

Race, Crime, and Media Bias

Alex Knecht, Ritam Mehta, Maverick Cook, Laura Vicinanza
Team 6

Background and Motivation

Media will always inherently include some form of authorial bias. Specifically, news stories covering crimes in America will often include unconscious, or even conscious, biases that further perpetuate a stereotype or have other harmful consequences. One area to compare and contrast how crimes are reported on, is those of white perpetrators versus those perpetrated by minorities. There are clear examples where the race of a perpetrator has significantly affected an article's headline, not to mention the content.

In order to test whether or not there is a real difference between media portrayals of people of color and white people, we decided to build a classifier that could take in the text of an article and identify, from the words used, the race or ethnicity of the perpetrator. This was meant to act as a demonstration of the extent of media bias -- that it was so heinous that one could reverse engineer it. We hope use this project to expose this bias.

Related Work

Although the interplay of crime and race & ethnicity has become increasingly researched in recent years, studies on how media and social media outlets portray the perpetrators of different races/ethnicities are sparse. To the best of our knowledge, a quantitative analysis on how media outlets discuss a crime given a particular race or ethnicity of a perpetrator has yet to be formally explored.

The first large study conducted on this topic was published by the Journal of Research in Crime and Delinquency in 2002. The study, concerned with the criminal typification of race and ethnicity in local television news stations, found that Blacks and Hispanics appear in more threatening contexts than Whites. This suggests that local TV news stations contribute to the social construction of threat in relation to Blacks and Hispanics, a condition that is associated with fear of crime, modern racism, and the mobilization of various social controls and exclusions." (Chicaros, T. & Eschholz, S., 2002).

A 2010 study at the University of California Berkeley explored this topic even further, using a multivariate logistic regression analysis on racial-ethnic portrayals of victims or perpetrators of violent crimes in television crime news. The study concluded that Black crime suspects are shown in more "threatening" contexts than White suspects. They also found that Blacks were shown 2.4 times more often as crime suspects than Whites (Bjornstrom, E. E. S., Kaufman, R. L., Peterson, R. D., & Slater, M. D., 2010).

Finally, a 2016 research study at the University of Nebraska used a Critical race feminism approach towards how media portrays differ for white and minority perpetrators, using data from USA Today. Using independent variables of race/ethnicity, gender, and offense type, the study categorized the content of articles with accompanying photographs into "mitigating themes" and "neutralizing themes." Cross analyzing these, they found that that stories about black men, Hispanic men, and Middle Eastern men were more likely to provide negative depictions than stories about white women. The study also found that in stories about terrorism, varying types of nonviolent crime, and illegal doping, the depictions of Hispanic male offenders tended to be more negative than story tones about white male offenders.

Hypothesis and Method

Hypothesis: News about crime is often unfairly biased to portray people of color as less sympathetic than white people. This difference is quantifiable and manifests itself in tone and word usage.

Part 1: Collection

For our corpus, we collected articles for crimes perpetrated by members three different ethnic groups: White, Black, and Hispanic. These articles consist of wanted suspects, or a suspect being charged with a crime, that included some identifying information about their race. The information identifying race could be from a picture, eye-witness description, or from the crime report. The types of crimes included fall under the category of violent crimes. We defined violent crimes as crimes where the perpetrator intentionally or unintentionally caused, or attempted to cause physical damage or intimidation to another life.

In order to collect the articles, we chose news outlets with national coverage, as well as outlets with local reporting. We manually went through digital articles on the outlet's website, and found potential articles based on headlines that related to crime. While looking at a potential article, we decided if it fit our definition of a violent crime, and if it had any information about the perpetrators race. If so, we downloaded the plain-text of the article, and tagged the file with the respective race.

Due to the manual collection of our articles, we encountered issues with identifying and categorizing articles. First, we identified ethnicity mainly by picture of the suspects/perpetrators. This was a "best guess" rather than a confirmed. However, we would not collect articles where we were not personally confident of the race. Skimmers were also subjectively applying our violent crime definition, and may have chosen articles that did not fit that definition properly.

