BASIC SENTIMENTAL ANALYSIS USING SHORT TEXT

A PROJECT REPORT

Submitted by

M SHAHUL HAMEED (230381172432121O1)

in partial fulfillment of requirements for the award of the course

AGB1252-FUNDAMENTALS OF DATA SCIENCE USING R

in

ARTIFICIAL INTELLIGENCE AND DATA SCIENCE

K. RAMAKRISHNAN COLLEGE OF TECHNOLOGY
(An Autonomous Institution, affiliated to Anna University Chennai and Approved by AICTE, New Delhi)

SAMAYAPURAM – 621 112

JUNE- 2025

K. RAMAKRISHNAN COLLEGE OF TECHNOLOGY (AUTONOMOUS)

SAMAYAPURAM – 621 112

BONAFIDE CERTIFICATE

Certified that this project report on **“BASIC SENTIMENTAL ANALYSIS USING SHORT TEXT”** is the bonafide work of **M SHAHUL HAMEED (23038117243212101)** who carried out the project work during the academic year 2024 - 2025 under my supervision.

SIGNATURE

Dr.T. AVUDAIAPPAN, M.E.,Ph.D.,

HEAD OF THE DEPARTMENT

ASSOCIATE PROFESSOR

Department of Artificial Intelligence

K.Ramakrishnan College of Technology
(Autonomous)

Samayapuram–621112.

SIGNATURE

Mrs. S.MURUGAVALLI, AP/AI

SUPERVISOR

ASSISTANT PROFESSOR

Department of Artificial Intelligence

K.Ramakrishnan College of Technology
(Autonomous)

Samayapuram–621112.

Submitted for the viva-voce examination held on **02.06.2025**

INTERNAL EXAMINER

EXTERNAL EXAMINER

DECLARATION

I declare that the project report on **“BASIC SENTIMENTAL ANALYSIS USING SHORT TEXT”** is the result of original work done by me and best of my knowledge, similar work has not been submitted to **“ANNA UNIVERSITY CHENNAI”** for the requirement of Degree of **BACHELOR OF TECHNOLOGY**. This project report is submitted on the partial fulfilment of the requirement of the completion of the course **AGB1252-FUNDAMENTALS OF DATA SCIENCE.**

.

Shahul hameed.M

Signature

M SHAHUL HAMEED

Place: Samayapuram

Date: 02.06.2025

ACKNOWLEDGEMENT

It is with great pride that I express our gratitude and in-debt to our institution "**K.Ramakrishnan College of Technology (Autonomous)**", for providing us with the opportunity to do this project.

I glad to credit honourable chairman **Dr. K. RAMAKRISHNAN**, **B.E.,** for having provided for the facilities during the course of our study in college.

I would like to express our sincere thanks to our beloved Executive Director **Dr. S. KUPPUSAMY, MBA, Ph.D.,** for forwarding to our project and offering adequate duration in completing our project.

I would like to thank **Dr. N. VASUDEVAN, M.Tech., Ph.D.,** Principal, who gave opportunity to frame the project the full satisfaction.

I whole heartily thanks to **Dr. T. AVUDAIAPPAN, M.E.,Ph.D.,** Head of the department, **ARTIFICIAL INTELLIGENCE** for providing his encourage pursuing this project.

I express our deep expression and sincere gratitude to our project supervisor **Ms.S.Murugavalli., M.E.,(Ph.D).,** Department of **ARTIFICIAL INTELLIGENCE,** for her incalculable suggestions, creativity, assistance and patience which motivated us to carry out this project.

I render our sincere thanks to Course Coordinator and other staff members for providing valuable information during the course.

I wish to express our special thanks to the officials and Lab Technicians of our departments who rendered their help during the period of the work progress.

INSTITUTE

Vision:

- To serve the society by offering top-notch technical education on par with global standards.

Mission:

- Be a center of excellence for technical education in emerging technologies by exceeding the needs of industry and society.
- Be an institute with world class research facilities.
- Be an institute nurturing talent and enhancing competency of students to transform them as all – round personalities respecting moral and ethical values.

DEPARTMENT

Vision:

- To excel in education, innovation, and research in Artificial Intelligence and Data Science to fulfil industrial demands and societal expectations.

Mission

- To educate future engineers with solid fundamentals, continually improving teaching methods using modern tools.
- To collaborate with industry and offer top-notch facilities in a conducive learning environment.
- To foster skilled engineers and ethical innovation in AI and Data Science for global recognition and impactful research.
- To tackle the societal challenge of producing capable professionals by instilling employability skills and human values.

