Afro-Asian Journal of Scientific Research (AAJSR)
المجلة الأفروآسيوية للبحث العلمي
E-ISSN: 2959-6505
Volume 3, Issue 4, 2025
Page No: 175-179
Website: https://aajsr.com/index.php/aajsr/index
SJIFactor 2024: 5.028 ISI 2025: 0.915 معامل التأثير العربي (AIF) 2025: 0.76

Analysis of Libyan Dialect Sentiment Using AraBERT and MARBERT: A Comparative Study

Souad Al-Haram[1*], Fatma Howedi[2]
[1]Department of Internet Technologies, Faculty of Information Technology, Al-Asmariya Islamic University, Zliten, Libya
[2] Department of Computer Science, Faculty of Information Technology, Al-Asmariya Islamic University, Zliten, Libya

تحليل مشاعر اللهجة الليبية باستخدام AraBERT و MARBERT: دراسة مقارنة

سعاد الهرم [1*]، فاطمة هويدي[2]
[1]قسم تقنيات إنترنت، كلية تقنية المعلومات، الأسمرية الإسلامية، زليتن، ليبيا
[2]قسم علوم حاسوب، كلية تقنية المعلومات، الأسمرية الإسلامية، زليتن، ليبيا

*Corresponding author: s.alharm@it.asmarya.edu.ly

Received: September 21, 2025 | Accepted: December 07, 2025 | Published: December 20, 2025

Abstract:
Because of the wide range of spoken dialects, the Libyan dialect is regarded as an uncommon digital resource and presents significant challenges for Arabic natural language processing. By providing an experimental comparative analysis of the capability of two cutting-edge language models based on this gap, this study seeks to address and evaluate it. Trained in the Libyan dialect to use the transformer architectures AraBERTvs and MARBERTv2 for sentiment analysis. A new dataset of 500 comments collected from social media sites was created and categorized. Both models were fine-tuned using the same parameters and evaluated on a specially designed test set. The AraBERTv2 model achieved 71. 0% accuracy, compared to the MARBERTv2 model's 70. 0%, demonstrating a slight advantage in the results. Notably, a study of training practices revealed that MARBERTv2 had early over fitting, whereas AraBERTv2 exhibited superior stability and enhanced learning ability. When applied to resource-constrained local vernaculars, the research demonstrates that models trained on various Modern Standard Arabic, such as AraBERT, may provide a more robust and generalizable foundation.

Keywords: Sentiment; Libyan Dialect; AraBERT, MARBERT, Arabic NLP.

الملخص:
تعتبر اللهجة الليبية من الموارد الرقمية النادرة فتواجه معالجة اللغات الطبيعية العربية تحديات كبيرة بسبب التنوع الهائل في اللهجات العامية، تهدف هذه الدراسة إلى معالجة وتقييم هذه الفجوة من خلال تقديم دراسة مقارنة تجريبية لقدرة نموذجين لغويين من نماذج اللغة المتقدمة القائمة على بنية المحولات، AraBERTvs و MARBERTv2، تم التدريب على أداء مهمة تحليل المشاعر باللهجة الليبية. تم بناء وتصنيف مجموعة بيانات جديدة مكونة من 500 تعليق تم جمعها من منصات التواصل الاجتماعي. تم صقل (Fine-tuning) كلا النموذجين باستخدام نفس الإعدادات وتقييمهما على مجموعة اختبار مخصصة. أظهرت النتائج تفوقًا طفيفًا لنموذج AraBERTv2 الذي حقق دقة بلغت 71.0%، مقارنة بـ 70.0% لنموذج MARBERTv2.والأهم من ذلك، كشف تحليل سلوك التدريب أن AraBERTv2 أظهر استقرارًا أكبر وقدرة أفضل على التعلم، بينما عانى MARBERTv2 من فرط تخصيص مبكر. تستنتج هذه الدراسة أن النماذج المدربة على اللغة العربية الفصحى المتنوعة مثل AraBERT قد توفر أساسًا أكثر قوة وقابلية للتعميم عند تكييفها للهجات الإقليمية قليلة الموارد.

الكلمات المفتاحية: المشاعر، اللهجة الليبية، ارابريت، مارابريت، البرمجة اللغوية العصبية الليبية.

