6-2023

Topics for He but not for She: Quantifying and Classifying Gender Bias in the Media

Tyler J. Lanni
The Graduate Center, City University of New York

TOPICS FOR HE BUT NOT FOR SHE: QUANTIFYING AND CLASSIFYING GENDER BIAS IN THE MEDIA

by

TYLER LANNI

A master's thesis submitted to the Graduate Faculty in Linguistics in partial fulfillment of the requirements for the degree of Master of Arts, The City University of New York

2023

APPROVAL

Topics for He but not for She: Quantifying and Classifying Gender Bias in the Media

by

Tyler Lanni

This manuscript has been read and accepted for the Graduate Faculty in Linguistics in satisfaction of the thesis requirement for the degree of Master of Arts.

Approved: April 2023

Sarah Ita Levitan, Advisor

Cecelia Cutler, Executive Officer

THE CITY UNIVERSITY OF NEW YORK

ABSTRACT

Topics for He but not for She: Quantifying and Classifying Gender Bias in the Media

by

Tyler Lanni

Advisor: Sarah Ita Levitan

In this study, we used computational techniques to analyze the language used in news articles to describe female and male politicians. Our corpus included 370 subtexts for male candidates and 374 subtexts for female candidates, gathered through the New York Times API. We conducted two experiments: an LDA topic analysis to explore the data, and a logistic regression to classify the subtexts as either male or female. Our analysis revealed some noteworthy findings that suggest the possibility of developing a gender bias classifier in the future. However, to create a more robust understanding of bias, additional research and data are needed, as well as clearer definitions of what constitutes bias.

ACKNOWLEDGEMENTS

I would like to express my gratitude to my advisor, Sarah Ita Levitan, and Kyle Gorman for their guidance and support throughout this study. Their patience and valuable feedback helped me to refine my ideas and complete this research.

I would also like to thank the members of the Second Language Acquisition Lab for the opportunity to collaborate with them on their fascinating projects. Despite the challenging circumstances we faced, it was a meaningful experience that provided a sense of community and connection.

Finally, I would like to express my gratitude to my friends and family for their tremendous support and encouragement throughout this journey. Whether it was meeting up in a cafe or catching up on a phone call, your presence and encouragement helped me stay motivated throughout the process.

Contents

List of Tables

List of Figures

Introduction

Gender bias in the media refers to the unequal and stereotypical representation of individuals based on their gender. This can manifest in various forms, including selective reporting, language usage, and presentation of information. Gender bias in the media can occur both intentionally and unintentionally, with varying levels of severity.

The issue of gender bias in media is pervasive and affects all forms of media, including news and entertainment. From sports to politics, bias can be observed in various areas of reporting, and it is prevalent across countries and organizations. Gender stereotypes are often reinforced by biased reporting, and the issue is not limited to news alone. Even entertainment media can unconsciously perpetuate gender bias, further perpetuating harmful stereotypes.

Several studies have applied computational methods to measure various forms of gender bias in the media. For example, Shor et al. (2019) found a lack of representation and equal coverage for women in similar positions and success compared to their male counterparts. Asr et al. (2021) found that only 29% of quotes in news articles over a two-year period came from women, compared to 71% from men, demonstrating a significant disparity in representation. Other research, such as Fu et al. (2016), has shown how gender-based language can unconsciously reinforce stereotypes and result in misrepresentation of women. These studies provide evidence of different types of gender bias in the media and offer a starting point for more focused analyses.

This paper features an analysis of gender bias in political articles that feature political candidates, using computational techniques. The study will explore the presence of bias and, if present, describe its nature. The examination of various forms of bias in media and how they arise will be used as a backdrop to understand the significance of this study.

Previous Works

Shor et al. (2019) and Asr et al. (2021) both examine gender-based disparities in media coverage. Shor et al. (2019) studied the media coverage of individuals compared to their level of public interest, as measured by Wikipedia hits. The results indicated that "women receive less media coverage compared to men with similar achievements or positions", and the disparity was statistically significant with women receiving only 80% of the coverage received by men. Asr et al. (2021) examined representation in Canadian media by analyzing the ratio of quotes from women to men in articles. The study found that men are overwhelmingly more quoted than women, with 62,564 quotes from men compared to 19,173 from women, which the authors attribute to a concentration of power in the upper echelons of each profession.

These two studies shed light on different facets of gender bias in the media: one focuses on the media's tendency to report on men over women, while the other highlights the media's preference for using men as sources for their reporting.

The media not only exhibits coverage discrepancies but also reinforces traditional gender stereotypes through language. Mañoso Pacheco (2018) argues that the language used in the media, including the choice of sources and portrayal of women, perpetuates long-standing stereotypes for women. Women have traditionally been described by their "irrationality, familial dependence, powerlessness, and sexual and physical excess," and their job functions were related to specific domestic situations, such as housewives and mothers (Fowler 1994, as cited in Mañoso Pacheco 2018). In contrast, men were more likely to be identified by positions of power in society, "outside the home and family" (Fowler 1994: 102, as cited in Mañoso Pacheco 2018). The media's perpetuation of gender roles further contributes to the underrepresentation of women in society, which limits them to specific roles and characteristics. As a result, readers may unconsciously assign gender stereotypes to sources of information. For example, even without explicit mention of their gender, readers might unconsciously assume a police officer’s gender if they are quoted in an article (Armstrong and Nelson 2005, as cited in Mañoso Pacheco 2018).

Stereotypes persist in various areas, including sports, politics, and beyond. Researchers, such as Fu et al. (2016) and McLoughlin (2021), have focused on gender and race biases in sports. Fu et al. (2016) analyzed post-match interview questions in tennis and found that women athletes were often asked questions that did not relate to their performance in the sport, while men were asked more performance-related questions. McLoughlin (2021) studied racial stereotypes in European football media coverage and found that darker-skinned athletes were more likely to be praised for their physical attributes while having their intelligence criticized, while players with lighter skin were more likely to be praised for their intelligence. This phenomenon is not restricted to European sports, as Hawkins (2002) discusses the practice of stacking minority players into less important positions, which created perceptions about which positions certain races could play. The quarterback position in American football is a notable example, where many people believe that black athletes are inferior to white athletes. Bigler and Jeffries (2008) argue that "NFL draft experts consistently rate African American quarterbacks higher than whites in the areas of physical abilities and lower in the areas of cognitive abilities, thus perpetuating racial stereotypes of African

Americans." Within American sports we also hear the stereotype that white athletes were forced to train hard for their abilities, whereas black athletes are naturally gifted. These studies illustrate how language is used to perpetuate stereotypes and emphasize the need for more diverse and inclusive representation in the media.

Studies on media bias highlight discrepancies that demand discussion. It's crucial to have open conversations about the causes and effects of bias, as it can influence how readers perceive and understand the world around them, consciously or unconsciously. This paper aims to initiate a discussion on how media agencies portray political candidates based on their gender. Specifically, we will analyze The New York Times and investigate if there is a difference in the language used to represent male and female political candidates. Our analysis will focus on identifying which topics are reported more frequently for each gender, the distinguishing words and phrases used, and whether there is evidence of bias. Additionally, we will explore whether a neural network can accurately classify articles based on the subject's gender. Through our analysis, we hope to shed light on the ways in which media outlets can perpetuate gender biases in their coverage of political candidates.

