

richardhoehn / mtsu.data.6550.team

🔍

🏠

📧

👤

<> Code

🔔 Issues 3

🔗 Pull requests

🔄 Actions

📁 Projects

📖 Wiki

🛡 Security

📈 Insights

⚙ Settings


Edit

New issue

📄

Module 2: Data Bias Project - Group Results (Disussion) #4

🔗 Open



richardhoehn

opened last week · edited by richardhoehn


Edits ▾ ⋮

Hi All,

Starting our discussion for this weeks **Data Bias Project**. I hope to be looking / googling for some ideas later on this evening or tomorrow during the day.

Create sub-issue ▾

😊



richardhoehn

last week · edited by richardhoehn

Edits ▾ Owner Author ⋮

A quick Chat GPT search yield the following that we might want to consider reviewing in more detail with the following prompt *Give me some datasets with known bias that I can pull from Kaggle*:

1. COMPAS Recidivism Dataset

- **Description:** Contains criminal recidivism risk scores used in the U.S. justice system. Studies show racial biases in risk score assignments.
- **Bias:** Racial bias in recidivism prediction.
- **Link:** [COMPAS Recidivism Dataset](#)

12. UCI Adult Income Dataset

- **Description:** Predicts whether an individual's income exceeds \$50K based on features like education, occupation, and race.
- **Bias:** Gender and racial bias in income predictions.
- **Link:** [Adult Census Income Dataset](#)

3. LendingClub Loan Dataset

- **Description:** Contains loan application and repayment information from LendingClub.
- **Bias:** Disparities in loan approvals based on race, gender, and income level.
- **Link:** [LendingClub Loan Dataset](#)

4. Gender Bias in IMDb Reviews

- **Description:** Includes IMDb reviews and ratings, often analyzed for gender disparities in movie ratings.
- **Bias:** Gender bias in film ratings.
- **Link:** [IMDb Dataset](#)

5. Healthcare Cost and Utilization Project (HCUP) Dataset

- **Description:** Hospitalization data used for analyzing racial and socioeconomic disparities in healthcare access.
- **Bias:** Healthcare access bias by race and socioeconomic status.
- **Link:** [HCUP Dataset](#)

6. Face Recognition Datasets (CelebA, UTKFace)

- **Description:** Used for facial recognition tasks, known for biases against darker-skinned individuals.
- **Bias:** Racial and gender bias in facial recognition models.
- **Links:**
 - [CelebA Dataset](#)
 - [UTKFace Dataset](#)

7. Amazon Hiring Dataset

- **Description:** Contains hiring decisions from Amazon's past recruitment processes, which showed biases against women in technical roles.
- **Bias:** Gender bias in hiring.
- **Link:** [Amazon Hiring Dataset](#)

These all came from Chat GPT - as a starting point for us to look at.



Kwoods132 last week

Collaborator ...

How do you all feel about using the UCI Adult Income dataset for our project? It seems like it would be a good one to evaluate for the biases the AI outlined and to do a report on. The dataset also seems a bit easier to work with than some of the other ones.



Hroge19007 last week

Collaborator ...

The UCI Adult Income dataset would be good for our project, but I fear it's not large enough for everyone to write a piece of code. It's just 15 columns.



richardhoehn last week

Owner Author ...

An option for all of us to write code would be creating visualizations? Those (visuals) usually take a lot of time but can be quite insightful.

Just an option - I'm open to other ideas.



IMOsbo last week

Collaborator ...

Yeah, I think the UCI dataset would work well (great find [@Kwoods132](#)!). Looking at the Kaggle description, it looks like there's some pretty obvious bias already (67% men, 33% women)

I think there should be enough to write code on? We could divide it up where each person focuses on one main variable and then shows the interaction of that variable across the dataset - that way we're all doing our own analysis and visualization? We'll probably find similarities and then we can bring those together in our paper conclusion.



Kwoods132 last week

Collaborator ...

I think there's enough to write code on as well. [@IMOsbo](#) has a good idea about splitting up this dataset so we all have some code we can write. As well, I imagine there are several different types of analysis and visualizations we can do that would work well for the given dataset.



richardhoehn last week

Owner Author ...

Hi All,

I got few of our needs done:

- Got the repository started under **Module-02**
- I pulled the dataset from Kaggle and placed it in the folder called: `/data`
- I created a super simple Jupyter notebook that loads the data into a dataframe called `data`

It might be good if you guys make a **copy** and rename with your first name?



richardhoehn last week

Owner Author ...

Side Note:

Please don't hesitate to reach out to me if you have any questions or issues getting GIT going on your PC.

