

محمد القحطاني

AE07.docx

 Midocean University

Document Details

Submission ID

trn:oid:::2945:381695354

Submission Date

Apr 30, 2026, 6:00 PM GMT+3

Download Date

Apr 30, 2026, 7:36 PM GMT+3

File Name

AE07.docx

File Size

1.9 MB

39 Pages

7,182 Words

41,708 Characters

*% detected as AI

AI detection includes the possibility of false positives. Although some text in this submission is likely AI generated, scores below the 20% threshold are not surfaced because they have a higher likelihood of false positives.

Caution: Review required.

It is essential to understand the limitations of AI detection before making decisions about a student's work. We encourage you to learn more about Turnitin's AI detection capabilities before using the tool.

Disclaimer

Our AI writing assessment is designed to help educators identify text that might be prepared by a generative AI tool. Our AI writing assessment may not always be accurate (i.e., our AI models may produce either false positive results or false negative results), so it should not be used as the sole basis for adverse actions against a student. It takes further scrutiny and human judgment in conjunction with an organization's application of its specific academic policies to determine whether any academic misconduct has occurred.

Frequently Asked Questions

How should I interpret Turnitin's AI writing percentage and false positives?

The percentage shown in the AI writing report is the amount of qualifying text within the submission that Turnitin's AI writing detection model determines was either likely AI-generated text from a large-language model or likely AI-generated text that was likely revised using an AI paraphrase tool or word spinner.

False positives (incorrectly flagging human-written text as AI-generated) are a possibility in AI models.

AI detection scores under 20%, which we do not surface in new reports, have a higher likelihood of false positives. To reduce the likelihood of misinterpretation, no score or highlights are attributed and are indicated with an asterisk in the report (*%).

The AI writing percentage should not be the sole basis to determine whether misconduct has occurred. The reviewer/instructor should use the percentage as a means to start a formative conversation with their student and/or use it to examine the submitted assignment in accordance with their school's policies.

What does 'qualifying text' mean?

Our model only processes qualifying text in the form of long-form writing. Long-form writing means individual sentences contained in paragraphs that make up a longer piece of written work, such as an essay, a dissertation, or an article, etc. Qualifying text that has been determined to be likely AI-generated will be highlighted in cyan in the submission, and likely AI-generated and then likely AI-paraphrased will be highlighted purple.

Non-qualifying text, such as bullet points, annotated bibliographies, etc., will not be processed and can create disparity between the submission highlights and the percentage shown.





Enhancing Arabic Sentiment Analysis via Domain-Adaptive Transformer Feature Extraction and Hybrid Machine Learning Classification

Prepared By

Name: Nasser Alshalwan

ID No.: 253000228

Name: Mohammed Alqahtani

ID No.: 253000264

Submitted in fulfillment of the requirements for the degree Master of
Science in Artificial Intelligence

At

Collage of Informatics

Midocean University

UAE

Supervisor Name: Dr. Wessam

Collage of Informatics, Midocean University, Comoros

Email:.....

Acknowledgements

Special thanks to Dr. Wessam.

Abstract

List of Publications

Table of Contents

List of Abbreviations	viii
List of Figures.....	ix
List of Tables	x
 Chapter 1 Introduction	 1
1.1 Overview	1
1.2 Problem Definition	2
1.3 Thesis Objectives.....	2
1.4 Thesis Contribution.....	2
1.5 Thesis Organization	3
Chapter 2 Background and Related Work	4
2.1 Introduction.....	4
2.2 Background.....	4
2.2.1 Introduction to Sentiment Analysis	4
2.2.2 Sentiment Analysis using Machine Learning.....	5
2.2.2.1 Logistic Regression	5
2.2.2.2 Support Vector Machine	6
2.2.2.3 XGBoost.....	7
2.2.3 Sentiment Analysis using Deep Learning	8
2.2.3.1 Convolutional Neural Network (CNN).....	9
2.2.3.2 Long Short-Term Memory (LSTM).....	9
2.2.3.3 Bidirectional Gated Recurrent Unit (BiGRU).....	10
2.2.4 Sentiment Analysis Using Transformer-Based Models	11
2.2.4.1 AraBERT	11
2.2.4.2 MARBERT.....	12
2.2.4.3 CAMeLBERT.....	13

2.2.5 Performance metrics	14
2.3 Literature Review	16
2.4 Comparative Analysis of Related Work.....	19
2.5 Chapter Summary	23
Chapter 3 Proposed Approach (Methodology).....	24
Chapter 4 Experimental Results and Discussion	25
Chapter 5 Conclusion and Future Work	26
References	27

List of Abbreviations

- **Acc.:** Accuracy
- **AraBERT:** Arabic Bidirectional Encoder Representations from Transformers
- **ASAD:** Arabic Sentiment Analysis Dataset
- **ASTD:** Arabic Sentiment Tweets Dataset
- **BiGRU:** Bidirectional Gated Recurrent Unit
- **CAMeLBERT:** Carnegie Mellon University Arabic BERT
- **CNN:** Convolutional Neural Network
- **F1-score:** F1-score
- **FN:** False Negative
- **FP:** False Positive
- **LSTM:** Long Short-Term Memory
- **MARBERT:** Masked Arabic Bidirectional Encoder Representations from Transformers
- **ML:** Machine Learning
- **NLP:** Natural Language Processing
- **Pr.:** Precision
- **R:** Recall
- **SVM:** Support Vector Machine
- **TF-IDF:** Term Frequency–Inverse Document Frequency
- **TN:** True Negative
- **TP:** True Positive

List of Figures

1	Logistic Regression S-shaped curve	14
2	SVM decision boundary with maximum margin	15
3	XGBoost sequential boosting process	16
4	CNN architecture for text classification	17
5	LSTM structure diagram	18
6	BiGRU architectural diagram	18
7	AraBERT feature extraction model architecture (MTL)	20
8	Fine-tuned MARBERT model architecture	21
9	CAMeLBERT architecture	21
10	Proposed Framework	33

List of Tables

1	Comparison of sentiment analysis approaches in Arabic NLP	22
2	Comparative Study of Related Work	29

Chapter 1 Introduction

1.1 Overview

Sentiment analysis is widely used in natural language processing because it helps identify the opinions, emotions, and attitudes expressed in text. Its importance has increased with the rapid growth of online communication, especially on social media platforms, review websites, and other digital spaces where users frequently share their views. In Arabic, sentiment analysis has received growing attention in recent years. However, recent survey studies still show that Arabic sentiment analysis is less mature than sentiment analysis in high-resource languages, particularly in modern deep learning settings [1]. One of the main challenges in Arabic sentiment analysis is the nature of the Arabic language itself. Arabic is used in different forms, including Modern Standard Arabic and many regional dialects, and these forms often differ in vocabulary, spelling, and style. This makes sentiment classification more difficult, especially when the data comes from informal environments such as social media. However, the challenge is bigger when sentiment data comes from different domains like tweets, reviews, or online discussions. Patterns from one domain may not work well in another. Also, recent studies showed that cross-domain and cross-dialect sentiment analysis is still difficult and needs better methods to deal with the variation in Arabic contexts [2]. Also, progress in this field is not only about making better models. It is also about using good and updated benchmarks to test them. Some recent studies showed that some Arabic sentiment benchmarks are still limited and do not always show the real power of new models. So, benchmark quality is still important for improving Arabic sentiment analysis research [3]. At the same time, recent research suggests that hybrid approaches are still promising for Arabic sentiment classification. Combining transformer-based contextual representations with traditional text features may be useful because sentiment in Arabic can appear through both contextual meaning and surface-level lexical patterns. Some recent studies on Arabic dialect sentiment classification also support this direction by showing the value of integrating transformer representations with traditional features such as TF-IDF [4]. Also, Arabic sentiment analysis is now focusing more on domain variation, benchmark quality, and combining different

features. So, it makes sense to look at a framework that uses domain-adaptive transformer feature extraction and hybrid machine learning classification in this study.