Data

To gather data for the analysis, we retrieved articles from the New York Times (NYT) website. We opted for this news agency for no reason other than their API[1] access. The NYT journal is meant to be a launching point for this study and not indicative of all journals. First, we created a list of candidates for different political positions, attempting to evenly split genders and political parties. We selected articles for examination if the name of a political candidate on our list appeared in the title. Since many articles contained multiple individuals, we created subtexts that included approximately 200 characters before and after the candidate's name. These subtexts were used for analysis, resulting in a corpus of 370 subtexts for male candidates and 374 subtexts for female candidates.

To ensure that the subtexts were specifically focused on the targeted candidates' gender and not in reference to them alongside another gender, we conducted a thorough analysis of the remaining subtexts. This analysis was essential to confirm that the language used in the subtexts exclusively referred to the target candidate and did not reference another entity, particularly someone of a different gender.

Table 1: Sample of Subtexts Before Preprocessing

Gender	Value	Subtext
Woman	0	ebate was a blowout, surely the most one-sided confrontation in American political history. Hillary Clinton was knowledgeable, unflappable and — dare we say it? — likable.

[1] https://developer.nytimes.com/apis

Woman	0	ague of United Latin American Citizens forum and then a campaign town hall in Las Vegas. Though Ms. Klobuchar spent months tailoring her presidential bid to Iowa caucusgoers — a rural Midwestern neighbor whose
Woman	0	heart out for everyone in Iowa and across the country.” In its editorial, the Register praised Ms. Warren’s approach to the economy, health care, climate change and other issues. “She says corporations sho
Male	1	n Jr. for president, while those who named the economy and jobs leaned toward re-electing President Trump. Reflecting a pervasive pessimism, nearly two-thirds of voters said they believed the country was h
Male	1	on Tuesday, when primaries were held in Florida, Illinois and Arizona. At Sunday’s fund-raiser, Mr. Biden said that a recreation room in his home in Delaware had been turned into a television studio, and t
Male	1	things, shell games that are played — we need to get rid of all that stuff.” Born into poverty, Mr. Carson was awarded a scholarship to Yale, and by age 33 he was named director of pediatric neurosurgery at

Preprocessing

The subtexts underwent preprocessing using scripts to eliminate stopwords, gendered language, and any human entities unrelated to the targeted candidate. Pronouns such as "he" or "she" were removed to avoid redundant classification. However, the words "man" and "woman" were retained to potentially analyze which gender is more likely to be associated with their gender. The subtexts were stored with their respective gender classification for the purpose of classification. Each subtext was tagged as either 1 or 0, representing male or female, respectively, for both experiments.

Methodology

In both experiments, we utilized the same corpus. Our pre-processing step involved removing human entities other than the focused candidate using Named Entity Recognition (NER) in Python's Spacy library (Honnibal & Montani 2017), as well as removing stopwords and applying tokenization using the Natural Language Toolkit, more widely known as NLTK (Bird et al. 2009). For the LDA topic analysis, we represented each article as a Bag of Words (BoW) representation using Gensim (Rehurek & Sojka 2011). Meanwhile, for the logistic regression experiment, we represented articles using Term Frequency-Inverse Document Frequency (TF-IDF) with SKLearn (Pedregosa et al. 2011).

LDA topic analysis

Latent Dirichlet Allocation (LDA) is a statistical model commonly used in natural language processing to uncover underlying topics in a text corpus. It accomplishes this by representing each document as a mixture of topics, with each topic represented as a distribution of words. The objective of this study is to identify the topics that occur across the corpora of both genders.

As an illustrative example, applying LDA to these corpora would reveal the prominent topics that characterize each group. Specifically, LDA may uncover within-group topics that reflect gender-specific interests, such as "suits," "ties," and "cuffs" within the male corpus and "dresses," "scarves," and "lipstick" within the female corpus.

For the LDA topic analysis, we first created a bag of words (BoW) representation for our corpus. To compare common topics for women and men respectively, we split the corpus into two groups before feeding it into the LDA models.

Table 2: LDA parameters

Number of topics	10
Random seed	100
Chunksize	100
Passes	300

Logistic Regression Binary Classifier

In this experiment, we developed a binary classifier using logistic regression to predict whether an article is about a man or woman. To train this model, we utilized the labels assigned during the pre-processing stage. We fed the same corpus of articles that we used in the LDA experiment to the model, but with each article's corresponding label (0 or 1 for man or woman, respectively) appended to it. Unlike in the LDA experiment, where we used a Bag of Words (BoW) representation, we opted for a Term Frequency-Inverse Document Frequency (TF-IDF) vector representation in this experiment. TF-IDF allows us to better understand each document by measuring the importance of each word within it relative to its frequency in other documents.

We divided our data into three sets: training, validation, and testing, using a 60/20/20 split. Each data point in our set consisted of a subtext that we had converted into a TF-IDF vector, and a binary value indicating the gender to which it referred. In total, we used 3,815 features, representing the unique words present in the subtexts. By applying TF-IDF, we assigned greater importance to the most relevant tokens for each subtext, reducing the computational burden of the model.

Results

To identify any potential biases in the language, we employ two different model types for data analysis. Our LDA topic model uncovers common topics within subtexts for both genders, while our binary classifier determines if any discernible differences exist within the subtexts that enable gender classification.

Experiment 1: LDA Topic Model

We present the results of our Topic Model in Tables 3 and 4, which show the top 10 words for each topic for both genders. These words were used to identify the themes and topics discussed in the subtexts related to each candidate.

Table 3: LDA Topics for Women

Topic #	Top Words Per Topic
1	Campaign, people, party, million, quick, care, health, year, women, president
2	Campaign, new, hampshire, senator, would, voters, state, show, little, candidates
3	New, president, campaign, senator, debate, party, candidate, woman, hampshire, voters
4	Campaign, would, former, woman, asked, senator, think, win, huge, executive
5	Would, race, campaign, woman, presidential, debate, election, court, president, primary
6	People, state, campaign, candidate, open, black, women, million, debate, backing
7	Would, party, past, last, presidential, election, race, time, campaign, attacks
8	Woman, would, campaign, senator, position,

	also, political, voters, message, last
9	Would, election, state, states, people, saying, including, senator, department, one
10	Political, senator, campaign, made, presidential, days, make, high, ground, seen

Table 4: LDA Topics for Men

Topic #	Top Words Per Topic
1	Campaign, senator, party, back, hampshire, candidate, mayor, could, debate, former
2	Campaign, support, poll, delegates, even, last, voters, appealing, online, aides
3	Senator, race, campaign, million, last, according, could, presidential, vote, return
4	Race, campaign, bid, called, week, people, president, gay, added, group
5	Campaign, president, former, would, new, year, another, mayor, pete, jr
6	Senator, would, city, campaign, president, candidates, former, term, national, million
7	Campaign, race, seemed, man, one, new, would, way, often, home
8	Debate, senator, city, told, campaign, friends, attack, party, terrorist, take
9	Campaign, state, would, publicly, billionaire, enough, black, home, housing, get
10	Like, told, candidates, campaign, time, choice, carried, preferred, acutely, aware

