

الجمهورية الشعبية الديمقراطية الجزائرية  
République Algérienne Démocratique et Populaire  
وزارة التعليم العالي و البحث العلمي  
Ministère de l'Enseignement Supérieur et de la Recherche Scientifique  
المدرسة العليا للإعلام الآلي .08 ماي 1945. بسيدي بلعباس  
École Supérieure en Informatique  
-08 Mai 1945- Sidi Bel Abbès



## THESIS

To obtain the diploma of **Engineer**  
Field: **Computer Science**  
Specialty: **Ingénierie des Systèmes Informatiques (ISI)**

### Theme

---

**Sentiment analysis in Algerian dialects considering  
Code-switching**

---

Presented by:  
**ZELLAGUI Iskander**

Submission Date: **July, 2023**  
In front of the jury composed of:

Mrs. <b>KLOUCHE</b> Badia	President
Mrs. <b>TAOULI</b> Amina	Examiner
Mrs. <b>ALIANE</b> Hassina	Supervisor
Mr. <b>KHALDI</b> Belkacem	Co-Supervisor
Mr. <b>ALIANE</b> Ahmed Amine	Co-Supervisor

*Academic Year : 2022/2023*

## Acknowledgement

First and before all, Alhamdulillah, all praises to Allah, for helping, guiding and giving us the patient and motivation to complete this humble work.

We offer our sincerest gratitude and thanks to our supervisor Dr ALIANE Hassina, the Head of Natural Language Processing and Digital Content team, Dr KHALDI Belkacem, and ALIANE Amine, for helping us during this journey, and guiding us through this study with their knowledge, support, patience, and motivation.

It was a great privilege and honor to work under their guidance.

This work could not have been achieved without them.

We would also like to offer our thanks to the jury members, for spending their precious time reviewing our work.

We would like to express our sincere thanks and appreciation to both our parents and siblings, for their support and encouragement throughout our academic journey. We would like to extend our thanks to our friends, especially BENSALAH Amdjed, NASRI Mohammed, and BOUCHOUKA Hidaya, for the great help they have assisted us with.

Last but not least, we would like to thank all the professors and staff of ESI-SBA for all their efforts.

# Abstract

Arabic sentiment analysis in the Algerian Dialect considering code-switching is a difficult classification task that can allow us to understand the sentiment between human-generated content in social media.

Although, the code-mixed type of content has been on the rise lately, with the continuous rise of social media use among all population's categories, research around it has been rare due to the difficulty of the problem, and the scarcity of the needed data.

Up till now, most of the solutions that have been adapted to deal with the problem need manually annotated data-sets, or pre-trained models with non-contextual word embeddings, due to the hard nature of the written Algerian dialects on social media, which makes their ability to deal with the problem at hand limited.

In our work, we present an architecture that aggregates the most suited BERT-Based models for our study of the Algerian dialect, using aggregation methods that help us solve the problem of code-switching in text.

The proposed architecture has been evaluated on the code-switched sub-data-set of the CERIST data-set, and on a data-set that has been manually collected, and the results achieved in both show that our architecture achieved good results.

Being tested on the CERIST, the manually collected data-sets, our architecture achieved an accuracy of 77% and 64% respectively.

*Keywords:* Sentiment Analysis, Social Media, Deep learning, Algerian Dialects, Code-switching, Arabic language, DziriBERT, MarBERT.