

DS-UA 2020 & DS-GA 1017, Responsible Data Science, Spring 2021

Course Project: Nutritional Labels for Automated Decision Systems

Objectives

In this project, you will work in **teams of two** to build an interpretability tool for an automated decision system (ADS) of your choice, based on the concept of a nutritional label. We suggest that you build a nutritional label for 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 message to "All Instructors in this site" through NYU Classes.

Detailed description and goals

Automated Decision System (ADS) are in widespread use in government and industry, and a number of efforts are currently underway to regulate them. New York City recently passed a law ([Local Law 49 of 2018](#)) that compels the development of procedures and recommendations that City agencies should follow when explaining the operation of an ADS to the public, and demonstrating that an ADS does not discriminate against individuals based on membership in protected groups. In this project, you will help NYC and other municipalities by designing a **nutritional label** for a system of your choice. Your nutritional label will be implemented by a Google Colaboratory notebook, and it will be complemented by your project report.

Unfortunately, we currently have very few examples of ADS that are in use by government agencies. (See <https://automating.nyc/> for information about some such systems.) For this reason, 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 Colab notebook(s), are the "nutritional label" you are developing in this project. We do not expect you to develop a UI or any other fancy data

presentation method. 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.

- The Automated Decision Systems task force report:
<https://www1.nyc.gov/assets/adstaskforce/downloads/pdf/ADS-Report-11192019.pdf>
- Automating NYC: <https://automating.nyc/>
- “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., arXiv 2018 <https://arxiv.org/abs/1803.09010>
- “Model cards for model reporting”, Mitchell et al., ACM FAT* 2019
<https://dl.acm.org/doi/10.1145/3287560.3287596>

Deliverables, grading, and submission instructions

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 proposal, due at 5pm EST on ~~Wednesday, March 17~~ **Wednesday, March 24.**

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. 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 summary, please add a brief (1-3 sentence) explanation of why you selected this specific ADS, in relation to the topics we study in the responsible data science 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.

2. Draft report, with Colab notebook, due at 5pm EST on ~~Friday, April 9~~ **Wednesday, April 14.** Refer to the reading list below, 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.

3. Final submission, due at 5pm EST on Friday, May 7. 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. You will also be required to present your project in class, during the week of May 10.

4. **Project presentation** will be scheduled during the week of May 10. Each team will have 3 minutes to present their project and results, with timing strictly enforced. We will circulate instructions about uploading your slides directly into a Google slides presentation. You will be required to finalize and **upload your slides by 9am EST on Monday, May 10.**

All portions of this assignment should be turned in on NYU Classes. Both project partners should turn in identical submissions for all deliverables. Your project proposal, draft report, and final report should be submitted as PDF files, created using LaTeX (we suggest to use Overleaf). You should submit a 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.

Structure of your report

The outline below may be refined in response to clarification questions. We will post announcements on NYU Classes 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 effectiveness (accuracy) of the ADS by comparing its performance across different subpopulations.
- b. Select one or several fairness or diversity measures, justify your choice of these measures for the ADS in question, and quantify the fairness or diversity of this ADS.
- c. Develop additional methods for analyzing ADS performance: think about stability, robustness, performance on difficult or otherwise important examples (in the style of LIME), or any other property that you believe is important to check for this ADS.

5. Summary

- 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?