VAWKUM Transactions on Computer Sciences
http://vfast.org/journals/index.php/VTCS@ 2024, ISSN(e): 2308-8168, ISSN(p): 2411-6335
Volume 12, Number 2, July-December 2024 pp: 178-190

A FINE GRAINED SENTIMENT ANALYSIS OF ARABIC LANGUAGE

Asim Tanveer[1*], **Mohibullah khan**[1], **Rehan Sarwar**[1], **Naeem Aslam**[1], **M Fuzail**[1]

[1]Department of Computer Science, NFCIET ,Multan, Pakistan

Keywords:*Sentiment Analysis,Machine Learning, AraBERT, Arabic Language.*

Journal Info:
Submitted:
October 05, 2024
Accepted:
November 30, 2024
Published:
December 04, 2024

Abstract

This work focuses on fine-grained sentiment analysis of Arabic text using recent Natural Language Processing methods. Arabic is a language rich in variation, spoken by over 400 million people, yet there is a significant lack of resources for sentiment analysis. To address these challenges, this study employs AraBERT, a model specifically fine-tuned for Arabic text. A corpus of one hundred thousand Arabic reviews across categories such as hotels, books, and movies was scraped and cleaned. These reviews were then categorized into positive, negative, and mixed sentiments. AraBERT was compared with traditional machine learning methods, including Logistic Regression, Decision Tree, Naïve Bayes, and Random Forest. AraBERT achieved superior accuracy of 88%, along with higher precision, recall, and F1 scores for both positive and negative sentiment classes compared to the other models. This work demonstrates that AraBERT effectively analyzes the syntactic and semantic structure of Arabic, making it a valuable tool for Arabic sentiment analysis across various applications. Future work will extend the model to handle neutral sentiments and include additional dialects to further improve its performance.

***Correspondence author email address:** adverbcollin786@gmail.com
DOI: 10.21015/vtcs.v12i2.1926

1 Introduction

Digital communication has continued to expand at a high rate, whereby users have produced massive content in the form of reviews, social media posts, and comments, among others. Utilizing this data to derive insights has become crucial for business communities, policymakers, and researchers. Probably the most popular type of such analysis is sentiment analysis, a method that seeks to classify opinions present in text, especially when categorized as positive, negative, or neutral [1]. Sentiment analysis is an area predominantly researched in languages such as English, while there is a pressing need for similar efforts in other languages like Arabic, which presents its own challenges and opportunities.

NLP is a branch of AI and computational linguistics dedicated to studying how to enable computers to process natural language of any kind. It encompasses several processes, such as language comprehension, language synthesis, and opinion mining. One of the most important areas traditionally associated with NLP is sentiment analysis, which focuses on identifying the emotional undertone of text. Arabic is among the most widely spoken languages globally, boasting over 400 million native speakers across twenty-two countries [2]. It is also one of the six official languages of the United Nations and one of the three languages used in the Catholic Church. However, despite its global significance, Arabic remains underrepresented in NLP studies, particularly in the field of sentiment analysis. Issues specific to Arabic in sentiment analysis are diverse, including the language's morphological complexity, dialectal variations, and unique writing system [3].

Given that Arabic-speaking markets are critically important to the global economy, there is an urgent need for advanced and effective tools for sentiment analysis of Arabic text. However, the utility of these tools varies depending on the machine learning algorithms and techniques they employ [4]. This brings us to the second major focus of this research: determining which method is most efficient for sentiment analysis of Arabic datasets.

The increasing use of the internet in Arab countries has led to rapid growth in the number of posts and shares on social networks, shopping platforms, and other websites containing user reviews. With this abundance of data, businesses, policymakers, and researchers have the potential to gain deeper insights into the moods and attitudes of Arabic-speaking communities. However, accurately evaluating the sentiment of Arabic content remains a challenge due to the unavailability of suitable tools [5].

The third objective of this study is to contribute to NLP by addressing the challenges faced in analyzing the Arabic language. Arabic is a highly inflected language with many roots and patterns, making traditional machine learning algorithms less effective for sentiment analysis. Current methods often face limitations when applied to Arabic, leading to compromised accuracy and suboptimal results [6]. To address this issue, this research aims to identify the most effective machine learning techniques for fine-grained sentiment analysis of Arabic text. By comparing the predictive capabilities of these models, this study seeks to enhance sentiment analysis methodologies and offer a robust solution for evaluating the sentiment of Arabic text data in various contexts.

The objectives of the study are listed below:

- Identify diverse trends within the vast dataset of Arabic language content that can be utilized for sentiment analysis.
- Propose a machine learning-based Arabic language text categorization model for fine-grained sentiment analysis.
- Compare the performance of the proposed model with that of traditional machine learning models used for sentiment analysis in Arabic.

