Ashlyn Bernard, Edgar Campos

MA 346: Midterm Project

Recidivism: Can a single number explain if a person will become a reoffender?

Recidivism, the tendency for a criminal to become a reoffender, represents an important challenge to the criminal justice system. It not only poses a threat to public safety but can also strain public resources. Predicting recidivism has become a pivotal area of study with policymakers and law enforcement, as it can have huge implications for rehabilitation efforts. One of the most popular recidivism risk assessment systems is the Correctional Offender Management Profiling for Alternative Sanctions tool (COMPAS). This tool has been developed by Northpointe and is popular with U.S. courts, which is the focus of the study.

To understand the accuracy of COMPAS, we pulled two datasets from the Broward County Sheriff's Office in Florida. The first dataset shows 28 variables on over 60,000 defendants including demographic characteristics and predictions based on the COMPAS tool. Our variables of interest are decile score, age, race, marital status, and their risk of recidivism. Decile score is a measurement of how likely a defendant is, on a scale from 1 to 10 on getting reincarcerated. The second dataset is based on a 2-year follow-up study, noting whether the defendants were recidivated.

Conducting our exploratory data analysis, as seen in Figure 1, it appears that decile score is associated with a higher rate of recidivism, however with lower decile scores the trend appears to disappear. Looking at the age of recidivists in the dataset, generally recidivists tend to be younger than non-recidivists. With marital status and race, there appears to be little correlation between the factors and recidivism.

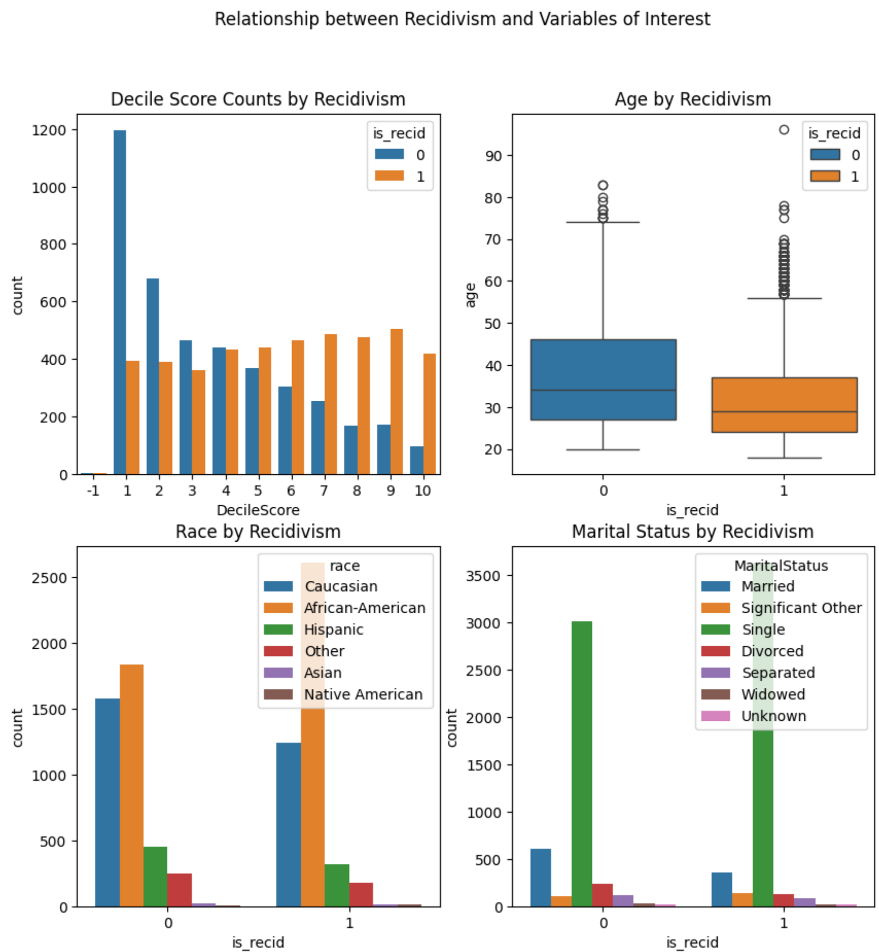

Figure 1: Relationships between response variable, recidivism, and predictor variables

To evaluate the accuracy of COMPAS predictions, we constructed two classification models using logistic regression to gauge the accuracies before and after isolating decile score. The models themselves predict the probability a defendant was recidivated. The first model uses decile score, race, marital status, and age as the independent variables. The second model uses only decile score as the independent variable. We evaluate the accuracy of the model using values from a confusion matrix including true positives, true negatives, false positives, and false negatives (also known as Type I and Type II errors). This comparison shows whether demographic variables make a difference in the model's accuracy, and what effect, if any, the variables of interest have on the probability of recidivism.

After running these two models we found that demographic variables have little effect on predicting if an individual will recidivate. The accuracy of our first model, which includes data on decile score, age, marital status, and race, is approximately 66.65%. When removing the three variables and isolating decile score, the accuracy of the model becomes 65.79%. Additionally, by fitting the models, we examined the coefficients for each feature. COMPAS decile score had a very high coefficient of .68, showing that COMPAS decile score is a significant predictor for recidivism. Coefficients for race and marital status were all below .05, meaning these factors were not very good predictors of recidivism. The only other relatively high feature was age at a coefficient of -0.19. When looking at the number of Type I and Type II errors, shown in Figure 2, we saw a 33% error in the model, which is high considering these two errors would lead to higher numbers of people who are inaccurately flagged as high/low risk at recidivism. These inaccuracies may lead to misjudgment and misallocation of rehabilitation resources.

