Media, Race, and Crime: A Classifier

Ritam Mehta, Alex Knecht, Maverick Cook, Laura Vicinanza

I. Background and Motivation

News articles often include some form of authorial bias. More specifically, news stories covering crimes in America exhibit implicit or even explicit biases that further perpetuate a stereotype or have other harmful consequences. There are clear examples where the race of a perpetrator has significantly affected an article's headline, not to mention the content. In March 2018, Mark Conditt, a man who sent bombs that killed multiple people in Austin, was described in the New York Times as "quiet and socially awkward." Nicolas Cruz, the Parkland Shooter, was described as having been "bullied." In contrast, Black victims of crime such as Trayvon Martin or Tamir Rice are routinely torn apart in the media. While this unfair treatment is not consistent across the ideological spectrum and can differ between media outlets, it is heavily covered in online discourse. This served as the motivation for our project; we wanted to quantify the difference in reporting in some way, and see if the differences in tone and word choice were consistent for various racial or ethnic groups.

To test whether a difference exists between media portrayals of People of Color and White people, we built a classifier that takes in articles on crime and identifies, based on the text content, the race or ethnicity of the perpetrator. This was meant to demonstrate the extent of media bias- that it was so heinous and omnipresent, that one would reverse engineer it. A high accuracy of our classifier would imply that news articles are indeed consistently and similarly racially biased when reporting crime. We hope to use this project to expose this bias.

Definitions:

- ***Bias:*** prejudice in favor of or against one thing, person, or group compared with another, usually in a way considered to be unfair.
- ***Explicit Bias***: attitudes and beliefs about a person or group on a conscious level
- ***Implicit Bias:*** the attitudes or stereotypes that affect our understanding, actions, and decisions in an unconscious manner. ***Characteristics of implicit biases***: pervasive, do not necessarily reflect stances we would explicitly endorse, tend to favor our own in-group (though possible to hold implicit biases against our in-group), malleable (can be gradually unlearned through debiasing techniques)
- ***Race***: physical differences that groups and cultures consider socially significant
- ***Ethnicity/ Ethnic Group***: shared culture, such as language, ancestry, practices, and beliefs
- ***Racial Bias***: An explicit or implicit bias against someone due to their race. Manifested in unequal treatment, which can have negative long-term effects on the victims' physical and mental health

We acknowledge that an issue as complicated as racial bias in society cannot be solved by one computer program. But it can be alleviated through conscious and intentional subversion and redefinition of stereotypes. Many newsrooms and media organizations hold bias training workshops. One application of our classifier could be to identify current biased practices to newsrooms in training. Another could be to create public awareness of the issue, enough to influence the way people interpret

reporting and hold news agencies accountable for their content.

II. Related Work

Although the interplay of crime and race / ethnicity has drawn more public attention in recent years, it is a growing field of research, and studies on how media outlets portray perpetrators of different racial/ethnic groups is sparse. To the best of our knowledge, our particular quantified classifier analysis has not been done before.

The first large recent study conducted on this topic was published by the Journal of Research in Crime and Delinquency in 2002. It concerned the criminal typification of race and ethnicity in local television news stations in Orlando, Florida. The study found that Blacks and Hispanics appear in more threatening contexts than Whites, suggesting 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," using the qualitative measures of 'shown in a mug shot', 'victim as stranger', and 'victim of different race or ethnicity'. They also found that Black people were shown 2.4 times more often as crime suspects than White people (Bjornstrom, E. E. S., Kaufman, R. L., Peterson, R. D., & Slater, M. D., 2010).

Another study conducted (Dixon, T. & Linz, D.) in 2000 compared how Latino, White, and Black people were represented in TV news coverage as lawbreakers or law defenders. The study concluded that Black people are overrepresented as perpetrators. In other words, TV news coverage represented Black people as perpetrators more than disproportionally when compared to other racial groups. Whites and Latinos were found to be more underrepresented in news than arrested. In line with this study, we would expect to find some type of quantifiable racial bias in written news articles as well.