2 Literature Review

A lot of research fields have an important component of identifying text sentiment. In this chapter, we will describe some related work where other researchers put their efforts into solving these challenges of sentiment analysis, and also, we briefly describe the approaches they used to solve them. In this review chapter, we will discuss previous researchers' findings and also identify research gaps that can lead to new discoveries and studies.

The study [7] identifies some of the challenges as being limited ASA sources, the Arabian dialects, and inadequate machine translation from the English language. It offers an overview of monolingual and bilingual strategies and their efficiency, particularly regarding dialectical differences and informal Arabic. The study reveals directions for further improvement of ASA and highlights gaps found when compared with the current state of the literature, such as the need for a better number and quality of sources and better adaptation of techniques to the specifics of the Arabic language.

The study [8] proposes a comprehensive framework for Arabic Language Sentiment Analysis (ALSA) that integrates six linguistic levels, including phonetics, morphology, syntax, lexicology, semantics, and the use of figurative language. It focuses on Arabic's rhetorical features and its hidden meanings, which are not considered by existing sentiment analysis programs. The study suggests several directions for further development, such as creating annotated corpora and conducting detailed phonetic and figurative analyses to improve sentiment detection accuracy. This framework aims to fill the gaps identified in present research, offering a more comprehensive approach to understanding sentiment in the Arabic language and enhancing Natural Language Processing applications for Arabic texts.

The study [2] shows that ANLP is gaining importance due to its use of ML algorithms. It emphasizes difficulties related to the Arabic language, such as letter combinations and the removal of diacritics, which hinder ANLP tasks. The survey highlights major ML algorithms employed in ANLP and their usefulness in solving intricate language projects without coding. Furthermore, this paper discusses the characteristics of ANLP, the requirements of ANLP tools, and the challenges in creating these tools, providing scientific reference values for future studies in ANLP tool development.

The study [4] explores Asian Arabic Sentiment Analysis by analyzing deep and machine learning models. Four classifiers—LSTM, SVM, LR, and KNN—are applied to the ARev dataset, which contains manually labeled Arabic comments from various sources. SVM and LR achieved the highest accuracy, with 92

The study [9] examines the difficulties inherent to ASA and the methods that may be employed using machine learning tools. Key challenges include the non-availability of resources like annotated corpora and the complexities posed by Arabic morphology and dialects. The investigation promotes new features suitable for Arabic, covering morphological, semantic, and stylistic analyses, and evaluates linear and nonlinear classifiers. It also examines the effect of negation on sentiment classification, showing that ASA performance improves with negation handling. The findings pave the way for future studies to explore ways of improving ASA by leveraging applications from other domains and advanced ML methodologies.

The study [6] provides a deep learning architecture for Arabic text sentiment classification, discussing four models based on deep belief networks, deep autoencoders, and recursive autoencoder models. The proposed model for Relevance and Acknowledgement Enhancement (RAE) addresses the lack of context handling in previous approaches, showing substantial improvement over current state-of-the-art methods. The study highlights the superior performance of the RAE model compared to other models while eliminating the dependency on expensive and lexicon-limited resources. Additionally, it supports the potential for improving Arabic sentiment analysis accuracy through context-aware deep learning models.

The study [10] the authors have proposed a detailed Sentiment Analysis framework for the Arabic tweets about COVID-19 using AraBERT and MARBERT models. Overall, the paper stresses the use of emojis in boosting the sentiment analysis and proposes a concept of the dynamically weighted loss function to minimize the effects of an imbalanced dataset. Evidently, the outcomes show that by applying the proposed method alongside the incorporation of emojis and DWLF enhances the accuracy of the multi-dialect Arabic sentiment analysis to get an F1-Micro of 0.73 on the same dataset known as SenWave. This work demonstrates the benefits of fine-tuning the state-of-the-art models for stereospecific affect analysis in multifaceted social media environments.

The study [11] identifies key challenges in Arabic sentiment analysis from two perspectives: there are some specific Arabic issues and some of the general linguistic problems. The difficulties peculiar to Arabic are as follows: Arabic is morphologically complex and has hardly any resources; there are also many dialects. Common language problems include polarity vagueness, strength of the positive or negative feelings conveyed, latent sentiments, irony, noise in the form of spam, quality of the reviews submitted, and dependencies more of the domain type. The study highlights that these challenges have a considerable impact on the effectiveness of the sentiment analysis on the Arabic content, which is highly important due to the growing amount of Arabic content available on the web. Mitigating these challenges is critical to enhancing sentential analysis depending on its application in business and government decisions.

The study [12] discusses the findings of various researches on Arabic as a Foreign Language-AFL learning websites focusing on their main issues and challenges. It also reveals that there are limited online resources specifically for the teaching of Arabic, hence the need for more resources to be posted online. Indeed, the analysis shows that the current base of websites is insufficient in its scientific content, utility, and adaptability to foreign language learning. In relation to this, the paper suggests the development of many additional websites offering the Arabic language with the integration of more educational institutions from the Arab world to improve the standards. This would improve the accessibility, functionality, and dissemination of Arabic language education, particularly for students who have to learn it as a second language.