PROGRAM EDUCATIONAL OBJECTIVES (PEO)

- **PEO1:** Compete on a global scale for a professional career in Artificial Intelligence and Data Science.
- **PEO2:** Provide industry-specific solutions for the society with effective communication and ethics.
- **PEO3** Enhance their professional skills through research and lifelong learning initiatives.

PROGRAM SPECIFIC OUTCOMES (PSOs)

- **PSO1:** Capable of finding the important factors in large datasets, simplify the data, and improve predictive model accuracy.
- **PSO2:** Capable of analyzing and providing a solution to a given real-world problem by designing an effective program.

PROGRAM OUTCOMES (POs)

Engineering students will be able to:

1. **Engineering knowledge:** Apply knowledge of mathematics, natural science, computing, engineering fundamentals, and an engineering specialization to develop solutions to complex engineering problems.

2. **Problem analysis:** Identify, formulate, review research literature and analyze complex engineering problems reaching substantiated conclusions with consideration for sustainable development.

3. **Design/development of solutions:** Design creative solutions for complex engineering problems and design/develop systems/components/processes to meet identified needs with consideration for the public health and safety, whole-life cost, net zero carbon, culture, society and environment as required.

4. **Conduct investigations of complex problems:** Conduct investigations of complex engineering problems using research-based knowledge including design of experiments, modelling, analysis & interpretation of data to provide valid conclusions.

5. **Engineering Tool Usage:** Create, select and apply appropriate techniques, resources and modern engineering & IT tools, including prediction and modelling recognizing their limitations to solve complex engineering problems.

6. **The Engineer and The World:** Analyze and evaluate societal and environmental aspects while solving complex engineering problems for its impact on sustainability with reference to economy, health, safety, legal framework, culture and environment.

7. **Ethics:** Apply ethical principles and commit to professional ethics, human values, diversity and inclusion; adhere to national & international laws.

8. **Individual and Collaborative Team work:** Function effectively as an individual, and as a member or leader in diverse/multi-disciplinary teams.

9. **Communication:** Communicate effectively and inclusively within the engineering community and society at large, such as being able to comprehend and write effective reports and design documentation, make effective presentations considering cultural, language, and learning differences.

10. **Project management and finance:** Apply knowledge and understanding of engineering management principles and economic decision-making and apply these to one's own work, as a member and leader in a team, and to manage projects and in multidisciplinary environments.

11. **Life-long learning:** Recognize the need for, and have the preparation and ability for i) independent and life-long learning ii) adaptability to new and emerging technologies and iii) critical thinking in the broadest context of technological change.

ABSTRACT

The project titled “Basic Sentiment Analysis on Short Texts” is a streamlined, web-based application developed using R and the Shiny framework, tailored to analyze and classify emotional tone in brief textual inputs. Designed for educators, marketers, researchers, and students, this tool allows users to input custom text or select from predefined samples for demonstration purposes. Emphasizing accessibility and simplicity, the interface adopts a clean, intuitive design with clearly segmented outputs and informative color coding to distinguish between positive, negative, and neutral sentiments. Powered by the tidytext and textdata packages, the app employs lexicon-based methods—such as the Bing or NRC sentiment lexicons—to identify sentiment-bearing words and compute aggregate sentiment scores. Users can explore visual outputs such as sentiment bars, word clouds, and polarity scores, with options to switch lexicons or adjust text preprocessing settings. Leveraging Shiny’s reactive programming, results update instantly with each new input, delivering immediate analytical feedback. This project showcases R’s capabilities in natural language processing, offering a practical tool for understanding public opinion, brand perception, or emotional trends in communication.

ABSTRACT WITH POs AND PSOs MAPPING

CO 5 : BUILD DATA SCIENCE USING R PROGRAMMING FOR SOLVING REAL-TIME PROBLEMS.

