

NAACP Media Research Project Proposal

-- CS506 Spark! Project

Lynn Jiang, lynnj@bu.edu | Naiwen Xu, xnw1127@bu.edu | Yuhao He, hey1@bu.edu | Manyuan Lu, manyuan@bu.edu

1. Motivation / Background

NAACP, the National Association for the Advancement of Colored People, was founded on February 12, 1909. It is the nation's foremost, largest, and most widely recognized civil rights organization. The vision of NAACP, the National Association for the Advancement of Colored People is to ensure a society in which all individuals have equal rights without discrimination based on race.

Our client, NAACP Boston seeks to understand the coverage of Boston Media in covering Black people and Black communities in Boston, specifically to assess if and how coverage has changed over time in different news outlets pertaining to specific neighborhoods and sub-neighborhoods.

2. Goals / Hypothesis / Outcomes

1) Main goals:

We are going to analyze the content and sentiment of articles from mainly Boston Herald and other printed media sources (using sentiment analysis), where the race of individuals is explicitly mentioned, and also compare the difference of articles between Black people and non-Black people. In addition, we will identify the frequent keywords (e.g. place, name, etc.) in the news about Black people from major Boston media sources.

2) Hypothesis:

- a. The articles about Black people in Boston are more negative than those about non-Black people.
- b. The volume of coverage based on Black people from major Boston media sources has decreased over time.
- c. The number of coverage of Black victims of homicides is greater than that of non-Black victims.

3) Expected outcomes:

Hopefully our project will help our client, NAACP Boston, to understand the facts about the news covering Black people from major Boston media sources, such

as the trend of the volume of news related to Black people and the specific keywords which may imply discrimination, so that they can take some actions to reduce the discrimination based on race.

3. Non-goals

In this project, we will not be covering

- a. the analysis of political attitudes towards Black people in different Boston media companies
- b. the sentiment analysis of news about Black people from non-major Boston media sources
- c. the analysis of news from over 5 years ago

4. End Results

The definitions of done for this project are to

- a. generate an analyzed file about the given data for users.
- b. modify or generate a better algorithm for sentiment analysis than *Dharm Sentiment*, which analyzes the data purely based on the happiness scales of the text conveying.
- c. show changes among times for news attitude for black people
- d. draw a conclusion of the whether there is a difference of sentiments between Black people related news and non-black people related news
- e. identify the keywords (place, name, etc.) of black people related news

5. Open questions

- a. Will the attitude of news outlets shape the public's opinions about black people and their community?
- b. Is there a difference in the number of obituaries of Black individuals vs. non-White individuals?
- c. How to measure the scale for other attitudes besides happiness delivered from the context?
- d. How to identify discrimination of Black communities from other negative reports like a bushfire in Australia?
- e. How to recognize the name of a Black person, if the passage didn't specify its subject's race?

6. General Approach

1) Data collection:

- ◆ Scrape the data from Boston Globe (or more sources) website

2) Data cleaning:

- ◆ Parse the text, e.g. standardizing letter case, standardizing punctuation and special character, etc.

3) Data visualization:

- ◆ Pie chart for the percentage of negative vs. positive attitude
- ◆ Line plot for the trend (how it changes over time)
- ◆ Map for the coverage area
- ◆ Bar plot for the difference among categories

4) Data analysis:

- ◆ Sentiment analysis to identify and extract subjective information
- ◆ Word2Vec library to vectorize words and discover with what words they form associations
- ◆ Model development to predict future data, e.g. How much negative news about black people will be covered in Boston?
- ◆ ...

5) Conclusion / proper suggestions