The study [13] proposes three new sentiment analysis datasets in Turkish and in English as well as in Arabic language. They assess the efficiency of the further pre-processing of the multilingual languages with different complexity levels and spelling mistakes, as well as the selected deep learning and machine learning algorithms for the pre-trained language models. Thus, the main finding suggests that some methods are more accurate in different languages because of their grammatical properties, such as the spelling, and other language peculiarity affects the results of the model. The theoretical contributions offer important implications for researchers using the technique of sentiment analysis by indicating which techniques are most appropriate for a given language and when text complexity becomes a key issue.

3 REEARCH METHODOLGY

This section discusses the method utilized in the construction as well as the assessment of the machine learning model for fine-tuning the sentiment analysis of Arabic text. The approach starts with data gathering, data preparation, and feature selection from the Arabic language dataset. For sentiment analysis, traditional as well as deep learning models are used as applied to the process to learn the required sentiment patterns. We also discuss how to deal with characteristics of Arabic, such as the morphological and dialectal variations. The result of the developed model is then compared to the existing models, in terms of accuracy and F1-score, to evaluate the efficiency of the model in sentiment classification.

This diagram shows the sequence of activities involved in fine-tuning the ARA-BERT for sentiment analysis using 100,000 Arabic reviews. The accumulation of the reviews is the first step, which is followed by the pre-processing of the textual data to feed into the BERT model. Some of the operations performed as part of pre-processing may include tokenization, filtering out unwanted characters, as well as addressing distinctive features in the Arabic language. The data goes through the pre-processing stage and is then fed into the BERT model, where ARA-BERT is used since it has been trained for the Arabic language. Token, segment, and position mappings are illustrated at the top of the diagram: token embedding represents the specific token, segment embedding represents the segment type, and position embedding represents the relative position of the token.

The input data is divided into three subsets: training, validation, and testing. The training subset is employed to help the model learn how patterns in the data appear. The validation set is used to adjust chosen hyperparameters and evaluate the model for overfitting during the training process. Finally, the testing set measures the model's performance on the new dataset, providing a realistic view of how well the model performs.

Finally, after the training and validation processes are to be finished, the fine-tuned BERT model is saved as the trained model. Subsequently, this model is then prepared to predict something, like the kind of new Arabic text reviews; is it a positive, negative or neutral polarity? The whole process is a general ML process for text classification using the highly effective model ARA-BERT for Arabic language reviews sentiment analysis.