ABSTRACT	POs MAPPED	PSOs MAPPED
This project involves building an interactive application using R and Shiny to perform basic sentiment analysis on short texts. Users can input their own text or choose from preloaded samples to explore emotional tone and polarity. The app utilizes lexicon-based methods from the tidytext and textdata packages to identify and classify words as positive, negative, or neutral. Visual outputs such as sentiment bars, word clouds, and polarity indicators are generated using ggplot2 for clear, immediate interpretation. Users can switch sentiment lexicons and adjust preprocessing options to refine the analysis. The app's reactive design enables real-time feedback, making it a responsive and engaging tool for exploring sentiments in short textual content. This application highlights R's strengths in natural language processing and provides a user-friendly platform for sentiment exploration in domains like marketing, education, and social research.	**PO1 -3** **PO2 -3** **PO3 -3** **PO4 -3** **PO6 -3** **PO9 -3** **PO11-3**	**PSO1 -3** **PSO2 -3** **PSO3 -3**

Note: 1- Low, 2-Medium, 3- High

TABLE OF CONTENTS

CHAPTER 1
INTRODUCTION

1.1 Objective

The main objective of the project titled *Basic Sentiment Analysis on Short Texts* is to develop an interactive web-based application that allows users to assess the emotional tone of brief text inputs. This platform aims to make sentiment analysis accessible and interpretable for users such as students, marketers, educators, and researchers. By replacing manual or opaque NLP methods with an intuitive and visually guided interface, the project seeks to simplify text analysis, enabling users to quickly determine whether a message conveys a positive, negative, or neutral sentiment. The tool provides immediate visual feedback, supports basic text preprocessing, and offers comparative lexicon options to enhance the flexibility and accuracy of sentiment interpretation.

1.2 Overview

In today's digital landscape, understanding the sentiment behind short text—such as social media posts, reviews, or feedback—is critical for making informed decisions across sectors. Traditional sentiment analysis tools often require advanced programming knowledge or are not user-friendly for beginners. Users can either input their own text or select from preloaded samples. The app utilizes lexicon- based approaches (such as Bing and NRC) from the tidytext and textdata packages to score and classify the text. Results are displayed using clear visualizations, including sentiment bar charts and word clouds generated with ggplot2. The app also offers customization options, allowing users to switch sentiment dictionaries or include/exclude stop words. Thanks to Shiny's reactive features,delivering a responsive and educational sentiment analysis experience.

1.3 R Programming Concepts Used

This project is built using the R programming language, leveraging its rich set of packages for text analysis, visualization, and interactive web applications. At its core is the **Shiny** framework, which powers the web interface and enables reactivity—ensuring that visualizations and sentiment scores update automatically based on user interaction. The **tidytext** package facilitates text tokenization and sentiment classification using lexicons such as Bing, NRC, and AFINN, while **dplyr** is employed for data wrangling tasks such as filtering, grouping, and summarizing tokenized words. Visual representations, including bar plots and word clouds, are created using **ggplot2** and **wordcloud2**, providing intuitive insights into the emotional tone of the input text. The application's design is structured using **shinydashboard** for organized UI components and **bslib** for enhanced visual theming. Together, these R tools create a user-centric platform that demonstrates the power and accessibility of R for natural language processing and sentiment analysis.

