

DSCI 552: MACHINE LEARNING FOR DATA SCIENCE

PROBLEM SET 4

Instructors: Dr. Kristina Lerman (lerman@isi.edu) and Dr. Keith Burghardt (keithab@isi.edu)

Deadline: Thursday, March 25, 2021, 10 A.M PDT

You can submit the report on Blackboard and code on GitHub Classroom. As long as the problem set is open, you will be able to upload multiple answers (the last attempt will be graded).

TASK (20 points)

You are a new hire in a mid-size company. You have just completed your third task and handed in your report. Your boss is meeting with officials from Chirper Inc. You eagerly wait near the coffee machine opposite to the boss's office waiting to be called in. You were not wrong. After the lunch, you were approached by your technical manager. "Our boss wants to see us - it's urgent. Let's go!", he commanded. Following your manager, you enter the meeting room. The CEO and the senior developer already wait inside.

Your CEO said: "Great job! The director of the hospital and the infectious diseases researchers were impressed by the results your model generated and are planning to deploy it on their systems to identify patients resistant to the more viral strain of the virus.

The interpretable linear model you designed for predicting used car prices has brought in several calls for collaboration. Chirper Inc., the social media giant, intends to analyze happiness among constituents in these testing times. They want to send out support messages to constituents in areas generating chirps with low average happiness. They want our help to build a linear model that can **fairly** predict these happiness scores and identify such areas. Egregious human rights violations involving race in the past year have rocked our nation. Keep in mind that **fairness of the model** is critical and unlike several modern AI systems our predictions should **not be biased**. Chirper Inc. does not want to be audited by the Federal Government and fined for violating federal anti-discrimination statutes. Hence, **the predictions should not reveal sensitive info about ethnicity**

You have two weeks. I want a detailed report describing your main findings, on my desk, on Thursday, March 25, at 10 am."

*Hint: This task is related to fairness and bias. You have to design a **linear model that can fairly predict happiness and identify areas with average happiness < 5.8**. Happiness also known as "valence" has been measured for a select few counties in the nation. Our model will help Chirper in predicting valence and the impact of socio-demographic features on it. In this study, mean square error is not the primary focus.*

Instead, we must focus on building a fair model. Because the primary focus is the measure of fairness, we would like you to build three models – *race aware, race blind and fair*.

Your Technical Manager said: “I have shared the data file (*chirper-happiness.csv*) with you. It consists of:

- *Id2* which gives the Region Id,
- *totalGroup1* is the number of people in ethnic group 1,
- *totalGroup2* is the number of people in ethnic group 2,
- *percent_bachelorPlus* is the percentage of population with at least a bachelor’s degree
- *households_meanIncome* gives us the mean income per household in the region and
- *meanvalence* gives us the mean happiness in the region.

Correlation between the *predicted outcome* and *independent predictors* can be used as a proxy for measuring fairness. Our goal is to ensure that valence predictions are not biased towards one of the ethnic group variables. We define *protected variables* as ones along which biases can be observed in the outcome. In our case, these are – *totalGroup1 and (or) totalGroup2*. *Your analysis report should consist of the following four models.”*

output (label)

1. **Ground Truth Model** – Generate correlations and scatterplots of *ground truth valence* with respect to *percent_bachelorPlus* and *households_meanIncome* with each of the scatterplots conditioned on *totalGroup1* and *totalGroup2*. *From the scatterplots state whether you see a bias to one ethnic group? Describe why you think so.*

*The prediction models are considered **fair** if the correlations and scatter plot of **predicted outcome** vs **independent variables** conditioned on **ethnic group variables** are similar to that of ground truth model.*

compare the ground truth model

2. **Ethnic Group Aware Model** (Prediction Model I) – Build a linear regression model using all predictors, on a percentage of the data provided. Predict the valence on the entire dataset. Analyze fairness of predicted valence by comparing the correlations and scatterplots of Group Aware Model with **Ground Truth Model**. Report your observations on the comparison and *state if there is bias to one of the two groups when the predicted valence threshold is set to 5.8.*
3. **Ethnic Group Blind Model** (Prediction Model II) – Build a linear regression model without using *totalGroup1* and *totalGroup2* as predictors, on a percentage of the data provided. Predict the valence on the entire dataset. Analyze fairness of predicted valence by comparing the correlations and scatterplots of Group Blind Model with Ground Truth

Model. Report your observations on the comparison. Does hiding these protected variables change the bias identified in the previous model change here? Describe either way.

4. **Fair Model** – Our R&D team has published a paper on debiasing linear models (available at: <https://arxiv.org/pdf/1910.12854.pdf>). On the basis of this work, we have written two helper functions that you have to use in order to debias the outcome variable. The python file consisting of these helper functions are available in *fair_var.py*.

Hint: If you run into doubts refer the Fairness lecture and slides from Week 7. Additionally, you can attend office hours or post questions on Piazza.

The Senior Developer took you aside and said: “My task will be to maintain your code. However, remember that I'm not a Data Scientist like you - so you have to be very careful when you are writing your code. Write comments and try to explain any nontrivial section.”

Data

The dataset to be used is *chirper-happiness.csv* and helper functions for debiasing the outcome variable are in *fair_var.py*.

Report

To help you, I prepared a template. See <https://www.overleaf.com/read/vnvhqxkpdhbk>. You are encouraged to use the template, but you are free to use other editors or make modifications. Just ensure that the final submission of the report has to be in PDF. Submission of report has to be done on Blackboard. As always submit your code on GitHub classroom and report on Blackboard.

Code Submission

We have created a GitHub Classroom where you can create private repositories. We will update the class on Piazza on how to go about uploading your solutions on this platform.

Grading Rules

In order to grade your work, we will role-play the following situation. We will assume, that you are a new-hire in our company. You are asked to provide a comprehensive technical report that illustrates your findings. We will evaluate it from the perspective of three people.

- Your CEO (she would like to hear high level stuff. She will probably only read the conclusions and look at main figure). (4 points for report).

- Your manager (he would like to see a detailed report; he might also look at some parts of the code). (6 points for report and 2 points for code).
- A senior developer (they would like to see the code and won't read the report at all). (8 points for code).

Your final score is: 10 points for report and 10 points for code.

Don't Panic

Don't panic. We understand that this is a large, open ended task. We also understand that this might be the very first technical report that you were asked to write. We are dedicated to help you do your best work all while keeping the standards high. We acknowledge that you have limited time and resources to complete the task. This report doesn't have to be perfect for 100% score. If something is not clear, ask your questions on Piazza or reach out to us during office hours.

Note: Cite any source you use (even if you adopt/copy a snippet of code). Failure to do so would amount to plagiarism.