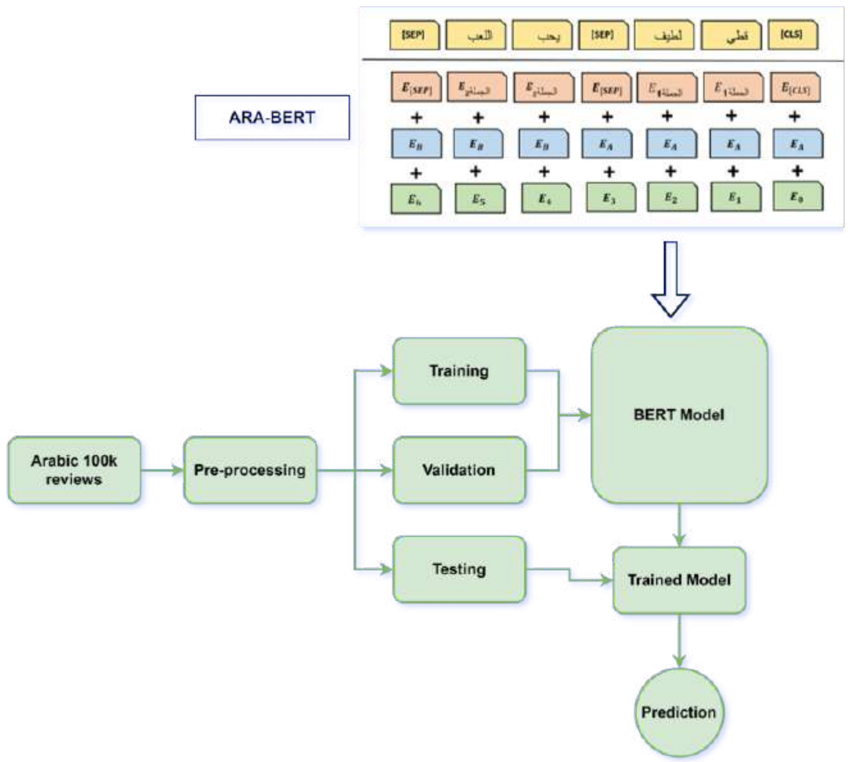

Figure 1. Proposed model for Sentiment Analysis for Arabic Language

3.1 Dataset

The dataset is a raw collection of 100,000 Arabic text reviews scraped from social media, blogs, and forums, mainly for classification tasks in the field of NLP. It includes reviews from different domains such as hotels, books, movies, products, and a few cases from airlines as well. The dataset is organized into three sentiment classes: Positive, Negative, and Mixed Transfer Orientation categories. These are classified in accordance with but not limited to the following category: Negative for reviews that are rated below 3, Mixed if the review is rated 3, and Positive for reviews that are rated above 3. Specifically, the removal of Arabic diacritical marks and non-Arabic characters is performed for better understanding. Compared with the Twitter dataset, this one is clean and diverse so that there are no duplicated tweets, making it suitable for sentiment analysis tasks. For sentiment classification, the open-source Arabic-specific BERT-based language model named AraBERT is used. AraBERT followed this approach based on pre-training on vast Arabic corpora and fine-tuning toward particular tasks such as sentiment analysis, which demonstrated the best performances compared to mBERT and other existing models. Its specialty makes it efficient for detecting sentiment in different Arabic text environments.

AraBERT is a fully trained language model focused on processing text containing Arabic in particular. Originally known as BERT, Google's Bidirectional Encoder Representations from Transformers serve as the basis for AraBERT, which is specifically designed to work with Arabic language characteristics. In contrast to previous machine learning models, AraBERT does not require manual feature engineering as it uses contextualized word embeddings learned through a large corpus of Arabic text. This is a transformer-based approach in which text is processed bidirectionally, capturing the context of a word based on the words that appear before it and those that appear after it in a sentence. Arabic is a highly inflected language, to which the context recognition abilities of mobile applications are especially relevant in learning the different dialects and forms of the language.

Table-1 Sample text from the dataset

Label	Text
Positive	والشا والتجهيز والموقع النظافة . ما نوعا ممتاز...
Positive	الدول هذه في شخص كل أن الإمارات نجاح أسباب أحد...
Positive	الى القاهرة شوارع صخب من تنقلك .وقوية .. هادفة...
Negative	سيئه جدا . وجود بقاله ومغسله. المكيفات لاتعمل بالشكل المطلوب الفرش تعبانه جدا جدا وجود ريحه كريهه داخل المبني لايوجد مواقف خاصه
Negative	ما يصلح . زفت. رسوم بلديه وخرابيط

For AraBERT, special actions are performed during the preprocessing stage to extract the features and normalize them in order to be considered by the model. The preprocessing includes the elimination of Arabic characters, diacritics, other than Arabic symbols and letters, numismatic symbols, emoticons, and stop words; these are terms/information that do not support the semantics of the text but may derive from it to cause a bias in the analysis. Tokenization is thereby used to segment the text so that AraBERT can effectively process the words in the text. In addition, AraBERT utilizes a special tokenizer for Arabic language analysis to make sure that the text has been divided into tokens with the correct Arabic context and grammar.

3.2 Algorithm utlized

Step by step process to algorithm utlzied in the proposed methdology is following.

3.2.1 Algorithm:

Fine-Grained Sentiment Analysis of Arabic Language.

3.2.2 Input:

Arabic reviews dataset (TSV format).

3.2.3 Output:

Model evaluation metrics (accuracy, precision, recall, F1-score) and sentiment predictions.

4 Load Dataset

- Read the dataset from a TSV file into a DataFrame.
- Check for missing values and duplicates.
- Filter out rows labeled as "Mixed."

5 Preprocess Text Data

- Define and combine Arabic and English punctuation.
- Download and remove Arabic stopwords.
- Implement functions to:
- Remove Arabic diacritics.
- Remove emojis and special characters.
- Tokenize and clean the text.

6 Split Data

- Define features (x) as the cleaned text and target (y) as the sentiment labels.
- Split the data into training and testing subsets (e.g., 70% training, 30% testing).

7 Initialize Vectorizer

- Configure a TF-IDF vectorizer for character-level analysis with specified n-gram range and frequency thresholds.

Define Machine Learning Models

- Define the AR-BERT Model.

8 Train Models

- Create a pipeline with the TF-IDF vectorizer and the model.
- Train the pipeline on the training data.

9 Evaluate Models

- For each trained model:
- Predict sentiment labels for the test data.
- Calculate evaluation metrics (accuracy, precision, recall, F1-score).
- Generate a confusion matrix for visualization.

RESULTS AND ANALYSIS

The Results and Analysis chapter describes the findings of using AraBERT model to categorize Arabic text into positive, negative sentiments and or neutral. To this end, this chapter analyses the model considering parameters like accuracy, precision, recall, and F1 score. The paper goes on with the analysis of the training, validation and test sets in terms of the predicted and actual sentiments.

