

Open source sentiment analysis software: an evaluative study of functions and capabilities

Research – English
Summary

Aya Medhat Shendy Mohamed

Assistant teacher, Department of Libraries and Information,
Faculty of Arts, Minia University, Egypt

aya.medhat@mu.edu.eg

ORCID: [0009-0002-3129-4008](https://orcid.org/0009-0002-3129-4008)

Copyright (c) 2024,
Aya Medhat Shendy
Mohamed



This work is licensed
under a Creative
Commons Attribution
4.0 International
License.

Abstract

Open-source sentiment analysis software is a powerful tool for understanding emotions and opinions, aimed at improving services and products. All organizations can benefit from these tools to gain insights and make informed decisions based on data analysis, in order to develop services or products that meet customer needs. Data has become a vast resource for organizations, and among the most important data are customer opinions and feedback. Interest in sentiment analysis methods has grown in importance with the increasing prevalence of social media, where users freely express and share their opinions, making these platforms a rich source of information on public sentiment regarding various issues and topics. As a result, numerous automated programs have emerged to analyze sentiment, aimed at understanding public opinion trends and improving products and services to provide a comprehensive picture and support decision-making. From this perspective, this study evaluates and compares these software tools to identify differences in features and assess the available capabilities. A comparative analysis was conducted on 38 open-source sentiment analysis programs, and it was found that AssemblyAI is the best among them, ranking first according to the evaluation criteria presented.

Keywords

Sentiment Analysis, Open-Source Software, Data Extraction, Opinion Mining, Social Media Platforms, Natural Languages

1. Introduction

Social networks have become invaluable sources of knowledge, serving as dynamic channels for communication, content sharing, and user interactions. These platforms generate vast amounts of data, which are challenging to process and analyze manually. Sentiment analysis plays a crucial role in addressing this challenge by determining public opinions and extracting subjective information from large, unstructured datasets. This field integrates data mining, machine learning, natural language processing (NLP), information retrieval, and knowledge management techniques to analyze sentiments expressed in text.

By leveraging NLP, sentiment analysis classifies textual data into positive and negative opinions. As a subfield of artificial intelligence (AI) and data science, sentiment analysis is widely used in text mining, particularly for analyzing user feedback on social media. Businesses and organizations rely on sentiment analysis tools to assess public sentiment toward their products, services, and initiatives.

2. Research Problem and Questions

With the growing interest in AI technologies, this study aims to highlight the role of artificial intelligence in scientific research, particularly in sentiment analysis. NLP algorithms have been developed to extract insights from unstructured social media data, enabling automated analysis of public sentiment. This study focuses on comparing and evaluating open-source sentiment analysis software to determine their effectiveness, identify their strengths and weaknesses, and provide a comprehensive assessment of their features and functionalities.

Key research questions include:

1. What are the essential criteria for evaluating open-source sentiment analysis software?
2. To what extent do these tools adhere to these criteria?
3. Which sentiment analysis software performs best?
4. What are the main characteristics, features, and functions of these tools?

3. Research Objectives

This study aims to:

1. Identify and analyze available open-source sentiment analysis software.
2. Compare and evaluate these tools to determine their efficiency, user-friendliness, and capabilities.
3. Assess the strengths, weaknesses, functionalities, and structural features of open-source sentiment analysis software.
4. Develop a comprehensive checklist of key evaluation criteria for sentiment analysis tools.

4. Significance of the Study

The significance of sentiment analysis research continues to grow due to the widespread use of social media. Unlike traditional public opinion polls, which often take weeks to conduct and face challenges such as sample selection biases and participant reluctance, social media provides real-time, continuous insights into public opinion.

As these platforms become primary arenas for opinion expression, sentiment analysis software plays a crucial role in tracking discussions, comments, and posts to extract meaningful insights. This study aims to highlight the importance of such software in improving products and services, supporting decision-making processes, and providing a structured approach to sentiment analysis in various fields.

Study Method and Data Collection Tools

The research adopts a **descriptive-analytical approach**, incorporating various measurement, analysis, and evaluation methods to align with the study's objectives. Additionally, an **evaluative approach** is employed to assess open-source sentiment analysis tools. The evaluation is based on predefined criteria and focuses solely on publicly available software.

Development of the Study Tool (Checklist)

A structured **checklist** was designed to compare the selected sentiment analysis tools. The checklist includes **19 main evaluation criteria**, each containing several specific requirements, totaling **59 evaluation parameters**.