Part 2: Analysis

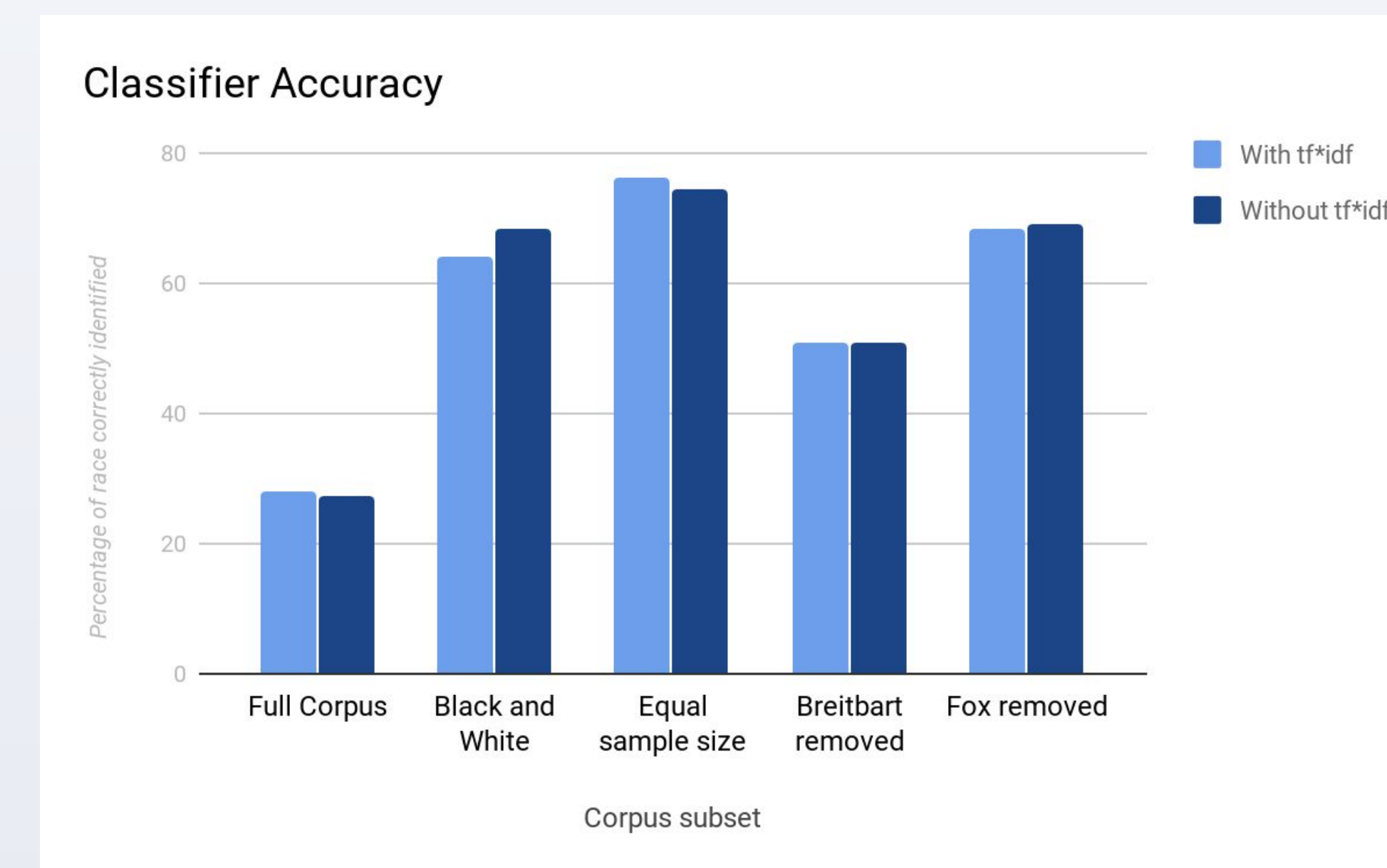
After considering our options, we ended up choosing to build a Naive Bayes classifier. This would classify documents based on probability of the whole document being in a specific class, as well as the probability of each word relating to specific classes. We would measure accuracy by individually testing each document in the corpus against a classifier trained on the rest of the corpus and averaging the results together.