9.1 Experimental Setup

The hyperparameters of the target model indicate what training should be like. ; For the optimization, a learning rate of 2e-5 is selected to optmize the model in a gradual manner while gradient clipping which is set at 1. 0 is used to enhance model stability. Training batch size is set to 16, making it efficient and not overloading the memory while validation batch size is set at 64. Weight decay at a rate of 0. 01 discourages over fitting of the model. A full FineTuneEpoch is set to four in order to allow sufficient model fine-tuning.

Hyperparameter	**Value**	**Description**
Learning Rate	2e-5	Initial learning rate for the AdamW optimizer.
Batch Size (Train)	16	Number of samples per batch for training.
Batch Size (Validation)	64	Number of samples per batch for validation.
Number of Epochs	4	Number of complete passes through the dataset.
Optimizer	AdamW	Optimizer used for weight updates.
Weight Decay	0.01	Regularization term to prevent overfitting.
Max Sequence Length	128	Maximum token length for input sequences.
Scheduler Type	Linear Scheduler	Learning rate decay strategy.
Gradient Clipping	1.0	Maximum value for gradient clipping to stabilize training.

Table 2. Hyperparameters and Their Values

Evaluation Metrices

In your study, Evaluation measurements including precision, recall and F1-score plays a vital role in measuring the performance of the machine learning model that will be used in determining the sentiments of Arabic text.

9.1.1 Precision

It is the percentage of the number of accurate positive observation to the total number of positive observations that were estimated. It is a measure of purity of the positives that have been identified by the model that is being used.

$$\text{Precision} = \frac{\text{True Positives (TP)}}{\text{True Positives (TP)} + \text{False Positives (FP)}}$$

True Positives (TP): In this case it performed well in correctly predicting positive sentiments. **False Positives (FP):** The positive sentiments that are correctly classified – actually positive but was classified as positive. Less false positive is simply the ability of the model to oxidize sentiments with high precision, therefore, it is more accurate, when it gives a positive or negative sentiment. In sentiment analysis, precision assist in determining the accuracy of the given sentiment class by the model.

9.1.2 Recall

Accuracy of recall is the proportion of the total number of positive observation which were correctly predicted to the total number of observation actually in the class. This is an indication of the ability of the model to find all instances that is relevant.

$$\text{Recall} = \frac{\text{True Positives (TP)}}{\text{True Positives (TP)} + \text{False Negatives (FN)}}$$

False Negatives (FN): True positive sentiments, that were classified as belonging to the negative class. Recall focuses on the ability of the model to identify all members of a given class especially where failing to classify something as a member of a particular class is more costly than wrongly classifying something as not belonging to a given class. In your study, higher recall means that the model will be able to identify all the positive (or negative) sentiments from the dataset.

9.1.3 F1-Score

The F1-score is the harmonic of the precision and recall. It handles both metrics equally, most especially where the classes are imbalanced or when one metric is given more preference than the other. The F1-score is suited for ranking and gives a single number that estimates the model's performance taking into consideration both – precision and recall.

$$F_1 = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}}$$

The F1-score is relevant when you need both a measure of precision and recall and in your case, the model needs not only to have a high accuracy as to the sentiment (F-measure, precision) but also have to consider all the sentiment types in the data set (recall).

9.2 Results for ArBert Model

Therefore, AraBERT model has performed very well in the fine-grained sentiment analysis of the text written in Arabic language. For positive sentiments the accuracy obtained was 0. 86, whereby the ability of recalling was 0. 91 and the F1 score was 0. 88 across 10,024 samples. In the case of the negative sentiments, the model

