Martin Luther King Speech – Text Mining Analysis

INTRODUCTION

One of the most famous speeches in American history is Martin Luther King Jr.'s "I Have a Dream" speech, which he gave on August 28, 1963, during the March on Washington for Jobs and Freedom. King's idea of a peaceful, integrated America is captured in this speech. The goal of this report is to better understand the significance and substance of King's remarks by analyzing the speech using text mining tools and identifying important terms and topics. I will investigate the frequency of particular terms using this analysis, depict these frequencies, and analyze how they represent the content and intent of the speech.

TEXT MINING PROCESS

Data Collection and Preparation

For analysis, the speech's text was gathered and added to a text corpus. A vast, organized collection of texts used for statistical analysis and hypothesis testing is called a text corpus. In this instance, King's speech is the only thing in the corpus.

Data Cleaning

Before the analysis, the text was carefully cleaned and normalized. Several crucial steps were taken in this crucial preprocessing stage to guarantee the data was in a state that would allow for insightful analysis. Below is a description of each action:

1. Change to Lowercase:
   Goal: Making all text lowercase guarantees that words are handled consistently. If this step is skipped, words like "dream" and "dream," which have the same capitalization, could be counted differently.
   Example: To ensure correct frequency counts and analysis, both "Freedom" and "freedom" would be transformed to "freedom" in the speech's context.
2. Removal of Numbers:
   Goal: Numerical data can contribute noise into analysis and is generally unrelated to the semantic meaning of speeches. Eliminating the numerals makes the meaningful material easier to read.
   Example: To keep the attention on the language content, dates and amounts mentioned in the speech would be removed. The year 1963, for example, would be eliminated because it adds nothing to the theme analysis.
3. Removal of Common Stopwords:
   Goal: Common words like "the," "and," and "is" that are often used but lack substantial significance on their own are known as stopwords. Eliminating these terms makes the text's more significant and important words stand out.
   Example: For instance, King would have eliminated terms like "the," "and," "is," "in," "of," and so on from his discourse. This makes it possible for the study to focus on terms like "freedom," "dream," and "justice" that convey the main ideas.
4. Removal of Punctuations:
   Goal: Because punctuation is appended to words and may be interpreted as a component of those words, it can cause issues with text analysis. Eliminating punctuation guarantees that words are not broken or misunderstood.
   Example: The phrase "I have a dream; that one day..." comes to mind. Because 'dream;' and 'day...' would be counted inaccurately, removing punctuation guarantees that 'dream' and 'day' are recognized as independent words.
5. Text Stemming
   Goal: By reducing words to their root forms, stemming combines several word grammatical forms into a single phrase. This aids in determining the root term's fundamental meaning and frequency.
   Example: The word "run" would replace words like "running," "runner," and "ran." This could imply that in King's speech, the terms "freedom," "freed," and "free" would all be stemmed to "free," thereby encapsulating the notion of freedom's general significance.

TERM DOCUMENT MATRIX AND FREQUENCY ANALYSIS

The frequency of terms that appear in a set of documents was described in a Term Document Matrix (TDM), which was made. In this instance, King's speech frequency is captured by the matrix. To find the most often used terms, the words were then arranged according to their frequency. This stage aids in identifying the terms and ideas that run throughout the speech the most.

VISUALIZATIONS

Word cloud: To give a visual depiction of the most frequently occurring terms in the speech, a word cloud was created. Words in a word cloud are arranged in different sizes based on how frequently they appear. Higher frequency words seem bigger and more noticeable.

Bar plot: To display the frequencies of the top 10 most often occurring words, a bar plot was also made. The most significant terms in the speech are shown in a straightforward, quantitative manner in the map

TERM ASSOCIATIONS

Examining the relationships between concepts was another aspect of the analysis. For instance, identifying terms that are commonly used with the word "freedom" can shed light on the context in which King talks about it.

RESULTS

Word Cloud

The most common terms in the speech are shown graphically in the word cloud below. Higher frequency words are larger and more prominent within the text.

