

DS-UA 202, Responsible Data Science, Spring 2024
Course Project: Technical Audit of an Automated Decision System
assigned on February 20, 2024; see description for due dates

Objectives

In this project, you will work in **teams of two** to conduct a technical audit of an automated decision system (ADS) of your choice. We suggest that you audit one of the systems developed in response to a Kaggle competition of your choice, but you should feel free to use other systems that are of interest to you. **Do not focus on Northpointe's COMPAS** in this assignment, since this tool was already covered extensively during class. Be sure to prominently cite your sources of code and data!

Both team members should work together on all parts of the project. You should not discuss your project submission or components of your solution with any students other than your project partner. If you have questions about this assignment, please send a private question to all instructors on Piazza.

Detailed description and goals

In this project, we encourage you to focus on examples from [Kaggle competitions](#), where the goals, the data, and one or several implementations are available for analysis. Select a Kaggle competition that has already finished, and for which you can find and successfully execute at least one solution. A list of solutions to Kaggle competitions is available [here](#), and you may be able to find solutions in other ways. If you decide to work with a system that's not from Kaggle, you should make sure that data and at least one implementation is available to you. **Once again: Be sure to prominently cite your sources of code and data!**

Background reading

Your report, and the corresponding Google Colab notebook(s), should be the result of our audit. We do not expect you to develop a UI or any other fancy data presentation methods. That said, it is important that the plots you produce are informative, and that they support your analysis.

This reading list should inspire you to think about interesting ways to analyze your ADS.

- “Closing the AI accountability gap: defining an end-to-end framework for internal algorithmic auditing”, Raji et al., ACM FAccT 2020, <https://dl.acm.org/doi/10.1145/3351095.3372873>
- “Towards Algorithm Auditing: A Survey on Managing Legal, Ethical and Technological Risks of AI, ML and Associated Algorithms”, Koshiyama et al., 2021 <https://www.ssrn.com/abstract=3778998>
- “Problematic Machine Behavior: A Systematic Literature Review of Algorithm Audits”, Bandy, ACM CHI / CSCW 2021 <https://dl.acm.org/doi/10.1145/3449148>

- “The algorithm audit: Scoring the algorithms that score us”, Brown et al., Big Data & Society 2021, <https://journals.sagepub.com/doi/10.1177/2053951720983865>
- “Resume Format, LinkedIn URLs and Other Unexpected Influences of AI Personality Prediction in Hiring: Results of an Audit”, Rhea et al., AIES 2022
<https://dl.acm.org/doi/10.1145/3514094.3534189>
- “Nutritional labels for data and models”, Stoyanovich and Howe, IEEE Data Engineering Bulletin Special Issue on Fairness, Diversity, and Transparency in Data Systems 42(3), 2019, <http://sites.computer.org/debull/A19sept/p13.pdf>
- “The imperative of interpretable machines”, Stoyanovich, Van Bavel, West, Nature Machine Intelligence 2, 2020, <https://rdcu.be/b57mr>
- “The dataset nutrition label: A Framework to drive higher data quality standards”, Holland et al., arXiv 2018, <https://arxiv.org/abs/1805.03677>
- “Datasheets for datasets”, Gebru et al., Communications of the ACM, 2021
<https://cacm.acm.org/magazines/2021/12/256932-datasheets-for-datasets/fulltext>
- “Model cards for model reporting”, Mitchell et al., ACM FAT* 2019
<https://dl.acm.org/doi/10.1145/3287560.3287596>

Deliverables and grading

The project is worth 30% of the course grade. Both partners will receive the same grade for the project. There are three deliverables, see below for description and due dates. **You may not use any late days towards the course project deliverables.**

1. **Project team formation, due at 11:59pm ET on Thursday, February 29.** Find a project partner and [fill out this form](#). We will assign a teaching assistant to shepherd your team, they will be your primary contact for any project-related questions.

If you have not identified a team partner by the deadline above, let us know by [filling out this form](#) by **11:59pm ET on Thursday, February 29**, so we can pair you up promptly.

