

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

Variable	Description
<i>ethnicity</i>	factor, Caucasian, African American, Asian, Hispanic, Native American or Other
<i>probability</i>	numeric, predicted probabilities for recidivism, ranges from 0 to 1
<i>predicted</i>	numeric, predicted values for recidivism, 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
```

AI Excel Project

Submitting: a file upload

Points: 70

INSTRUCTIONS:

- Work directly on the attached Excel file [here](#).
- When finished, save the file as "MAT2326-AIExcelProject-FirstnameLastname.xlsx".

Correctional Offender Management Profiling for Alternative Sanctions (COMPAS) is a tool used by US courts to assess the likelihood of a defendant reoffending. COMPAS was developed by Northpointe in the 1990s and is used by jurisdictions such as New York, Wisconsin, and California. Some have raised concerns that COMPAS exhibits bias against Black defendants. A 2016 study by ProPublica found that Black

This material was developed by Dr. Bao Anh T. Maddux, as a part of Winston-Salem State University's Center for Applied Data Science (CADS) Faculty Adopter Program 2023-2024

defendants were more likely to be incorrectly judged at a higher risk of recidivism. Researchoutreach.org also concluded that COMPAS discriminates against Black defendants because its algorithm produces a much higher false positive rate for Black defendants than White defendants.

The data set stored in COMPASData.xlsx gives us a glimpse into a portion of COMPAS data.

1. Number of Priors: The number prior criminal records of defendants
2. Age Above 45: Binary variable indicating whether a defendant is above 45 years old. (1 if the defendant is more than 45 years old, 0 otherwise)
3. Age Below 25: Binary variable indicating whether a defendant is below 25 years old. (1 if the defendant is less than 25 years old, 0 otherwise)
4. Female: Binary variable indicating whether a defendant is a female. (1 if the defendant is a female, 0 otherwise)
5. Misdemeanor: Binary variable indicating the charge degree of defendants. (1 if it's a misdemeanor, 0 if it's a felony)
6. Ethnicity: Categorical variable for a defendant's ethnicity.
7. Predicted Probability (with Ethnicity): the predicted value for the probability that a defendant would commit a further offend within 2 years after their release, taking into account their ethnicity in the prediction algorithm.
8. Predicted Probability (without Ethnicity): the predicted value for the probability that a defendant would commit a further offend within 2 years after their release, disregarding their ethnicity in the prediction algorithm.
9. Actual Two-Year Recidivism: Binary variable indicating whether a defendant actually reoffended within the span of 2 years. (1 if they did and 0 if they did not)

Use Excel to compute the following statistics for the data.

1. (5pts) Use IF function in Excel to categorize the Predicted Probability (with Ethnicity). If the predicted probability (with Ethnicity) for a defendant is greater than 0.5, then label them 1 (i.e. the defendant would reoffend within the span of 2 years after their release), otherwise label them 0. Place the labels in column H.
2. (5pts) Use IF function in Excel to categorize the Predicted Probability (without Ethnicity). If the predicted probability (without Ethnicity) for a defendant is greater than 0.5, then label them 1 (i.e. the defendant would reoffend within the span of 2 years after their release), otherwise label them 0. Place the labels in column J.
3. (5pts) In the prediction model using Ethnicity as a predicting factor, count the number of defendants that were predicted not to reoffend and actually did not reoffend, using COUNTIFS function in Excel. Place your answer in cell O3.

4. (5pts) In the prediction model using Ethnicity as a predicting factor, count the number of defendants that were predicted not to reoffend and actually reoffended, using COUNTIFS function in Excel. Place your answer in cell P3.
5. (5pts) In the prediction model using Ethnicity as a predicting factor, count the number of defendants that were predicted to reoffend and actually did not reoffend, using COUNTIFS function in Excel. Place your answer in cell O4.
6. (5pts) In the prediction model using Ethnicity as a predicting factor, count the number of defendants that were predicted to reoffend and actually reoffended, using COUNTIFS function in Excel. Place your answer in cell P4.
7. (5pts) Calculate the accuracy rate for the prediction model using Ethnicity as a predicting factor. Place your answer in cell Q2.
8. (5pts) Calculate the false positive rate for the prediction model using Ethnicity as a predicting factor. Place your answer in cell R2.
9. (20pts) Complete questions 3 through 6 for the prediction model not using Ethnicity as a predicting factor, using COUNTIFS function in Excel. Place your answers in cell O7, P7, O8 and P8 respectively.
10. (5pts) Calculate the accuracy rate for the prediction model not using Ethnicity as a predicting factor. Place your answer in cell Q6.
11. (5pts) Calculate the false positive rate for the prediction model not using Ethnicity as a predicting factor. Place your answer in cell R6.
12. Save the file as "MAT2326-ExcelProject2-FirstnameLastname.xlsx" then submit

Assessment

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