1.2 Problem Definition

Arabic sentiment analysis still has many challenges because Arabic texts are different in dialects, domains, and writing styles. Also, this makes sentiment classification difficult, especially when the data is not the same as the training data. Also, working well on one model does not mean all results will be good. It can give poor results when the text is noisy, informal, or from special domain. Also, the problem is more complex in social media data, where there are many dialects, non-standard spelling, and informal words.

1.3 Thesis Objectives

This study's primary objective is to explore some more effective method for the task of Arabic sentiment analysis. For this purpose, it employs domain-adaptive transformer features and hybrid machine learning for the classification task. Additionally, it aims at enhancing the quality of sentiment classification for Arabic language texts, especially when these texts are from divergent domains and writing styles. Meanwhile, it pertains to how the combination of transformer features with hybrid classification could be more effective. Moreover, it aims at constructing a more enhanced and balanced method for the task of Arabic sentiment analysis particularly in more complex context.

1.4 Thesis Contribution

This study is expected to contribute to Arabic sentiment analysis in the following ways:

- It shows a framework, which applies domain-adaptive transformers to get the features and hybrid machine learning to do the classification. The model (MARBERT, CAMeLBERT)

is responsible for the feature extraction, then the (feature) is transferred to the classification for their sentiment prediction.

- It applies the proposed framework on ASAD dataset in order to evaluate its performance to classify sentiment of Arabic tweets.
- It is along with the bag of words performance using TF-IDF baseline models, namely Logistic Regression and SVM. In order to demonstrate the proposed framework is an improvement over traditional approaches.
- It find if using transformer based features and machine learning classification together can improve and stabilize the results compared to only use one model or only one feature type.

1.5 Thesis Organization

The remainder of the thesis is structured as follows:

Chapter 2: provides the background and related work about Arabic tweet sentiment analysis. Also, it talks about machine learning, deep learning, and transformer-based methods. In addition, it shows some recent studies about Arabic sentiment analysis.

Chapter 3: outlines the proposed approach of the study, including the dataset, preprocessing steps, feature extraction process, classification models, and evaluation metrics.

Chapter 4: presents the experimental results and discussion of the proposed framework, including the performance of the models and the comparison between the proposed approach and the baseline methods.

Chapter 5: concludes the study by summarizing the main findings, highlighting the limitations, and presenting directions for future work.

Chapter 2 Background and Related Work

2.1 Introduction

This chapter provides the background for Arabic tweet sentiment classification. It begins by explaining sentiment analysis and the main challenges that appear in Arabic sentiment analysis, especially in social media and dialectal contexts. It then reviews the main approaches used in sentiment classification, including machine learning, deep learning, and transformer-based models. It also presents the evaluation metrics used in this study and discusses recent related work to clarify the research gap behind the proposed framework.

2.2 Background

2.2.1 Introduction to Sentiment Analysis

Sentiment analysis, or opinion mining, is a common task in Natural Language Processing (NLP). It is mainly used to identify the sentiment expressed in text, such as whether the text is positive, negative, or neutral. In some studies, more detailed sentiment classes are also used depending on the dataset and the aim of the research. So, these days have become more important because people share a lot of text every day on social media, review sites, forums, and other places online. So, sentiment analysis became more important. Moreover, it is also beneficial for many aspects such understanding customers' comments, following the public opinions, checking products and watching brands. It definitely plays a significant role in the social media, whereby the people are giving their views directly and very transparently. Yet, it is not easy to conducting a sentiment analysis; as the meaning of the utterance could be changeable due to the context, sarcasm, negation, ambiguity or single word. But for Arabic language, it is more complex because of the existence of Modern Standard Arabic and many dialects that are differing by the word choice, spellings and style; whereas on the social media, it will be more extreme as the users usually confuse the dialects, encrypt the words by using repetitive alphabet, EMOJIs or others. Also, sentiment analysis can be done at document-level,

sentence-level, or aspect-level. In tweet classification, short-text sentiment analysis is particularly important because each tweet is usually treated as one unit for prediction. Arabic sentiment analysis may also use binary, three-class, or multi-class settings depending on the dataset and the purpose of the study. In this study, a hierarchical classification strategy is used. In the first stage, tweets are classified into Neutral and Polar. In the second stage, tweets classified as Polar are further divided into Positive and Negative. This gives a more focused way to classify Arabic tweet sentiment. Over time, sentiment analysis methods have moved from traditional feature-based approaches to machine learning, deep learning, and transformer-based models. The following sections review these approaches and explain their relevance to Arabic sentiment analysis [5][6].

2.2.2 Sentiment Analysis using Machine Learning

Machine learning is used in sentiment classification by training models on labeled data. Also, the text is converted into features, and TF-IDF is often used in this step. After that, machine learning models like Logistic Regression, SVM, and XGBoost are used for sentiment classification [7].

2.2.2.1 Logistic Regression

Logistic Regression is a common machine learning model for sentiment analysis. Also, before using it, the text data is changed to numbers, for example by using TF-IDF. Then, the model is trained on labeled data to find patterns and predict the sentiment. Researchers usually include Logistic Regression because it is straightforward and gives reasonable results, especially as a baseline model, as shown in Figure 1. Even so, it mainly depends on the quality of the extracted features, and this makes it less capable of understanding contextual meaning compared with newer models [8].

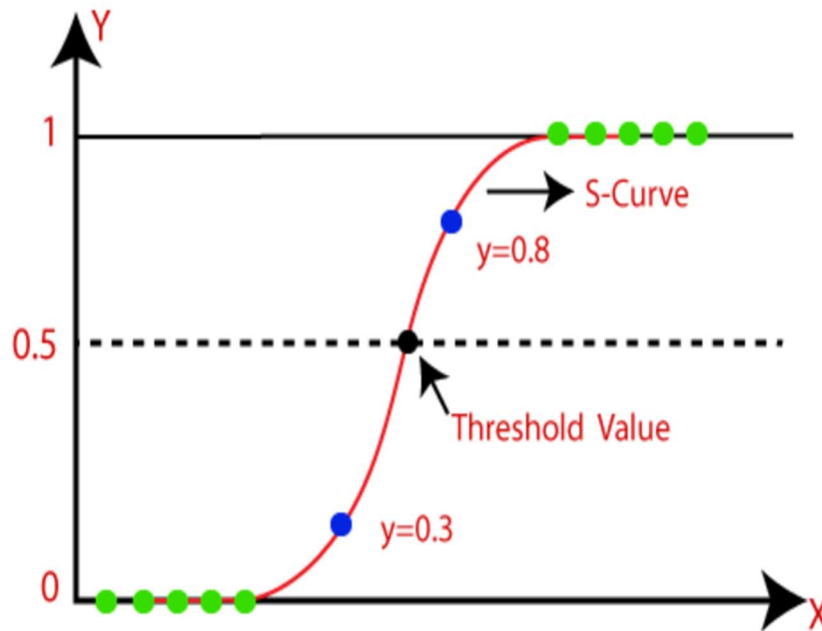


Figure 1. Sample Logistic Regression Function S-shaped curve [9].