2. **Project proposal, due at 11:59pm ET on Thursday, March 28.** Submit a 1-page summary of your proposed project, listing the names of both project partners and the ADS you propose to analyze in the project. Be explicit about where you’ll get the data and the code implementing the ADS: cite all sources properly in your project proposal. Leading up to the submission of your project proposal, you should make sure that the data is available, and that you are able to run the code on that data.

As part of your project proposal, include a brief (1-3 sentence) explanation of why you selected this specific ADS, in relation to the topics we study in this course. We are still early in the course, but we encourage you to look at the schedule / syllabus for a full list of topics when answering this question.

3. **Draft report, with Colab notebook, due at 11:59pm ET on Thursday, April 18.** Refer to the reading list above, and to the report structure. Submit a draft of your project report,

filling in the “Background” and “Input and Output” sections. Also develop a detailed plan for the other sections, and describe this plan in your draft. Submit a PDF of your draft, and a Colab notebook used for the computation.

4. **Final submission, due at 11:59pm ET on Thursday, May 9.** Submit your project report, implementation, and slides. You will be graded on your execution of the project (with a Colab notebook), and on the quality of the project report and presentation. You should submit a Colab notebook implementing your project, an accompanying written report in PDF format (up to 10 pages), and a PDF version of your slides.
5. **Project presentation, due at 11:59pm ET on Thursday, May 9.** Submit a 5 minute video of your presentation, in which you discuss your findings, and the slides you used in your presentation.

Submission instructions

Both students should submit all project deliverables on Bright Space.

- Submissions from the team partners should be identical with the exception of the brief **project contributions** document that each partner should submit individually as part of the final submission (due on May 9). In that document, each partner should discuss their own and their partner’s contributions to the project.
- Your project proposal, draft report, and final report should be submitted as PDF files, created using LaTeX (we suggest using Overleaf). You should submit a Google Colab notebook, or a collection of notebooks, that support the computation in your report. Finally, submit slides for your 3-minute presentation in PDF format.
- If you have trouble submitting your presentation video to Bright Space, upload it to [this folder](#) on NYU’s google drive.

Report structure

The outline below may be refined in response to clarification questions. We will post announcements on Bright Space if and when changes are made.

You may use any of the methods we discussed in class, as well as additional methods you find in the literature, for your analysis.

1. Background: general information about your chosen ADS
 - a. What is the purpose of this ADS? What are its stated goals?
 - b. If the ADS has multiple goals, explain any trade-offs that these goals may introduce.
2. Input and output
 - a. Describe the data used by this ADS. How was this data collected or selected?
 - b. For each input feature, describe its datatype, give information on missing values and on the value distribution. Show pairwise correlations between features if

appropriate. Run any other reasonable profiling of the input that you find interesting and appropriate.

- c. What is the output of the system (e.g., is it a class label, a score, a probability, or some other type of output), and how do we interpret it?
3. Implementation and validation: present your understanding of the code that implements the ADS. This code was implemented by others (e.g., as part of the Kaggle competition), not by you as part of this assignment. Your goal here is to demonstrate that you understand the implementation at a high level.
 - a. Describe data cleaning and any other pre-processing
 - b. Give high-level information about the implementation of the system
 - c. How was the ADS validated? How do we know that it meets its stated goal(s)?
4. Outcomes
 - a. Analyze the accuracy of the ADS by comparing its performance across different subpopulations, with respect to different accuracy metrics. Carefully justify your choice of accuracy metrics.
 - b. Analyze the fairness of the ADS, with respect to different fairness metrics. Carefully justify your choice of fairness metrics.
 - c. Develop additional methods to analyze ADS performance: think about stability, robustness, performance on difficult or otherwise important examples, or any other property that you believe is important to check for this ADS. Carefully justify your methodology.
5. Summary: reflect on the following points in your report.
 - a. Do you believe that the data was appropriate for this ADS?
 - b. Do you believe the implementation is robust, accurate, and fair? Discuss your choice of accuracy and fairness measures, and explain which stakeholders may find these measures appropriate.
 - c. Would you be comfortable deploying this ADS in the public sector, or in the industry? Why so or why not?
 - d. What improvements do you recommend to the data collection, processing, or analysis methodology?