Introduction:

In recent years, the field of Natural Language Processing (NLP) has witnessed a real revolution driven by tremendous advances in deep learning models. driven by the tremendous advancements in deep learning models. Traditional language models heavily relied on statistical methods or Recurrent Neural Networks (RNNs) and Convolutional Neural Networks (CNNs), but they faced challenges in understanding long-range context and complex relationships between words. The emergence of the Transformer architecture fundamentally changed this landscape, introducing the Attention Mechanism, which allowed models to process long sequences of data with unprecedented efficiency and effectiveness [1].

Social media platforms such as Facebook and Twitter have become a repository for big data and a primary space for expressing opinions and feelings about various political, social, and economic issues in the Arab world [2]. Representing this vast amount of textual data produced by users daily is a valuable treasure that can be analyzed to understand public opinion trends, measure customer satisfaction, and monitor important events. Sentiment analysis, a branch of natural language processing, is the main tool for converting these unstructured texts into measurable insights [3].

Sentiment analysis applications for Arabic content encounter unique challenges, primarily stemming from the vast diversity of colloquial dialects. These dialects differ significantly from Modern Standard Arabic (MSA), which is used in formal texts [4]. The Libyan dialect is a case in point; as a regional dialect, it possesses its own distinct vocabulary, syntax, and specific grammatical rules. Consequently, language models trained on MSA are incapable of comprehending it with high accuracy.

This issue is compounded by a severe scarcity of digital linguistic resources and annotated datasets for the Libyan dialect, which has led to a dearth of research into its computational analysis, thus classifying it as a low-resource dialect. However, a new opportunity to address this gap has emerged with the advent of pre-trained Transformer-based language models, such as AraBERT and MARBERT. These models, trained on massive volumes of Arabic texts encompassing various dialects, possess the capability to understand complex contexts and can be adapted through fine-tuning to perform specific tasks on dialects for which they were not directly trained.

Accordingly, this study aims to investigate the capability of two leading language models, AraBERT and MARBERT, to comprehend the nuances of the Libyan dialect for sentiment analysis. To this end, the paper makes two primary contributions: We construct and annotate a novel dataset for Libyan sentiment analysis, comprising 500 comments. We conduct a comparative experimental study to evaluate and determine which model delivers superior and more stable performance when fine-tuned for this low-resource dialect.

Literature Review:

Evolution of Arabic Sentiment Analysis Models:

The Support Vector Machine (SVM), a traditional machine learning method, has made amazing progress in the field of Arabic sentiment analysis. and Naive Bayes, which was based on manual feature engineering, to sophisticated linguistic models. Many studies, like [5,6], have shown that deep learning models are more accurate in sentiment categorization than conventional machine learning models.

Emergence of Transformer Models and Their Adaptation to Dialects:

With the advent of Transformer models, a paradigm shift occurred in the field. These models have proven their effectiveness not only with Modern Standard Arabic but also in their ability to adapt to colloquial dialects. For instance, the foundational study of the AraBERT model [5] demonstrated its efficacy in handling the Egyptian dialect. This success has extended to North African dialects, with a recent study [5-7] confirming AraBERT's superiority over other models like CAMeLBERT in Moroccan dialect sentiment analysis.

Performance Comparisons Among Different Models:

The selection of the optimal model is not always straightforward, necessitating comparative studies. In a study by [8] on English language data, the RoBERTa model showed superiority over other models such as BERT and XLNet. In the context of the Arabic language, a study by [9] revealed variations in the performance of AraBERT and MARBERT depending on the nature of the task. Other studies, such as that by [10-12], have shown that models like XLNet may outperform AraBERT in certain contexts. This variability underscores the absence of a one-size-fits-all solution and highlights the urgent need for customized experimental evaluations for each dialect.

Recent Trends in Model Development:

Recent research aims to further enhance the performance of these models. On one hand, studies such as [9] emphasize the importance of fine-tuning hyperparameters and using techniques like early stopping to prevent overfitting. On the other hand, innovative studies like [10] propose integrating

AraBERT outputs with other models such as Long Short-Term Memory (LSTM) to capture long-term dependencies in text, which has led to achieving exceptional accuracy exceeding 97% in specific tasks.