Software Selection and Evaluation

The study focused exclusively on **open-source sentiment analysis tools** available online. Findings revealed:

- **15 tools** required no login for access.
- **23 tools** required user authentication (username and password).
- **5 tools** supported local installation, allowing users to run them on personal devices or servers.

The evaluation was conducted between **November 2023 and May 2024**. The study aligns with previous research (e.g., Georgiou et al., 2015), which found that open-source sentiment analysis tools generally outperform commercial alternatives.

Evaluation Method

The **binary scoring method** was used, assigning:

- **(√) = 1** if a requirement is met.
- **(X) = 0** if a requirement is absent.

This method, commonly used in automated system evaluations, ensures a structured assessment of each software's features. Given the diversity in classification methods, supported data formats, and language capabilities, a pure percentage-based evaluation was deemed inadequate. Instead, scores were assigned based on pre-defined criteria, ensuring a fair comparison.

Formula for calculating program efficiency per requirement:

$$\text{Efficiency} = \left(\frac{\text{Number of programs meeting the criterion} \times 100}{\text{Total number of programs}} \right) \%$$

This formula was used to calculate individual and overall software evaluation scores, allowing for a comprehensive assessment of each tool's performance.

13. Study Results

This study has yielded several key findings:

1. **Best Performing Software:** AssemblyAI ranked first according to the proposed evaluation criteria. This is primarily due to its extensive support for various data formats and natural languages.

2. **Second Place:** Mentionlytics secured second place, excelling in its support for the highest number of natural languages among all tools. Additionally, it leads in technical support and display features.
3. **Lowest Ranked Software:** AFINN Sentiment Analysis and Sentiment Analyzer ranked last due to their lack of data import options, inadequate technical support, weak data security measures, and limited display features.
4. **Online Accessibility:** All sentiment analysis tools support online usage, enabling easy access from any internet-connected device.
5. **Operating System Compatibility:** Most sentiment analysis tools are web-based and compatible across all operating systems. Only 10 tools require installation on specific operating systems.
6. **Programming Languages:** Python is the most widely used programming language for sentiment analysis tools, followed by Java, JavaScript, PHP, and TypeScript—the latter being used in only one tool.
7. **Technological Leadership:** The Nocodefunctions tool outperforms others in terms of advanced technological features.
8. **Supported Data Patterns:** Text data is the most commonly supported format, with all tools (except one) accommodating it. Image and video data follow, while audio data is the least supported.
9. **Text Data Handling:** The best tools for handling diverse text data formats are Nocodefunctions, Speak Ai, and IBM Watson Natural Language. In contrast, AssemblyAI does not support any text data formats.
10. **Analysis Output Formats:** Speak Ai leads in offering analysis results in the highest number of data formats, followed by Mention, Brand24, Mentionlytics, and Social Mention.
11. **Data Import Sources:** Most tools import data from social media platforms, recognizing their role as primary channels for public opinion, sentiment, and discourse.
12. **Software Availability:**
 - 20 sentiment analysis tools are available for free.
 - 15 tools offer direct trials without requiring login.

- 18 tools provide time-limited trial versions.

13. **Machine Learning Dominance:** The majority of sentiment analysis tools employ machine learning techniques for sentiment classification and analysis.

Recommendations

Based on these findings, the study presents the following recommendations:

- **Ensuring Quality Standards:** Sentiment analysis tools should meet essential quality standards, including comprehensive technical capabilities, advanced analysis features, and visually engaging result presentations using charts and graphs.
- **Workshops & Awareness:** Conducting workshops to emphasize the importance of sentiment analysis tools and the latest technological advancements in supporting information institutions.
- **AI & Big Data Research:** Encouraging further studies on the application of AI in real-time big data analysis.
- **Security & Privacy Focus:** There is a need for in-depth research on security challenges related to big data analytics in AI-driven sentiment analysis.
- **Investment & Innovation:** Increased funding and support for research in sentiment analysis, fostering innovation in new tools and techniques.
- **Practical Applications:** Sentiment analysis should be leveraged across various fields to enhance customer experience, improve products and services, and detect emerging trends. Additionally, it can be used for early complaint identification and resolution.
- **Interactive Development:** Sentiment analysis tools should evolve to be more interactive, allowing users to modify displayed results and provide feedback to enhance analysis accuracy.