We began by training our classifier on our full corpus. Our three classes were Black, White, and Hispanic. The first issue we ran into was underflow -- multiplying such small probability values for each word caused numerous zero values. This was remedied by converting the multiplication of small probabilities into the addition of large ones, using log base 10. When performing classification with the base algorithm, our accuracy was abysmal -- we were expecting something higher than 33%, but were left with 30% accuracy. We quickly determined that while the Black and White sample sizes were large, our Hispanic sample size was too small, and that the probability of classifying anything as a Hispanic-centered article was very low. We decided to repeat the experiment with just the Black and White classes and data, and were greeted with promising results - a higher than expected accuracy.

After exploring some of the data and seeing the kinds of words that were unique in each set, we decided to add a weighting scheme to our work. We decided on tf-idf, because it made sense that less frequent terms should be more relevant -- language relating to all violent crime would probably not help us, as all the articles were about that topic. We performed all of our tests with and without tf-idf.

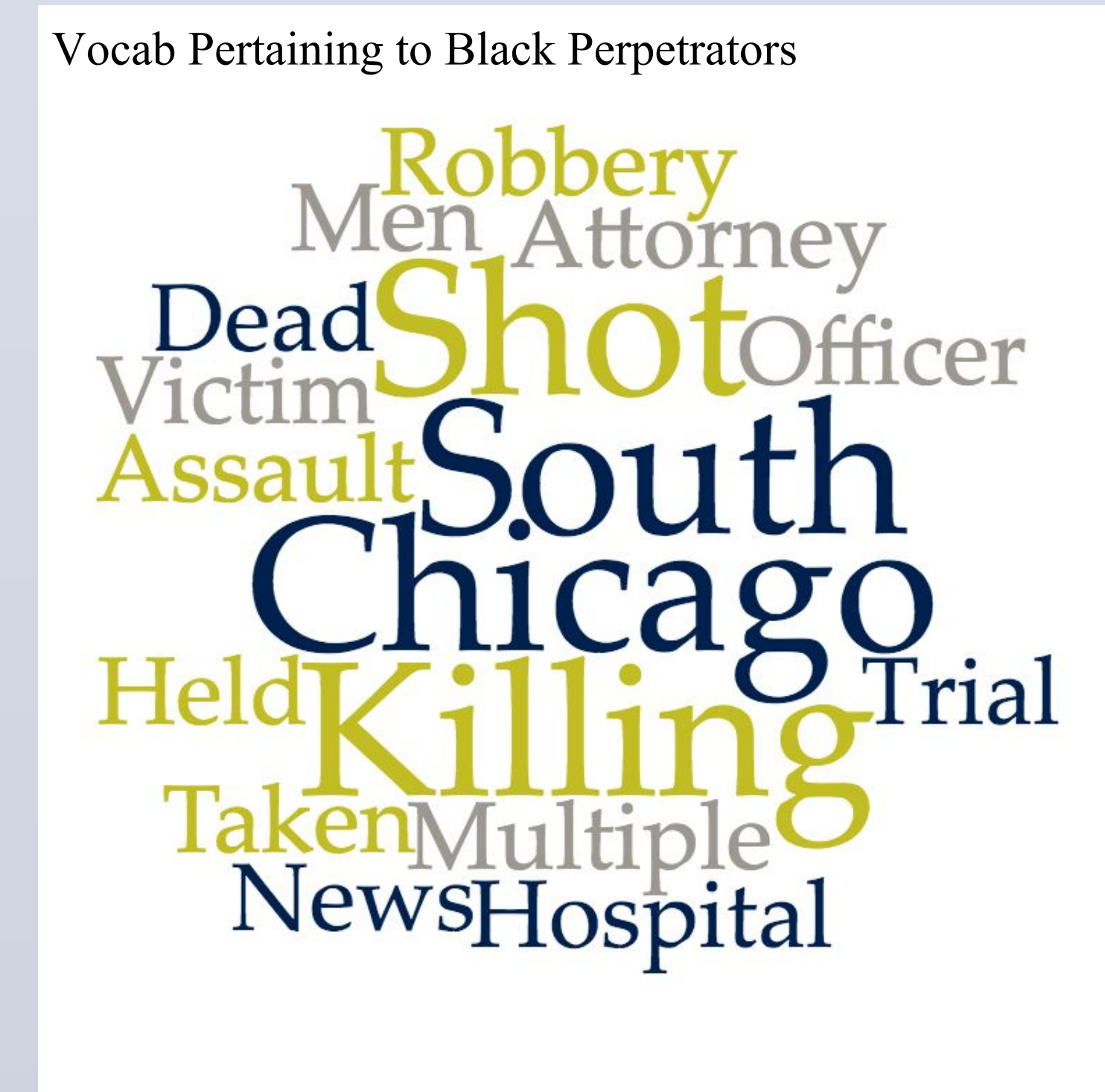
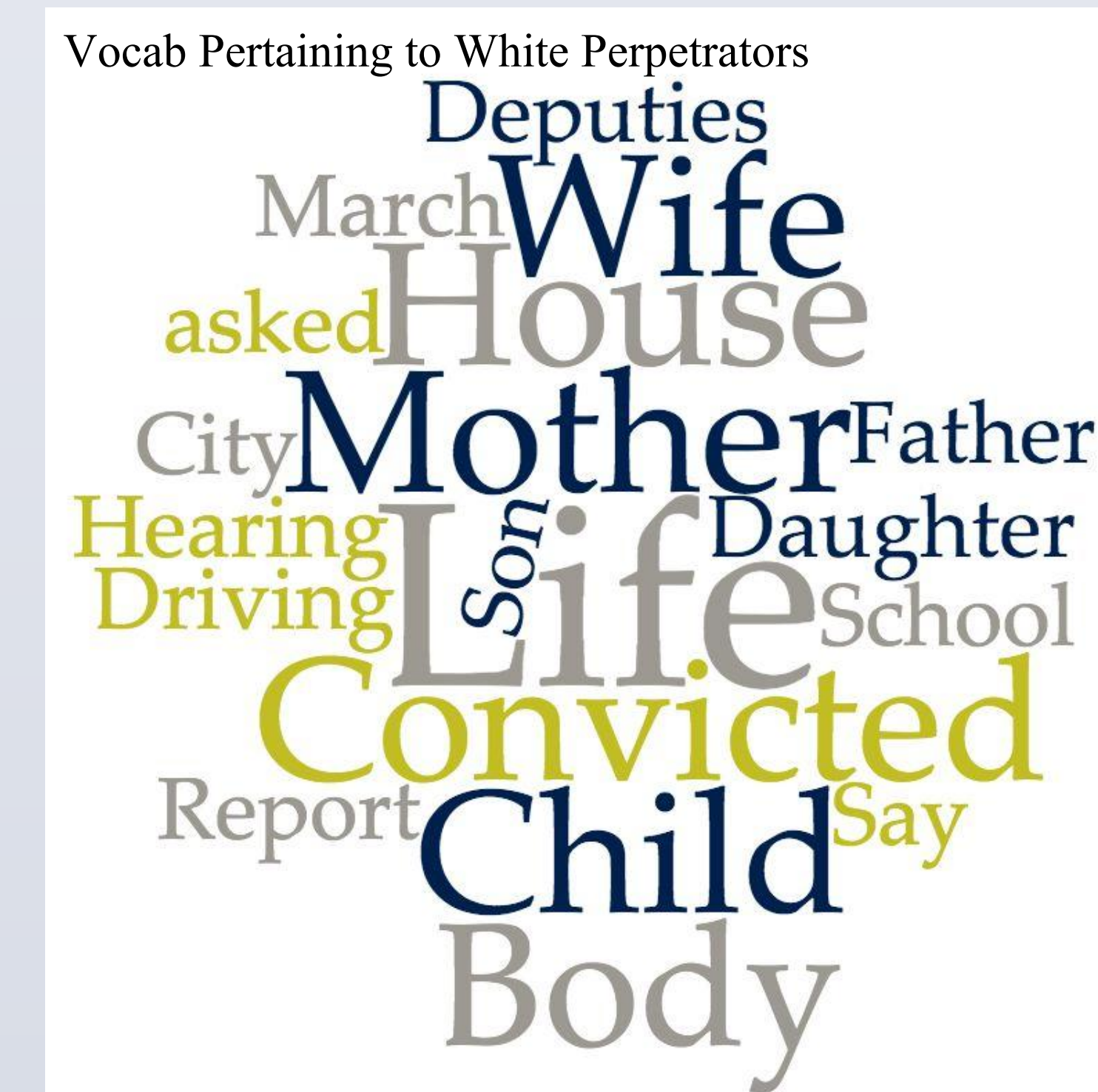
Since Breitbart and Fox News are politically polarized outlets with a blatant agenda, we decided to run some tests while excluding them from the corpus. And we ran a test where we randomly culled the Black and White data to be the same sample size as the Hispanic data, and re-ran the test with the Hispanic data.

