AI, Ethics, and Society

Final Project

Spring 2022

Contents

[Step 0: Data preprocessing 2](#_Toc102153484)

[Step 1: 4](#_Toc102153485)

[Step 2 5](#_Toc102153486)

[Step 3 6](#_Toc102153487)

[Step 4 7](#_Toc102153488)

[Step 5 8](#_Toc102153489)

## Step 0: Data preprocessing

1. Data: <https://www.kaggle.com/datasets/takumafujiwara/2020-census-data>
2. Rows with null/empty/? characters are removed
3. The data has one dependent variable ‘income-90k-threshold’. Additionally, ‘coded-income’ column provides 41 different codes.

A picture containing calendar

Description automatically generated

1. We further process these 41 codes in 6 tax backets provided by IRS. <https://www.irs.gov/newsroom/irs-provides-tax-inflation-adjustments-for-tax-year-2021>

Tax brackets per IRS in 2021.

A computer screen capture

Description automatically generated with low confidence

Our formula:

|  |  |  |
| --- | --- | --- |
| coded-income range | Tax bracket | Comment |
| 0 | 0 | No income |
| 1-4 | 1 |  |
| 5-17 | 2 |  |
| 18-35 | 3 |  |
| 36-41 | 4 |  |

With above formula, we device additional column, tax-bracket and add to our data-source.

Final data: Based on this discussion (<https://piazza.com/class/kwb93smsa9n1hw?cid=426_f32>) we do not need to submit find copy of data. If required data can be obtained from here: <https://github.com/vikbehal/Explore/blob/master/GT-AIES/census_2020-v33.csv>

Screenshot of find data with tax-backet.

A screenshot of a computer

Description automatically generated with medium confidence

## Step 1:

1. Which dataset did you select?
   1. Selected 2020 Census Data from Kaggle. This data is collected by 'University of Texas' students in 2020.
   2. Link: <https://www.kaggle.com/datasets/takumafujiwara/2020-census-data>
2. Which regulated domain does your dataset belong to?
3. How many observations are in the dataset?
   1. 36252
4. How many variables in the dataset?
   1. 15. (14 excluding target variable ‘income-90k-threshold’)
5. Which variables did you select as your dependent variables?
   1. income-90k-threshold.
   2. tax-backet
6. How many and which variables in the dataset are associated with a legally recognized protected class? Which legal precedence/law (as discussed in the lectures) does each protected class fall under?
   1. 4 variables.

|  |  |  |
| --- | --- | --- |
| Variable | Protected class | Legal precedence/law |
| age | Age | Age Discrimination in Employment Act of 1967(over 40) |
| race | Race | Civil Rights Act of 1964, 1991 |
| gender | Sex | Equal Pay Act of 1963; Civil Rights Act of 1964, 1991 |
| native-country | National Origin | Civil Rights Act of 1964, 1991 |

## Step 2

1. Table documenting the relationship between members and membership categories for each protected class variable (from Step 2.1)

**Age:**

|  |  |
| --- | --- |
| Age range | Group name |
| 17 until 29 | (16, 29] |
| 30 until 43 | (29, 43] |
| 44 until 57 | (43, 57] |
| 58 until 71 | (57, 71] |
| 72 until 85 | (71, 85] |

**Race**:

Text

Description automatically generated

**Gender**: Female & Male

**native-country**: Each of the 156 countries as-is. The list is printed in Jupyter notebook.

We also created a new column per **(**[https://piazza.com/class/kwb93smsa9n1hw?cid=426\_f34](/class/kwb93smsa9n1hw?cid=426_f34)) for all questions to limit the values for native-country.

|  |  |
| --- | --- |
| native-country | Group name |
| United-States | United-States |
| All other 155 countries | Others |

1. Table documenting the relationship between values and discrete categories/numerical values associated with your dependent variables (from Step 2.2)

**income-90k-threshold:**

|  |  |
| --- | --- |
| Income | Group name |
| Income <= $90,000 | <=90k |
| Income > $90,000 | >90k |

