem beyond a particular example. I have also had women, particularly Black women, so inspired by Joy Buolamwini that they have decided to pursue a flavor of computer science for their major.

3. Reflections and Conclusions

I want to reiterate that despite the discussion here of a "module," the entire course is structured in such a way such that this discussion about marginalization and computing and data science are not a one class event, but part of a broader quantitative critical inquiry arc in the class. The implementation of inclusive pedagogies such as the use of open sources software and educational resources, co-construction of knowledge, and attention to small group power dynamics are meant to cultivate a supportive community learning journey that can foster this discourse.

Most of the attention on the benefits of Open Educational Resources are on the redistributive properties of social justice - shifting financial burdens on students away to gain full participation in the course experience. Lambert 2018 points out two additional axes of social justice for the classroom, recognitive justice, the intentional inclusion of diverse voices and viewpoints, and representational justice, which allows marginalized people to speak for themselves. While Bergstrom and West are entertaining and easy for students to follow, their worldviews are that of white (cis)men. The series of Joy Buolamwini videos to help accomplish a course experience in which recognitive and representation justice is also present, where the Black woman is both computer scientist and social justice activist, as well as telling her story in her own words.

In preparation to facilitate such discussions in class, I point readers to scholars and scholarship in science and technology studies in addition to the resources above. Books such as Race after Technology (Benjamin, 2019), Algorithms of Oppression (Noble, 2018), and Weapons of Math Destruction (O'Neil, 2016) discuss these topics at a greater detail. I also recommend discussing class facilitation around topics relating to race, gender, and sexuality with colleagues who teach courses in the social sciences and/or one's office of equity and inclusion or intercultural education where staff are professional facilitators of such conversations on campus.

Students will not leave this course with the ability to perform machine learning techniques. However, I argue that creating the foundation to have critical discussions of such techniques will lead to deeper insight and better science when they reach courses which do teach these techniques. In addition, all corners of our data-driven economy is increasingly dependent on the results of such techniques. Therefore we have an obligation to be able to teach computational approaches such as machine learning to audiences beyond computer science and data science students and in our general education quantitative literacy courses. Likewise, these ethical discussions should not be relegated to a stand alone ethics course, but should be infused throughout the computer science curriculum at all levels.

References

BBC News Reality Check Team. Homosexuality: The coun-

tries where it is illegal to be gay, 2021. https://www.bbc.com/news/world-43822234 [Accessed: 6-23-21].

Benjamin, R. *Race after Technology: abolitionist tools for the new Jim code*. Polity, Medford, MA, 2019. ISBN 1509526404;9781509526390;9781509526406;1509526390;.

Bergstrom, C. and West, J. Calling Bullshit, 2021. https://callingbullshit.org/ [Accessed: 6-22-21].

Boulamwini, J. Ain't I a Woman?, 2021a. https://youtu.be/QxuyfWoVV98 [Accessed: 6-22-21].

Boulamwini, J. Ted talk: Coded gaze, 2021b. https://www.ted.com/talks/joy_buolamwini_how_i_m_fighting_bias_in_algorithms [Accessed: 6-22-21].

Boulamwini, J. Gendershades, 2021c. https://youtu.be/TWWsW1w-BVo [Accessed: 6-22-21].

Diaz Eaton, C., Highlander, H. C., Dahlquist, K. D., Ledder, G., LaMar, M., and Schugart, R. C. A "rule-of-five" framework for models and modeling to unify mathematicians and biologists and improve student learning. *PRIMUS*, 29(8):799–829, 2019.

Freire, P. and (Firm), P. *Pedagogy of the oppressed*. Continuum, New York, 30th anniversary edition, 2000. ISBN 150130531X;9781501305313;.

Garg, N., Schiebinger, L., Jurafsky, D., and Zou, J. Word embeddings quantify 100 years of gender and ethnic stereotypes. *Proceedings of the National Academy of Sciences*, 115(16):E3635–E3644, 2018. ISSN 0027-8424. doi: 10.1073/pnas.1720347115. URL https://www.pnas.org/content/115/16/E3635.

Google. Jamboard, 2021. https://jamboard.google.com/ [Accessed: 6-22-21].

Lambert, S. R. Changing our *dis* course: A distinctive social justice aligned definition of open education. *Journal of Learning for Development*, 5(3):225–244, 2018.

Noble, S. U. *Algorithms of oppression: How search engines reinforce racism*. NYU Press, 2018.

O'Neil, C. *Weapons of math destruction: How big data increases inequality and threatens democracy*. Broadway Books, 2016.

Perusall. Perusall, 2021. https://perusall.com/ [Accessed: 6-22-21].

Wikipedia. Lgbt rights in the United States: Summary of state protections, 2021. https://en.wikipedia.org/wiki/LGBT_rights_in_the_United_States#Summary_of_state_protections [Accessed: 6-23-21].