Research Gap and the Position of the Current Study:

Despite this significant progress, a clear gap is observed in the literature regarding dialects suffering from resource scarcity, particularly the Libyan dialect. Most of the mentioned studies focused either on more common dialects (such as Egyptian and Gulf) or on English or Modern Standard Arabic data. From this standpoint, the current study addresses this notable research deficiency by presenting the first comparative experimental evaluation of AraBERT and MARBERT models on a Libyan dataset specifically built and classified for this purpose, focusing not only on final accuracy but also on analyzing training behavior and each model's generalization capability.

Research Methodology:

This section is divided into three parts to explain how this work was accomplished:

Data Collection and Annotation:

The data collection stage is a crucial and fundamental part of any analytical study. In the shadow of the digital revolution, social media platforms like Facebook and Twitter have become rich repositories of data reflecting public opinion and community interactions. Given the scarcity of available datasets for sentiment analysis in the Libyan dialect,a specialized dataset was built manually from social media platforms, Facebook and Twitter, during July and August 2025. The focus was on comments related to the topic of "my salary is late" (a local economic issue), to ensure genuine and organic opinions. The aim of data collection is the direct review and extraction of content, whether a comment or a post, without relying on automation tools or programming interfaces for applications. This highlights its importance in understanding the Libyan dialect and deciphering sentiments correctly. In this section, we will focus on explaining the meticulous methodological framework for data collection, **including:**

- **Defining Selection Criteria:** Relevant hashtags (#) or groups and accounts pertinent to the event on each platform were identified.
- **Units of Analysis:** A dataset of 500 comments representing various Libyan dialects was constructed.
- **Collection Procedures:** The raw data was documented, recorded, and organized using Google Sheets. Each comment was then classified into one of three categories: "positive," "negative," or "neutral."
- Short, unclear comments were excluded.
- Finally, the overall dataset was randomly divided into 80% for training (400 samples) and 20% for testing (100 samples), while maintaining the same distribution ratio for classifications in both sets.

Models Used:

Language models refer to the pre-trained models that leverage existing linguistic knowledge. This approach, known as transfer learning, is currently very popular due to its effectiveness and speed, and its need for rich data to train deep neural networks [5].

In this study, the latest Arabic language processing models based on the transformer architecture were evaluated and compared, **specifically:**

AraBERTv2 is an advanced model specifically designed for the Arabic language. It is considered an evolution of the original AraBERT model and has shown significant improvements in performance and accuracy. This model relies on the Transformer bidirectional encoder (BERT) architecture. It is one of the most powerful pre-trained models for handling various Arabic tasks, such as sentiment analysis, automatic summarization, and other natural language processing tasks [13].

AraBERT:

Unlike Multilingual BERT, AraBERT was developed exclusively for the Arabic language. Its pretraining corpus primarily consists of Modern Standard Arabic news articles sourced from various Arabic media outlets. The initial version of AraBERT (Version 1) was trained on approximately 23 gigabytes of text, encompassing 77 million sentences and 2.7 billion tokens, which significantly exceeded the size of the Arabic pretraining dataset used for Multilingual BERT by a factor of 17. Subsequent iterations, specifically the most recent version, leveraged an even larger dataset of 77 gigabytes of text, representing a 3.5-fold increase in pretraining data. Architecturally, AraBERT shares similarities with Multilingual BERT, featuring 12 transformer blocks, each with 768 hidden units. It also incorporates 12 self-attention heads, culminating in a total of 110 million trainable parameters [5].