CHAPTER 2
PROJECT METHODOLOGY

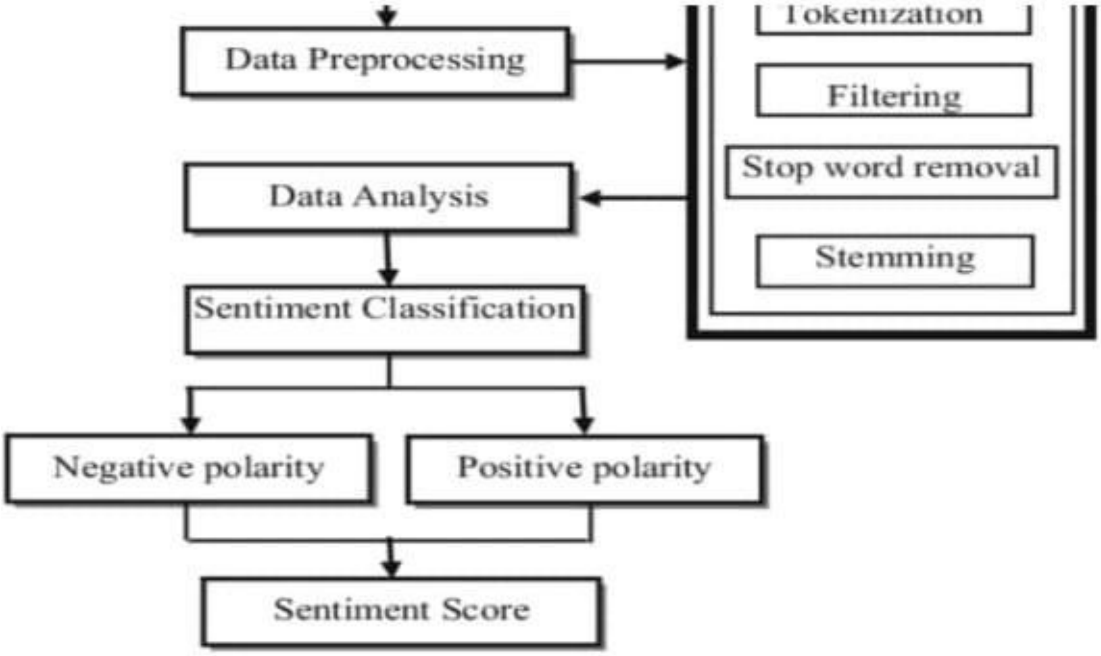
2.1 Proposed Work

The proposed project aims to build a lightweight, web-based sentiment analysis tool using R and the Shiny framework, designed to process and interpret the emotional tone of short texts. The application will feature a modular, user-friendly interface organized into multiple tabs, including options for text input, sample text selection, sentiment results, visual analysis, and settings customization.

Users will be able to input custom text manually or choose from a library of predefined examples. Upon submission, the application will leverage reactive programming to tokenize the text, apply sentiment scoring using lexicon-based methods, and immediately display the results through clear visualizations. These will include sentiment polarity bar charts, categorized word frequency graphs, and dynamic word clouds to highlight emotional keywords.

Users can switch between sentiment lexicons (e.g., Bing, NRC, AFINN), enable or disable stop word filtering, and view summaries of detected sentiment types. This flexible analysis environment is supported by a clean, responsive frontend and efficient backend text processing. The tool is designed for accessibility and ease of use, empowering individuals with limited technical backgrounds to explore and understand sentiment in textual data. By providing real-time emotional insights, the system supports applications in education, marketing, feedback analysis, and social media monitoring.

2.2 Block Diagram



CHAPTER 3
MODULE DESCRIPTION

3.1 Login Module

In this project, the Login Module is intended to manage user access to the sentiment analysis application. While the current prototype may not implement complete authentication, a full version would include user credential verification to restrict access to authorized users. After logging in successfully, users are directed to the main interface where they can input or select text samples for sentiment evaluation. This module sets the stage for a secure and personalized analytical experience.

3.2 Home Module

The Home Module serves as the landing page of the application, providing users with an overview of the system's capabilities. It introduces key features such as entering custom text, selecting sentiment lexicons, and viewing emotional breakdowns through visualizations. With intuitive navigation elements, this module helps users quickly understand how to interact with the app and directs them to core analytical functions like input submission and sentiment result viewing.

3.3 Dashboard Module

The Dashboard Module is the primary workspace of the sentiment analysis application. Here, users can submit short texts or select predefined examples to analyze. The module displays real-time outputs including sentiment classification (positive, negative, or neutral), emotional tone distribution, and word frequency visualizations. Users can switch between different visualization formats and sentiment lexicons to compare results. This dynamic feedback allows for fast and effective interpretation of the emotional content within the text.

3.4 Text Input Module

The Text Input Module allows users to manually enter short text samples for sentiment analysis. This module includes a clean input area, along with options for text preprocessing such as converting to lowercase, removing punctuation, and excluding stop words. While a demo version may include sample texts for testing, the full implementation supports fully custom user input. Once the text is submitted, it is tokenized and analyzed in real time, with results automatically displayed in the Dashboard Module.

3.5 Sentiment Visualization Module

The Sentiment Visualization Module is a central feature of the application, transforming processed textual data into insightful graphical outputs. This module presents sentiment scores through polarity bar charts, emotion distribution plots (if using NRC), and dynamic word clouds of impactful words. Built using **ggplot2** and **wordcloud2**, these visuals update instantly in response to user input or lexicon changes. The module allows users to visually grasp the emotional weight of the text, making sentiment analysis more intuitive and actionable across use cases such as brand feedback, educational texts, or social media content.

