# 

Yuting Li

200434800

ENEL 489 Social and economic impacts of Artificial Intelligence

Project Report

July 28th, 2022

1.0 Introduction

Recidivism is a term used to describe the tendency of a prisoner to commit another crime after being released. When people count the recidivism rate, they usually associate it with a given time period. For instance, the two-year recidivism rate refers to the likelihood of an offender committing another crime within the two years of being released. This requirement of timeliness has also added a lot more difficulties to the recidivism prediction.

NorthPointe is a private company that is responsible for developing artificial intelligence applications for judicial purposes in the U.S. particularly to examine the recidivism rate of prisoners (Julia 2016). Its most well-known algorithm COMPAS has been widely used by New York, Wisconsin, California, Florida's Broward County, and other jurisdictions. In the early version of this algorithm, it accepts “more than 100 factors, including age, sex and criminal history. Notably, the race is not used” (Corbett 2016). CAMPAS will then assign him/her a score, from 1 to 10, the higher the score, the higher the probability of re-offending in the future and the court will take this score as an important factor to assign the offender a prison time, the higher the score the longer the prison time.

However, in research led by a non-profit publisher ProPublica, researchers have collected information for over 10 thousand people that had committed a crime and were arrested in Florida’s Broward County. These prisoners are being scored by COMPAS and researched checked to see how many of them were charged with further crimes within two years (Julia 2016). The outcome of this research is shocking, the result showed that black defendants were twice as likely to be incorrectly labelled as having a higher risk than white defendants.

This project is intended to investigate the machine bias within the U.S. AI recidivism predictor COMPAS using mathematical-statistical methods as well as to improve the overall accuracy of the model and thereby eliminate the potential racism within that system.

2.0 Analysis of Social and Economic Issues

Crime can happen in every society. Its consequences may affect everyone within that society to some degree. The costs and influences of crimes could be widely varied. In addition, some costs are short-term while others may last a lifetime.

The increase in crime rate & recidivism rate could result in increased security expenses. For example increase of locks, CCTVs, security guards or even police. Considerable money is spent to avoid being victimized. Direct economical losses due to crimes may include medical expenses, property damage, or even funeral expenses.

Some consequences of crimes are less tangible. For instance, mental damage. Mental damage can be ranged from depression and suffering all the way to committing suicide. Intangible damages are not as obvious or as measurable as direct financial losses but they certainly do have negative impacts on people’s quality of life. Moreover, mental damage also has traumatic impacts on friends and the disruption of the family of both the offender and the victims. Mentality or even behaviours can be forever changed and shaped by crime, whether it be never trusting people or even the fear of making new friends.

As technology develops, the U.S. government seeks new opportunities to evaluate offenders’ recidivism rates on large scale and in an automated way by adopting the latest technologies in artificial intelligence. With this need, NorthPointed developed the COMPAS algorithm, if the algorithm gives a prisoner an overall score of 1, the recidivism rate was only 22% while the prisoner who has an overall score of 10 has an 81% of recidivism rate(Corbett 2016). On average, COMPAS is having a good overall accuracy rate of about sixty per cent.

However, when specifically counting cases of no re-offending but being assigned with a score of 5 or more. For the same prisoner who did not re-offend later, blacks were 42% more likely to be given high scores by the algorithm, while whites were only 22% likely.

Hence, COMPAS is doing good but yet it can not fully fulfil the needs of saving the judicial budget, eliminating biases as well as protecting the overall safety of society.

According to Dr. Wagner's research, in the year 2022, the U.S. has nearly two million prisoners, and the country is spending $182 billion U.S. dollars annually on the prison system. Among the two million people, about half of them are imprisoned in state jails; ten per cent are in federal jails and the rest are in private jails because the country just does not have enough amount of jails to handle them. Therefore, it is important to evaluate the recidivism rate of a prisoner, if the rate is low, the prisoner might as well be released early to decrease the expenses. Whereas if the rate is too high, it is the judicial system’s job to ensure the safety of the entire society and not to release the prisoners too early.

However, racism has long been a problem in the U.S. judicial system. Many black people do believe that they are being systematic suppressed. As a result of that, many of them do not believe that they are being treated fairly by society and therefore they have also given up on their obligation of obeying the law. With that being said, there is a possibility that the crime rate might decrease if people are feeling that they are being treated equally.