MARBERTv2:

represents a significant advancement in natural language processing for the Arabic language, specifically engineered to interpret both vernacular dialects and Modern Standard Arabic (MSA) as used on social media. The model was pre-trained on a substantial corpus of one billion tweets, amounting to 128 GB of text and approximately 15.6 billion tokens. This pre-training dataset is notably twice the size

of that used for AraBERTv2, establishing it as the largest among a cohort of nine comparable models. While architecturally similar to multilingual BERT, MARBERTv2 diverges by excluding the Next-Sentence Prediction (NSP) task, a design choice justified by the developers on the basis of the brief, fragmented nature of Twitter data. The model comprises approximately 160 million trainable parameters, positioning it as a robust tool for Arabic-centric NLP research and applications. [13-15]

Experimental Setup:

This section describes the practical settings and encodings used to evaluate the performance of the language models that underwent the fine-tuning process for both models (AraBERTv2:MARBERTv2). The same settings were used to ensure a fair comparison. The dataset, consisting of 500 samples, was split into 80% for training and 20% for testing. The Hugging Face Transformers library and the Google Colab environment, equipped with a Graphics Processing Unit (GPU), were used. The following hyperparameters were set: number of training epochs = 3, batch size = 8, and learning rate = 5e-5. Accuracy and F1-score (weighted) were used as evaluation metrics to assess the models' performance on the test set of 100 samples. These are common practices in similar research.

Results:

Two models, AraBERTv2 and MARBERTv2, were tested in the task of sentiment analysis for the Libyan dialect using the test dataset. Table 1 shows the final results for both models.

Table (1): Final Results for Both Models

SN.	Model Type	Accuracy	F1-Score (Weighted)
1	MARBERTv2	0.7000	0.6737
2	AraBERTv2	0.7100	0.6751

As shown in the table, the results indicated that AraBERTv2 slightly outperformed MARBERTv2 in both Accuracy and F1-Score. Training behavior revealed substantial differences, as illustrated in Figure 1. In the first training epoch, MARBERTv2 reached its peak performance, after which its performance began to decline, indicating overfitting. In contrast, AraBERT showed more stable behavior, continuing to improve until the second epoch before starting a slight decline, indicating better learning and generalization from the available data.

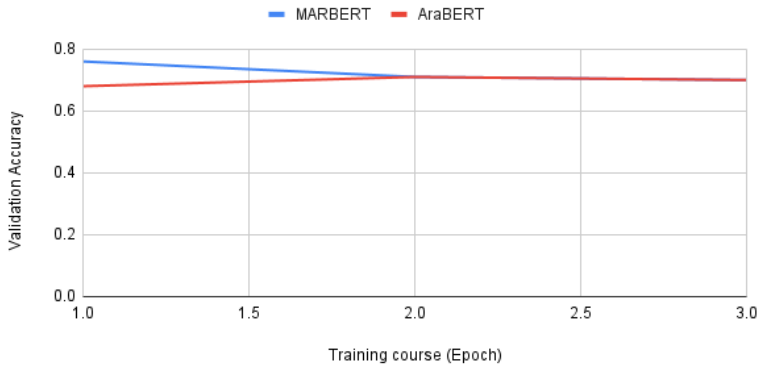

Figure (1): Comparison of verification accuracy behavior between MARBERT and AraBERT.

Discussion:

This study focused on the performance of AraBERTv2 and MARBERTv2 models in the context of the low-resource Libyan dialect. The most prominent finding was the slight superiority of AraBERTv2 with an accuracy of 71.0% compared to MARBERTv2's 70.0%. Although this difference might seem small, further analysis of the training epochs provided a deeper understanding. AraBERTv2 demonstrated a better ability for continuous learning and benefiting from additional training cycles, while MARBERTv2 showed clear signs of early overfitting. This suggests that the foundational AraBERT model, trained on diverse Arabic texts, inherently offers greater robustness and adaptability for generalization, particularly when compared to MARBERTv2, which may be more susceptible to noise in smaller, specialized datasets due to its intensive training on dialectal data.

These findings partly align with what others [13-15] have found, who indicated that while MARBERT excels in Twitter-like data, our study demonstrated that AraBERT is more stable when dealing with diverse local and limited datasets. Our results also support the conclusions of other research [6] regarding AraBERT's ability to adapt to various dialects.Despite the promising results, certain limitations of this study must be acknowledged. Firstly, the main constraint is the size of the dataset; although it represents an initial contribution (500 samples), it still does not fully cover the diversity and linguistic

nuances of the Libyan dialect. Secondly, the study was limited to a tripartite classification task (positive, negative, neutral) and did not delve into the analysis of finer emotions (e.g., anger, joy, sadness). Finally, standard default settings were used for the models, which suggests that conducting extensive hyperparameter tuning could further improve the results

Conclusion:

In conclusion, this study conducted an experimental comparison to evaluate the ability of both AraBERT and MARBERT models in sentiment analysis for the Libyan dialect. A new dataset was built and classified for this purpose. The study demonstrated that both models are capable of achieving good performance, with AraBERTv2 showing a slight superiority and greater stability. These results underscore the potential of pre-trained language models as an effective tool for bridging the resource gap in Arabic dialects and highlight the importance of analyzing model training behavior for generalization capability.