We sought to identify the most significant words in the subtext collection by examining the weights assigned to each word by the model during topic modeling. Specifically, we reviewed the top 100 words in each of the ten topics and calculated their weights across all topics. Based on this analysis, we compiled a list of the most influential words for men and women, which is presented below:

Table 5: Word Weights of Top 50 Terms across all Topics for both Genders

Top 50 most weighted words for Women	Top 50 most weighted words for Men
1. campaign: 0.086 2. senator: 0.053 3. would: 0.052 4. new: 0.04 5. woman: 0.038 6. president: 0.037 7. party: 0.035 8. candidate: 0.032 9. people: 0.032 10. presidential: 0.031 11. debate: 0.03 12. voters: 0.025 13. last: 0.024 14. state: 0.024 15. race: 0.023 16. one: 0.022 17. also: 0.02 18. million: 0.02 19. election: 0.02 20. candidates: 0.019 21. women: 0.019 22. political: 0.018 23. hampshire: 0.018 24. could: 0.017 25. national: 0.015 26. win: 0.015 27. even: 0.014 28. time: 0.014 29. like: 0.013 30. states: 0.013 31. another: 0.012 32. black: 0.012	1. campaign: 0.118 2. senator: 0.056 3. new: 0.04 4. race: 0.04 5. would: 0.039 6. former: 0.036 7. president: 0.033 8. debate: 0.032 9. state: 0.03 10. could: 0.028 11. presidential: 0.026 12. voters: 0.026 13. city: 0.025 14. party: 0.024 15. like: 0.023 16. last: 0.023 17. time: 0.021 18. hampshire: 0.021 19. million: 0.02 20. mayor: 0.019 21. candidates: 0.018 22. get: 0.018 23. people: 0.018 24. support: 0.018 25. vote: 0.016 26. also: 0.016 27. candidate: 0.015 28. south: 0.015 29. though: 0.014 30. back: 0.014 31. many: 0.014 32. supporters: 0.014

33. called: 0.012 34. supporters: 0.012 35. fund: 0.012 36. former: 0.012 37. primary: 0.012 38. think: 0.012 39. might: 0.011 40. democratic: 0.011 41. back: 0.011 42. vice: 0.011 43. still: 0.01 44. saying: 0.01 45. including: 0.01 46. year: 0.01 47. stage: 0.01 48. many: 0.009 49. yet: 0.009 50. since: 0.009	33. percent: 0.013 34. political: 0.013 35. still: 0.012 36. told: 0.012 37. day: 0.012 38. delegates: 0.011 39. housing: 0.011 40. week: 0.011 41. country: 0.011 42. office: 0.011 43. bend: 0.011 44. another: 0.011 45. delegate: 0.011 46. saying: 0.01 47. rivals: 0.01 48. contest: 0.01 49. friends: 0.01 50. often: 0.01