Here comes a long debating problem, whether we should give severe punishment or lenient punishment to the criminals. many conservatives believe that we should always take the public’s safety into first consideration. In order to achieve this goal, it is preferred to give severe punishment to the criminals. On the contrary, another point of view is that giving criminals lenient punishment is also to give them a chance to rehabilitate, criminals should also be treated with respect (Basic, 1900).

From consequentialism’s perspective, the harm that a criminal can do to society is inestimable whereas judicial injustice can only make limited harm to specific individuals, therefore it is okay to give severe punishment. From deontology’s perspective, both kinds of punishment have their own reasons and therefore it is hard to determine.

In my personal opinion, I prefer neither severe nor lenient punishment, instead, I prefer flexible punishment. This is because different people have different motivations when committing crimes, for the majority of them, we want them to be punished for what they have done but we also want them to never make any crime in the future and get back to their normal lives as soon as possible. On the other hand, impenitent criminals especially professional criminals, are bounded to commit crimes one after the other and they deserve heavy punishment. Therefore, it is important to have a tool that can evaluate criminals’ recidivism rate and thereby differentiate between these two kinds of criminals and in this scenario, the accuracy rate will become the most important part of the prefectorial system.

3.0 Design of prototype system

There are two goals that I would attempt to achieve in this project. The first is to prove the machine bias within the COMPASS system using the statistical method. The second is to improve the recidivism predictor by using logistic regression and thereby reduce the overall racism in the old system. This is because each prisoner that was assigned with a high re-offending score but did not commit another crime means that they are wronged, that extra prison time is a waste of their lifetime and money of the judicial system. Yet, each prisoner that was assigned with a low score but still committed another crime after being released represents a false-negative error of the system and this could result in severe harm to the entire society.

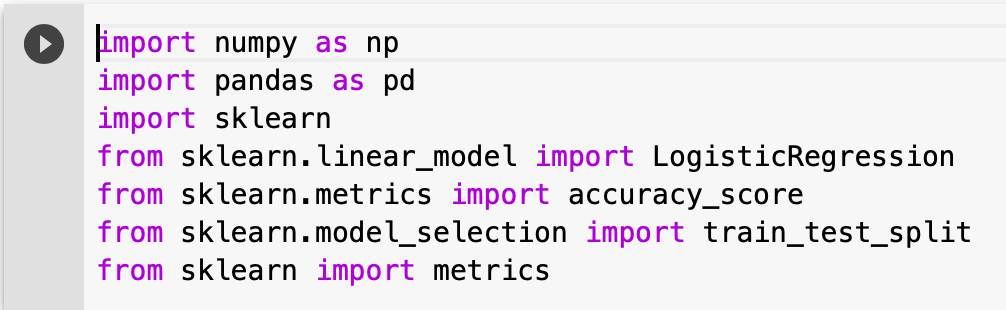
I have been considering both using linear regression and logistic regression to build my predictor but eventually, I decided to use logistic regression. The reason is that:

1. Linear regression is good for solving regression problems whereas logistic regression is good for classification problems and recidivism is a classification problem.
2. Linear regression can predict the value of continuous variables but when it comes to recidivism rate, every input is independent of the other
3. Logistic regression is relatively easier to implement,
4. Logistic regression can interpret model coefficients as indicators of feature importance.

4.0 Implementation of Prototype System

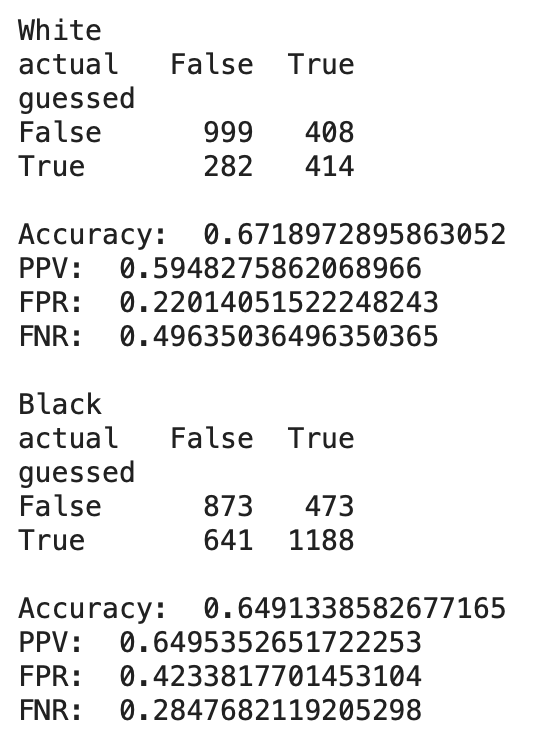
First of all, the language that I plan to use is Python, the programming platform is Google Colaboratory.The dataset and inputs of my model are provided by ProPublica. ProPublica’s journalists Julia Angwin and Jeff Larson have been studying the judicial inequality of the U.S. government since the year 2016. Their dataset consists of more than 7000 entries.

