Project 2: Hidden Bias in COMPAS

What is COMPAS?

1. Across the nation, judges, probation and parole officers are increasingly using algorithms to assess a criminal defendant's likelihood of becoming a recidivist – a term used to describe criminals who re-offend.
2. Northpointe Inc. developed COMPAS (which stands for Correctional Offender Management Profiling for Alternative Sanctions)
3. When most defendants are booked in jail, they respond to a COMPAS questionnaire. Their answers are fed into the COMPAS software to generate several scores including predictions of "Risk of Recidivism" and "Risk of Violent Recidivism."

What is wrong with COMPAS?

Its findings concluded that the popular risk assessment tool COMPAS discriminates against Blacks because its algorithm produces a much higher false positive rate for Blacks than Whites, meaning that it overpredicts high risk of reoffending for Blacks. COMPAS overpredicts the risk for women to reoffend, therefore leading to unfair sentencing of female offenders.

COMPAS Dataset Description

A data frame with 6172 rows and 9 variables:

Variable	Description
Two_yr_Recidivism	factor, yes/no for recidivism or no recidivism. This is the outcome or target in this dataset.
Number_of_Priors	numeric, number of priors, normalized to mean = 0 and standard deviation = 1
Age_Above_FourtyFive	factor, yes/no for age above 45 years or not
Age_Below_TwentyFive	factor, yes/no for age below 25 years or not
Female	factor, female/male for gender
Misdemeanor	factor, yes/no for having recorded misdemeanor(s) or not

Variable	Description
ethnicity	factor, Caucasian, African American, Asian, Hispanic, Native American or Other
probability	numeric, predicted probabilities for re-cidivism, ranges from 0 to 1
predicted	numeric, predicted values for recidi-vism, 0/1 for no/yes

A Glimpse at COMPAS Data

```
> head(compas)
   Two_yr_Recidivism Number_of_Priors Age_Above_FourtyFive Age_Below_TwentyFive
4                 no       -0.6843578                   no                   no
5                yes        2.2668817                   no                   no
7                 no       -0.6843578                   no                   no
11                no       -0.6843578                   no                   no
14                no       -0.6843578                   no                   no
24                no       -0.6843578                   no                   no
   Female Misdemeanor         ethnicity probability predicted
4    Male          yes            Other  0.3151557         0
5    Male           no        Caucasian  0.8854616         1
7  Female          yes        Caucasian  0.2552680         0
11   Male           no African_American  0.4173908         0
14   Male          yes         Hispanic  0.3200982         0
24   Male          yes            Other  0.3151557         0
```

Tasks to Complete

In this project, students are given the outcomes (predicted recidivism probabilities) of 2 prediction models: (1) the model using Ethnicity as a predicting factor, provided by COMPAS algorithm and (2) the model developed by the instructor without using Ethnicity as a predicting factor. Students need to complete the following tasks:

1. For both models, using 0.5 as the threshold, classify the recidivism probabilities into 2 categories: high risk (code 1) and low risk (code 0).
2. Summarize the accuracy of Model (1) by completing the confusion matrix (count the number of false positive, false negative, true positive, true negative cases).
3. Summarize the accuracy of Model (2) by completing the confusion matrix (count the number of false positive, false negative, true positive, true negative cases).
4. Calculate the accuracy rate and false positive rate in Model (1).
5. Calculate the accuracy rate and false positive rate in Model (2).

Assessment

The learning outcome of this assignment aligns with the Evaluation of Different Ethical Perspectives/Concepts rubric of AACU (aacu.org) and for that reason, this assignment will be evaluated on the Evaluation of Different Ethical Perspectives/Concepts rubric.