According to a paper from Duke Law, the effects of racial biases in news are more profound than one might think. "These images matter because they are a central component in a circular process by which racial and ethnic misunderstanding and antagonism are reproduced, and thus become predictable influences in the criminal-justice process". The paper discusses how continued news coverage affects juries, and the outcomes in court. Being able to quantifiably identify bias in an article, which our project aims to do, could potentially help break the vicious cycle described (Entman, Robert & A. Gross, Kimberly).

Kassia Kulaszewicz, a researcher at St. Catherine University, collected articles from national news outlets after conducting searches including racial identifiers. Kulaszewicz found that "in the event of an incident with a police officer, such as a shooting, Black men are often criminalized and represented as violent." In her analysis, she discusses how this representation of Black men in the media both creates and enforces negative racial biases, creating a feedback loop. (Kulaszewicz, Kassia E., 2015).

Finally, a 2016 research study at the University of Nebraska used a Critical race 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.

III. Methodology

1) Collection

For our corpus, we collected 222 articles for crimes perpetrated by members of three different racial/ethnic groups: White, Black, and Hispanic. These articles consist of wanted suspects or suspects being charged with crimes, that included identifying information about their race. These articles did not include textual information about the race/ethnicity of the perpetrator (which could easily skew our classifier results), rather a mugshot, which we used to tag the articles. To further limit confounding variables, we only used articles where the types of crimes included fell under the category of violent crimes. We define violent crimes as *crimes where the perpetrator intentionally or unintentionally caused, or attempted to cause physical damage or intimidation to another life.*

The articles we choose came from both national and local news outlets that report on crime. Our team manually looked up news articles on the outlets' websites, and found potential articles based on headlines that related to violent 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.

Of the 222 articles, 93 were about Black perpetrators, 90 White, and 39 Hispanic. Many of these articles were a few sentences describing the crime, and when the charged person would be seen in court.

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 notion. While this may seem problematic, our justification lay in the fact that racial bias stems not from any scientific definition of race, but from sheer visual identification. If we were identifying people based on the way they looked, so would potential members of the media. However, we would not collect articles where we were not confident our categorization. We were also subjectively applying our violent crime definition, and may have chosen articles that did not fit that definition properly, depending on who was collecting the data.

One of many interesting pieces of data in our corpus is a story of a White teenager who killed multiple people after driving drunk. The headline looked like a standard drunk driving headline, but most of the article focused on how the criminal was a great track star with only one real leg. It continued to report on his journey as a track star, instead of the crime he committed.

2) Analysis

After considering various types of classifiers, we opted for a Naive Bayes classifier to categorize documents based on the probability of the whole document being in a specific class, as well as the probability of each word relating to specific classes. We measured 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 with the leave-one-out strategy, using our full corpus minus one document, and testing our classifier with the remaining document. 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 the data even further and seeing the kinds of words that were unique in each set, we decided to add a weighting scheme to our work to prevent common words from being the strongest determinant. To address this, we decided on tf-idf, which makes less frequent terms more relevant, as

language relating to all violent crime would probably not help us. We performed all our tests with and without tf-idf.

Since Breitbart and Fox News are known to be politically polarized outlets with conservative agendas, we decided to run tests that both included and excluded these sources from the corpus. We also ran tests 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.

Finally, we extracted the top words from our binary classification scheme of White and Black perpetrators. After noticing similar top words for both classifications, we wondered what a set difference operation of the top words might reveal. Thus, we analyzed the top words from each classification using set difference operations to get the unique top-words for each race/ethnicity, which resulted in a powerful findings.

IV. Results

We ran our classifier with five variations of our original corpus, as described below.

Note: Each test was run with the leave-one-out strategy, so our classifier was trained on as much data as possible.

