

MM325 Project:

Text mining is a thriving avenue of qualitative data analysis. The process of transforming unstructured text into a structured format to glean patterns and new insights is part of improving decision-making of any organization. Its applications extend broadly, for example, from ads using popular idioms, or slang, targeted to consumers in particular regions of the country to graphical illustrations of frequently expressed themes in a sequence of social media posts.

Perspectives on our changing demographics can be studied as we have seen in previous discussions. Cultural perspectives of social inequities can be gathered in various forms. Sometimes social inequities are experienced first-hand, covered in news and media, and often depicted in art. Some broad examples include racial bias in the news coverage, gender inequality in certain professions, absence of diverse representation in various fields, racial/gender/age bias in patient healthcare, and lack of basic services in lower socioeconomic areas. How a society confronts inequities largely depends on public opinion and is often influenced by news.

For this project, which will culminate in Unit 9, you will act as an analyst for a non-profit, social-justice think tank. Your organization has put you to the task of investigating recent articles from media sources to extract verbiage patterns in their publications. You will implement techniques to statistically mine text in order to complete this project.

You will submit incremental assignments as noted in Units 3, 5, 7, and 9.

Unit 3 Dropbox (Part 1):

- **Introduction:** Identify a social inequity or inequality that you believe is an issue that needs more public attention. Identify at least three articles on this topic published in the past four weeks; these may be taken from the same news media company or multiple sources. Some potential news media sources to consider are provided in this [list](#). In a Word document, state the social issue in 2-3 sentences. List the URLs of the three articles you have chosen.
- **Prepare Plain Text File:** Copy and paste the text of all articles for your text mining analysis into a Word document or Notepad file, deleting any extra text or images generated by advertisements or other content from the webpage which is not a part of the article. Your final combined text content should be at least 2,000 words. Save the text to a plain text file (.txt).
- Submit two files in the Unit 3 dropbox. Be sure to include the following:
 - Word document with social issue description and URLs of three articles.
 - Plain text file containing 2000+ words from the three sources.

Unit 5 Dropbox (Part 2):

- **Plain Text Mining:** Research *R* coding community websites to find some of the suggested code and packages for text mining. Some places to begin your search are listed in the [R coding community websites](#). In your research and implementation, you must discover and implement three different text mining techniques which result in at least three different visualizations of textual trends. The visualizations may not all be simple reformulations of word counts or frequencies, such as a pie chart, bar chart, or word cloud. These visualizations must depict at least three distinct types of text mining results; for example, you could include one visualization of sentiments, one graphic related to word frequencies, and one or more illustrations of other unique results. All *R* code must be clearly labeled and organized with the use of comments. For any block of code taken from a source, proper citation of the source is required.

Note: You may participate in the Open Forum Discussion if you want to share code, and work together to find and edit code.

- Submit a plain text file (.txt) in the Unit 5 dropbox. Be sure to include the following:
 - A plain text file containing all code and citations.

Unit 7 Dropbox (Part 3):

- **Implement text mining:** Using the *R* code you researched from your Unit 5 dropbox (Part 2), revise the code as needed to analyze your selected articles, using the plain text file from your Unit 3 dropbox (Part 1).
- Submit two files in the Unit 7 Dropbox. Be sure to include the following:
 - Word document:
 - Write explanations for how text mining was done including citing the code sources.
 - Cut/Paste visualizations and explanations.
 - Code file used to implement text mining.

Unit 9 Dropbox (Part 4):

- **Project Report:** Write a report of your analysis, introducing your readers (i.e., other staff at your non-profit) to the topic, to the articles you sourced, and from what sources they were obtained. In the report, you should:
 - Write an introduction which includes the article information listed above.
 - Describe how the text mining analysis was constructed, citing the sources used to build your understanding of the analysis and to develop your code.
 - Explain what the output for each of the three (or more) unique text mining techniques is intended to reflect about the text.

- Interpret each of the visualizations resulting from the analysis of your text file.
- Conclude the report with one or two paragraphs, summarizing the textual trends discovered from the articles. The conclusion should also include reflections on how text mining was practical for this analysis and how it may be a useful analysis technique in other contexts.
- Cite all news articles and *R* coding resources on a separate References page, and include all graphical content on separate pages at the end of your report, following APA guidelines.

You may save your report as a Word document (.docx) or PDF file (.pdf).

- Submit the Word document or PDF file in the Unit 9 Dropbox. Be sure to include the following:
 - If changes were made to any of the previous files submitted in Units 3, 5, and 7, this should be noted at the end of the Word document on a separate Updates page so grades may be adjusted accordingly.
 - Final Project Report

Project Guide

Professional, written reports should include an introduction, body, and conclusion, as well as citations to referenced material. As you plan your written content for this project, it is recommended that you first outline the organization of the content, including at least three specific ideas you want to discuss in the body of your report. As you develop graphics and output from your analysis in *R*, you can write a first draft of the introductory paragraph, at least three body paragraphs, and your concluding paragraph(s). Then, once you finalize your analysis using *R*, you may revise your first draft to clearly express your analysis and address all of the objectives for the Project Report, as outlined in the Project rubric.

Your report should include a highly developed purpose and viewpoint; it should also be written in Standard English and include appropriate content, organization, style, grammar, and mechanics. **There should be no evidence of plagiarism.** If you are unsure about what constitutes plagiarism, please review the plagiarism policy. Your report, including references, must use proper APA format. You can find APA resources in the Academic Success Center Reference Library on the Research, Citation, and Plagiarism [page](#).