Secondly, I am planning to import libraries NumPy and Pandas. NumPy stands for Numeric Python, it provides functions for linear algebra, matrices and etc. I will use it mainly for mathematical operations as well as statistical purposes. Pandas is a Python library for data sets.

It has functions for analyzing, cleaning, exploring, and manipulating data and I will use it for most of my dataset operations such as data cleaning, making dummy variables, etc.

Lastly, the algorithm that I am planning to use is logistic regression. Logistic regression is usually used for classification and prediction purposes and is a form of supervised learning. It can estimate the probability of the occurrence of an event based on past records and a set of variables.

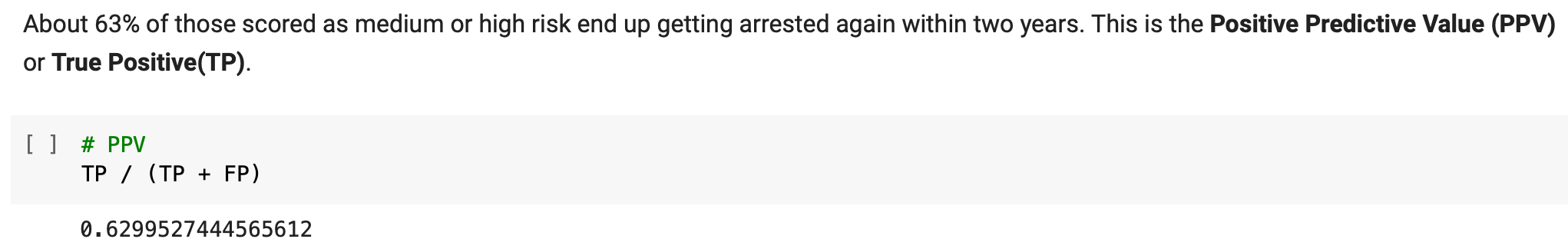
For the first part of my project, I proved the machine bias within the COMPAS system by calculating the PPV(Positive Predictive Value) TN(True Negative), FPR(False Positive Rate) and FNR(False Negative Rate) values. Positive Predictive Value predicts the prisoner will re-offend in two years after being released and he/she really did. False Positive is also known as the Type I Error, it means that we predict the prisoner will re-offend in two years after being released and he/she did not, which also means that the offender has been wronged by the system. On the other hand, False Negative is also known as the Type II Error, it means that we predict the prisoner will not re-offend in two years after being released but he/she did re-offend. The result are shown below:



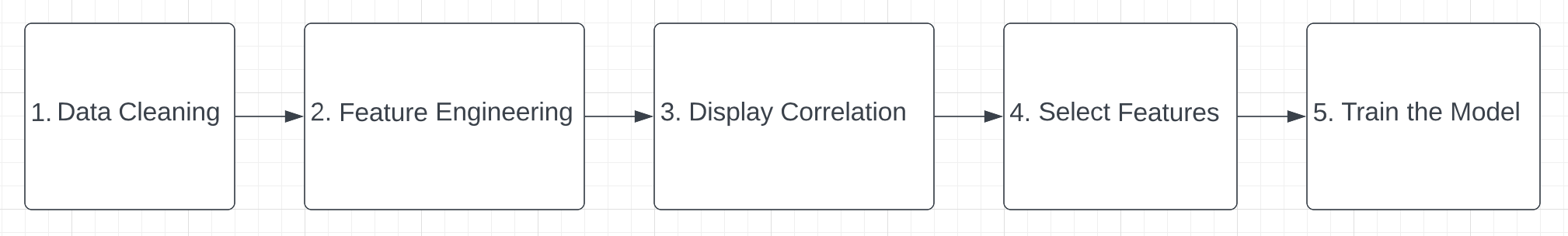
From this image we can tell:

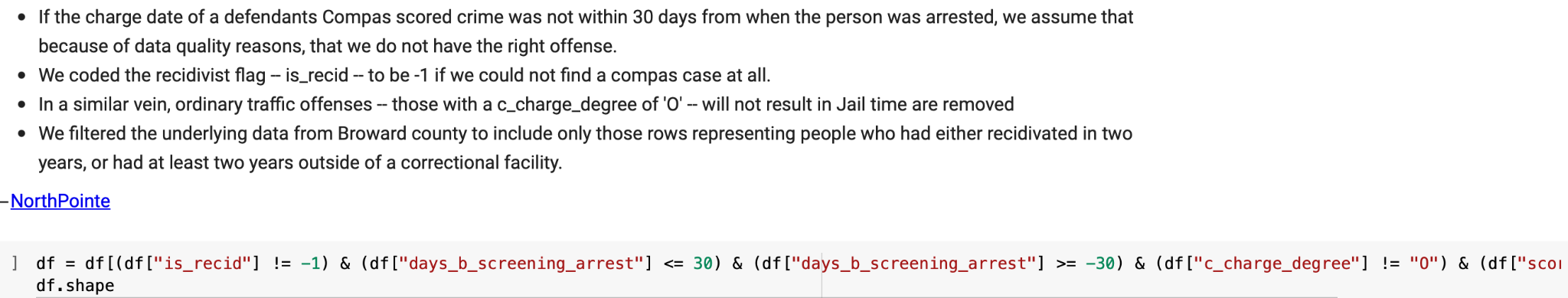
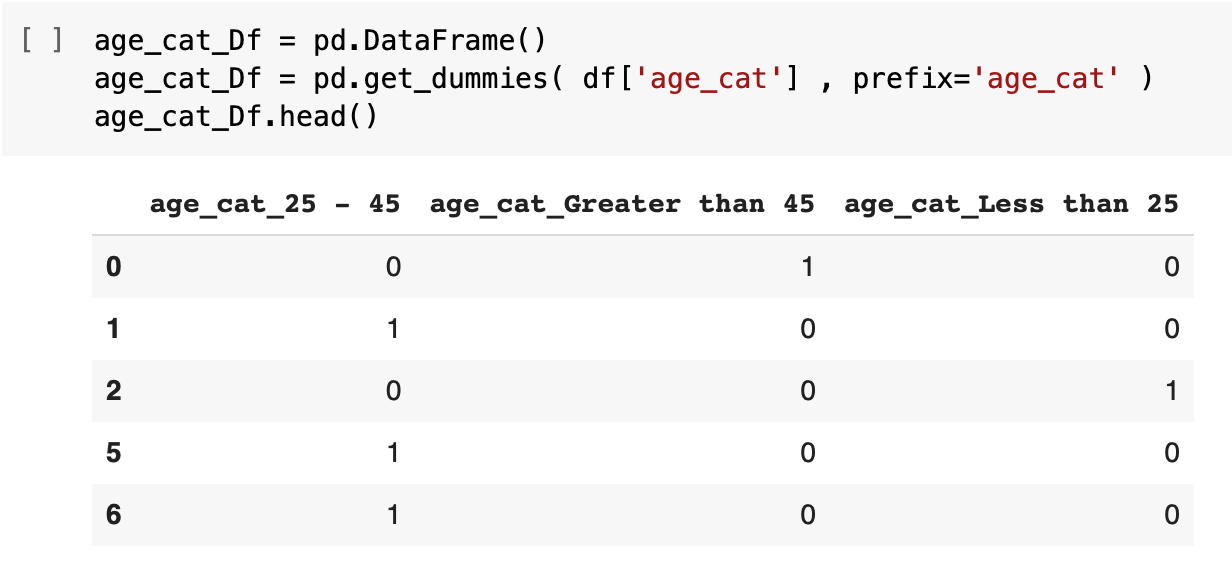
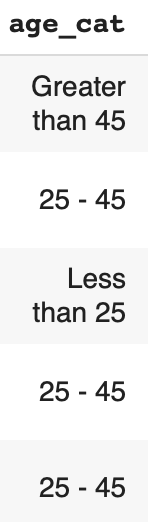
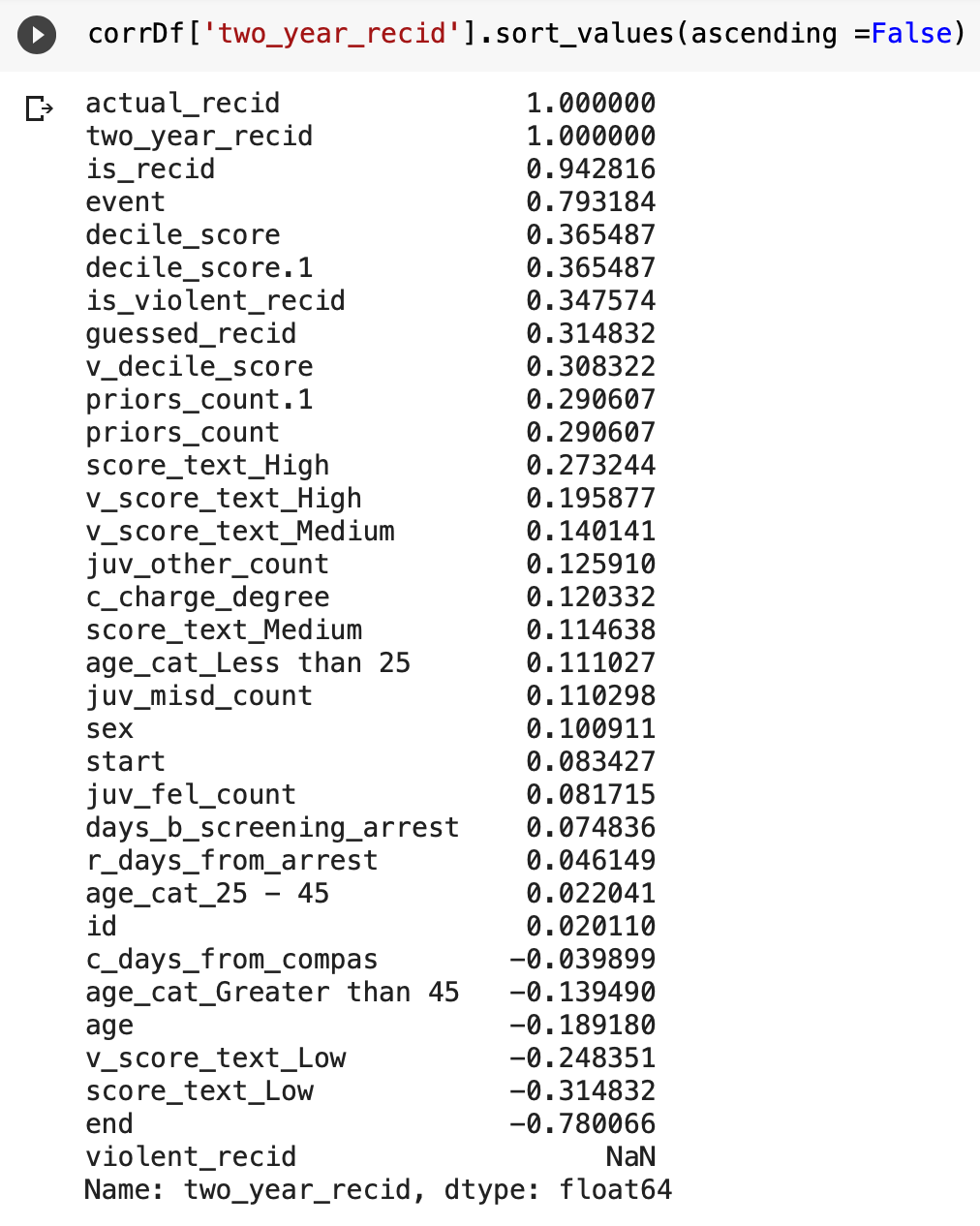
1. The overall accuracy for both the white (67%) and black (64%) offenders are pretty close to each other.
2. 42% of the black offenders and 22% of the white offenders are being assigned to an overly high score.
3. 28% of the black offenders and 49% of the white offenders are being assigned to an overly low score.

