**AI, Ethics, and Society**

**Homework Project #5**

**Xiaoka Xiang**

**Xxiang31**

**903000686**

Step 2:

I chose Taiwan Credit Dataset.

There are 30000 observations and 24 attributes.

3 Protected classes, they are:

1. X2- Sex - (Equal Pay Act of 1963; Civil Rights Act of 1964, 1991)
2. X5 – age – (Age Discrimination in Employment Act of 1967) (over 40))
3. X4 – Marital Status - (Civil Rights Act of 1968)

Step 3:

* Outcome variable(s) that relates to the creditworthiness of a customer include:

1. Y – Default payment
2. X6 - X11: History of past payment. X6 = the repayment status in September, 2005; X7 = the repayment status in August, 2005; . . .;X11 = the repayment status in April, 2005. The measurement scale for the repayment status is: -1 = pay duly; 1 = payment delay for one month; 2 = payment delay for two months; . . .; 8 = payment delay for eight months; 9 = payment delay for nine months and above.

I chose default payment as variable of credit worthiness, and the max delay month for the 6 months.

If max is less 0 then score is 100, max = 0 score is 90, max = 1 score is 80, … max = 8 score is 10.

I chose to study age, and have older (age >= 40) as the unprivileged group and Young (age < 40) as the privileged group. Let older be 0, younger be 1.

Random split with seed = 1234, we have:

Training:

|  |  |
| --- | --- |
| age group | Count |
| 0 | 4578 |
| 1 | 10422 |

Testing:

|  |  |
| --- | --- |
| age group | Count |
| 0 | 4566 |
| 1 | 10434 |

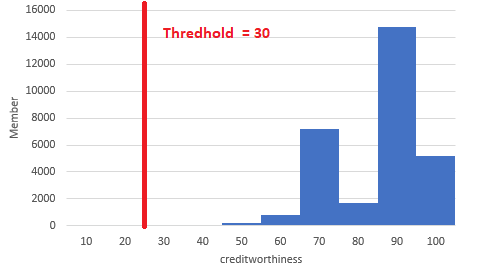
Outcome variable: Creditworthiness derived from History of past late payments and Y

Formula used to score members creditworthiness from 0 to 100 is [If max is less 0 then score is 100, max = 0 score is 90, max = 1 score is 80, … max = 8 score is 10.]

Protected Class Attribute: Age. Privileged group: Young (age < 40); Number of Members in Training Set: 10422; Number of Members in Testing Set: 10434 Unprivileged group: Older (age >=40); Number of Members in Training Set: 4578; Number of Member in Testing Set: 4566

Step 4:

My threshold turns out to be 30 and max profit is 200,380.



|  |  |  |
| --- | --- | --- |
|  | Privileged | Unprivileged |
| Favorable | 20798 | 9110 |
| Unfavorable | 58 | 34 |

Step 5: Default threshold is 30.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Total | Favorable | Rate | Disparate Impact |
| Privileged | 4578 | 4563 | 0.996723 | 1.000399322 |
| UnPrivileged | 10422 | 10392 | 0.997121 |  |
|  |  |  |  |  |
| Privileged |  |  |  |  |
| Predict\Actual | ND | D | TPR | Equal Opportunity Difference |
| Favorable | 3522 | 1041 | 0.999432 | 7.5653E-05 |
| Unfavorable | 2 | 13 |  |  |
| Unprivileged |  |  |  |  |
| Predict\Actual | ND | D | TPR |  |
| Favorable | 8128 | 2266 | 0.999508 |  |
| Unfavorable | 4 | 25 |  |  |

There is very minimal bias between groups, Disparate Impact is extremely close to 1, and Equal Opportunity Difference is extremely close to 0, it is a fair model.

Step 6/Step 7:

My formula does not have bias, however it also has low accuracy due to high false positive…

I am not sure what other people choose as their indicator of Creditworthiness, to me summing up the number of month does not seem to be a good idea as the max length of late payment shall be a good indicator. Maybe should have included # of bill statement or amount of previous payment.