2.2.2.2 Support Vector Machine

Support Vector Machine is a well-known machine learning model in sentiment analysis. It is commonly applied after converting text into numerical form using feature extraction methods such as TF-IDF. The main idea of this model is to classify the text by identifying a separating boundary between different sentiment categories. Many studies use Support Vector Machine because it is effective in handling text classification problems and can produce good performance in many cases, as shown in Figure 2. However, its results are still influenced by the quality of the extracted features, which means it may not represent contextual meaning as effectively as more recent deep learning and transformer-based models [10].

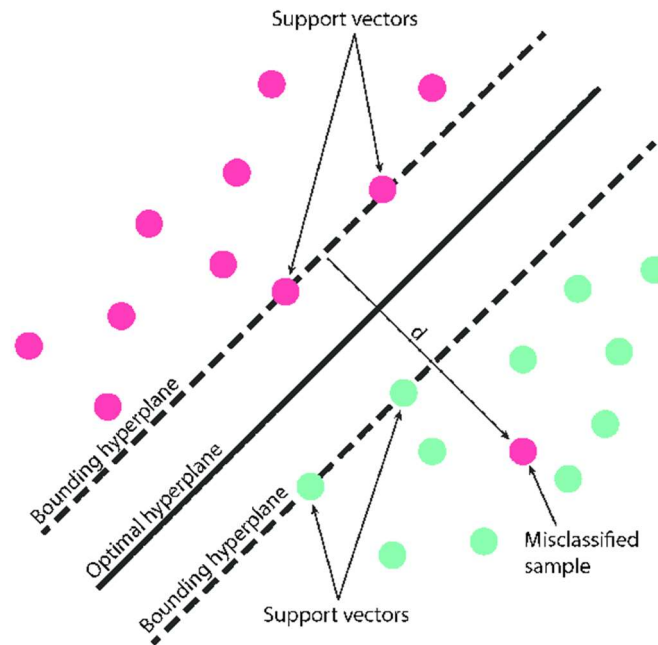


Figure 2. Support Vector Machine (SVM) decision boundary with maximum margin [11].

2.2.2.3 XGBoost

XGBoost is also used in sentiment analysis as a feature-based machine learning model. In most cases, the text is not used directly, but is first changed into numerical representation using methods such as TF-IDF. These features are then used to train the model so it can distinguish between sentiment classes. Also its included in some sentiment analysis studies cause it applies a tree-based boosting strategy, so that can be improve classification performance in feature-driven tasks, as shown in Figure 3. However, its performance still depends on the numerical representation of the text, which makes it less effective in understanding context when compared with deep learning and transformer-based approaches [12].

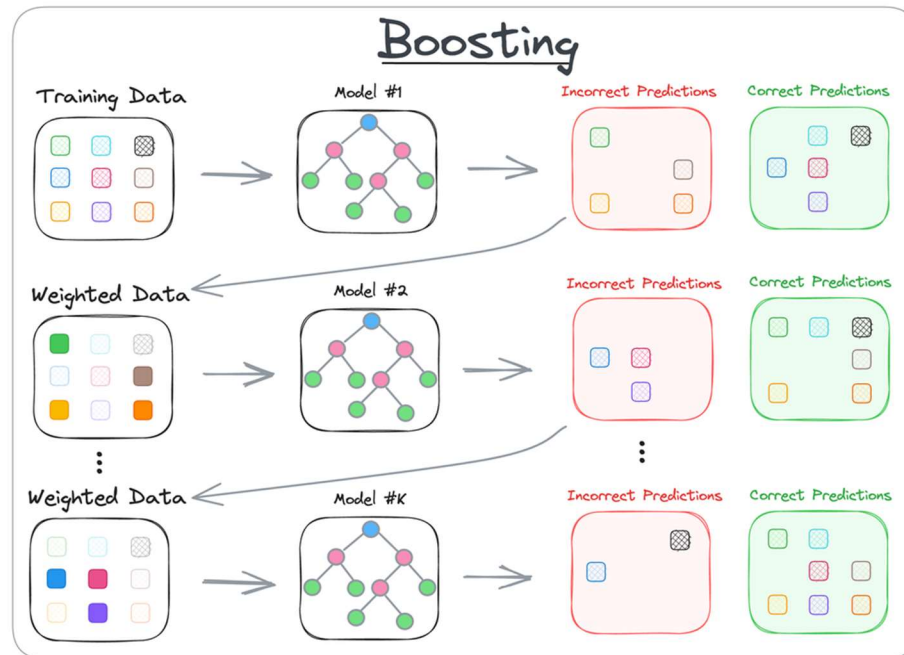


Figure 3. XGBoost sequential boosting process [13].

However, machine learning models are still important in sentiment analysis because they provide clear baseline results. Their effectiveness is closely related to the preprocessing stage and the quality of feature extraction. Also, their ability to represent contextual meaning is more limited than that of newer techniques. Also, in Arabic sentiment analysis, this problem is more difficult. There are many dialects and complex morphology. Also, the informal style is common, especially in tweets [14].

2.2.3 Sentiment Analysis using Deep Learning

Deep learning is a branch of machine learning used in sentiment classification. These models do not depend on feature extraction methods such as TF-IDF as in machine learning (ML), because they can learn text representation automatically from the input data. On the other hand, models such as CNN, LSTM, and BiGRU are often used to classify text based on the patterns learned during training [15].

2.2.3.1 Convolutional Neural Network (CNN)

Convolutional Neural Network is also used in sentiment analysis as a deep learning model. In text classification, it learns useful patterns from the input text and uses them in the classification process. CNN is often applied because it can detect important parts of the text that support the sentiment decision, as shown in Figure 4. However, it may be less effective when the text requires understanding longer word relations compared with some other deep learning models [16].

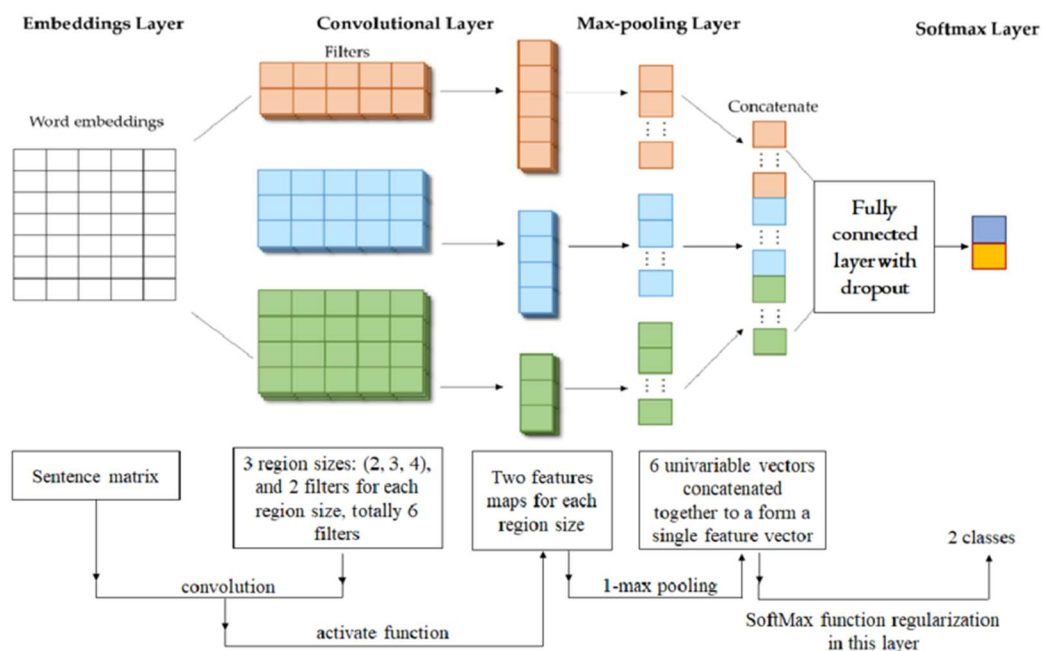


Figure 4. The architecture of a CNN for text classification [17].