I am glad to help.

IMOsbo last week

Collaborator

...

Here's the link to our Word doc: https://mtmailmtsu-my.sharepoint.com/w:/g/personal/imo2d_mtmail_mtsu_edu/Edq8QUgVyQpCoCwHB5cl4PsB_IP6LmP_N6EACSQ4dZIFlw?e=b6N3Ml

IMOsbo 5 days ago · edited by IMOsbo

Edits

Collaborator

...

I was checking the guidelines for the Group Project and we have to create a separate CitiHub repo for this project... I'll create it and invite y'all. (Never mind, we're just using this repo)

3. One group member will create a GitHub repository. The repository should be named DATA6550-Bias. You will like the URL to the D2L Dropbox. Again – only one submission per group, please. The file structure MUST following the directory structure below:

databias_report.docx – Word document with tracking enabled to identify contributions. Create this using one drive – share it with all team members and the professor.

readme.md – Overview of the repository.

Code/

- Lastname1/
- Lastname2/
- Lastname3/
- Lastname4/

Data/

- Dataset file or files

Collaboration/

- WeekA.docx – AI-generated summary of the first week's discussions.
- WeekB.docx – AI-generated summary of the second week's discussions.

Analysis/

(Optional) Place graphics and intermediate results here.

richardhoehn 5 days ago via email

Owner

Author

...

I called Dr. Wallin to ask if we really need to do that; it might be easier for us to keep all our discussions and all our data in one place rather than making a new repos each time. Richard Hoehn(615) 319-2171Sig: iPhoneOn Feb 7, 2025, at 17:35, IMOsbo ***@****> wrote:

I was checking the guidelines for the Group Project and we have to create a separate GitHub repo for this project... I'll create it and invite y'all.

image.png (view on web)

—Reply to this email directly, view it on GitHub, or unsubscribe.You are receiving this because you authored the thread.Message ID: ***@****>

IMOsbo 5 days ago

Collaborator

...

Oh perfect - if he's cool with it, then I think that would be much easier!

richardhoehn 5 days ago via email

Owner

Author

...

Well guys. Dr Wallin wants a new GitHub repo for each project. Sorry I tried to convince him but he would like us to create a new one. Let's use the one that was created. Thanks. Richard Hoehn(615) 319-2171Sig: iPhoneOn Feb 7, 2025, at 17:43, IMOsbo ***@****> wrote:

Oh perfect - if he's cool with it, then I think that would be much easier!


—Reply to this email directly, view it on GitHub, or unsubscribe.You are receiving this because you authored the thread.Message ID: ***@****>

Kwoods132 3 hours ago

Collaborator

...

Where exactly are we supposed to put the AI summary for our group discussions?





IMOsbo 2 hours ago

Collaborator

...

I think there's supposed to be a folder called "Collaboration" in the repo - I think they all go in there...







HrogeI9007 2 hours ago

Collaborator

...

I thought our group discussion was going to be in the repository that IMOsbo created. Did I put my notebook in the wrong place?







IMOsbo 2 hours ago

Collaborator

...

No, I think you're good [@HrogeI9007](#) - that's where everybody else's code is :)







Kwoods132 18 minutes ago

Collaborator

...

I think the Word doc is done unless there is something else that we should add.





richardhoehn 13 minutes ago


Owner


Author

...

I like it. I will take care of submitting if ok.

[@Kwoods132](#) Please keep in mind that we move our code to the new repository and also have a thread discussion going on there. This is the repo that Isaiah setup. Link: <https://github.com/IMOsbo/DATA6550-Bias>





richardhoehn 10 minutes ago


Owner


Author

...

All - I have my zoom link open if you guys want to chat about were we are. Feel free to jump on so we can talk thorough any late stage updates

<https://mtsu.zoom.us/j/84343780740>






Add a comment

H B I | ≡ <> 🔗 | ≡ ≡ ≡ | @ 📎 ↶ 📎

Write Preview

Use Markdown to format your comment

 Paste, drop, or click to add files

✓ Close issue

▼

Comment

Metadata

Assignees

No one - [Assign yourself](#)

<https://github.com/richardhoehn/mtsu.data.6550.team/issues/4>

4/5

Labels

No labels

Projects

No projects

Milestone

No milestone

Relationships

None yet

Development

[Create a branch](#) for this issue or link a pull request.

Notifications

Customize

Unsubscribe

You're receiving notifications because you're subscribed to this thread.

Participants



- Transfer issue
- 🔒 Lock conversation
- 📌 Pin issue
- 🗑 Delete issue