

Pojeto Final - Análise do COMPAS

Henrique Urban*
Thiago Mohallem*

ACM Reference Format:

Henrique Urban and Thiago Mohallem. 2019. Pojeto Final - Análise do COMPAS. In *Proceedings of ACM Conference (Conference'17)*. ACM, New York, NY, USA, 3 pages. <https://doi.org/10.1145/nnnnnnnn.nnnnnnnn>

1 INTRODUÇÃO

Em múltiplos estados dos EUA, são usados por juízes algoritmos para avaliar a probabilidade de reincidência criminal, ou seja, de um preso realizar uma nova ofensa. Dentre esses algoritmos, entrou em destaque um em específico, uma ferramenta comercial desenvolvida pela Northpointe, Inc., chamado COMPAS (Correctional Offender Management Profiling for Alternative Sanctions).

Esse algoritmo usa 3 escalas para classificar cada indivíduo, a partir de conceitos comportamentais e psicológicos de suposta alta relevância para a decisão de reincidência criminal. As métricas são: Liberação Pré-julgamento, Reincidência Geral e Reincidência Violenta. A primeira diz respeito ao risco de haver um novo crime ou falha de comparecer ao julgamento em casos nos quais o ofensor é liberado antes da conclusão de seu julgamento. Já as outras duas métricas determinam o risco de o indivíduo cometer um novo crime após ser solto pelo seu julgamento atual, diferenciando apenas na natureza desse novo crime [3].

2 MOTIVAÇÃO

Sistemas de pré-julgamento que usando técnicas de aprendizado de máquina podem apresentar um viés racista por assim dizer dependendo da cor de um indivíduo. No caso do COMPAS, esse sistema é utilizado efetivamente para ranquear, de uma maneira indireta, o nível de perigo que certa pessoa representa ao ser solta. Esse perigo não deveria levar em conta o tom de pele de uma pessoa, e sim seu passado e sua condição mental, porém, há relatos nos Estados Unidos de pessoas brancas que cometeram crimes utilizando armas de fogo e foram presos sendo classificadas como menor risco de reincidência do que negros que cometeram delitos menores quando eram jovens[1].

Esses estudos são de vital importância para que injustiças possam parar de serem cometidas, ainda mais quando levam em conta liberdade de indivíduos. A escolha do COMPAS, especificamente, foi feita por ele representar um dos algoritmos de ranqueamento mais populares usados nos Estados Unidos, e que tem uma popularidade ascendente entre os sistemas prisionais e de julgamento.

*Both authors contributed equally to this research.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted by ACM, Inc., provided that the fee of \$15.00 is paid directly to ACM. This permission is granted without fee or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.
Conference'17, July 2017, Washington, DC, USA
© 2019 Association for Computing Machinery.
ACM ISBN 978-x-xxxx-xxxx-x/YY/MM...\$15.00
<https://doi.org/10.1145/nnnnnnnn.nnnnnnnn>

2019-11-29 22:01. Page 1 of 1-3.

3 TRABALHOS RELACIONADOS

Dada a dimensão e a sensibilidade do COMPAS, a ProPublica, organização de jornalismo investigativo não lucrativa e independente, realizou um estudo [2] com o objetivo de descobrir a acurácia do algoritmo de reincidência e testar se ele era enviesado contra certos grupos e concluiu que o esse algoritmo é tendencioso. Foi concluído que, de forma geral, ofensores negros tinham uma chance maior de serem classificados como alto risco de reincidência de forma equivocada que os brancos, que por sua vez tinham uma chance maior de serem classificados como risco baixo erroneamente.

O trabalho da ProPublica foi feito em cima de dois riscos que o COMPAS relata em seu algoritmo: Risco de reincidir e risco de violência, eles não trataram os dados de risco de não aparecer, isto é, o risco da pessoa não aparecer para o julgamento.

As principais conclusões feitas a partir da análise do COMPAS por esse estudo foram: 1- Réus negros eram marcados como maior risco de reincidência do que realmente eram para ter. 2- Réus brancos eram marcados como menor risco de reincidência do que realmente eram para ter. 3- Réus negros também eram duas vezes mais propensos a serem classificados como tendo um risco maior de reincidência violenta do que os réus brancos, enquanto os réus brancos eram 63% mais prováveis de terem um score errôneo de Baixo Risco, ao comparar com os Negros.

As Figura 1 a seguir mostra a diferença encontrada pelo estudo da ProPublica dos scores do sistema COMPAS ao comparar as cores dos indivíduos, é nítido como que negros são muito mais classificados como alto risco (8-10) do que os brancos que têm uma concentração de baixo risco muito maior.