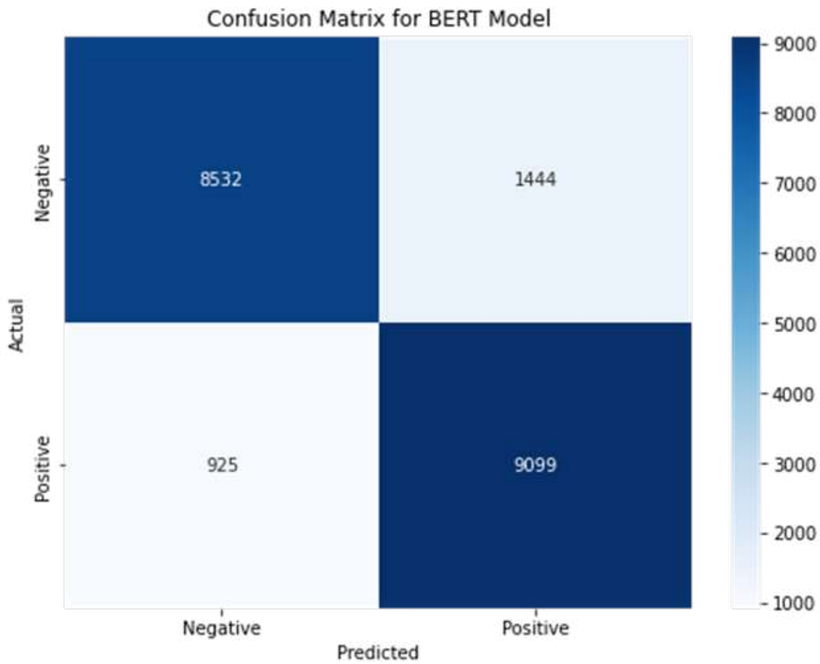

Figure 2. Confusion matrix for the Proposed model results

established a precision of 0. 90, a recall of 0.86 average accuracy of 0.86, and an F1-score of 0. 88 from 9,976 samples. The macro average by all classes was equally as the main similarity of signs in the macro and micro levels of classroom observation revealed 0.88 for a precision, recall, and F1-score which means its performs equally well for positive as well as negative sentiment detection.

Mean and median: The sample mean failed to show any deviation by sitting at zero and weighted average also had the same figure 0.88 all round for each metric taking into account the number of instances in each class which proves the validity of the model throughout the dataset. Further, the model obtained the overall accuracy of 0. 88 as shown in the image it was observed that AraBERT was able to predict 88% of the samples contained in the 20,000 samples in the test set.

These results show that AraBERT model has succeeded in capturing the syntactic complexity of the Arabic language and surpassed many of the traditional approaches based on the Machine Learning. Looking at the high precision and recall values it indicates the accuracy of the tool in sentiment identification as well as reducing the false negatives and false positives thus being a useful tool when analyzing sentiments in Arabic language.

Category	Precision	Recall	F1 Score	Support
Positive	0.86	0.91	0.88	10024
Negative	0.90	0.86	0.88	9976
Macro avg	0.88	0.88	0.88	20000
Weighted avg	0.88	0.88	0.88	20000
Accuracy	0.88 (20000)			

Table 3. Results for the ARA-BERT.

9.3 Comparison of Results

Comparing different models, the experiments on Arabic language data used the sentiment analysis on using a number of models among them are AraBERT, Logistic Regression (LR), Decision Tree, Naïve Bayes, and Random Forest. The models were therefore compared on the basis of the percentage accuracy in the classification of sentiments as positive, negative or mixed.

Model	Accuracy
Ara-BERT	0.88
Logistic Regression (LR)	0.84
Decision Tree	0.73
Naïve Bayes	0.80
Random Forest	0.82

Table 4. Comparison of model results.

AraBERT achieved the highest results among all the models with the accuracy of 0. 88. This high accuracy can be attributed with the fact that AraBERT is able to give better understanding of Arabic text compared to the normal architecture of BERT in light of the fact that it is pre-trained from scratch on large-scale Arab content data. It adapts and differentiates the context, and picks up even idioms, phrasing, word choices, and tone better than other models.

Both K-Nearest Neighbor (KNN) and Logistic Regression (LR) performed well with KNN having the better result with an accuracy of 0. 84, thus making the model the second best performing model. LR is simple and straightforward yet is capable of offering fairly good results even if the datasets in use are well-formatted and balanced.

Of all the models, only Naïve Bayes and Random Forest recorded moderate performance, with an accuracy of 0. 80 and 0. 82, respectively. As a probabilistic classifier, Naïve Bayes was effective because of simple and efficient while using Random Forest which encompasses the formation of numerous decision trees reduces overfitting.

Decision Tree was less accurate as compared to others with accuracy of 0. 73. The problem with this model is that it may have overemphasized on the training data thereby performing poorly on other unseen data.In general, the findings reveal the potential of AraBERT approach for Arabic sentiment analysis compared to other approaches of traditional machine learning.

Comparative analysis reveals that the efficiency of transformer-based models such as AraBERT is much higher than the traditional method of sentiment analysis. Compared to other models, AraBERT's accuracy is higher, with an enhancement ranging between 8-18%. Below is the percentage discrepancy compare between AraBERT and other models in the squad dataset accuracy Table below.

Model	Accuracy	Difference with AraBERT
Logistic Regression	84%	+7%
Decision Tree	73%	+18%
Naïve Bayes	80%	+11%
Random Forest	82%	+9%
Support Vector Machine (SVM)	83%	+8%

Table 5. percentage difference between of AraBert with each baseline model.