2.2.3.2 Long Short-Term Memory (LSTM)

Long Short-Term Memory is used in sentiment analysis to classify sequential text data. Instead of treating words independently, it follows the order of the sequence during learning. This allows the model to make use of information that appears across the text when predicting the sentiment class, as shown in Figure 5. However, it often requires more training effort than simpler models [18].

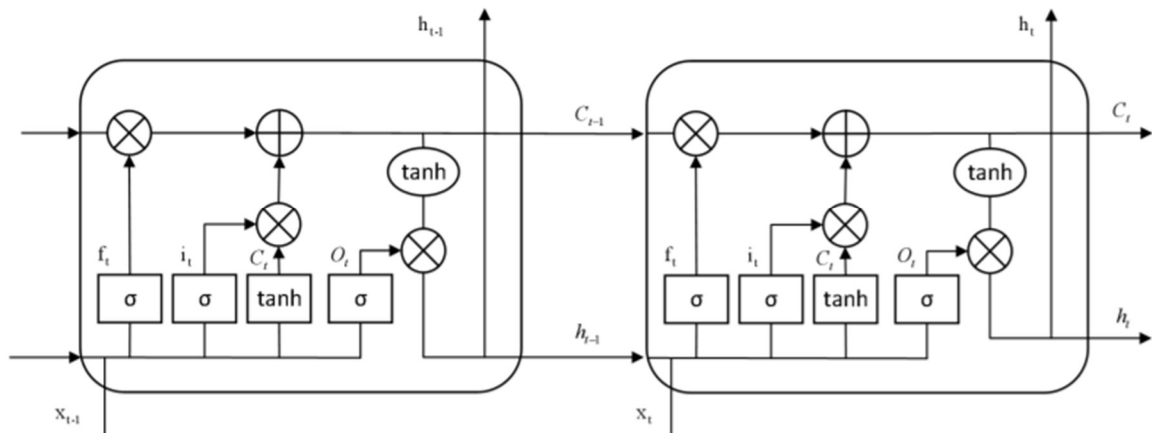


Figure 5. LSTM structure diagram [19].

2.2.3.3 Bidirectional Gated Recurrent Unit (BiGRU)

Bidirectional Gated Recurrent Unit is also applied in sentiment analysis as a deep learning model. It reads the sequence from both directions, which gives the model access to information from the beginning and the end of the text. This can support sentiment classification by providing a better view of the sequence, as shown in Figure 6. However, the model result is still affected by the training data and parameter settings [20].

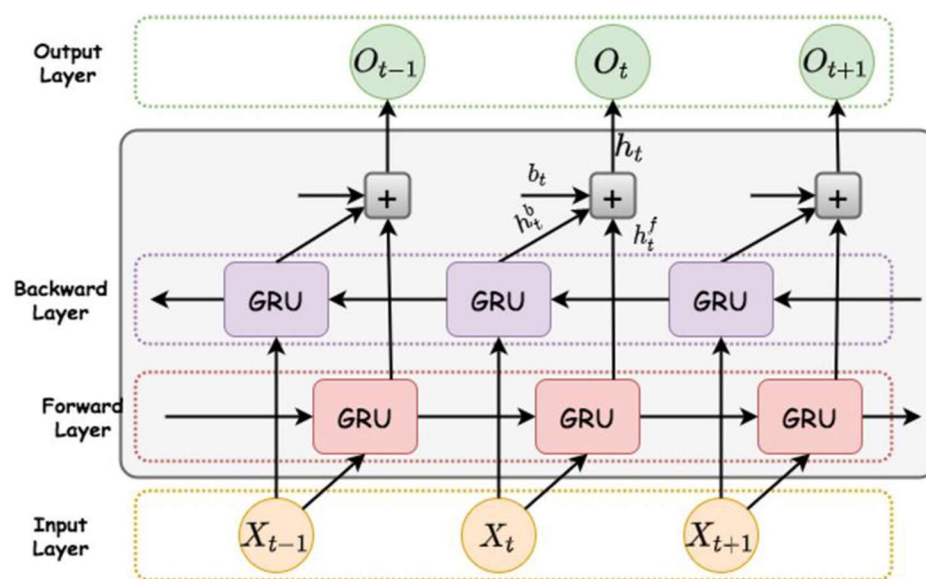


Figure 6. Architectural diagram of the BiGRU model [21].

Deep learning models are widely used in sentiment analysis. CNN, LSTM, and BiGRU are common examples in text classification. At the same time, their results still depend on the quality of the data, and how the model is designed, especially for Arabic text [22].

2.2.4 Sentiment Analysis Using Transformer-Based Models

Transformer-based models are used in sentiment analysis. These models help represent the meaning of text based on context. Also, they have a lot of advantages, one of them is the ability to understand the same word differently depending on the surrounding words. However, in Arabic sentiment classification, it is common to use models such as AraBERT, MARBERT, and CAMeLBERT [23].

2.2.4.1 AraBERT

AraBERT is a transformer-based model introduced for Arabic language understanding. It was pre-trained on large Arabic text corpora and presented as a model can support different Arabic NLP tasks. Cause of this, it has been used in tasks such as sentiment analysis and text classification, as shown in Figure 7 [24].

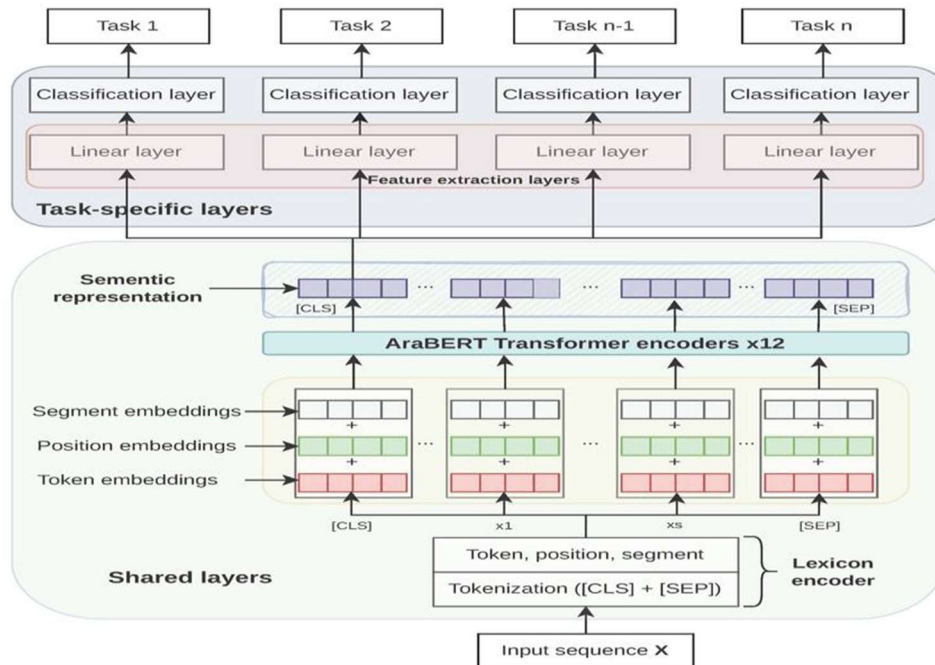


Figure 7. The feature extraction model architecture based on MTL and AraBERT [25].