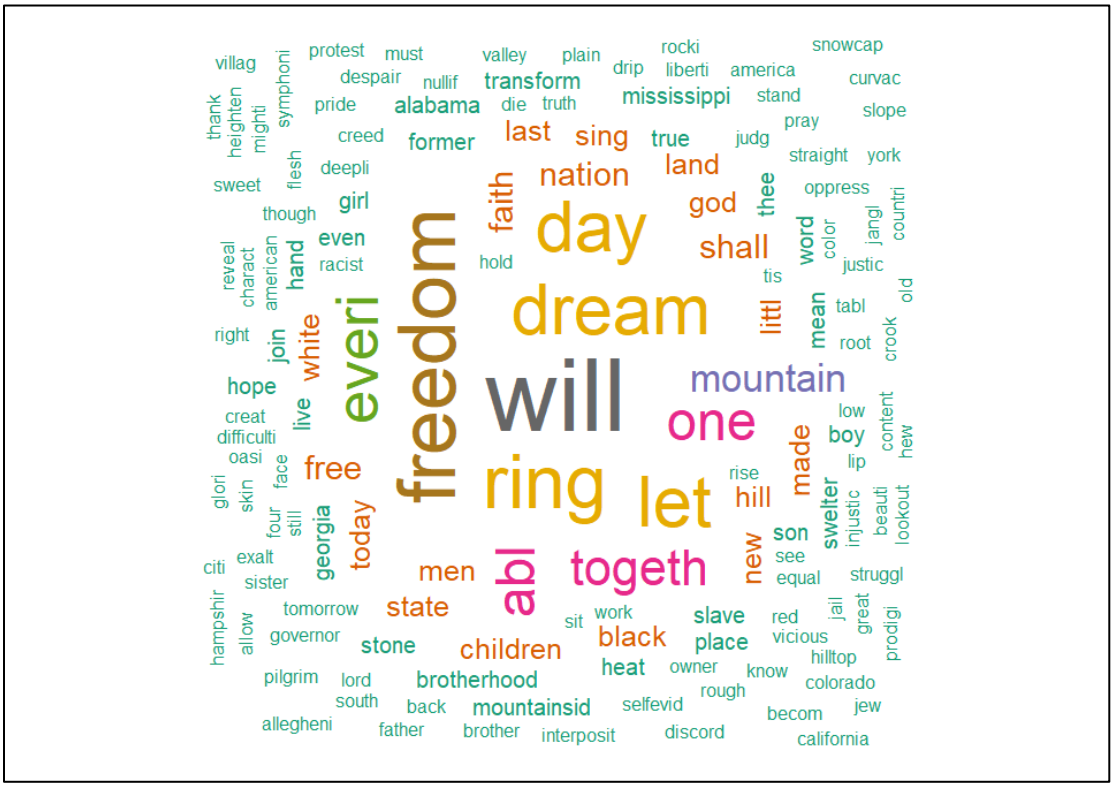

Fig1: Word cloud

Bar Plot

The significance of these important terms is shown by the bar plot below, which displays the frequencies of the top 10 most often occurring words in the speech.