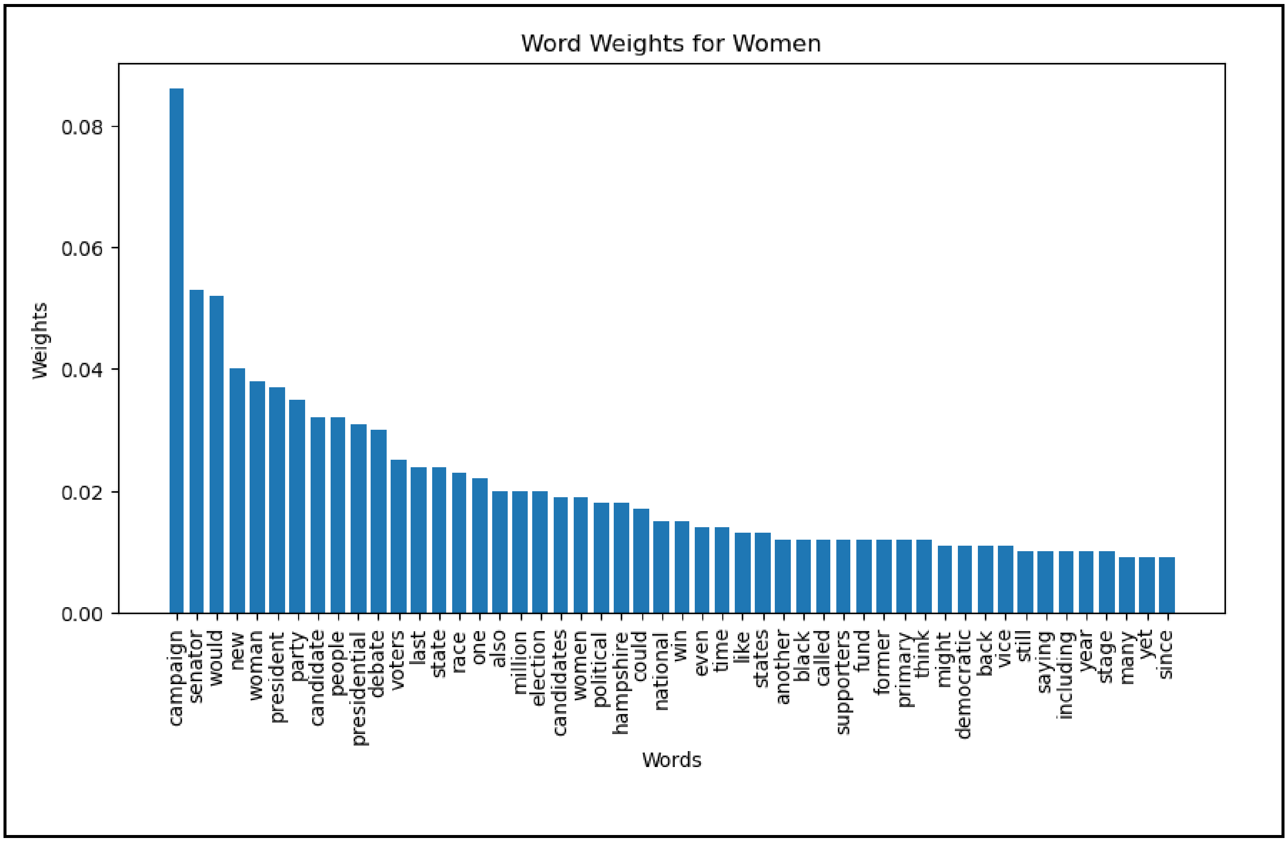

Figure 1: Top Weighted Words for Women across all Topics

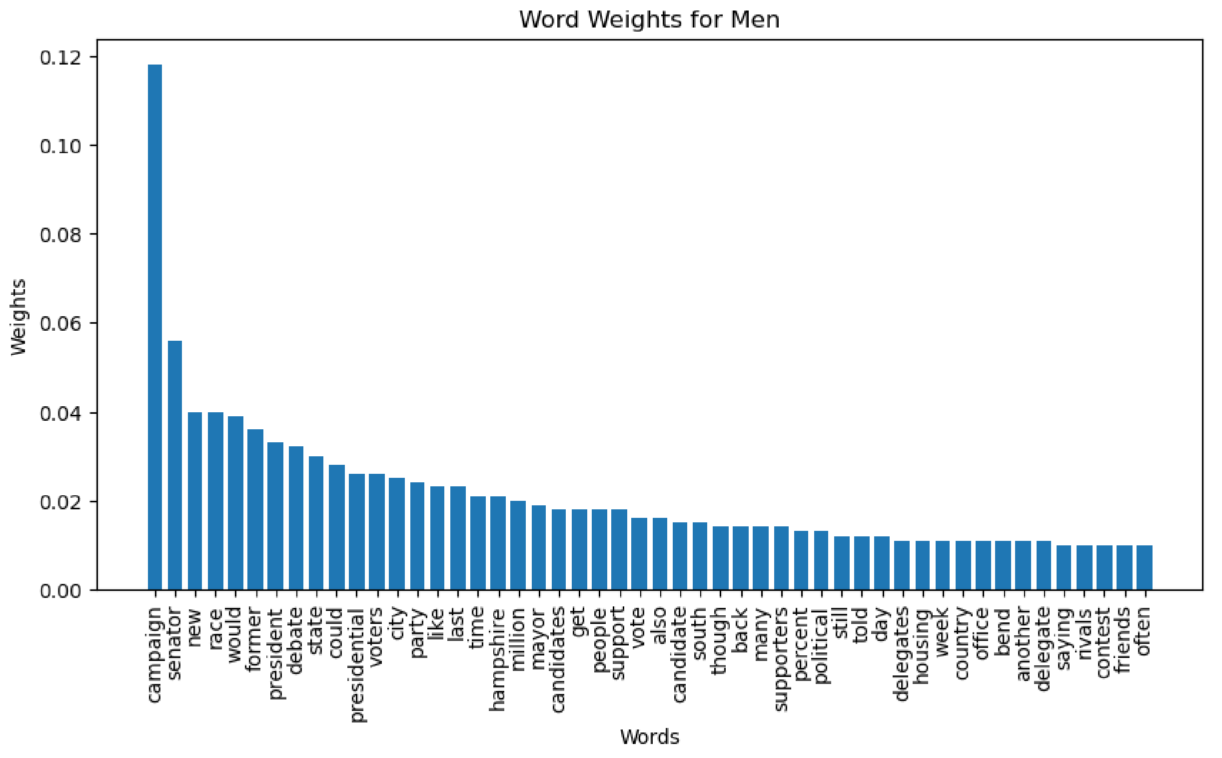

Figure 2: Top Weighted Words for Men across all Topics

We also conducted an analysis of the model's topics to identify the unique words that were heavily weighted in subtexts related to women and men. First, we generated two separate sets of words that only occurred in the subtexts related to women and men. Then, for each topic, we searched for these unique words within the corresponding gender's model. By examining the top 100 words for each topic, we determined the weight assigned by the model to each unique word across all topics. The resulting list presents the unique words and their respective weights, which are aggregated across all topics:

Table 6: Top Weighted Unique Terms for both Genders

Top 50 most weighted unique words for	Top 50 most weighted unique words for Men

Women	
1. grooming: 0.007 2. care: 0.007 3. pushed: 0.006 4. companies: 0.005 5. billion: 0.005 6. advised: 0.005 7. ground: 0.005 8. pre: 0.005 9. suggests: 0.005 10. hearing: 0.005 11. incumbent: 0.005 12. cartoon: 0.005 13. neither: 0.004 14. keeping: 0.004 15. fighter: 0.004 16. approval: 0.004 17. saw: 0.004 18. deport: 0.004 19. details: 0.004 20. model: 0.004 21. corruption: 0.004 22. education: 0.004 23. mate: 0.004 24. board: 0.004 25. legislators: 0.004 26. closest: 0.003 27. free: 0.003 28. appeared: 0.003 29. dismissed: 0.003 30. prohibits: 0.003 31. career: 0.003 32. endorsing: 0.003 33. meredith: 0.003 34. employee: 0.003 35. extend: 0.003 36. spoiler: 0.003 37. script: 0.003 38. foundation: 0.003 39. favored: 0.003 40. table: 0.003 41. great: 0.003 42. dislike: 0.003 43. conservatives: 0.003	1. delegates: 0.011 2. housing: 0.011 3. agency: 0.007 4. solutions: 0.006 5. side: 0.006 6. self: 0.006 7. rival: 0.006 8. millions: 0.005 9. appealing: 0.005 10. fewer: 0.004 11. precincts: 0.004 12. head: 0.004 13. virtual: 0.004 14. campaigning: 0.004 15. seeks: 0.004 16. cutting: 0.004 17. redlining: 0.003 18. businessman: 0.003 19. dog: 0.003 20. rubio: 0.003 21. de: 0.003 22. leader: 0.003 23. misogynistic: 0.003 24. channels: 0.003 25. slurs: 0.003 26. pig: 0.003 27. sensible: 0.003 28. stomach: 0.003 29. reporting: 0.003 30. withdrew: 0.003 31. urban: 0.003 32. supposed: 0.003 33. everything: 0.003 34. combined: 0.003 35. mostly: 0.003 36. worst: 0.003 37. cities: 0.003 38. leveling: 0.003 39. promises: 0.003 40. based: 0.003 41. moments: 0.003 42. schwartz: 0.003 43. doug: 0.003

44. stance: 0.003 45. leaning: 0.003 46. gabbard: 0.003 47. rich: 0.003 48. estate: 0.003 49. expressed: 0.003 50. grandmother: 0.003	44. play: 0.003 45. intellect: 0.003 46. chose: 0.003 47. charisma: 0.003 48. nature: 0.003 49. worried: 0.003 50. golf: 0.003