From the three points above, we can tell that the COMPAS recidivism predictor has a tendency to assign black offenders with high scores and assign white offenders with low scores. This means that this system is biased.

For the second part of my project, I intended to reduce racism by increasing the overall accuracy of the system. By calculating the True Positive among all defendants, we can get that the overall accuracy of the COMPAS is about 63% and my goal is to achieve a higher number than that. 

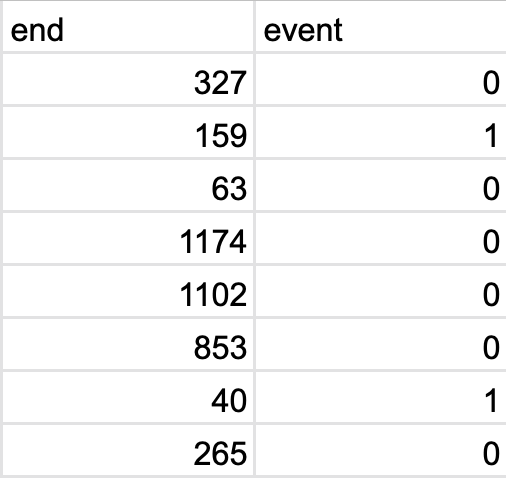
In order to achieve this goal I decided not to use “race” as an input of my system, and I tried to use only the features that show strong positive/negative correlation to the two-year recidivisms. The whole process can be split into five steps:



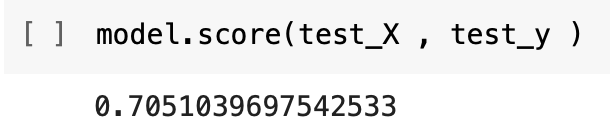
1. Data Cleaning: this step is to correct the structural errors in data sets. I did this step by reading the documentation provided by ProPublica and eliminated illegal entries accordingly. 
2. Feature Engineering: This step is to process the data into a way that the machine can work with it in better ways. In this step, binary form and decimal form are preferred.
   1. For example instead of display “Male” and “Female”, I made the sex to display “1” and “0”. “1” representing “Male” and “0” representing “Female”.
   2. For inputs that can be categorized into groups, I have created dummy variables for them. For example: for the input column “age\_categories”, we used to have inputs of “Less than 25”, “25-45” and “Greater then 45”. I created three dummy variable each representing one category and used “1” and “0” to represent true and false.
3. Display Correlation: in this step, I displayed all the features that show correlation with the two year recidivism. The numbers range from positive one to negative one, the greater the number, the stronger the correlation; positive number means positive correlation and negative number means negative correlation. Positive means that if the offender has this feature, his probability of re-offend will increase; negative correlation means if the offender has this feature, the reoffend rate will be decreases. 
4. Select Features: this is the most important part of the entire project, the quality of the features will directly affect the overall accuracy rate and will directly affect whether or not my project will succeed. I sticked to a set of standards during this process:

1. I eliminated inputs that are in string type (names, charge description), this is not because they cannot contribute to the overall accuracy, in fact, I do think charge description could be very useful. However, due to the fact that I do not have the knowledge to correctly process them I have to eliminate them.

2. I eliminated all inputs that are redundants to each other. For example, inputs “actual\_recid”, “is\_recid” and “two\_year\_recid” are showing extremely high correlation with each other (>0.9) so I shall only keep only one of them.

3. I eliminated all inputs that I cannot understand. Unfortunately, when ProPublic provided this dataset, it was not well-documented, therefore I cannot completely understand all of them. For example, input “end” and “event” all shows extremely strong correlation (-0.78 & 0.79) to the two year recidivism rate but I still decided to eliminate them because I cannot justify my AI with inputs that I cannot understand.

4. I have kept all variables that had left after the three elimination standards, I did not pick a particular range because there are not many of them left.

1. Train the model: I used all the inputs that are left in step 4 and trained my model using logistic regression, the split of my model is 80% for training and 20% for testing and the overall accurate is about 70% which is higher than the 63% accuracy rate of the COMPAS algorithm:

5.0 Discussion

I have achieved my two goals of my project: first to prove the bias existed in COMPAS and second to create a better predictor while using less variables.

I stick with the idea of assigning flexible punishment to different offenders as I stated earlier. From the result of my project, I can tell that a 7% increase in accuracy could result in huge amounts of money being saved and a bit less national crime rate.

Using logistic regression is also the right decision to make. It is easy to implement and has made my project much easier than I thought.

By reducing the amount of factors, the system has also became more accountable and transparent as the system becomes simpler and therefore we can justify the system’s decision easier.