1. **Full corpus**

Used the full corpus consisting of 93 articles with Black perpetrators, 90 White, 39 Hispanic, for a total of 222 articles

2. **Black and White**

Used only the Black and White articles from the full corpus, for a total of 183 articles

3. **Equal sample size**

Removed Black and White articles from the corpus so that all ethnic groups had 39 articles

4. **Removal of outlet: Breitbart**

Removed all articles from the Breitbart news outlet, totaling 159 articles

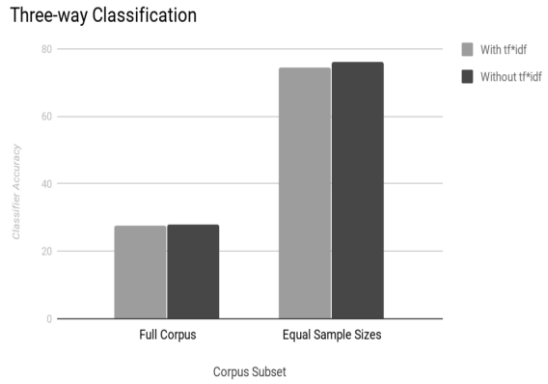
5. **Removal of outlet: Fox**

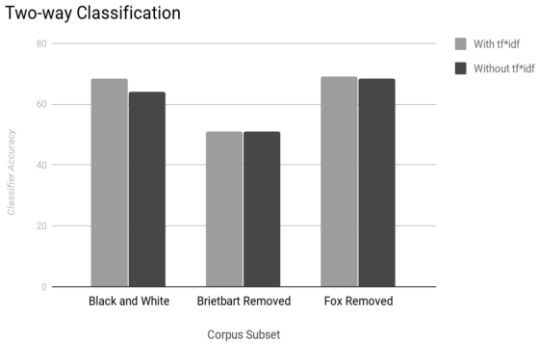
Removed all articles from the Fox news outlet, totaling 124 articles

All the tests were repeated but using an additional tf-idf weighting scheme, resulting in ten different accuracy measures.

The accuracy of these tests is determined by how many articles the classifier correctly identifies as a given ethnic group. The highest accuracy test was using the Equal Sample Size collection, which resulted in 76.1% accuracy without tf-idf weighting, while the lowest accuracy came from the Full Corpus collection, at 27.9%.

Test	*Accuracy without tf-idf*	*Accuracy with tf-idf*
Full corpus	0.279279279279	0.274774774775
Black & White	0.639344262295	0.68306010929
Equal samples	0.760683760684	0.74358974359
Brietbart removed	0.509433962264	0.509433962264
Fox removed	0.682692307692	0.692307692308





Tests one and three were used with data pertaining to all three ethnic races, so this was a three-way classification. The other three tests were conducted on only data pertaining to Black and White perpetrators, so these were two-way classifications. The

baseline for the three-way classifications is 33%, and the baseline for the two-way classification was 50%. For all the two-way classification tests, our results were above this benchmark. Only one of our tests did not meet the benchmark accuracy, which was test one. However, it was very close, only missing the baseline by a few percent.

Finally, in our analysis of the top words using the set difference operator to find the unique top 100 words for White and Black classifications, we found a stark contrast between the types of words used. We found that articles on White perpetrators were more likely to mention humanizing terms like "body" and "life' as well as familial terms like "child," "mother," "father," "son," and "wife." On the other hand, articles on Black perpetrators were more likely to mention specifics about a violent crime, such as "robbery," "killing," "shot," "assault," and "victim."

Top unique words for White perpetrators	Top unique words for Black perpetrators
house	attorney
mother	shot
hearing	dead
say	victim
city	shooting
father	chicago
convicted	hospital
deputies	officer
body	taken
life	men
child	multiple
report	killing
son	assault
daughter	trial
driving	robbery
wife	south

One concern to the quality of our analysis is that these results are based on a relatively small sample size. To get more accurate and meaningful results, we would need to scale this experiment to include a much larger corpus. This experiment is easily scalable with a few modifications. If articles are preprocessed beforehand and term frequency and inverse document frequency scores are collected before any analysis occurs, the experiment could scale to much larger input numbers. The most challenging aspect of scaling the project would be the data collection, as race/ethnicity must be verified visually.