2.2.4.2 MARBERT

MARBERT is an Arabic transformer-based model developed with a stronger focus on dialectal Arabic and text from social media. It was trained on a large amount of Arabic data, including informal content, cause of this its become a useful for tasks that involve noisy and dialectal text. For this reason, it has been used in Arabic sentiment classification, as shown in Figure 8 [26].

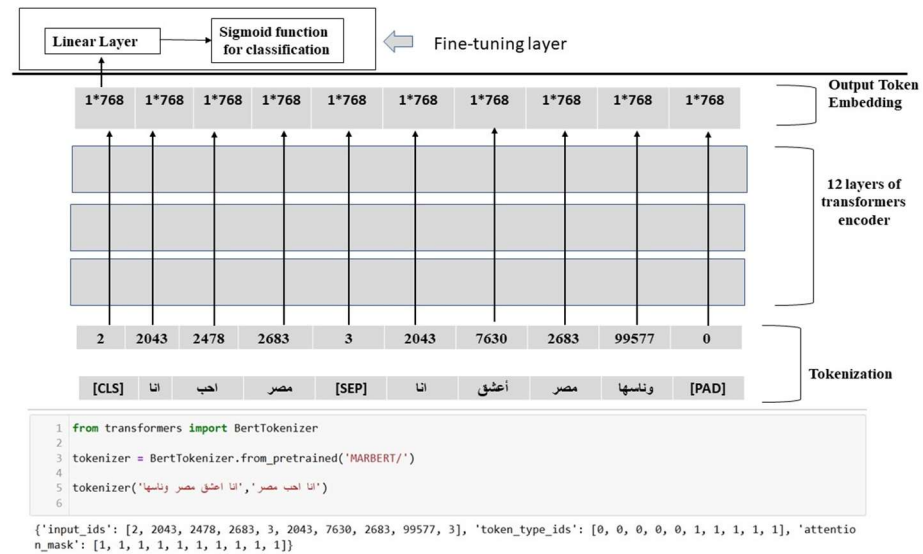


Figure 8. The fine-tuned MARBERT model architecture [27].

2.2.4.3 CAMELBERT

CAMELBERT is one of the Arabic pre-trained language models introduced for Arabic NLP tasks. The related paper focused on the effect of language variant, model size, and task type on model performance. Based on this, CAMELBERT was presented as a model that can support Arabic text classification under different Arabic forms, as shown in Figure 9 [28].

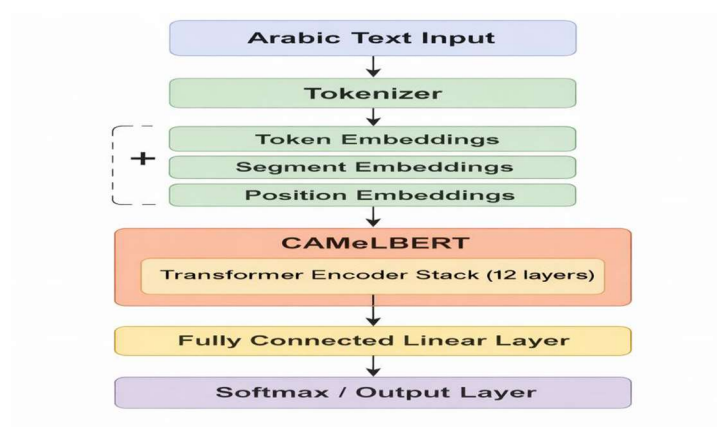


Figure 9. CAMELBERT architecture.

Transformer-based models have shown strong results in recent Arabic sentiment analysis studies. So, this is especially clear in Arabic models that were applied to sentiment tasks and dialectal or noisy text. Although these models usually need more computation and careful fine-tuning, they are still among the most suitable choices for Arabic sentiment classification. Cause of this, they are an important part of the proposed framework [29]. As shown in Table 1, compares the main sentiment analysis approaches used in Arabic NLP.

Table 1. Comparison of sentiment analysis approaches in Arabic NLP.

Approach	Common Models	Suitable Data	Main Advantage	Main Limitation
Machine Learning	Logistic ,SVM ,Regression XGBoost	Structured or relatively clean text represented using features such as TF-IDF	,Strong baseline models ,interpretable ,efficient and useful for comparison	Limited ability to capture semantic ,word order ,relations and contextual meaning
Deep Learning	,LSTM ,CNN BiGRU	Sequential text with semantic ,patterns including noisy text with word dependencies	Better representation learning than traditional machine learning and improved ability to capture local and sequential patterns	Requires more more ,training data and ,tuning effort higher computational cost
Transformer-Based Models	,ArABERT ,MARBERT CAMELBERT	,Arabic tweets ,dialectal text and noisy social media data	Strong contextual modeling and better suitability for complex Arabic sentiment analysis tasks	Requires more computational resources and careful fine-tuning

2.2.5 Performance metrics

To evaluate the performance of the sentiment classification models used in this study, four standard performance metrics were used: accuracy, precision, recall, and F1-score. These metrics are commonly used in classification tasks because they help assess model performance clearly. In classification tasks, the evaluation is usually based on four possible outcomes. True Positives (TP) are the cases that are correctly classified as positive, while True Negatives (TN) are the cases that are correctly classified as negative. False Positives (FP) are the cases that are incorrectly classified as positive, whereas False Negatives (FN) are the cases that are

incorrectly classified as negative. Accuracy measures the proportion of correctly classified instances out of the total number of instances. It gives a general indication of the model's overall performance. The accuracy is calculated using Eq. (1) [30].

$$\text{Accuracy} = (TP + TN) / (TP + TN + FP + FN) \quad (1)$$

Precision measures the proportion of correctly predicted positive instances out of all instances predicted as positive. It shows how accurate the positive predictions are. The precision is calculated using Eq. (2) [31].

$$\text{Precision} = TP / (TP + FP) \quad (2)$$

Recall measures the proportion of actual positive instances that are correctly identified by the model. It shows how well the model can detect positive cases. The recall is calculated using Eq. (3) [31].

$$\text{Recall} = TP / (TP + FN) \quad (3)$$

F1-score combines precision and recall into a single value by taking their harmonic mean. It is useful when both precision and recall need to be considered together. The F1-score is calculated using Eq. (4) [31].

$$\text{F1-score} = 2 \times (\text{Precision} \times \text{Recall}) / (\text{Precision} + \text{Recall}) \quad (4)$$

2.3 Literature Review

This section introduces recent related studies in Arabic sentiment analysis and focuses on Arabic tweets, dialectal content, noisy user-generated text, and transformer-based sentiment classification. The purpose of this review is to identify the main research directions in recent literature, underline the strengths and limitations of prior work, and clarify the academic position of the proposed study. The reviewed studies also help establish the research limitations that motivate the use of a hybrid machine learning approach with a transformer-based framework on the ASAD dataset.

