HOMEWORK 8 – PART 3

ROSE DUNDERDALE

INFS 691 – PRINCIPLES OF ANALYTICAL PROGRAMMING

APRIL 22, 2025

**INSTRUCTIONS**

You will prepare a summary report (3-4 pages) addressing the following:

1. Summary report for Part 1
2. Summary report for Part 2
3. Compare the sentiment values (positive words, negative word, sentiment scores, and sentiment ratios). Note that the number of reviews used in the text mining in Part I and Part II is not the same. Therefore, comparing these two documents (LUC and RMP) based on the number of positive/negative words and the sentiment score may not be appropriate. Therefore, you need to use a ratio the percent positive ratio as given below:  
   Percent positive score =
4. If the sentiment ratios are different for the reviews by LUC and RMP, what can you possibly conclude about these two reviews? What might cause this difference?

You will submit the summary report document (prepared in MS word, pdf, or any other format that I can read) in windows system.

**More on Part 3.**You are expected to present your findings in Part I and Part II to an audience. Consider that the Dean’s office has asked you to do the text mining analysis for Dr. Guder’s evaluations. Administrators in the Dean’s office are not familiar with the text mining techniques, but they are interested in hearing your results (analysis). After completing your analysis, the Dean’s office asks you to present your findings. The report will be the presentation of your findings to the outside audience.

Specifically,   
You will prepare a summary report (3-4 pages) addressing the following:

1. Summary report for Part 1    
   What type of analysis did you do (data used, results observed which may include some visual displays, the sentiment results, etc.)?
2. Summary report for Part 2  
   What type of analysis did you do (data used, results observed which may include some visual displays, the sentiment results, etc.)?
3. **(You need to calculate the percent positive scores (Degree of Positivity) for each part and use them in the comparison)** Compare the results of your findings for these two evaluations (LUC and RMP). The percent positive scores will be the appropriate measures for comparison in this case since the number of reviews used in the text mining in Part I and Part II is not the same.
4. What might you conclude about the students’ sentiments based on these results? What possibly could be causing the difference?
5. **SUMMARY OF REPORT for PART 1**

On the most fundamental, numerical level working under the assumption that:

*Degree of positivity/negativity = countPos/(countPos + countNeg)*

Degree of positivity = 398/453 = 87.86%, meaning that 87.86% of the sentiment words in the reviews are positive. This is a high percentage of positive words.

OR

*Degree of positivity = countPos/countNeg*

This ratio is an example equal to 398/55 = 7.24, meaning that the reviews in this example included 7.24 positive words for each negative word. Again, this ratio indicates a very strong positive sentiment.

The type of analysis I conducted was to use a text file containing unorganized text data to reach a mathematical understanding surrounding the attitude students had towards a particular class, this being INFS 691 – Principles of Analytical Programming (taught by F. Guder). The results observed after cleaning, extracting and transforming the data using wordclouds and barplots indicate that 87.86% of the sentiment words in the reviews are positive; additionally, indicating an elevated level of satisfaction for the professor (in this case, professor Faruk Guder). The top 10 words the highest frequencies (ranked in descending order) are: well, great, organized, questions, material, helpful, homework, students, online, really, easy, learning, available, enjoyed, provided.

Additionally, to improve the results in this dataset and in this sentiment, analysis would be to exclude these words: student, really. As a more developed strategy, one could conduct outside research to reveal any R packages that are designed to handle the significance (if any at all) of non- English-common-stopwords and to weigh them according to the previous and following word(s) – up to 1-2 words.

1. **SUMMARY of REPORT 2**

The number of positive and negative sentiments, and the sentiment score depends on the number of reviews. The number of positive and negative words tend to increase together for a longer review. Therefore, it might be helpful to use the ratios of positive and negative word counts if you are comparing the sentiment scores for 2 different products or services with different number of reviews (texts), it might be helpful to use the ratios of the number of positive and negative words.

On the most fundamental, numerical level working under the assumption that:

Degree of positivity/negativity = countPos/(countPos + countNeg)

Degree of positive = 122/157 = 77.7%, meaning 77.7% of the sentiment words

in the reviews are positive. This is a high percentage of positive words. Only, 33.33% of the words were negative. To explore this figure in depth, we can seek to find the top 10 negative words for this individual professor in order to understand how to use the negative figures as a means to implement or not implement a teaching strategy.

OR

Degree of positivity = countPos/countNeg

This ratio is an example equal to 122/35 = 3.49, meaning that the reviews in

this example included 3.49 positive words for each negative word. Again, this

ratio indicates a strong positive sentiment.

1. **PERCENTAGE CALULATION & COMPARISON**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | Positive Words | Negative Words | Total Sentiment Words | Sentiments score | Degree of Positivity | Degree of Negativity |
| LUC | 398 | 55 | 453 | 343 | 87.86 | 12.14 |
| RMP | 122 | 35 | 157 | 87 | 77.71 | 22.29 |
| Total | 520 | 90 | 610 | 430 | 85.25 | 14.75 |

LUC has a higher percentage positive score (87.86%) compared to RMP (77.71%). This suggests that reviews collected through LUC’s internal evaluations were more positive in tone than those found on RateMyProfessors. The difference in about 10% may reflect a more lenient and supportive tone in institutional feedback or a more critical tone on platforms like RMP where students may feel freer to express dissatisfaction.

1. CONCLUSION

On Sentiments

Students express more positive sentiments in the LUC evaluations than on RMP. The sentiment score also supports this, with LUC showing a much higher net positive score (343 vs. 87). This suggests students are generally more favorable in their feedback when completing internal (LUC) evaluations compared to external, anonymous platforms like RateMyProfessors

On Causes of the Difference

*Anonymity & Audience* – RMP is public and anonymous, which may encourage students to be more candid, critical, or even emotionally expressive LUC evaluations may feel less anonymous or more official, potentially making students more cautious or generous in their comments.

Sample Bias - Students who go out of their way to post on RMP might represent a smaller, more opinionated group, while LUC evaluations typically capture a larger, more representative portion of the class.

*Purpose & Context* – LUC evaluations are designed for improvement and may ask targeted, structured questions that steer toward constructive feedback. RMP is more of a venting or recommendation tool, where extreme opinions (positive or negative) are more common.

*Timing of reviews* – LUC reviews are often completed during or just before finals, possibly while emotions are still high but tempered by the fact that students are still interacting with the course or instructor. RMP reviews might be written after grades are posted, influenced by disappointment, frustration, or relief.

*Incentives & Framing* - At LUC, students might believe their comments can affect course/instructor development, so they might be more thoughtful. On RMP, the motivation might be to warn or recommend based on personal experience, leading to more polarized sentiments.