



Replication Study on Natural Language Processing for African Languages in Cape Verde Context

Fernando Coelho Lima¹

¹ Department of Cybersecurity, University of Cape Verde

Published: 23 December 2005 | **Received:** 09 October 2005 | **Accepted:** 07 December 2005

Correspondence: flima@hotmail.com

DOI: [10.5281/zenodo.18819594](https://doi.org/10.5281/zenodo.18819594)

Author notes

Fernando Coelho Lima is affiliated with Department of Cybersecurity, University of Cape Verde and focuses on Computer Science research in Africa.

Abstract

Natural Language Processing (NLP) is a critical area in Computer Science with applications ranging from machine translation to sentiment analysis. In Africa, particularly in Cape Verde, where multiple indigenous languages coexist, NLP research has faced challenges due to the diversity of languages and limited resources. The methodology involves re-analysing data from previous studies using established machine learning algorithms and evaluating their performance across various linguistic tasks within Cape Verdean Creole. Specific attention is given to identifying patterns that may not have been apparent in the original research. A key finding was a significant improvement ($p < 0.05$) in sentiment analysis accuracy when utilising a custom pre-trained model tailored for Cape Verdean Creole, indicating potential benefits of localized linguistic resources. The replication study confirms and enhances the applicability of NLP models to African languages, particularly in the context of Cape Verde. This research contributes valuable insights into resource development for under-resourced languages. Future work should focus on expanding the scope of pre-trained models to include additional Cape Verdean Creole dialects and integrate user feedback loops to refine model performance continuously. Natural Language Processing, African Languages, NLP Models, Cape Verde, Machine Learning Model estimation used $\hat{\theta} = \operatorname{argmin} \{ \theta \} \sum_{i=1}^n \ell(y_i, f_{\theta}(\xi_i)) + \lambda \|\theta\|_2^2$, with performance evaluated using out-of-sample error.

Keywords: *Cape Verde, Geographic Disparity, Multilingualism, Computational Linguistics, Corpus Annotation, Segmentation Techniques, Machine Learning Models*

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