Abdul-Mageed et al. [32] presented the NADI 2022 shared task for dialect identification and dialectal sentiment analysis on Arabic tweets. The study did not focus on a single standalone model but rather established a benchmarking framework to evaluate different systems on nuanced Arabic tweet data. The results showed that the winning system achieved an F1 score of 75.16 in the sentiment subtask. This study is important because it provides a clear benchmark for Arabic tweet sentiment analysis and demonstrates the performance level that strong systems achieved in a competitive setting. However, as a shared-task paper, its main contribution is evaluation and comparison rather than proposing a reusable sentiment framework.

Abdel-Salam et al. [33] developed a system for dialect and sentiment identification in nuanced Arabic tweets using an ensemble of prompt-based, fine-tuned, and multitask BERT models on the NADI 2022 dataset. Their system got first place in the sentiment subtask with F1-PN score 75.15. Also, it showed that using different BERT-based methods can work well for Arabic tweet sentiment analysis, especially when there are small dialect differences. However, the system was made mainly for a shared-task, so it is not easy to use in normal research work.

Bayrak and Issifu [34] explored domain-adapted BERT-based models for Arabic dialect identification and sentiment analysis on tweets. The authors used MARBERT and other Arabic BERT models with special preprocessing for Arabic on NADI 2022 tweets. The best sentiment model got 74.29 Macro F1-PN. Also, it shows that domain adaptation and preprocessing are important when working with Arabic tweets that have many dialects. And, the performance was still lower than the best recent ensemble systems. And that means domain adaptation alone may not be enough to get the best results.

Qaddoumi [35] presented an ensemble system for Arabic tweet sentiment analysis. The system was a simple averaging of the Arabic BERT-based models provided on NADI 2022 Subtask 2. They placed third place, with 72.49 Macro-PN-F1 in the development set, and 69.86 F1-PN in the test set. After present their system, it can be seen that a simple ensemble can be strong for Arabic tweet sentiment analysis. However, this system performance was still inferior to the top system in the same benchmark.

Belbachir et al. [36] proposed a sentiment analysis system for Algerian Arabic tweets. It used MARBERT with lexical and emoji-based filtering on the SemEval-2023 Task 12 dataset. Also, the filtering step improved the F1 score from 0.732 to 0.736. However, the improvement was small. On other hand, it shows that targeted preprocessing can help sentiment classification in dialect Arabic. Also, the study is important because it works on Algerian Arabic, which is not common in Arabic sentiment analysis. However, the main limitation is that the results are only for one dialect, so it is hard to use them for all Arabic tweets.

Alosaimi et al. [37] proposed a hybrid model for Arabic sentiment analysis by combining AraBERT with LSTM. Also, they tested this model on four Arabic sentiment benchmark datasets. And for improving classification performance, they need to use a transformer and sequential learning as the main goal. Also, the model achieved accuracy above 97%, which shows good potential for combined models in Arabic sentiment analysis. However, the experiments used only benchmark datasets, not Arabic tweets, so maybe the results do not apply directly to noisy or dialectal social media text.

Baniata and Kang [38] developed a Switch Transformer model for Arabic dialect sentiment analysis with Mixture of Experts. The model was tested on Arabic dialect datasets with three-polarity and five-point settings. The result showed 84.02% accuracy. Also, it showed that advanced transformer models can help with sentiment classification in Arabic, cause of dialects and non-standard writing make it difficult. However, the model needs more computation, so on the other hand, it may not be practical for lightweight academic use.

Almaqtari et al. [39] developed a hybrid model for Arabic sentiment analysis using AraBERT, multi-channel CNN, and BiGRU on three Arabic review datasets. The accuracy reached 96.92% for ACRD, 96.68% for HARD, and 92.93% for LABR. The study also reported that the advantage of combining contextual, local, and sequential features can improve

classification performance. Meanwhile, the addressed sentiment analysis of reviews rather than tweets.

Alawneh et al. [40] examined the effect of preprocessing and emoji encoding on Arabic sentiment analysis by using a CNN-LSTM framework on the Arabic Sentiment Twitter Corpus (ASTC). And the best result was 91.85% accuracy. This was after several preprocessing steps. In addition, the steps were to keep non-Arabic words, remove punctuation, use Snowball stemming, change emojis to Arabic text, and use Keras embeddings. Meanwhile, the main importance is that tweet preparation is very important for Arabic sentiment classification, especially in social media, where people use informal language and emojis. However, the main limitation is that the experiments were only done on one tweet dataset.

Mansour et al. [41] proposed a transformer-based ensemble model for dialectal Arabic sentiment classification. Also, the study compared TF-IDF baselines, machine learning models, deep learning models, and transformer models like CAMeLBERT, MARBERT, and XLM-RoBERTa on ASTD, ASAD, and TEAD datasets. The ensemble yielded some of the highest performances, with 90.4% mean accuracy, 87.7% F1 and 95% F1 on ASAD. As a result, it is presently one of the reference systems for Arabic tweet sentiment analysis. Due to that, the best system was still an ensemble, and in consequence it is more complex and less easy to interpret than a single system.

Aljomah et al. [42] explored Arabic sentiment analysis on noisy text by using a pre-trained transformer model. Also, the authors fine-tuned CAMeLBERT and compared it with AraBERTv02-twitter and MARBERT on the MARSAS dataset. Fine-tuned CAMeLBERT got 92% accuracy and 91.5% F1-score. This is important because noisy Arabic tweets are hard to classify because of spelling variation, informal words, and dialects. However, the experiments were only on one noisy-text dataset. Also, the study showed that fine-tuning transformers is useful for noisy Arabic social media sentiment tasks.

Jaber et al. [43] developed an ensemble machine learning approach for Arabic sentiment analysis using pre-trained embeddings and ensemble voting on ArTwitter and Syria_Tweets. The results showed strong performance, including 92.43% accuracy and 92.00% F1-score on ArTwitter. This study is useful because it shows that traditional ensemble machine learning can still achieve high results when supported by strong embeddings. It also reminds us that strong Arabic sentiment performance does not always require a fully end-to-end transformer

pipeline. Nevertheless, the method remains more classical in design than recent transformer-based sentiment frameworks.

Alharbi et al. [44] presented the AHaSIS shared task on sentiment analysis for Arabic dialects using a multi-dialect hospitality review dataset. One of the top-performing system achieved a Micro-F1 score of 0.81. So, the study a useful reference for dialect-sensitive evaluation in Arabic sentiment analysis. But, still cause the dataset is based on hospitality reviews rather than tweets, its relevance to tweet sentiment classification remains indirect.

Alghamdi et al. [45] examined sentiment analysis in Saudi-English code-switching text and introduced the Saudi-English Sentiment Analysis Code-Switching Dataset (SESA-CSD). Several small language models were evaluated, and the best result came from XLM-RoBERTa, which achieved 95.50% accuracy and 95.53% Macro-F1. These results are strong and recent, and they are also connected to Saudi social media language use. However, the study focuses on code-switching rather than fully Arabic-only tweets, so it is closer to a supportive reference than a directly matching one.

Alsahafi et al. [46] performed sentiment analysis using deep learning methods on Arabic tweets during Riyadh Season. For this, the authors created a dataset out of 33,300 tweets from X and compared CNN, LSTM, CNN-BiLSTM, and BiLSTM with attention with AraBERT embeddings. The best performance was obtained by BiLSTM with attention with AraBERT embeddings, with an accuracy of 83.37%. This is interesting as a more recent work, it concerned tweets and it was linked to the Saudi social media context. However, it is hard to generalize the results to other Arabic tweet dataset due to the fact that the dataset is linked to one event domain only.