6.0 Conclusion

Overall, I do think that I have accomplished the most parts of my project that I intended to. However, there are many things that I can improve if I have more time and more knowledge. First of all, if I have the knowledge to process strings and extract valuable information from it, I can certainly improve the overall accuracy of my project. Secondly, more background knowledge about crime. From this project, I have learned that as software engineers, we have to know both the software technology that we are using and also the problem that we intend to solve. I had very little background knowledge about crimes and recidivism rate beforehand and I learned most of the knowledge that I need while working on this project. I wish I could have more insights on the problem and so that I might be able to solve similar problems using different approaches in the future.

7.0 Future Work

7.1 Better User Interface

In future, a better user interface could be added to the system to allow people easily operate the system without having to change the code. By doing so, this system might be able to be deployed for large-scale usage and by non-technical people.

7.2 More Diverse Dataset

In this project, I have only used a dataset from the Broward County of Florida, which means that this AI system might not be useful for defendants from elsewhere. This is because recidivism rate is affected by the family, regional or even cultural background. U.S. as the country which has the most prisoner on the world and yet keeping up a high recidivism rate(Frederick 1999)

8.0 References

Julia Angwin and Jeff Larson, Bias in Criminal Risk Scores Is Mathematically Inevitable, Researchers Say, ProPublica, Dec. 30, 2016

In this article, Julian and Jeff had studied the COMPASS AI application for predicting recidivism rate that was developed by a private company called NorthPointe. They did an experiment using the COMPAS scores for more than 10,000 people arrested for crimes in Florida’s Broward’s County. They have found that this AI application has a rate of accuracy of 60 percent for both the black and white people. However, when they looked at the people who did not commit new crimes, they found a racial disparity. The data showed that black defendants were twice as likely to be incorrectly labeled as higher risk than white defendants.

Jeff Larson, J. A. (2016, May 23). How we analyzed the compas recidivism algorithm. ProPublica. Retrieved June 7, 2022, from https://www.propublica.org/article/how-we-analyzed-the-compas-recidivism-algorithm

In this article Jeff explained how ProPublica analyzed the COMPASS algorithm: they first examined some of the most influential and effective thesis about recidivism rate prediction. Next, they collected data of 34,000 federal offenders using tools that was collected by the federal court and compared the result with COMPASS. Eventually, they had found that Black people and Hispanic people had a relatively higher chance to get a higher crime score compared to the White people.

Jon Kleinberg, Sendhil Mullainathan, Manish Raghavan, Inherent Trade-Offs in the Fair Determination of Risk Scores, arXiv:1609.05807 [cs.LG]

Corbett-Davies, S., Pierson, E., Feller, A., & Goel, S. (2021, December 7). A computer program used for bail and sentencing decisions was labeled biased against blacks. it's actually not that clear. The Washington Post. Retrieved June 7, 2022, from https://www.washingtonpost.com/news/monkey-cage/wp/2016/10/17/can-an-algorithm-be-racist-our-analysis-is-more-cautious-than-propublicas/

This report stated the study result of the ProPublica, it also provided the COMPAS input in the questionnaire form.

Frederick, B. (1999). *Factors contributing to recidivism among youth placed with NYS DFY*. Retrieved July 31, 2022, from https://www.criminaljustice.ny.gov/crimnet/ojsa/dfy/dfy\_research\_report.pdf

In this report, Frederick stated that there are three factors that would affect the recidivism rate the most: crime history, age and community characteristics.

"Economic and Social Effects of Crime ." Crime and Punishment in America Reference Library. . Retrieved July 28, 2022, from Encyclopedia.com: https://www.encyclopedia.com/law/encyclopedias-almanacs-transcripts-and-maps/economic-and-social-effects-crime

This article talks about the economic and social impacts of crimes, both the tangible ones and the intangible ones, it also talked a lot about the psychological impacts on crimes which is something that I had ignored.

Wagner, P. (2022). United States profile. United States profile | Prison Policy Initiative. Retrieved July 25, 2022, from https://www.prisonpolicy.org/profiles/US.html

*Basic principles for the treatment of prisoners - OHCHR*. Basic Principles for the Treatment of Prisoners. (1990). Retrieved August 1, 2022, from https://www.ohchr.org/sites/default/files/basicprinciples.pdf

This article stated eleven points of the basic principles for the treatment of prisoners.