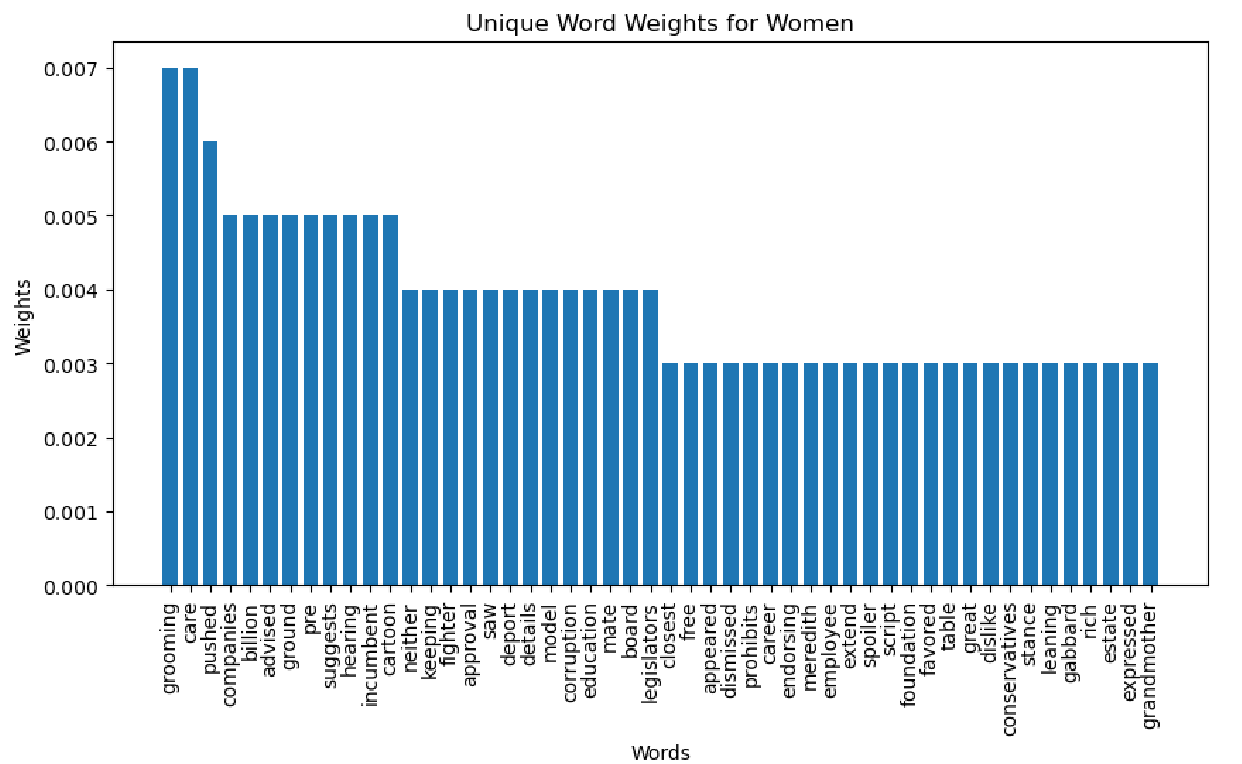

Figure 3: Top Weighted Unique Terms for Women

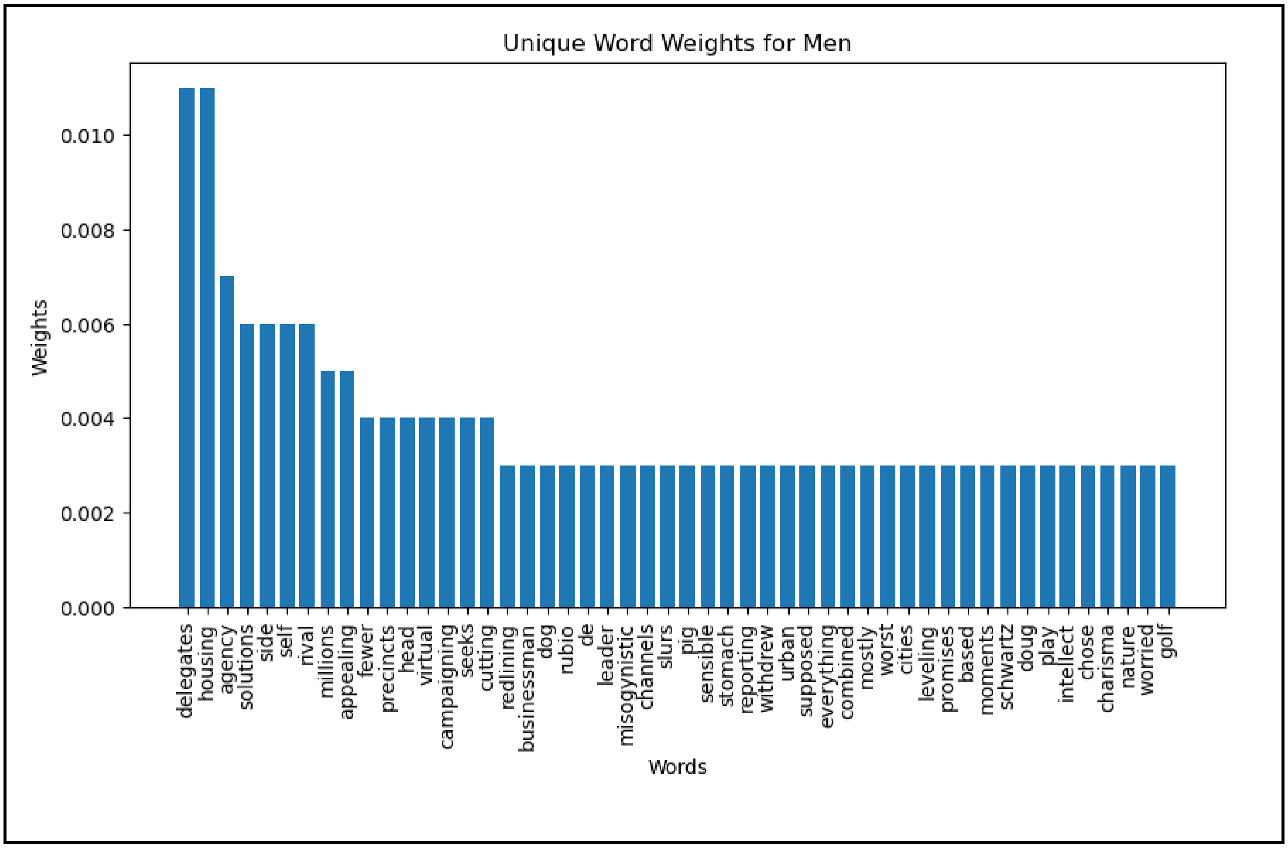

Figure 4: Top Weighted Unique Terms for Men

Experiment 2: Logistic Regression

The logistic regression model was used to predict the gender (man or woman) of an article based on the language used within it. The goal of this experiment was to investigate whether an article's language could reveal any gender bias. Figures 5 and 6 display the weights of the trained model used to classify a subtext as either about a man or a woman. Specifically, Figure 5 presents the most important features, in our case words, associated with subtexts about men, while Figure 6 shows the least important features associated with men.