2.4 Comparative Analysis of Related Work

From the literature reviewed, it appears that Arabic sentiment analysis has advanced quite far in recent years, and these studies focus much on shared tasks for benchmark-oriented research, such as competitions designed around analyzing a corpora of Arabic Tweets (and other micro-blogs and SMS) that tend to be found in dialect-rich environments and for this specific task.

In addition, these experiments are useful in providing baselines in terms of performance, and are a reminder that sentiment classification is challenging in the domain of Arabic social media. However, some of the experiments are designed to be used in challenge settings and as such their systems may not be easily used as straightforward research application systems. However, much of the work on Tweets and dialectal text is very relevant to this work project while in addition, some experiments relied on Arabic Tweets or noisy social media texts. That showed in better result because of preprocessing, dealing with emojis, and the use of dialects. Regarding to best result in sentiment analysis of Arabic tweets, it should not be only depending on the model, but also depend on dataset and on text preparation before classification. This is relevant in this work because dataset ASAD contained social media content. Thus, use of informal language and dialects is predictable. Also, there is switch from classical ML to deep learning, and now to transformer-based models. Despite the fact that classical ML still can be used as a baseline, particularly through the use of tf-idf features. There are greater numbers of studies, using neural networks based and transformer-based models. To reach better understanding. On the other hand, hybrid deep learning models also performed well, particularly those that employ contextual embeddings (such as in Trans and CapBERT) using an LSTM or a bi-GRU. Moreover, many high-performing deep learning models were tested on general benchmark or review datasets, and not specifically on Twitter data; thus, a direct comparison for classifying sentiment of Arabic social media content becomes difficult. Furthermore, models that base their predictions using transformer architecture are the state of the art for recent classifying sentiment in Arabic. Studies with the MARBERT, CAMELBERT, XLM-RoBERTa, and similar transformers outperform traditional approaches, particularly where the input contained dialect or noise in the Arabic language. In the work of Mansour et al., the machine learning, deep learning, and transformer models were compared across three datasets: ASTD, ASAD, and TEAD. They attained performance measures such as accuracy (90.4%), F1-score (87.7%), and precision (95%) on the ASAD dataset. Their best classifier was an ensemble system. Therefore, this approach is more complicated and difficult to understand than applying one single model. Moreover, a solution still needs to be broken down. The authors of most recent state-of-art research work combine ensemble system, treat them as benchmark comparisons, or try using different datasets which are not exactly same as the conducted task. Hence, the proposed project departs to the alternative framework through the

application of domain-adaptive transformer feature extraction with hybrid machine-learning classifier to the ASAD dataset. The project applied TF-IDF within Logistic Regression or SVM as the benchmark, and MARBERT as the proposed main transformer model, and CAMELBERT as the control model. The working and methods of the studies discussed above are summarized in Table 2, Comparative Study of Related Work.

Table 2. Comparative Study of Related Work

Authors	Year	Objective	/ Dataset Domain	Model / Technique	Key Result	Limitation
Abdul-Mageed et al.	2022	To benchmark dialect identification and dialectal sentiment analysis on Arabic tweets	/ NADI 2022 Arabic tweets	Shared-task benchmark systems	Winning system achieved 75.16 F1 in sentiment subtask	Benchmark not a ,paper single standalone model
Abdel-Salam et al.	2022	To identify dialect and sentiment in nuanced Arabic tweets	/ NADI 2022 Arabic tweets	fine- ,Prompt-based and ,tuned multitask BERT ensemble	Achieved 1st place with 75.15 F1-PN	Designed mainly for shared-task setting
Bayrak and Issifu	2022	To improve nuanced Arabic dialect and sentiment identification through domain adaptation	/ NADI 2022 Arabic tweets	Domain-adapted MARBERT and Arabic BERT models	Best sentiment model achieved 74.29 Macro F1-PN	Lower than stronger recent ensemble systems
Qaddoumi	2022	To build an ensemble for Arabic tweet sentiment analysis	/ NADI 2022 Arabic dialect tweets	Average ensemble of Arabic BERT models	Macro- 72.49 PN-F1 on 69.86 ,dev F1-PN on ranked ,test 3rd	Simpler ensemble performed below top systems
Belbachir et al.	2023	To improve sentiment analysis of Algerian Arabic tweets	SemEval-2023 Task 12 Algerian / Arabic tweets	MARBERT with lexical and emoji-based filtering	Improved from 0.732 F1 to 0.736 F1	Focused on one dialect only
Alosaimi et al.	2024	To improve Arabic sentiment analysis using a hybrid model	Benchmark Arabic sentiment datasets	LSTM + AraBERT	Reported accuracy above 97%	General benchmark not ,setting tweet-specific
Baniata and Kang	2024	To address Arabic dialect sentiment analysis using advanced transformers	Arabic dialect sentiment datasets	Switch Transformer with Mixture of Experts	Achieved 84.02% accuracy in one setting	High computational cost

Almaqtari et al.	2024	To improve Arabic sentiment classification in consumer reviews	,ACRD ,HARD / LABR Arabic reviews	Multi- + AraBERT + channel CNN BiGRU	96.92% on ,ACRD 96.68% on ,HARD 92.93% on LABR	Review- not ,based tweet-based
Alawneh et al.	2024	To study the effect of preprocessing and emoji encoding on Arabic tweet sentiment analysis	/ ASTC Arabic Twitter corpus	CNN-LSTM with preprocessing and emoji encoding	Best setup achieved 91.85% accuracy	Evaluated on one tweet dataset only
Mansour et al.	2025	To improve dialectal Arabic sentiment classification with transformer ensembles	,ASTD ,ASAD / TEAD Arabic tweets	Ensemble of ,CAMELBER XLM- ,MARBERT RoBERTa and others	90.4% average ,accuracy ,87.7% F1 95% F1 on ASAD	Ensemble is heavier and less interpretable
Aljomah et al.	2025	To classify noisy Arabic sentiment text using transformers	/ MARSA noisy Arabic text	Fine-tuned CAMELBER vs AraBERTv02- twitter and MARBERT	92% ,accuracy 91.5% F1- score	One noisy- text dataset only
Jaber et al.	2025	To improve Arabic sentiment analysis using embeddings and ensemble ML	,ArTwitter Syria_ Tweets	Pre-trained + embeddings ensemble voting	92.43% ,accuracy 92.00% F1- score on ArTwitter	More classical than recent transformer frameworks
Alharbi et al.	2025	To evaluate sentiment analysis for Arabic dialects in hospitality domain	Multi-dialect hospitality reviews	Shared-task systems	Top system achieved 0.81 Micro- F1	Domain is not ,reviews tweets
Alghamdi et al.	2026	To analyze Saudi-English code-switching sentiment and introduce a new dataset	/ SESA-CSD Saudi- English code- switching social media text	Small language best with ,models XLM-RoBERTa	95.50% ,accuracy 95.53% ,Macro-F1 with 3%– 11% improvement over baseline	Code- switching not ,focus fully Arabic- only tweets
Alsahafi et al.	2026	To analyze sentiment of Arabic tweets during Riyadh Season	tweets 33,300 X (Twitter) /	,LSTM ,CNN ,CNN-BiLSTM + BiLSTM attention with AraBERT embeddings	Best model achieved 83.37% accuracy	Domain- specific to Riyadh Season