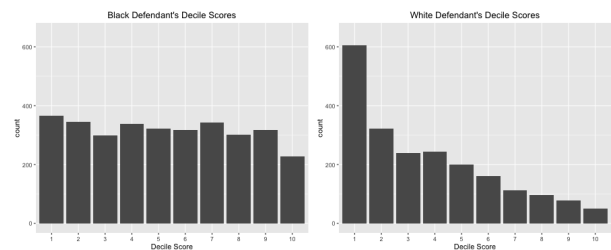


Figure 1: Resultado da Análise do ProPublica - Risk of Recidivism

Neste trabalho, pretende-se determinar se os resultados do estudo da ProPublica de conclusão enviesados ainda são válidos se tomados a partir de uma perspectiva diferente: Utilizando o "Risk of Failure to Appear", igual está escrito no sistema.

4 OBJETIVOS

O trabalho foi inteiramente baseado no estudo feito pela ProPublica, porém enfatizado em uma linha em que eles citam:

“We analyzed the COMPAS scores for “Risk of Recidivism” and “Risk of Violent Recidivism.” We did not analyze the COMPAS score for “Risk of Failure to Appear.”[2]

Basicamente, o estudo, muito bem feito, pela ProPublica, focou no Risco de reincidência de presos e no risco de reincidência cometendo atos mais violentos, ambos baseados na cor de pele dos indivíduos, porém não avaliaram se o “score” dado pelo COMPAS para “Risco de falhar para aparecer” também tem o viés racista que os outros obtiveram. O trabalho entra exatamente nesse ponto, a dupla de autores analisou os dados do COMPAS fornecidos pelo estudo do trabalho relacionado, mais precisamente do condado de Broward, que fica no estado da Flórida.

Como a dupla não têm muitos conhecimentos em Machine Learning e Data Science, a metodologia a ser seguida foi fundamentada na metodologia dos autores da ProPublica, todos os passos do estudo foram replicados, porém com modificações nos dados analisados.

5 METODOLOGIA

O Condado de Broward liberou dados referentes a dois anos de ranqueamentos dados pelo COMPAS, no total foram 18.610 pessoas analisadas pelo algoritmo, e esse score é utilizado nesse condado para saber se solta ou detem um réu antes de seu julgamento.

Tal qual os autores da ProRebublica, os dados referentes as pessoas selecionadas estão filtradas para atenderem certas condições: Foram descartados os dados de pessoas que estavam em liberdade condicional ou outras etapas de sistemas de justiça criminal. Os scores do compas variam de 1 a 10, sendo que de 1 a 4 fornece o rótulo de “Baixo Risco”, 5-7 “Médio Risco” e de 8-10 “Alto Risco”.

Aproveitamos de toda essa filtragem ja feita pelo estudo da ProRepublica em que eles disponibilizaram um arquivo no formato csv com os dados das pessoas que precisavamos, porém nesse dado não estava incluído o “Risk of Failure to Appear”, então a dupla foi atrás dos dados originais, filtrou os dados para coincidir com as mesmas pessoas e adicionou na tabela já existente os scores de “Risk of Failure to Appear”, e o label correspondente ao score atribuído. No total foram cerca de 6172 pessoas analisadas em que 57.2% tinham entre 25 e 45 anos, 20.94% tinham mais de 45 anos e 21.86% tinham menos de 25 anos. Desse total de 6172 pessoas 51.44% eram classificadas como negras, e 80.96% eram homens.

6 RESULTADOS

Em contraste com os resultados da Figura 1, que demonstram as tendências de classificação de risco de reincidência criminal, os resultados obtidos em relação ao “Risk of Failure to Appear” demonstraram resultados muito menos tendenciosos, como pode ser observado na Figura 2 abaixo.

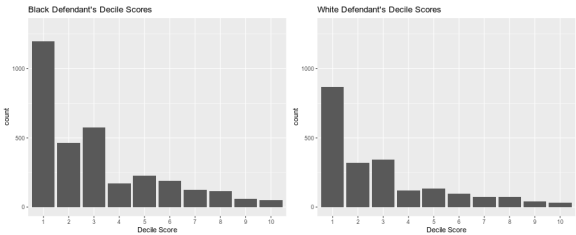


Figure 2: Resultado da Análise da Dupla - Risk of Failure to Appear

Observa-se que a distribuição dos scores é semelhante para ambas as etnias, concentrado nos scores mais baixos. Em seguida, de maneira semelhante à metodologia usada pela ProPublica, foi feita uma regressão logística com intuito de comparar os scores a partir de raças, gêneros e idades.

Os resultados obtidos aqui foram no sentido oposto daqueles obtidos na análise de recidivismo. Enquanto os scores de recidivismo para negros, mulheres e pessoas com menos de 25 anos eram respectivamente 45%, 19.4% e 250% mais prováveis de serem altos do que para brancos, homens e pessoas de meia idade, nos de failure to appear, as mesmas categorias foram 12%, 15.5% e 60% menos prováveis de ter scores altos. As figuras abaixo mostram o sumário das regressões logísticas usadas.