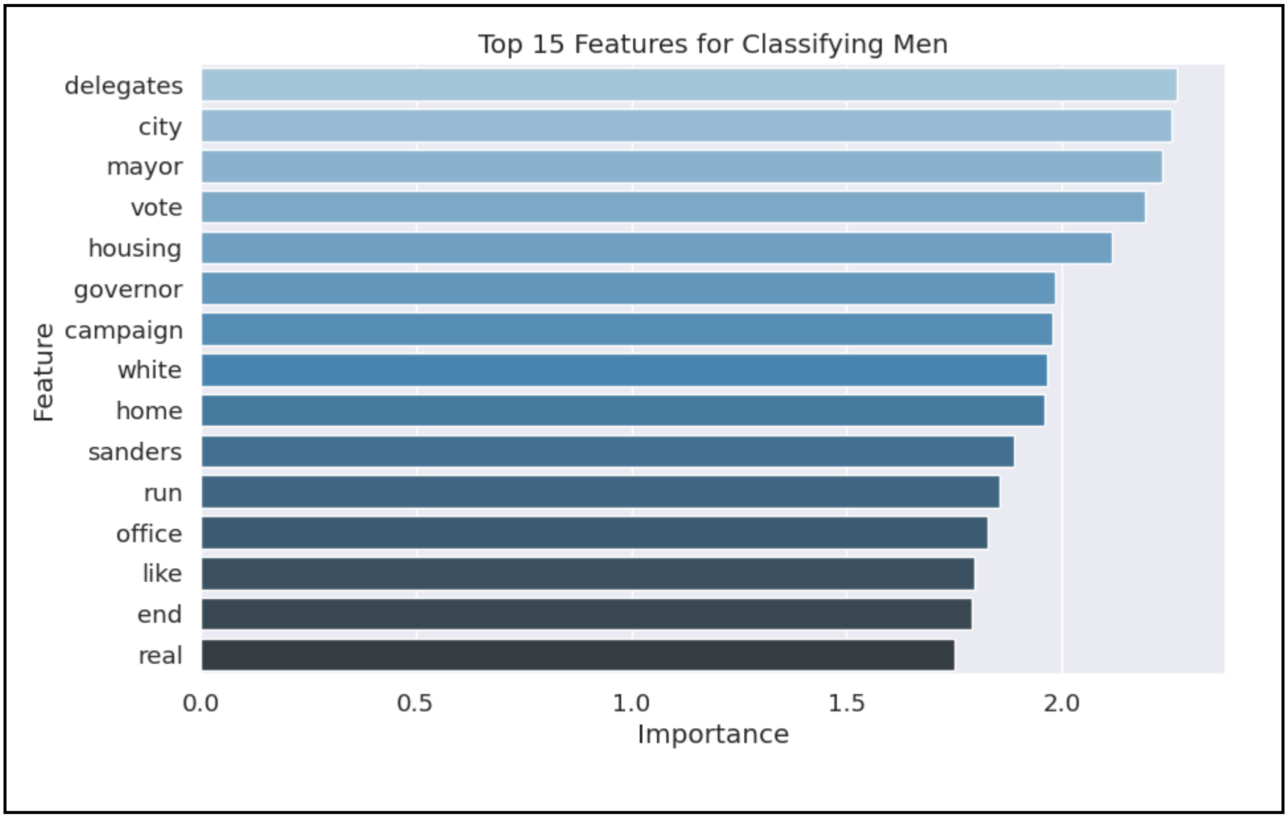

Figure 5: Top Weighted Features

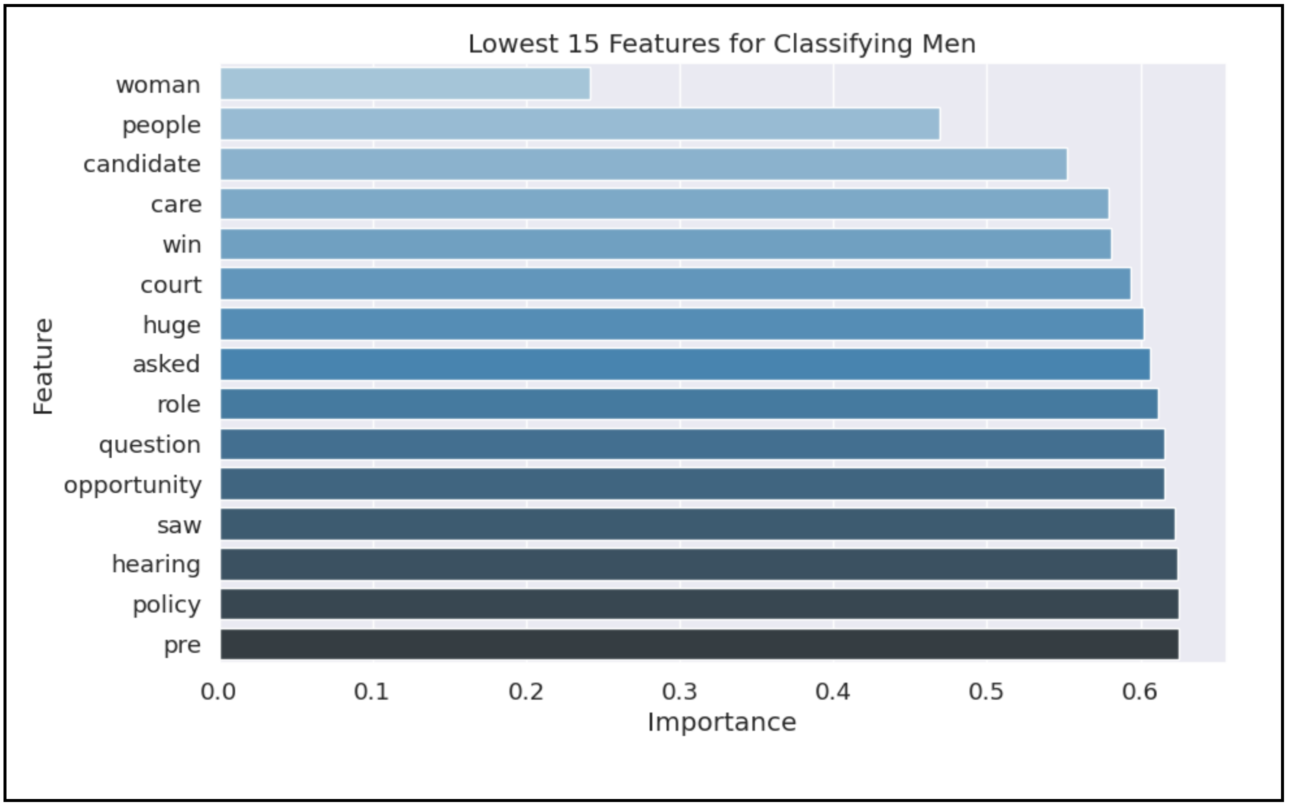

Figure 6: Lowest Weighted Features

Our model's performance scores are as follows:

Accuracy: Our model's accuracy score is 0.6912, meaning that out of all the predictions made, 69.12% of them were correct. Accuracy measures the proportion of correct predictions to the total number of predictions.

Precision: Our model's precision score is 0.7000, indicating that out of all the predictions it made for men, 70.00% of them were actually about men. Precision measures the proportion of true positives (correctly predicted articles about men or women) out of all the positive predictions made by the model.

Recall: Our model's recall score is 0.6621. Recall measures the proportion of true positives (correctly predicted articles about men) out of all the actual positive cases in the data. In other words, it is the ratio of the number of true positives to the sum of true positives and false negatives. In our model's case, recall represents its ability to correctly label all subtexts that are actually about men as such, out of all the subtexts that are actually about men and those that are misclassified as about women (false negatives).

Therefore, the recall score of 0.6621 indicates that our model correctly identified 66.21% of the subtexts that are actually about men, but it missed 33.79% of them (false negatives).

To evaluate the performance of our logistic regression model and gain insights into its decision-making process, we utilized a confusion matrix. This matrix provides a visual representation of the classification results of our model for the two target classes, man or woman.

The confusion matrix highlights the model's performance for each class. The model had a higher tendency to classify subtexts as belonging to the female category, with 79 subtexts predicted to be about women compared to 70 predicted to be about men. The accuracy of the model in predicting subtexts about women was 68.35%, while it was 70.00% for predicting subtexts about men.

		Predicted Values	
		Negative	Positive
True Values	Negative	True Negative (TN)	False Positive (FP)
	Positive	False Negative (FN)	True Positive (TP)

$$Accuracy = \frac{True\ Positive + True\ Negative}{All\ Predictions}$$

$$Precision = \frac{True\ Positive}{True\ Positive + False\ Positive}$$

$$Recall = \frac{True\ Positive}{True\ Positive + False\ Negative}$$

Figure 7: Explanation of Confusion Matrix Values

		Predicted Values	
		Woman	Man
True Values	Woman	True Woman 54	False Man 21
	Man	False Woman 25	True Man 49

Figure 8: Confusion Matrix of Experiment 2 Results

Despite being trained on an equal number of subtexts about men and women, the model demonstrated a bias towards predicting that a given subtext was about a woman. While the accuracy of the model's gender predictions was fairly balanced overall, it exhibited a slightly stronger tendency to predict male subtexts correctly by 1.65%. Nevertheless, the model made more false predictions for women, but it also made more true predictions for women as well. This suggests that the model had a tendency to hesitate to classify a subtext as positive or, in our context, about a man.