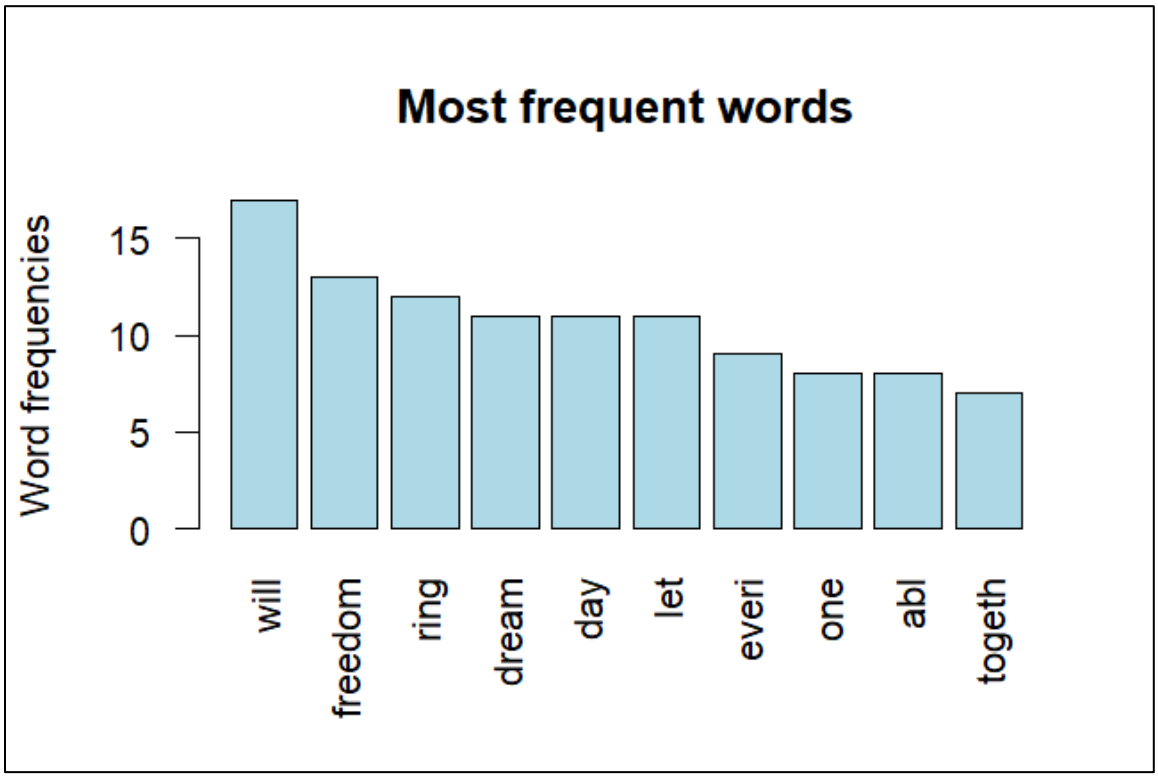

Fig2: Bar Plot of top 10 most frequent words spoken

ANALYSIS OF FREQUENT WORDS

King used terms like "will," "freedom," "ring," "dream," and "day" a lot in his address; these words draw attention to the main ideas and points he was trying to make. King's emphasis on "will" is a reflection of his forward-thinking outlook and his unwavering faith in the achievement of equality and freedom. The word "freedom" keeps coming up, emphasizing the main demand of the civil rights movement, of which King was a prominent member.

Because it appears in the phrase "let freedom ring," which is used several times in the speech to highlight how universal and inclusive the freedom King envisioned, the word "ring" has special significance. Throughout the speech, the word "dream" is used frequently to express King's inspiring optimism for a better future in which racial equality will prevail.

VISUALIZATION INSIGHTS

The word cloud presents an impressive pictorial overview of the speech's essential vocabulary, with the terms "freedom," "dream," and "will" taking center stage. The speech's thematic emphasis on ambitions for the future and the pursuit of freedom is reinforced by this graphic aid.

A more accurate quantitative view is provided by the bar plot, which reveals that "will" is the most often appearing word, closely followed by "freedom." This demonstrates King's focus on the likelihood of attaining equality and freedom in the future.

TERM ASSOCIATIONS

A closer look at the associations with "freedom" reveals its relationship to other important terms in the speech, like "dream," "ring," and "let." This demonstrates how King's concepts of liberty, hope, and action are interrelated. For instance, the connection between "freedom" and "ring" highlights the notion that freedom is cherished and proclaimed all around the country.

REFLECTION ON THE SPEECH'S CONTEXT

The important points of King's "I Have a Dream" speech are succinctly conveyed by the text mining analysis. The speech's setting and the historical period in which it was given are nicely matched by the common words and their meanings. King's address served as a call to action, imploring the country to uphold its principles of justice and equality for all. His persuasive and upbeat message is shown in the way that he emphasizes "freedom," "dream," and "will" repeatedly.

By highlighting the importance of these ideas, the visualizations improve our comprehension even more. The visually arresting word cloud makes it easier to immediately understand the main ideas of the speech. In the meantime, the bar plot offers a thorough analysis of word frequencies, verifying the prominence of phrases like "freedom" and "dream."

CONCLUSION

The "I Have a Dream" speech of Martin Luther King Jr. can be better understood in light of its main ideas and messages thanks to text mining research. Through an analysis of the most commonly used terms, their correlations, and graphic depictions using word clouds and bar plots, we can recognize the speech's potent rhetoric and lasting importance. The study demonstrates that King's word choice and frequency successfully convey the speech's central themes of freedom, equality, and hope. This study illustrates the significance of the speech historically as well as the usefulness of text mining methods for the analysis and interpretation of textual material.