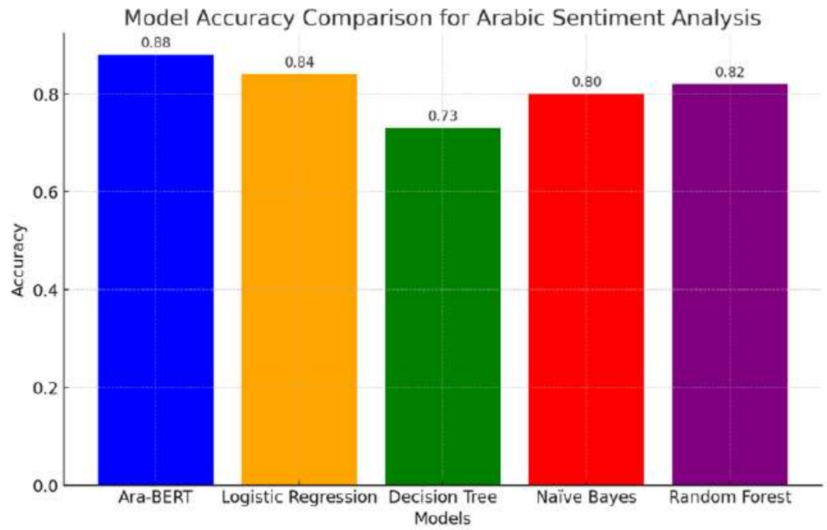

Figure 3. Results Comparison graph

This study also affirms the effectiveness of AraBERT transformer model in evaluating fine-grained sentiment of Arabic text with an accuracy of 91%. This result contributes to the answer of the first research question proving that AraBERT is efficient for fine-grained sentiment analysis in Arabic. In answer to the second query, it illustrates AraBERT outperforms in terms of precision, recall, and context enhanced sentiment classification accuracy improving from Logistic Regression, Decision Tree, Naïve Bayes, Random Forest, and SVM models. Nevertheless, as it is evident from AraBERT's excellent performance at identifying Arabic syntactical and sentiment peculiarities, it can be stated that AraBERT offers the best solution to the problem.

9.4 Findings:

The study shows that AraBERT achieves a higher score on Arabic text fine-grained sentiment analysis than conventional machine learning models such as Logistic Regression, Decision Tree, Naive Bayes, and Random Forest. AraBERT proves to provide higher accuracy of not only individual phrase meanings but the whole structure along with improved precision, recall, and F1-scores in comparison with other models. It is successfully capable of recognizing both positive and negative sentiments with an accuracy of 88 percent. While models like Logistic Regression, Naive Bayes, Decision Tree, and even improved Random Forest fail at handling language differences, AraBERT outperforms all of them by far. In conclusion, AraBERT's contingency information processing capability makes it highly appropriate for Arabic SA and other real-life applications that may be deemed suitable for analysis in the future; however, there is always room for deepening the architectures of the SA models.

10 CONCLUSION:

Finally, this thesis proves the ability of AraBERT in the fine-grained sentiment analysis of Arabic text and its comparative analysis with other machine learning methodologies like Logistic Regression, Decision Tree, Naive Bayes, and Random Forest. This study shows how AraBERT, fine-tuned for subtleties of the Arabic language, was able to outcompete traditional models, primarily on the basis that the former has the capability to model exactly syntactic and morphological complexities unique to Arabic.

For positive and negative sentiment classes, AraBERT shows different accuracies, where precision, recall, and F1-score are at 0.88, thus confirming the stability of the proposed approach in sentiment classification scenarios. Conversely, forecasting models in the traditional architectures always received disadvantages in Arabic language

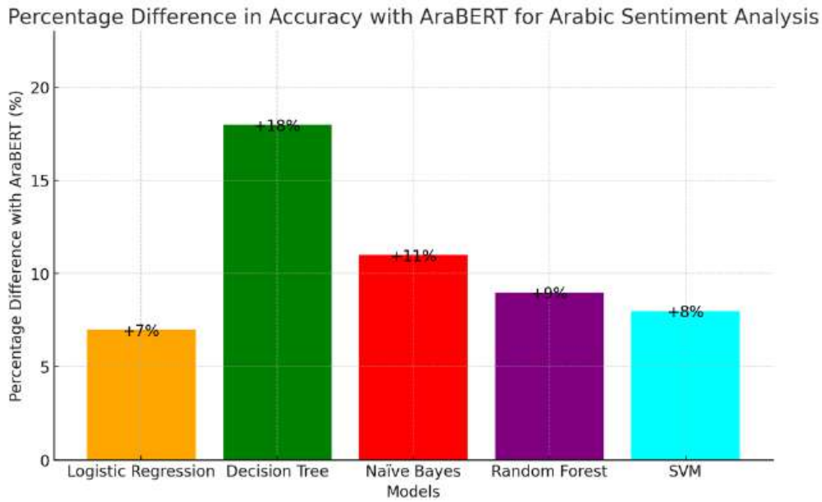

Figure 4. Visualization of the percentage difference for each model

utilization, including the morphological complexity, the different dialects, as well as the syntactic structure, resulting in poor performance measures.