Discussion

This study aimed to explore whether gender bias can be detected and quantified in news articles covering political candidates, using machine learning techniques applied to a dataset of New York Times articles featuring both men and women. While the results were inconclusive regarding the presence of bias, they provide an important starting point for the discussion of gender and reporting on political candidates.

Experiment 1 shed light on the words that were most commonly attributed to each gender in the dataset. Tables 3 and 4 displayed the most influential words used by the model when creating topics for men and women, revealing some interesting findings. The most weighted words for men did not appear to have outliers when considering the political context of news articles, whereas women were more likely to be associated with words that could be perceived as perplexing or limiting, such as words directly related to their gender or the word 'black', which was a heavy determining factor for topic creation. This could suggest a subtle form of bias in the language used to describe female candidates, as they are often associated with their gender. However, in our data, it appears to be coincidental reporting, particularly in the case of Kamala Harris being the first Black Vice President.

Table 6 also provided insight into the association of words and gender in the dataset. The words found in this table were unique to either women or men in subtexts and highlighted the correlation between gender and certain words. Men were uniquely associated with words such as misogynistic, pig, dog, slurs, intellect, charisma, leader, and golf. These words were not typical of political language and were solely used in subtexts about men. Women, on the other hand, were uniquely associated with the words grandmother, care, rich, and fighter. While some of the unique words used to describe political candidates may be coincidental, a study by Van der Pas & Aaldering (2020) notes a recurring theme in which women are often associated with characteristics such as sensitivity, honesty, gentleness, and compassion, while men are attributed with being objective, competitive, strong, tough, intelligent, and ambitious. These findings suggest that gender stereotypes and societal expectations may influence the language used to describe political candidates within the subtexts.

Previous research has established that certain words associated with gender can reinforce stereotypes (Wood 1994; Mañoso Pacheco 2018; Asr et al. 2021). The prevalence of such associations warrants further discussion about what constitutes bias and how it can impact readers. As noted earlier, such associations can unconsciously affect readers' perceptions of the subjects of articles. The issue of bias and stereotype reinforcement is complex and requires further investigation. While it is not common for news networks to display overt bias towards a gender, unconscious bias can manifest through compliments, criticisms, or general remarks.

Fu et al. (2016) define perplexity as a metric that measures the level of surprise of a trained language model when encountering a word based on the context of a given text (an interviewer asking about boyfriends in a post-match tennis interview). In our study, we use perplexity as a means to identify statements or words that seem out of place within the context of politics, in simpler terms. The perplexing language shown in figure 9 prompts consideration of how to approach instances of potential bias. It is not clear from this study whether the detected bias is the result of circumstantial reporting or of stereotypes being reinforced. Criticizing the New York

Times for reporting on a candidate's gender or other circumstantial information, such as being the first woman to run for a position, would be unreasonable. While such information could be distracting or harmful, it is still a factual component of the story. However, it can be argued that both conscious and unconscious bias are the responsibility of the media.

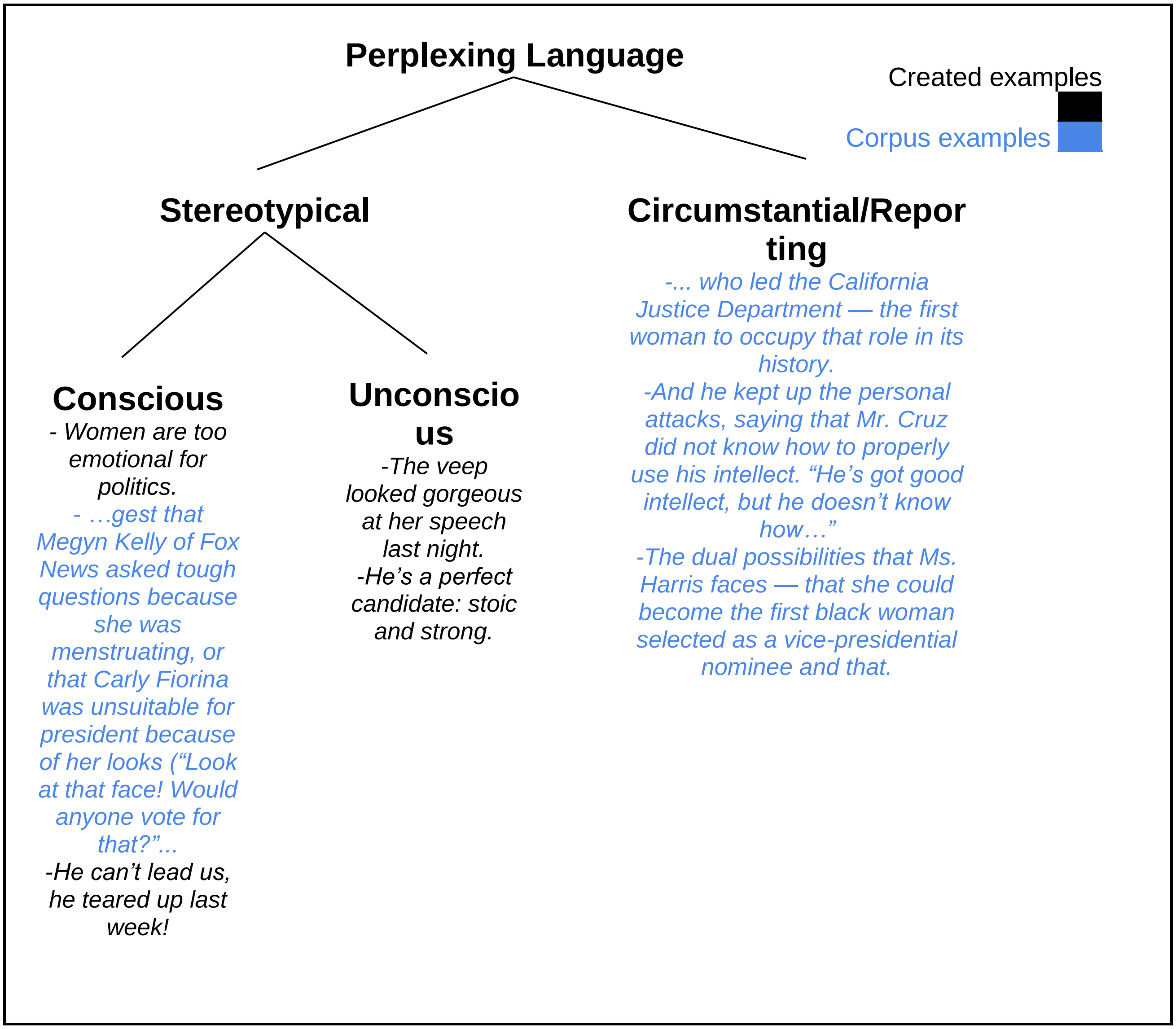

Figure 9: Differentiating Bias

Our attempt to create a binary classifier to determine whether subtexts are about a man or woman was minimally successful. Although our accuracy scores were not ideal, the fact that our test pool had a nearly equal distribution of men and women and our model still achieved over 50% accuracy, precision, and recall suggests that some measurable bias may exist. Interestingly, the model was slightly more accurate at predicting texts about men, which could shed light on how the New York Times discusses male candidates, or vice versa.