Risk of General Recidivism Logistic Model	
Dependent variable:	
Score (Low vs Medium and High)	
Female	0.221*** (0.080)
Age: Greater than 45	-1.356*** (0.099)
Age: Less than 25	1.308*** (0.076)
Black	0.477*** (0.069)
Asian	-0.254 (0.478)
Hispanic	-0.428*** (0.128)
Native American	1.394* (0.766)
Other	-0.826*** (0.162)
Number of Priors	0.269*** (0.011)
Misdemeanor	-0.311*** (0.067)
Two year Recidivism	0.686*** (0.064)
Constant	-1.526*** (0.079)
Observations	6,172
Akaike Inf. Crit.	6,192.402
Note: *p<0.1; **p<0.05; ***p<0.01	

Figure 3: Sumário de regressão logística - Risk of Recidivism

Risk of Failure to Appear Logistic Model

Dependent variable: Score (Low vs Medium and High)	
Female	-0.192*(0.093)
Age: Greater than 45	0.491*** (0.082)
Age: Less than 25	-0.996*** (0.117)
Black	-0.145 (0.077)
Asian	0.226 (0.452)
Hispanic	-0.478** (0.147)
Native American	-0.151 (0.841)
Other	-1.273*** (0.233)
Number of Priors	0.146*** (0.008)
Misdemeanor	-0.119 (0.074)
Two year Recidivism	0.417*** (0.073)
Constant	-1.822*** (0.085)
Observations	6172
Akaike Inf. Crit.	5437.6
Note: **p<0.1; ***p<0.05; ****p<0.01, 'p>0.1, 'p<1	

Figure 4: Sumário de regressão logística - Risk of Failure to Appear

Devido a tamanha disparidade na probabilidade de scores altos de acordo com a idade, foi feito uma comparação entre scores para pessoas com menos de 25 anos e aquelas com mais de 45. As figuras abaixo ilustram graficamente essa disparidade. As consequências desse resultado envolvem quesitos sociais e estão fora do escopo desse estudo.

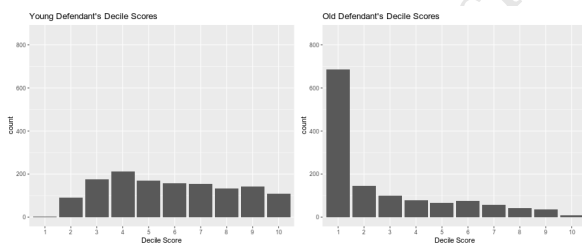


Figure 5: Resultado de comparação entre scores por idade - Risk of Recidivism

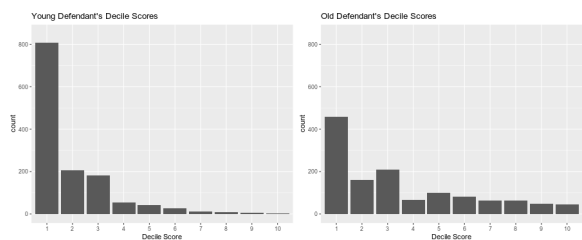


Figure 6: Resultado de comparação entre scores por idade - Risk of Failure to Appear

7 CONCLUSÃO

É indiscutível que o estudo realizado pela ProPublica a respeito do COMPAS mostrou uma vertente racista tendenciosa do algoritmo ao analisar o algoritmo de julgamento e seus resultados para diversas pessoas do condado de Broward, feito esse de extrema importância em prol de seres humanos que estavam sendo erroneamente classificados com índices que os classificam como risco "Alto" de reincidência.

O estudo feito pelos autores desse trabalho não visou desmerecer nenhum trabalho já feito sobre o tema, e sim adicionar uma variável nova para análise e verificar se a tendência racista continua sendo válida para esse novo caso, que é o "Risk of failure to appear".

Ao analisar os dados tirados e as análises logísticas feitas foi possível concluir que o score de falha para aparecer não segue o mesmo ideal tendencioso que os outros seguiram, mostram resultados completamente opostos, onde negros têm menor chance de serem classificados como maior risco, o que foi surpreendente dado os resultados anteriores. Além disso, outro fato interessante descoberto foi o de pessoas mais velhas serem classificadas com maior risco de "Failure to Appear" do que os mais novos, explicações para esse fato precisam ser devidamente analisadas e qualquer suposição sobre isso seria outra forma de preconceito.

É de fundamental importância a continuidade de análises feitas como essas para assegurar a igualdade de julgamento independente de qualquer fator social, e qualquer empenho em prol desse objetivo é bem-vindo.

REFERENCES

- [1] Julia Angwin, Jeff Larson, Surya Mattu, and Lauren Kirchner. 2016. Machine bias risk assessments in criminal sentencing. *ProPublica*, May 23 (2016).
- [2] Lauren Kirchner Jeff Larson, Surya Mattu and Julia Angwin. 2016. How We Analyzed the COMPAS Recidivism Algorithm. *ProPublica* (May 2016). <https://www.propublica.org/article/how-we-analyzed-the-compas-recidivism-algorithm>
- [3] Wikipedia. 2019. COMPAS (software). (Oct. 2019). [https://en.wikipedia.org/wiki/COMPAS_\(software\)](https://en.wikipedia.org/wiki/COMPAS_(software))