Diving a little deeper into the COMPAS scores and the distributions across our variables of interest, as shown in Figure 3 of the Appendix, we find that there are some strong correlations between COMPAS decile scores and factors like race and marital status. With race, for example, we see that African American defendants were associate with much higher COMPAS scores than their Caucasian counterparts. Other races will be omitted from this analysis as there are not enough observations in this dataset to create a meaningful analysis. When looking at marital status, we find married individuals are associated with a much lower decile score compared to single people. Similarly, other marital statuses will be omitted from this analysis as there are not enough observations in the dataset to create a meaningful analysis.

While there are many factors that may contribute to a person's risk of recidivism, this study found that COMPAS scores statistically created a good predictor for modeling risk of recidivism. This result shows that COMPAS scores are a good start, but with the underlying inaccuracies we suggest that stakeholders seek better methods. If COMPAS becomes a large factor in evaluating defendants, this creates a very risky situation. With our data, we saw higher predicted scores correlated with African American defendants as well as individuals who are younger and individuals who are single. This has the potential to be dangerous if COMPAS generated scores that show a high risk of recidivism have an anchoring effect on policymakers and law enforcement. Without diving deeper into the underlying factors that contribute to COMPAS scores, this analysis may suggest some bias in COMPAS scores assigned to individuals. Recidivism is a result of many different factors that currently cannot be measured solely by data.

This study is set to understand the accuracy of COMPAS scores in predicting recidivism. Using machine learning, a logistic regression model was used to better understand the effect of COMPAS scores along with demographic factors through comparing the different model accuracies. With this method, we found the models identical for the sake of meaningful differentiation. Considering our model comparison, we found that demographic factors were not good predictors for recidivism. Through evaluating the proportion of false positives and false negatives in our model, the accuracy of the model was very low considering the use case of the COMPAS tool. This study suggests that the COMPAS decile score itself is unreliable to judge an individual despite finding it to be the most effective measure. Uncovering some of the distributions of COMPAS decile score among our variables of interest suggests that there may be some underlying biases with assigned COMPAS scores, but further analysis is needed to better understand these relationships. Our suggestion is for stakeholders to hold off on including

COMPAS scores in their evaluation of individuals as its inaccuracy with predicting individuals is very poor, especially those who do not fall on the high-risk end.

Appendix

To build a machine learning model with categorical data, one must transform our categorical variables into dummy variables. Dummy variables are a way to express categorical data in ones and zeros to represent scoring true or false for each category. An example of this is shown below where the categories are single, married, and divorced and how they are shown in a table.

Interpretation	Single	Married
Is Single	1	0
Is Married	0	1
Is Divorced	0	0

Zero is the new value if the row is false for that category and one if it is true. The last category, divorced, does not have a column because this is represented by a row of false values for each category. We first remove the original categorical columns and replace them with the respective dummy columns. Once a fully numeric data frame is created, independent variables and the dependent variable become a subset. We normalized the data to prevent miscalculations when training the data. The age and decile score columns contain large numbers compared to the other columns. The data frame next split into training data and testing data. We chose to allocate 30% of our data to testing data. Finally, we build a logistic regression model and fit it with the training data. This gives us coefficients for each column and the intercept of the regression line. To evaluate the model, we made a confusion matrix which is shown in Figure 2. This highlights the number of true positives, true negatives, false positives and false negatives when running the model. Overall accuracy is calculated below based on the first confusion matrix where we take the sum of true positives and true negatives divided by the total number of values.

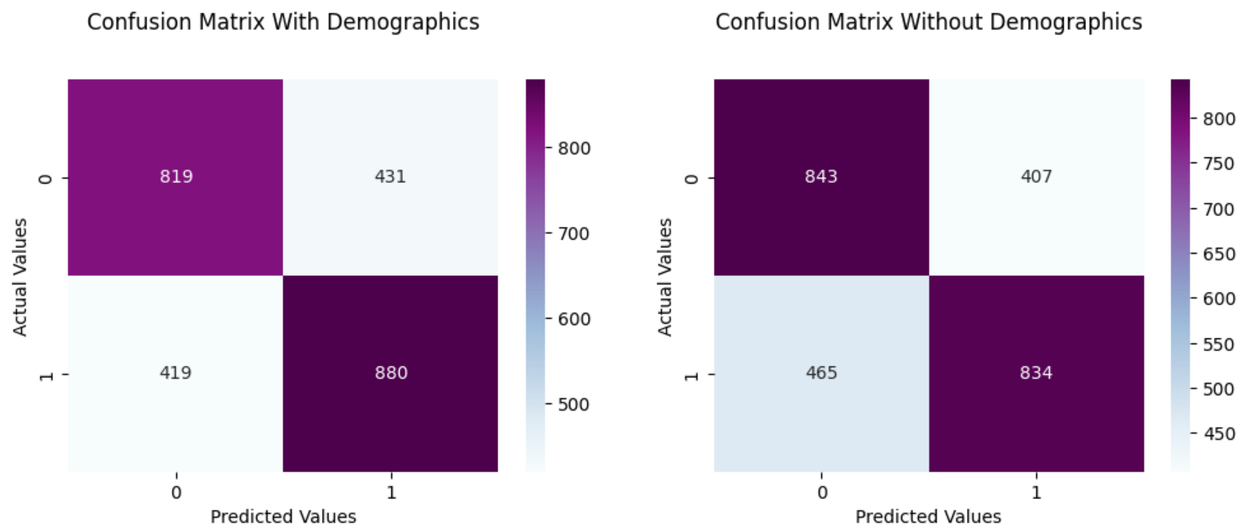

Figure 2: Confusion matrices for our two models

$$\frac{880 + 819}{819 + 431 + 419 + 880} \approx 66.65\%$$

Figure 3: Pairwise Relationships Diagram with Variables of Interest