Manchester Metropolitan University

e-space

Manchester Metropolitan University's Research Repository

**Please cite the Published Version**

Adepeju, Monsuru (2021) R-Opitools – An Opinion Analytical Tool for Big Digital Text Document (DTD). The Journal of Open Source Software, 6 (64). p. 3605.

**DOI:** https://doi.org/10.21105/joss.03605

**Publisher:** Open Journals

**Version:** Published Version

**Downloaded from:** https://e-space.mmu.ac.uk/628309/

**Usage rights:** CC BY Creative Commons: Attribution 4.0

**Additional Information:** This is an Open Access article published in The Journal of Open Source Software, published by Open Journals, copyright The Author(s).

**Enquiries:**
If you have questions about this document, contact openresearch@mmu.ac.uk. Please include the URL of the record in e-space. If you believe that your, or a third party's rights have been compromised through this document please see our Take Down policy (available from https://www.mmu.ac.uk/library/using-the-library/policies-and-guidelines)

The Journal of Open Source Software

# R-Opitools – An Opinion Analytical Tool for Big Digital Text Document (DTD)

**Monsuru Adepeju**[1]

**1** Crime and Well-being Big Data Centre, Manchester Metropolitan University, United Kingdom

**DOI:** 10.21105/joss.03605

**Software**

- Review
- Repository
- Archive

**Editor:** Andrew Stewart

**Reviewers:**

- @cjbarrie
- @justinchuntingho
- @jaeyk

**Submitted:** 29 July 2021
**Published:** 31 August 2021

**License**
Authors of papers retain copyright and release the work under a Creative Commons Attribution 4.0 International License (CC BY 4.0).

## Statement of Need

Since the year 2000, various computational intelligence techniques have been developed for analyzing sentiments of users in the field of natural language processing (NLP). To date, the majority of the techniques as deployed across various fields, including social sciences (Ansari et al., 2020; Nikolovska & Ekblom, 2020; Somasundaran & Wiebe, 2010) and market research (Al-Otaibi et al., 2018; Feldman et al., 2011), have focused largely on detecting subjectivity, and/or extracting and classifying sentiments and opinions in a text document. Building on this existing work, the current paper advances an opinion impact analytical tool, named `Opitools`, that not only extracts inherent themes from within a digital text document (DTD), but also evaluates the extent to which a specified theme may have contributed to the overall opinions expressed by the document. Based on this advancement, `Opitools` has wider applications in the aforementioned application fields. For example, in law enforcement, the package can be deployed to understand factors (themes) that drive public perception of police services (Adepeju & Jimoh, 2021); and in product marketing, to identify factors that underlie customers satisfaction in a product.

## Implementation

Having extracted a set of thematic keywords from a digital text document, the goal is to computationally classify the sentiments expressed in each text record into positive, negative or a neutral sentiment, using a lexicon-based classification approach (Adepeju & Jimoh, 2021; Nielsen, 2011). The resulting sentiment scores are combined in order to estimate the overall opinion score of the document. To assess the impacts of a selected theme (or a subject) on the estimated opinion score, we simply ask the question; *If expected opinion scores were generated under the null hypothesis, how likely would we be to find a score higher than the estimated score?*. The question is answered by employing a non-parametric randomization testing strategy (Fisher, 1935; Good, 2006) which involves random re-assignment of sentiment labels of the original text document to derive the expectation distribution, which is then compared with the observed score to obtain the statistical significance of the impacts.

## Key Functionalities

The package contains text exploratory functions for extracting themes from a digital text document. In order to conduct impact analysis, a user can draw on a number of interrelated functions to compute the required measures, such as the observed opinion score, the expectation distribution, and the statistical significance of impacts. Whilst different types of opinion

score functions are embedded in the package, there is also a provision that allows a user to integrate his/her own pre-defined user score function. This feature is to further facilitate the uptake of the package in more application fields.

## Acknowledgment

We gratefully acknowledge the Economic and Social Research Council (ESRC), who funded the Understanding Inequalities project (Grant Reference ES/P009301/1) through which this research was conducted.

## References

Adepeju, M., & Jimoh, F. (2021). An analytical framework for measuring inequality in the public opinions on policing – assessing the impacts of COVID-19 pandemic using twitter data. *Journal of Geographical Information System*, *33*, 122–147. https://doi.org/10.4236/jgis.2021.132008

Al-Otaibi, S., Alnassar, A., Alshahrani, A., Al-Mubarak, A., & al., et. (2018). Customer satisfaction measurement using sentiment analysis. *International Journal of Advanced Computer Science and Applications*, *9*.

Ansari, MZ., Aziz, M. B., Siddiqui, M. O., Mehra, H., & Singh, K. P. (2020). Analysis of political sentiment orientations on twitter. *Procedia Computer Science*, *167*, 1821–1828. https://doi.org/10.1016/j.procs.2020.03.201

Feldman, R., Rosenfeld, B., Bar-Haim, R., & Fresko, M. (2011). The stock sonar - sentiment analysis of stocks based on a hybrid approach. *Proceedings of the Twenty-Third Conference on Innovative Applications of Artificial Intelligence*.

Fisher, R. A. (1935). *The design of experiments*. New York: Hafner.

Good, P. (2006). *Resampling methods*. Birkhauser, 3rd Ed.

Nielsen, F. (2011). A new ANEW: Evaluation of a word list for sentiment analysis in microblogs. *In Proc. ESWC-11, arXiv Preprint arXiv:1103.2903*.

Nikolovska, S., M.and Johnson, & Ekblom, P. (2020). "Show this thread": Policing, disruption and mobilisation through twitter. An analysis of UK law enforcement tweeting practices during the covid-19 pandemic. *Crime Science*, *9*, 20. https://doi.org/10.1186/s40163-020-00129-2

Somasundaran, S., & Wiebe, J. (2010). Recognizing stances in ideological online debates. In D. Kurlander, M. Brown, & R. Rao (Eds.), *CAAGET '10: Proceedings of the NAACL HLT 2010 workshop on computational approaches to analysis and generation of emotion in text* (pp. 116–124). ACM Press.