CHAPTER 4
CONCLUSION & FUTURE SCOPE

Conclusion

This project successfully developed an R Shiny web application for performing basic sentiment analysis on short texts. Utilizing core R programming concepts such as reactive programming, text preprocessing with tidytext, and visualizations through ggplot2 and wordcloud2, the application delivers an intuitive and interactive platform for understanding emotional tone in written content. The combination of polarity charts, word frequency plots, and word clouds enables users to explore sentiment dynamics in a clear and engaging manner. The use of shinydashboard ensures a structured and user-friendly layout, while real- time feedback mechanisms allow users to analyze and interpret sentiment instantly. Overall, this tool provides a practical solution for educators, marketers, and researchers looking to extract meaningful emotional insights from short text data with minimal technical overhead.

Future Scope

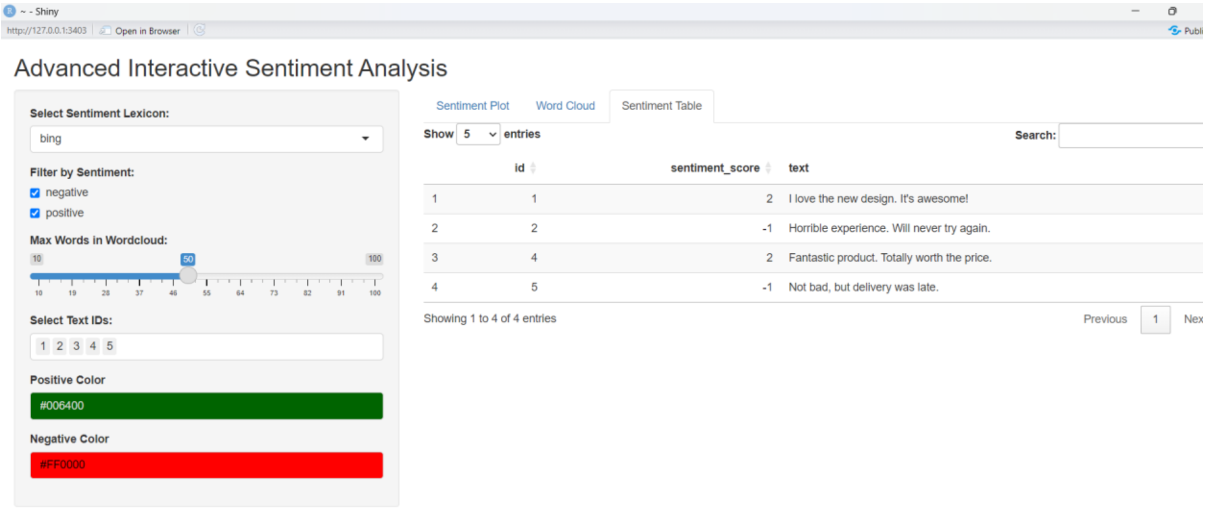
While the current version offers a solid foundation for sentiment analysis, several future enhancements can further improve functionality and user experience:

- **Expanded NLP Techniques**: Incorporating machine learning or deep learning-based sentiment models (e.g., using text or keras) can provide more nuanced and context-aware analysis.
- **Emotion Classification**: Adding support for multi-emotion classification (anger, joy, sadness, etc.) using lexicons like NRC or external APIs can enrich the emotional insights.