Frequencies:

Text

Description automatically generated

**tax-backet:** This column was devised during data pre-processing, so no additional categorization was required. Below table is same as shows in data pre-processing step (step 0).

|  |  |  |
| --- | --- | --- |
| coded-income range | **tax-bracket** | Comment |
| 0 | 0 | No income |
| 1-4 | 1 |  |
| 5-17 | 2 |  |
| 18-35 | 3 |  |
| 36-41 | 4 |  |

Frequencies**:**

Text, table

Description automatically generated

1. Table providing the computed frequency values for the membership categories each protected class variable (from Step 2.3)

**Age:**

Text

Description automatically generated

**Race**:

Text

Description automatically generated with medium confidence

**Gender**:

Text

Description automatically generated with medium confidence

**native-country**: Each of the 156 countries as-is. The list is printed in Jupyter notebook.

For revised column as per step 1 earlier, frequency is below.

Text

Description automatically generated

1. Histograms derived from Step 2.4
2. Histogram for each of the protected class variable that graphs the frequency values of its membership categories as a function of the dependent variable ‘income-90k-threshold’

Chart, bar chart, waterfall chart

Description automatically generated

1. Histogram for each of the protected class variable that graphs the frequency values of its membership categories as a function of the dependent variable ‘tax-bracket’

Chart, bar chart

Description automatically generated

## Step 3

1. Provide the resulting code (can be as an additional .ipynb file if submitting a PDF)
2. Provide a table documenting the protected class variable selected, the privileged/unprivileged groups/values, the pre-processing bias mitigation function selected, and the fairness metrics/resulting values computed in Step 3.2 and Step 3.4

Protected class variables:

|  |  |  |
| --- | --- | --- |
| Variable | Privileged group | Unprivileged group |
| age | (43,57] | (16,29] |
| race | White | Black |
| gender | Male | Female |
| native-country | United-States | Others |

## Step 4

Option A

1. Provide the resulting code (can be as an additional .ipynb file if submitting a PDF)
2. Document 1) the privileged/unprivileged groups, 2) the dependent variable, 3) the quantitative results from applying the two fairness metrics on the classifier output associated with the original and transformed dataset, 4) a table documenting whether there was positive, negative, or no change in each of the fairness metrics after transforming the dataset, after training the classifier on the original dataset, and after training the classifier on the transformed dataset.

Option B

1. Provide the resulting code (can be as an additional .ipynb file if submitting a PDF)
2. Document 1) the privileged/unprivileged groups, 2) the dependent variable, 3) the quantitative results from applying the two fairness metrics associated with the original testing dataset and transformed testing dataset after bias mitigation, 4) a table documenting whether there was positive, negative, or no change in each of the fairness metrics on the transformed dataset from Step 3.4, the original testing dataset (Step 4.5), and on the transformed dataset after applying your bias mitigation algorithm (Step 4.5)

## Step 5

* 1. List the members of your project team

|  |  |
| --- | --- |
| Name | GTID |
| Vikrant | V33 |
| Anuradha | TBD |
| Sudeshna | TBD |

* 1. Graph the results from applying the two fairness metrics on your privileged/unprivileged groups as derived from Step 3.2, 3.4, and 4.5
  2. Explain which fairness metric (if any) is best and provide a justification for your answer
  3. Each team member must provide a separate answer to the following questions in no-more than a one-paragraph response (this is to be included in the submitted group report, with a reference to the student author). Note: If a group member fails to provide a response, the team is free to indicate that in the final report submission with No Response and a reference to the student author.
     1. § Did any of these approaches seems to work to mitigate bias (or increase fairness)? Explain your reasoning. Did any group receive a positive advantage? Was any group disadvantaged by these approaches? What issues would arise if you used these methods to mitigate bias?
     2. (v33) Vikrant’s response

* + 1. (TBD) Anuradha’s response
    2. (TBD) Sudeshna’s response