Sentiment Analysis on Tweets using R

Name: Garvit Sharma

University ID: 2301730190

Project Topic: Sentiment Analysis on Tweets using R

Date of Submission: 19 April 2025

GitHub Repository Link: [https://github.com/Garvit7878/R-Programming-Projects/tree/main]

Introduction

Sentiment analysis is a Natural Language Processing (NLP) technique used to identify and extract subjective information from textual data. In this project, we analyze sentiments of tweets related to 'R programming' using R. This helps in understanding public opinion or emotion towards a specific topic.

Prerequisites

The following R packages are used in the project:
- rtweet: to fetch Twitter data
- tidytext: to process and analyze textual data
- dplyr: for data manipulation
- ggplot2: for data visualization
- tidyr: for cleaning and restructuring data

Step 1: Fetch Twitter Data

Using the rtweet package, we extract 500 English tweets containing the keyword 'R programming'. Retweets are excluded to ensure originality and variety in data.

Step 2: Text Cleaning and Sentiment Analysis

The tweets are cleaned by removing stop words, links, and special characters. Then, sentiment analysis is performed using two lexicons: Bing and NRC. The Bing lexicon classifies words as positive or negative, while the NRC lexicon associates them with emotions like joy, anger, fear, etc.

Step 3: Visualizing Sentiment Distribution

Bar charts are used to visualize sentiment distribution. The first chart shows the number of positive and negative words, while the second provides an emotion-wise breakdown of sentiments using the NRC lexicon.

Conclusion

The project successfully demonstrates how to use R for extracting and analyzing the sentiment of tweets. This basic analysis can be expanded further by integrating machine learning models for more accurate classification or analyzing larger datasets.