V. Conclusions

In our analysis, we found a weak correlation between race and the language used to describe the perpetrator by news outlets. The accuracy of our classifier and its ability to (with 75% accuracy) predict the race/ethnicity of a perpetrator suggest that there indeed exists a difference in the way violent crime is depicted in the news based on the race or ethnicity of the perpetrator.

Although this finding is significant, our most telling data came not from the classifier, but rather from the term frequency differences between Black and White perpetrators. The difference in the top vocabulary used to describe Black versus White perpetrators suggests an implicit or even explicit racial bias in the way crime is talked about by media news outlets. Articles on White perpetrators were more likely to contain familial and humanizing language while articles on Black perpetrators were more likely to contain aggressive and criminal language. These stark differences, usually subtle in the articles themselves, shed light on a type of racial media bias, perpetuating negative stereotypes.

There are some changes that we would need to make for future work- the first being a change to the type of classifier that we used. Using a classifier that does not treat words independently might lead to some different results. Also, classifying by news outlet would likely produce a different outcome. Based on our results, we already discovered that some news outlets are biased towards some ethnic groups. We expect that this bias would be more evident if we sorted news articles by their outlet and then classified them.

It was also apparent that not having equal sample sizes of articles from each ethnic group impacted the accuracy of our classifier. When we ran the classifier using our full corpus, we saw an accuracy that was below the benchmark of 33%. This was because we had significantly less articles about Hispanic perpetrators (only 39, compared to 93 and 90 for Black and White perpetrators, respectively). We should also gather data from other ethnic groups to increase our sample size. Only looking at three ethnic groups may not lead to the most accurate results.

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 naturally leads to a lower accuracy, as this creates 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.

Finally, it would be interesting to analyze sentiment as well to test if there exists a different tone of voice used for different racial/ethnic groups. We have already seen that there is a difference in the words that some news outlets use to describe perpetrators of different ethnic groups, so we could expand this analysis to include sentiment as well.

VI. Team Member Contributions

Laura Vicinanza:
Active in the project's development, implementation, and analysis. More specifically, she created a word graph visual representation of the data, researched and wrote definitions, gathered 96 data sources, and researched related works in addition to playing a key role in editing.

Ritam Mehta:
Organized the group meetings and delegated tasks, came up with the idea for the project. Wrote and edited a significant amount of the first and final paper. Wrote the base classifier algorithm in python and collected 40 relevant samples from Breitbart. Formatted final paper. Presented poster.

Alex Knecht:
Edited and formatted the first paper, contributed to the methodology and conclusion of the final paper, and formatted and prepared the poster. Brainstormed ideas for statistics that we wanted to focus on. Collected relevant news articles from the Chicago SunTimes. Worked with the rest of the team to come up with a suitable classifier algorithm.

Maverick Cook:
Found related works, wrote project background and data collection section, added tf-idf to classifier algorithm, made graph of data, formatted classifier output and collected ~50 articles.

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. 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/10.1177/0022427022377286

Dixon, T. and Linz, D. (2000), Overrepresentation and underrepresentation of African Americans and Latinos as lawbreakers on television news. Journal of Communication, 50: 131-154. doi:10.1111/j.1460-2466.2000.tb02845.x

Entman, Robert & A. Gross, Kimberly. (2008). Race to Judgment: Stereotyping Media and Criminal Defendants. Law and Contemporary Problems. 71. 10.2307/27654685.

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

Kirwan Institute. Understanding Implicit Bias. (n.d.). Retrieved April 16, 2018, from http://kirwaninstitute.osu.edu/research/understanding-implicit-bias/

Kulaszewicz, Kassia E., "Racism and the Media: A Textual Analysis" (2015). Master of Social Work Clinical Research Papers. Paper 477. http://sophia.stkate.edu/msw_papers/477

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/

Perception Institute. Explicit Bias Explained. (n.d.). Retrieved April 16, 2018, from https://perception.org/research/explicit-bias/

Sweet, Lynn, et al. “Chicago Sun-Times.” Chicago Sun-Times, chicago.suntimes.com/. https://chicago.suntimes.com/

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