CHAPTER 5
APPENDIX A – SOURCE CODE

```
library(shiny)
library(tidytext)
library(dplyr)
library(ggplot2)
library(plotly)
library(wordcloud)
library(tm)
library(DT)
library(colourpicker)
#   Sample text dataset
sample_data <- data.frame(
id = 1:5,
  text = c(
    "I love the new design. It's awesome!",
    "Horrible experience. Will never try again.",
    "Okay service, nothing special.",
    "Fantastic product. Totally worth the price.",
    "Not bad, but delivery was late."
  )
)

#  UI
ui <- fluidPage(
  titlePanel("Advanced Interactive Sentiment Analysis"),

  sidebarLayout(
    sidebarPanel(
      selectInput("lexicon", "Select Sentiment Lexicon:",
                  choices = c("bing", "nrc", "afinn"), selected = "bing"),

      checkboxGroupInput("selected_sentiment", "Filter by Sentiment:",
                         choices = NULL, selected = NULL),

      sliderInput("max_words", "Max Words in Wordcloud:", min = 10,
max = 100, value = 50),

      selectInput("doc_select", "Select Text IDs:",
                  choices = sample_data$id, selected = sample_data$id,
multiple = TRUE),

      colourInput("pos_color", "Positive Color", value = "darkgreen"),
      colourInput("neg_color", "Negative Color", value = "red")
    ),

    mainPanel(
      tabsetPanel(
        tabPanel("Sentiment Plot", plotlyOutput("sentimentPlot")),
        tabPanel("Word Cloud", plotOutput("wordCloud")),
        tabPanel("Sentiment Table", DTOutput("textTable"))
      )
    )
  )
)

# Server
server <- function(input, output, session) {

  observe({
    sentiments <- switch(input$lexicon,
                         "bing" = unique(get_sentiments("bing")$sentiment),
                         "nrc" = unique(get_sentiments("nrc")$sentiment),
                         "afinn" = c("positive", "negative"))

    updateCheckboxGroupInput(session, "selected_sentiment",
                             choices = sentiments,
                             selected = sentiments)
  })

  tidy_data <- reactive({
    lex <- switch(input$lexicon,
                  "bing" = get_sentiments("bing"),
                  "nrc" = get_sentiments("nrc"),
                  "afinn" = get_sentiments("afinn"))

    text_df <- sample_data %>%
      filter(id %in% input$doc_select) %>%
      unnest_tokens(word, text) %>%
      anti_join(stop_words)
    if (input$lexicon == "afinn") {
      inner_join(text_df, lex, by = "word") %>%
        mutate(sentiment = ifelse(value > 0, "positive", "negative"))
    } else {
      inner_join(text_df, lex, by = "word") %>%
        filter(sentiment %in% input$selected_sentiment)
    }
  })

  output$sentimentPlot <- renderPlotly({
    data <- tidy_data() %>%
    count(sentiment, sort = TRUE)

    p <- ggplot(data, aes(x = sentiment, y = n, fill = sentiment)) +
      geom_col(show.legend = FALSE) +
      scale_fill_manual(values = c("positive" = input$pos_color,
                                   "negative" = input$neg_color)) +
      theme_minimal() +
      labs(title = "Sentiment Counts", x = "Sentiment", y = "Count")

    ggplotly(p)
  })

  output$wordCloud <- renderPlot({
    words <- tidy_data() %>%
    count(word, sentiment, sort = TRUE)

    wordcloud(
      words = words$word,
      freq = words$n,
      colors = ifelse(words$sentiment == "positive", input$pos_color,
input$neg_color),
      max.words = input$max_words
    )
  })

  output$textTable <- renderDT({
    summary <- tidy_data() %>%
    group_by(id) %>%
      summarise(sentiment_score = sum(ifelse(sentiment == "positive", 1,
                                             ifelse(sentiment == "negative", -1, 0)))) %>%
      left_join(sample_data, by = "id")

    datatable(summary, options = list(pageLength = 5))
  })
}

# 🚀 Launch App
shinyApp(ui, server)
```

Appendix B – Screenshots



Advanced Interactive Sentiment Analysis

Select Sentiment Lexicon: bing

Filter by Sentiment:

☑ negative

☑ positive

Max Words in Wordcloud: 50 (10–100)

Select Text IDs: 1 2 3 4 5

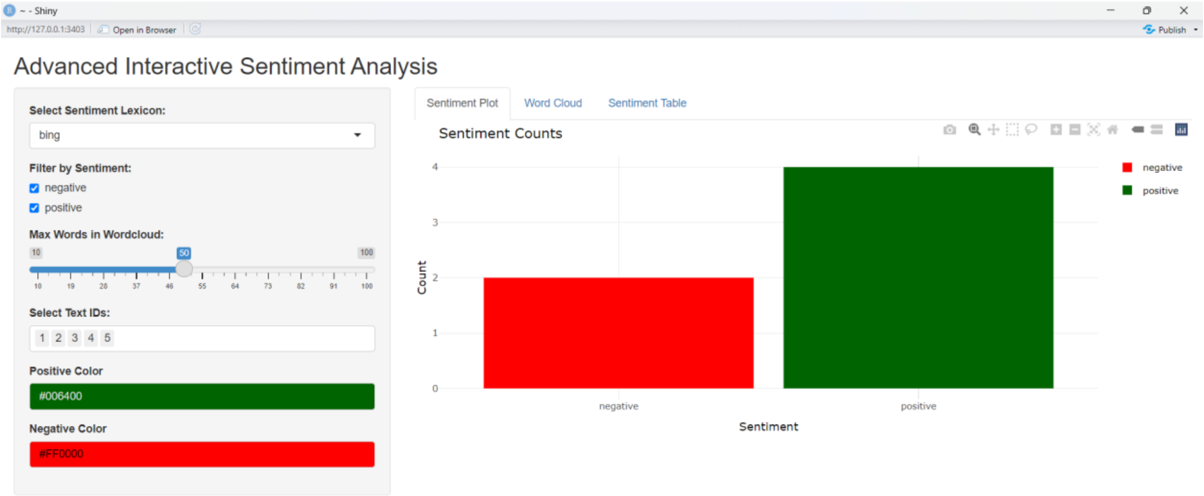
Positive Color: #006400

Negative Color: #FF0000

Show 5 entries Search:

	id	sentiment_score	text
1	1	2	I love the new design. It's awesome!
2	2	-1	Horrible experience. Will never try again.
3	4	2	Fantastic product. Totally worth the price.
4	5	-1	Not bad, but delivery was late.

Showing 1 to 4 of 4 entries



Advanced Interactive Sentiment Analysis

Select Sentiment Lexicon: bing

Filter by Sentiment:

☑ negative

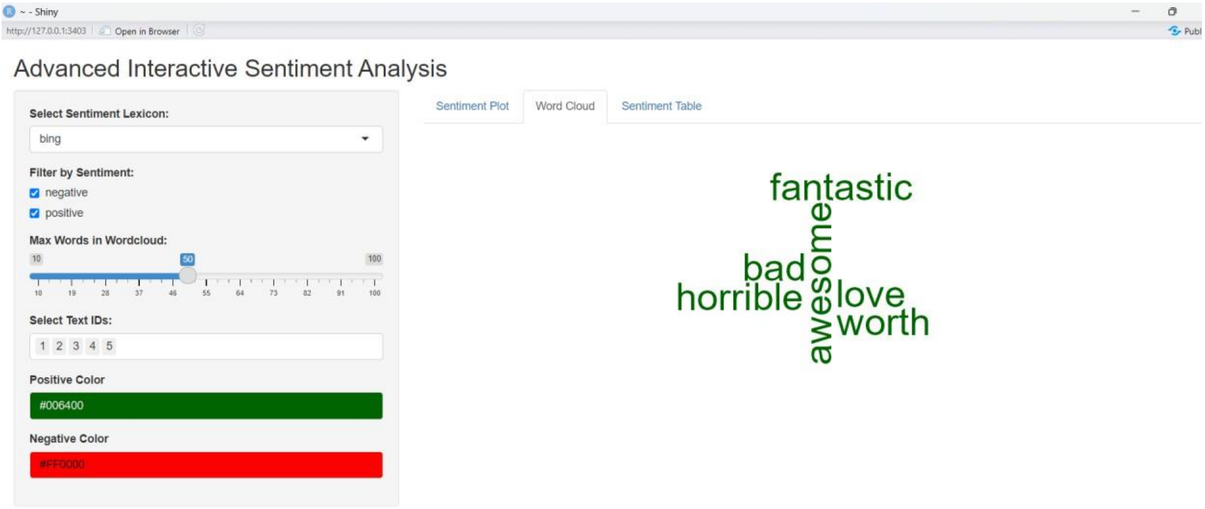
☑ positive

Max Words in Wordcloud: 50 (10–100)

Select Text IDs: 1 2 3 4 5

Positive Color: #006400

Negative Color: #FF0000


~ - Shiny

http://127.0.0.1:3403 | Open in Browser

Advanced Interactive Sentiment Analysis

Select Sentiment Lexicon:

bing

Filter by Sentiment:

☑ negative

☑ positive

Max Words in Wordcloud:

10 50 100

Select Text IDs:

1 2 3 4 5

Positive Color

#006400

Negative Color

#FF0000

References

1. Chang, W., Borges Ribeiro, B., & Silva, D. (2024). *shinydashboard: Create Dashboards with 'Shiny'*. R package version 0.7.2. https://CRAN.R-project.org/package=shinydashboard
2. Chang, W., Cheng, J., Allaire, J., Xie, Y., & McPherson, J. (2024). *Shiny: Web Application Framework for R*. R package version 1.8.0. https://CRAN.R-project.org/package=shiny
3. Silge, J., & Robinson, D. (2023). *tidytext: Text Mining using 'dplyr', 'ggplot2', and Other Tidy Tools*. R package version 0.4.2. https://CRAN.R-project.org/package=tidytext
4. Wickham, H., François, R., Henry, L., & Müller, K. (2024). *dplyr: A Grammar of Data Manipulation*. R package version 1.1.4. https://CRAN.R-project.org/package=dplyr