Classifier Results



Word Association Results

For our Black and White corpus subset, we listed the words with the highest conditional probability for the vocabulary from articles pertaining to Black and White perpetrators.



Conclusion

We were able to find a weak correlation between race and the language used to describe the perpetrator, but we cannot say with certainty that there exists any strong correlation between race and language. If we had access to a crime database providing us with a larger corpus then we might have been able to produce stronger results. Many articles on crime use the same vocabulary, which leads to a lower accuracy, as this naturally leads to less of a distinction between articles. In addition, many of the articles in our training data simply described the crime rather than the background and description of the suspect, which makes it difficult for our classifier to correctly predict a suspect's race.

Our most telling data came not from the classifier, but rather from the term frequency differences between Black and White perpetrators. We found that articles on White perpetrators were more likely to mention familial terms like "child," "mother" and "wife," while articles on Black perpetrators were more likely to mention specifics about a violent crime, such as "robbery," "killing," "shot" and "victim." This finding suggests an implicit racial bias in the way crime is talked about by media news outlets..

References

- Bjornstrom, E. E. S., Kaufman, R. L., Peterson, R. D., & Slater, M. D. (2010). Race and ethnic representations of lawbreakers and victims in crime news: A national study of television coverage. *Social Problems*, 57(2), 269-293. <http://dx.doi.org/10.1525/sp.2010.57.2.269> Retrieved from <http://proxy.lib.umich.edu/login?url=https://search.proquest.com/docview/216935952?accountid=14667>
- Chiricos, T., & Eschholz, S., (2002). The Criminal Typification of Race and Ethnicity in Local Television News. *Journal of Research in Crime and Delinquency*, Vol 39, Issue 4, pp. 400 - 420. Retrieved from <http://journals.sagepub.com/doi/abs/10.1177/002242702237286>
- Epinger, Ebonie. (2016). Visual and Narrative aspects of Front-Page Crime Stories for Male and Female Offenders: Does Race and Ethnicity Matter? Thesis. Retrieved from <https://search-proquest-com.proxy.lib.umich.edu/docview/1830439473?pq-origsite=summon>
- Mitchell, A., Gottfried, J., Barthel, M. and Shearer, E. 2018. *3. Loyalty and source attention*. [online] Pew Research Center's Journalism Project. Retrieved from <http://www.journalism.org/2016/07/07/loyalty-and-source-attention/>
- Wing, N. 2017. When The Media Treat White Suspects And Killers Better Than Black Victims. Retrieved March 15, 2018, from https://www.huffingtonpost.com/entry/when-the-media-treats-white-suspects-and-killers-better-than-black-victims_us_59c14adbe4b0f22c4a8cf212

Contact

aknecht@umich.edu
ritmehta@umich.edu
mavcook@umich.edu
lavici@umich.edu