

CHILLAX - at Arabic Hate Speech 2022: Detect Hate Speech using Logistic Regression and Random Forest

Prof Ensaf Hussein

Kirollos Hany, Kirollos George, Malak Emad,
Shady Zekry, Seif Hesham, Fady Fayek

Abstract

Hate speech and offensive language became a crucial problem nowadays due to the large usage of social media by people from different gender, nationality, religion and other types of characteristics allowing anyone to share their thoughts and opinions. In this research paper we experiment the result of using the Arabic pretrained Bert language model MARBERT for feature extraction of the Arabic tweets in the dataset provided by OSACT2022 shared task then feeding the features to classic machine learning models (Logistic Regression, Random Forest). The best results achieved were by the Logistic Regression model with accuracy, precision, recall, and f1-score of 0.8, 0.78, 0.78, and 0.78 for the offensive tweet detection task respectively and for the hate speech tweet detection task results achieved were 0.89, 0.72, 0.8, and 0.76 these results were achieved on the OSACT2022 test dataset.

Keywords: keyword1, keyword2, keyword3

1. Full Paper Submission

Each full paper submission should be submitted on white A4 paper and must be between 4 and 8 pages in length (plus more pages for references if needed). The fully justified text should be formatted in two parallel columns. The new LREC paper dimensions are now aligned with ACL layout to ensure quick integration within the ACL Anthology. These dimensions are as follows:

- The paper is in A4-size format, that is 21 x 29.7 cm.
- The text height is 24.7 cm and the text width 16.0 cm in two columns separated by a 0.6 cm space.
- The font for the main body of the text must be Times New Roman 10 pt with interlinear spacing of 11 pt.
- The use of LREC2022.sty will ensure the good formatting.

2. Final Paper

Each final paper should be submitted on white A4 paper. The fully justified text should be formatted according to LREC2022 style as indicated for the Full Paper submission. As indicated above, the font for the main body of the text should be Times New Roman 10 pt with interlinear spacing of 11 pt. Papers must be between 4 and 8 pages in length, including figures (plus more pages for references if needed), regardless of the mode of presentation (oral or poster).

2.1. General Instructions for the Final Paper

The unprotected PDF files will appear in the on-line proceedings directly as received. **Do not print the page number.**

3. Page Numbering

Please do not include page numbers in your Paper. The definitive page numbering of papers published in the proceedings will be decided by the Editorial Committee.

4. Headings / Level 1 Headings

Level 1 Headings should be capitalised in the same way as the main title, and centered within the column. The font used is Times New Roman 12 pt bold. There should also be a space of 12 pt between the title and the preceding section, and a space of 3 pt between the title and the text that follows.

4.1. Level 2 Headings

The format of Level 2 Headings is the same as for Level 1 Headings, with the font Times New Roman 11 pt, and the heading is justified to the left of the column. There should also be a space of 6 pt between the title and the preceding section, and a space of 3 pt between the title and the text that follows.

4.1.1. Level 3 Headings

The format of Level 3 Headings is the same as Level 2, except that the font is Times New Roman 10 pt, and there should be no space left between the heading and the text as in 4.1.1. There should also be a space of 6 pt between the title and the preceding section, and a space of 3 pt between the title and the text that follows.

5. Citing References in the Text

5.1. Bibliographical References

All bibliographical references within the text should be put in between parentheses with the author's surname followed by a comma before the date of publication, (Strötgen and Gertz, 2012). If the sentence already includes the author's name, then it is only necessary to

put the date in parentheses: Strötgen and Gertz (2012). When several authors are cited, those references should be separated with a semicolon: (Strötgen and Gertz, 2012; Castor and Pollux, 1992). When the reference has more than three authors, only cite the name of the first author followed by “et. al.”, e.g. (Superman et al., 2000).

5.2. Language Resource References

5.2.1. When Citing Language Resources

As you may know, LREC introduced a separate section on Language Resources citation to enhance the value of such assets. When citing language resources, we recommend to proceed in the same way as for bibliographical references. Please make sure to compile your Language Resources stored as a .bib file **separately** (Bib-Tex under pdfLaTeX). This produces the required .aux et .bbl files. Thus, a language resource should be cited as (?) and (?) .

6. Figures & Tables

6.1. Figures

All figures should be centred and clearly distinguishable. They should never be drawn by hand, and the lines must be very dark in order to ensure a high-quality printed version. Figures should be numbered in the text, and have a caption in Times New Roman 10 pt underneath. A space must be left between each figure and its respective caption.

Example of a figure enclosed in a box:



Figure 1: The caption of the figure.

Figure and caption should always appear together on the same page. Large figures can be centred, using a full page.

6.2. Tables

The instructions for tables are the same as for figures.

Level	Tools
Morphology	Pitrat Analyser
Syntax	LFG Analyser (C-Structure)
Semantics	LFG F-Structures + Sowa's Conceptual Graphs

Table 1: The caption of the table

7. Footnotes

Footnotes are indicated within the text by a number in superscript¹.

8. Copyrights

The Language Resources and Evaluation Conference (LREC) Proceedings are published by the European Language Resources Association (ELRA). They are available online from the conference website.

ELRA's policy is to acquire copyright for all LREC contributions. In assigning your copyright, you are not forfeiting your right to use your contribution elsewhere. This you may do without seeking permission and is subject only to normal acknowledgement to the LREC proceedings. The LREC 2020 Proceedings are licensed under CC-BY-NC, the Creative Commons Attribution-Non-Commercial 4.0 International License.

9. Conclusion

Your submission of a finalised contribution for inclusion in the LREC Proceedings automatically assigns the above-mentioned copyright to ELRA.

10. Acknowledgements

Place all acknowledgements (including those concerning research grants and funding) in a separate section at the end of the paper.

11. Providing References

11.1. Bibliographical References

Bibliographical references should be listed in alphabetical order at the end of the paper. The title of the section, “Bibliographical References”, should be a Level 1 Heading. The first line of each bibliographical reference should be justified to the left of the column, and the rest of the entry should be indented by 0.35 cm. The examples provided in Section 12 (some of which are fictitious references) illustrate the basic format required for papers in conference proceedings, books, journal articles, PhD theses, and books chapters.

11.2. Language Resource References

Language resource references should be listed in alphabetical order at the end of the paper.

Appendix: How to Produce the .pdf

In order to generate a PDF file out of the LaTeX file herein, when citing language resources, the following steps need to be performed:

- Compile the .tex file once
- Invoke bibtex on the eponymous .aux file
- Compile the .tex file twice

¹Footnotes should be in Times New Roman 9 pt, and appear at the bottom of the same page as their corresponding number. Footnotes should also be separated from the rest of the text by a 5 cm long horizontal line.

12. Bibliographical References

- Castor, A. and Pollux, L. E. (1992). The use of user modelling to guide inference and learning. *Applied Intelligence*, 2(1):37–53.
- Strötgen, J. and Gertz, M. (2012). Temporal tagging on different domains: Challenges, strategies, and gold standards. In Nicoletta Calzolari (Conference Chair), et al., editors, *Proceedings of the Eight International Conference on Language Resources and Evaluation (LREC'12)*, pages 3746–3753, Istanbul, Turkey, may. European Language Resource Association (ELRA).
- Superman, S., Batman, B., Catwoman, C., and Spiderman, S. (2000). *Superheroes experiences with books*. The Phantom Editors Associates, Gotham City, 20th edition.

13. Language Resource References