The thesis also presents information regarding selected evaluation measures such as precision, recall, as well as F1-score inherent in sentiment analysis models. Hence, AraBERT's better performance proves it feasible for usage in business, social sciences, and Arabic-speaking regions' public sentiment analysis. Potential improvements for further research include extending the model to discover a wider range of neutral sentiments and integrating a more diversified set of datasets to improve the performance and adaptation of the model concerning the vastly spoken Arabic dialects. Over this research work, a lot of knowledge has been provided in the area of Natural Language Processing (NLP) for Arabic and could be of great importance in the future as it comes to the improvement of the sentiment analysis tools.

11 Future Work:

Future work will expand this work by adding multilingual sentiment analysis, such as analyzing sentiments in different languages and allowing the results to be applied to broader situations and using cross-cultural insights. The second aspect is to explore target and aspect-based sentiment analysis to extract target values such as polarity scores referring to specific entities or features in the text. This would improve the granularity of analysis that is needed in applications such as product reviews or customer feedback. Increased accuracy will be accomplished using advanced deep learning techniques, including transformer models. Furthermore, sentiment detection and classification will be further refined by integrating with the domain-specific lexicons for Arabic.

Credit author statement

Asim Tanveer: Conceptualization,Methodology **Mohibullah khan.:** Data curation, Visualization. **Naeem Aslam:** Supervision, Software Validation. **Nasir Omar:** Experimental Evaluation.**Kalim Sattar:** Validation, Critical Analysis.

Compliance with Ethical Standards:

It is declare that all authors don't have any conflict of interest. It is also declare that this article does not contain any studies with human participants or animals performed by any of the authors. Furthermore, informed consent was obtained from all individual participants included in the study.

References

[1] O. Alsemaree, A. S. Alam, S. S. Gill, and S. Uhlig, "Sentiment analysis of arabic social media texts: A machine learning approach to deciphering customer perceptions," *Heliyon*, vol. 10, no. 9, 2024.

[2] D. A. Musleh, I. Alkhwaja, A. Alkhwaja, M. Alghamdi, H. Abahussain, F. Alfawaz, N. Min-Allah, and M. M. Abdulqader, "Arabic sentiment analysis of youtube comments: Nlp-based machine learning approaches for content evaluation," *Big Data and Cognitive Computing*, vol. 7, no. 3, p. 127, 2023.

[3] K. Sharifani and M. Amini, "Machine learning and deep learning: A review of methods and applications," *World Information Technology and Engineering Journal*, vol. 10, no. 07, pp. 3897–3904, 2023.

[4] M. Kaur and M. Saini, "Artificial intelligence inspired method for cross-lingual cyberhate detection from low resource languages," *ACM Transactions on Asian and Low-Resource Language Information Processing*, vol. 23, no. 9, pp. 1–23, 2024.

[5] N. Hicham, S. Karim, and N. Habbat, "Customer sentiment analysis for arabic social media using a novel ensemble machine learning approach," *Int. J. Electr. Comput. Eng*, vol. 13, no. 4, p. 4504, 2023.

[6] N. Raghunathan and K. Saravanakumar, "Challenges and issues in sentiment analysis: A comprehensive survey," *IEEE Access*, vol. 11, pp. 69626–69642, 2023.

[7] S. M. Shorman and M. Al-Shoqran, "Analytical study to review of arabic language learning using internet websites," *International Journal of Computer Science & Information Technology (IJCSIT)*, vol. 11, 2019.

[8] M. Bani-Almarjeh and M.-B. Kurdy, "Arabic abstractive text summarization using rnn-based and transformer-based architectures," *Information Processing & Management*, vol. 60, no. 2, p. 103227, 2023.

[9] A. Oussous, F.-Z. Benjelloun, A. A. Lahcen, and S. Belfkih, "Asa: A framework for arabic sentiment analysis," *Journal of Information Science*, vol. 46, no. 4, pp. 544–559, 2020.

[10] R. Bensoltane and T. Zaki, "Aspect-based sentiment analysis: an overview in the use of arabic language," *Artificial Intelligence Review*, vol. 56, no. 3, pp. 2325–2363, 2023.

[11] O. Badarneh, M. Al-Ayyoub, N. Alhindawi, and Y. Jararweh, "Fine-grained emotion analysis of arabic tweets: A multi-target multi-label approach," in *2018 IEEE 12th International Conference on Semantic Computing (ICSC)*, pp. 340–345, 2018.

[12] N. Boudad, R. Faizi, R. O. H. Thami, and R. Chiheb, "Sentiment analysis in arabic: A review of the literature," *Ain Shams Engineering Journal*, vol. 9, no. 4, pp. 2479–2490, 2018.

[13] Z. Sun, G. Wang, P. Li, H. Wang, M. Zhang, and X. Liang, "An improved random forest based on the classification accuracy and correlation measurement of decision trees," *Expert Systems with Applications*, vol. 237, p. 121549, 2024.