Natural Language Processing Frontiers in Djiboutian African Languages

Mohamed Ali[1], Hassan Ahmed[1]

[1] University of Djibouti

Author notes

Mohamed Ali is affiliated with University of Djibouti and focuses on Computer Science research in Africa.
Hassan Ahmed is affiliated with University of Djibouti and focuses on Computer Science research in Africa.

Abstract

Natural Language Processing (NLP) is a field within Computer Science that aims to enable machines to understand and process human language effectively. Despite its widespread application in English and other major languages, NLP for African languages has been underexplored. The methodology involves an extensive literature review to identify existing NLP research for African languages, followed by an empirical analysis using state-of-the-art machine learning models tailored for NLP tasks such as sentiment analysis and named entity recognition. A key aspect of the work is the development and evaluation of custom language-specific models. A significant finding was the variability in performance across different Djiboutian African languages, with some models achieving up to 85% accuracy in sentiment classification for Tigrinya, while others struggled with lower accuracies around 60%. This highlights the need for tailored approaches and increased data availability. The study concludes that while there are substantial challenges in developing NLP systems for African languages, these can be overcome through targeted research and collaboration. The development of custom models is crucial to achieving high accuracy across diverse linguistic contexts. Future work should prioritise the collection of more annotated data for underrepresented languages and explore hybrid approaches combining traditional machine learning with contemporary deep learning techniques. Natural Language Processing, African Languages, Djibouti, Custom Models, Sentiment Analysis Model estimation used $\hat{\{\theta\}} = argmin\{\theta\} sumiell(yi, f\theta(\xi)) + \lambda lVert\theta rVert 2^2$, with performance evaluated using out-of-sample error.

Keywords: *African languages, Computational linguistics, Data-driven methods, Morphology, Syntax, Translation systems, Word representation*

ABSTRACT-ONLY PUBLICATION

This is an abstract-only publication. The complete research paper with full methodology, results, discussion, and references is available upon request.

✉ REQUEST FULL PAPER

Email: info@parj.africa

Request your copy of the full paper today!

SUBMIT YOUR RESEARCH

Are you a researcher in Africa? We welcome your submissions!

Join our community of African scholars and share your groundbreaking work.

Submit at: app.parj.africa

Scan to visit app.parj.africa

Open Access Scholarship from PARJ

Empowering African Research | Advancing Global Knowledge