Previous studies have highlighted the differences in how men and women are portrayed in the media, with stereotypes being reinforced through the association of certain character traits and policies with each gender (Van der Pas & Aaldering 2020). For example, women are often associated with traits such as sensitivity, honesty, and caring, while men are characterized as competitive, intelligent, and strong. Women also tend to receive more coverage when they introduce policies related to education and healthcare. These findings are consistent with our Experiment 1 results and lend support to the idea that gender bias exists in media coverage of political candidates.

Table 6 displays the top weighted unique words for both men and women. The word “care” had the highest weight for women, along with “education” being another unique word for women. This aligns with what Van der Pas & Aaldering (2020) found, as they stated that women receive more coverage when they introduce policies regarding education and healthcare. For instance, a subtext from the corpus refers to Cecile Richards, the president of Planned Parenthood, introducing Mrs. Clinton as a “fighter” for health care and education programs. In contrast, men were uniquely associated with the words “sensible”, “intellect”, and “charisma”. An example from the corpus for men reads, "…before questioning the ethics of such a move. And he kept up the personal attacks, saying that Mr. Cruz did not know how to properly use his intellect. 'He's got good intellect, but he doesn't know how to…". A noteworthy avenue of research would be to examine the number of female versus male candidates making education and healthcare a cornerstone of their candidacy, and compare the amount of media coverage each gender receives on those topics. In our study, however, these words are unique to each gender.

These examples demonstrate the challenge in distinguishing between instances where the media is reporting on biased or stereotypical remarks made by others versus creating them themselves. Some instances of circumstantial reporting may involve acknowledging that a candidate is the first black woman to run for a position, while others may involve reporting on another candidate displaying bias, which can further perpetuate stereotypes.

Future Work

To achieve a more precise delimitation of the different types and severity of bias, the inclusion of more data is required. This paper and the data gathered provides an effective starting point, but it is not conclusive of the media as a whole. Before a bias classifier can be created with reasonable success, several issues need to be addressed. As we observed in this study, there are blurred lines between the various levels of bias and misrepresentation.

Moreover, increased data will address the limitation of few politicians representing the many. Similar to the problem encountered in Fu et al. (2016), our corpus focused on a select few candidates due to unequal reporting on candidates by the journal. This inequality was even more skewed for women. In fact, the list of candidates had to be extended over positions and years to find a somewhat fair representation of men and women.

To further advance research on identifying gender bias in political reporting, a larger corpus is necessary that includes data from multiple organizations. This study was limited by the inclusion of only the New York Times. Therefore, obtaining more data from a broader range of sources is necessary to gain a more comprehensive understanding of how media bias is perpetuated.

Overall, our results suggest that machine learning can be used to detect and quantify gender bias in media coverage of political candidates. However, further research is needed to explore the nuances of gender bias and the extent to which media outlets contribute to its perpetuation. Our study achieved reasonable success but requires more data and attention to differentiate circumstantial reporting from bias.

Conclusion

Our study, which used machine learning techniques, provides evidence that gender bias exists in the New York Times' coverage of political candidates. While the results from our binary classifier were only minimally successful, they do suggest the presence of some kind of measurable bias. The LDA experiment also revealed gendered stereotypes that are consistent with previous research on media coverage of men and women in politics. Further research is needed to better understand the nuances of this bias and the underlying factors contributing to it.

Our study also highlights the potential for machine learning to assist in identifying and quantifying gender bias in media coverage. While our results were not perfect, they demonstrate that technology can be used to analyze and detect gender bias in media. Future studies can expand upon our work by developing a classifier that can differentiate between conscious, unconscious, and circumstantial bias.

Overall, this study adds to the ongoing conversation about media bias and its effects on society. By providing evidence of gender bias in media coverage of political candidates, we hope to encourage further research and discussions on this important topic.

References

Asr FT, Mazraeh M, Lopes A, Gautam V, Gonzales J, Rao P, et al. (2021) The Gender Gap Tracker: Using Natural Language Processing to measure gender bias in media. PLoS ONE 16(1): e0245533. https://doi.org/10.1371/journal.pone.0245533

Bigler, M., & Jeffries, J. L. (2008). "An Amazing Specimen": NFL Draft Experts' Evaluations of Black Quarterbacks. Journal of African American Studies, 12(2), 120-141. http://ezproxy.gc.cuny.edu/login?url=https://www.proquest.com/scholarly-journals/amazing-specimen-nfl-draft-experts-evaluations/docview/228062722/se-2

Bird, S., Klein, E., & Loper, E. (2009). Natural language processing with Python: analyzing text with the natural language toolkit. " O'Reilly Media, Inc."

Fu, L., Danescu-Niculescu-Mizil, C., & Lee, L. (2016). Tie-breaker: Using language models to quantify gender bias in sports journalism. arXiv preprint arXiv:1607.03895.

Hawkins, B. (2002). Is stacking dead? A case study of the stacking hypothesis at a Southeastern Conference (SEC) football program. International Sports Journal, 6(2), 146.

Honnibal, M., & Montani, I. (2017). spaCy 2: Natural language understanding with Bloom embeddings, convolutional neural networks and incremental parsing.

Mañoso Pacheco, L. (2018). Gender asymmetries in news reports. Miscelánea, 57, 121-139. Retrieved from http://ezproxy.gc.cuny.edu/login?url=https://www.proquest.com/scholarly-journals/gender-asymmetries-news-reports/docview/2189509489/se-2

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., ... & Duchesnay, É. (2011). Scikit-learn: Machine learning in Python. Journal of Machine Learning Research, 12(Oct), 2825-2830

Rehurek, R., & Sojka, P. (2011). Gensim–python framework for vector space modelling. NLP Centre, Faculty of Informatics, Masaryk University, Brno, Czech Republic, 3(2).

Saradhambika, K. (2022). Non Verbal Cues and Gender Bias in Select Indian TV Advertisements. Advances in Language and Literary Studies, 13(2), 42-45. http://ezproxy.gc.cuny.edu/login?url=https://www.proquest.com/scholarly-journals/non-verbal-cues-gender-bias-select-indian-tv/docview/2759877755/se-2 indian commercials non verbal

Shor, Eran, Arnout van de Rijt, and Babak Fotouhi. 2019. “A Large-Scale Test of Gender Bias in the Media.” Sociological Science 6: 526-550.

Van der Pas, D. J., & Aaldering, L. (2020). Gender differences in political media coverage: A meta-analysis. Journal of Communication, 70(1), 114-143.

Vijayarani, S., Ilamathi, M. J., & Nithya, M. (2015). Preprocessing techniques for text mining-an overview. International Journal of Computer Science & Communication Networks, 5(1), 7-16.

Wood, J. T. (1994). Gendered media: The influence of media on views of gender. Gendered lives: Communication, gender, and culture, 9, 231-244.