Recommendations:

Based on the above, we recommend expanding the current dataset to include thousands of instances and organizing it into diverse sub-topics and dialects within Libya. We also recommend applying these models to more specific tasks such as Aspect-Based Sentiment Analysis to gain a deeper understanding of user opinions. Finally, comparative studies can be conducted with the latest state-of-the-art models in this field. We anticipate increased confidence with larger training datasets

References

1. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., ... & Polosukhin, I. (2017). Attention is all you need. Advances in neural information processing systems, 30.
2. Hamdi, A., Shaban, K. & Zainal, A. (2016). A Review on Challenging Issues in Arabic Sentiment Analysis. Journal of Computer Science, 12(9), 471-481. https://doi.org/10.3844/jcssp.2016.471.48
3. AlOtaibi, S., & Khan, M. B. (2017). Sentiment analysis challenges of informal Arabic language. International Journal of Advanced Computer Science and Applications, 8(2).
4. M. Abdeldaiem Mahboub and T. Gopi Krishna, “An identification model used for Arabic Libyan Dialects based on machine learning approach,” *Int. J. Adv. Res. (Indore)*, vol. 12, no. 03, pp. 889–902, 2024.
5. El Jundi O, Antoun W, El Droubi N, Hajj H, El-Hajj W, Shaban K. hULMonA: the universal language model in Arabic. In:Proceedings of the Fourth Arabic Natural Language Processing Workshop. 2019.p. 68–77.
6. Antoun, W.; Baly, F.; Hajj, H. Arabert: Transformer-based model for arabic language understanding. arXiv 2020, arXiv:2003.00104.
7. Abdul-Mageed, M.; Zhang, C.; Elmadany, A.; Ungar, L. Toward micro-dialect identification in diaglossic and code-switched environments. In Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing, Online, 8–12 November 2020; pp. 5855–5876.
8. Areshey, A., & Mathkour, H. (2024). Exploring transformer models for sentiment classification: A comparison of BERT, RoBERTa, ALBERT, DistilBERT, and XLNet. Expert Systems, 41(11), e13701..
9. Mathew, L., & Bindu, V. R. (2021, August). Efficient classification techniques in sentiment analysis using transformers. In International Conference on Innovative Computing and Communications: Proceedings of ICICC 2021, Volume 1 (pp. 849-862). Singapore: Springer Singapore
10. Alosaimi, W., Saleh, H., Hamzah, A. A., El-Rashidy, N., Alharb, A., Elaraby, A., & Mostafa, S. (2024). ArabBert-LSTM: improving Arabic sentiment analysis based on transformer model and Long Short-Term Memory. Frontiers in Artificial Intelligence, 7, 1408845
11. El Karfi, I., & El Fkihi, S. (2022). An ensemble of Arabic transformer-based models for Arabic sentiment analysis. International Journal of Advanced Computer Science and Applications, 13(8).
12. Alduailej, A., & Alothaim, A. (2022). AraXLNet: pre-trained language model for sentiment analysis of Arabic. Journal of Big Data, 9(1), 72.
13. Alammary, A. S. (2022). BERT Models for Arabic Text Classification: A Systematic Review. Applied Sciences, 12(11), 5720. https://doi.org/10.3390/app12115720
14. Alturayeif, N., & Luqman, H. (2021). Fine-grained sentiment analysis of arabic covid-19 tweets using bert-based transformers and dynamically weighted loss function. Applied Sciences, 11(22), 10694
15. Areshey, A., & Mathkour, H. (2024). Exploring transformer models for sentiment classification: A comparison of BERT, RoBERTa, ALBERT, DistilBERT, and XLNet. Expert Systems, 41(11), e13701