2.5 Chapter Summary

This chapter presented the main background of Arabic sentiment analysis, particularly on social media text and dialectical Arabic. It first described the term sentiment analysis, the challenges faced by researchers, the importance of sentiment analysis, and the settings used for sentiment classification task in previous research. Also, the traditional machine learning approaches, deep learning approaches, and the transformer-based approaches. In addition, this chapter illustrated the usage of the latest transformer-based approaches in modern Arabic sentiment analysis researches on social media, and dialectical Arabic. It also discussed about the metrics of performance in sentiment classification as they are important to evaluate the performance of models and compare between different approaches. Moreover, the studies reveal that there are many recent approaches achieved better results by using deep learning and transformer-based models. However, most of these studies do not exactly fit with the current task, so a focused and practical framework for Arabic tweet sentiment classification is still requiring. Based on the review, the domain-adaptive transformer feature extraction and hybrid machine learning classification applied on the ASAD dataset is used in the proposed project. Traditional machine learning models are used as baselines, MARBERT is the main model and CAMELBERT is added as a comparison model. Overall, this chapter provides the backgrounds for the methodology in the next chapter.

Chapter 3 Proposed Approach (Methodology)

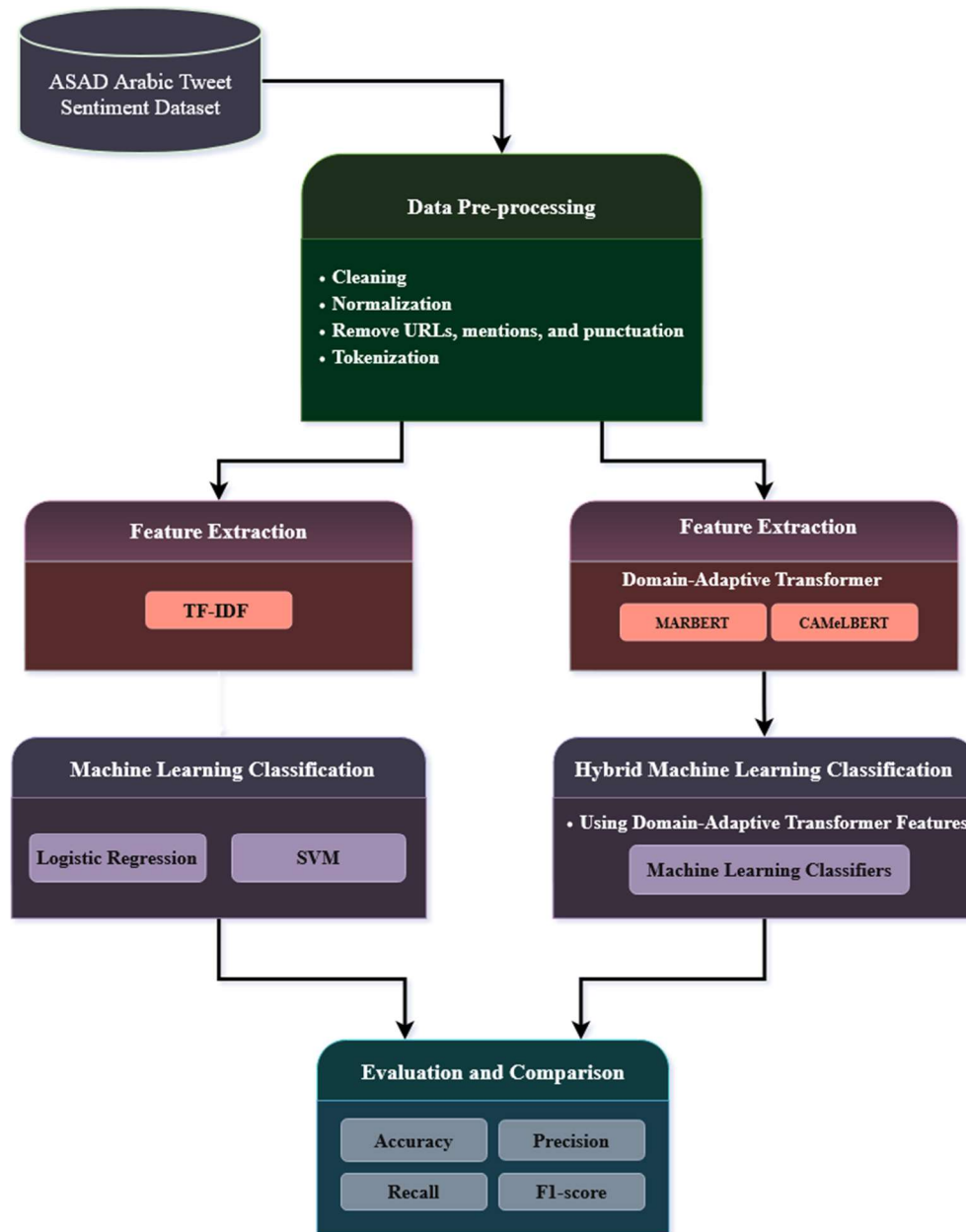


Figure 10. Proposed Framework.

Chapter 4 Experimental Results and Discussion

Chapter 5 Conclusion and Future Work

References

- [1] Shi, Z., AlShenaifi, N., Al-Refaie, H., Abdul-Mageed, M., & Ungar, L. (2025). *A comprehensive survey of contemporary Arabic sentiment analysis: Methods, challenges, and future directions*. In *Findings of the Association for Computational Linguistics: NAACL 2025*. Association for Computational Linguistics.
- [2] El Mekki, A., El Mahdaouy, A., Berrada, I., & Khoumsi, A. (2021). *Domain adaptation for Arabic cross-domain and cross-dialect sentiment analysis from contextualized word embedding*. In *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*. Association for Computational Linguistics.
- [3] Fang, Y., Xu, C., Guan, S., Yan, N., & Mei, Y. (2024). *Advancing Arabic sentiment analysis: ArSen benchmark and the improved fuzzy deep hybrid network*. In *Proceedings of the 28th Conference on Computational Natural Language Learning*. Association for Computational Linguistics.
- [4] Alfugaha, R., & AL-Smadi, M. (2025). *A hybrid transformer-based model for sentiment analysis of Arabic dialect hotel reviews*. In *Proceedings of the Shared Task on Sentiment Analysis for Arabic Dialects*. INCOMA Ltd.
- [5] Liaqat, M. I., et al. (2022). *Sentiment analysis techniques, challenges, and opportunities*.
- [6] Matrane, Y., et al. (2023). *A systematic literature review of Arabic dialect sentiment analysis*.
- [7] Taher, H. A. (2024). *Dual-Language Sentiment Analysis: A Comprehensive Evaluating SVM, Logistic Regression, XGBoost, and Decision Tree Using TF-IDF on Arabic and English Dataset*.
- [8] Aljabri, M., Chrouf, S. M. B., Alzahrani, N. A., Alghamdi, L., Alfehaid, R., Alqarawi, R., Alhuthayfi, J., & Alduhailan, N. (2021). *Sentiment Analysis of Arabic Tweets Regarding Distance Learning in Saudi Arabia during the COVID-19 Pandemic*.
- [9] Jha, U. M., & Guhan, B. (2024). *A comparison of the classification models of machine learning for predicting the risks of diabetes: An analysis of National Family Health Survey data of Delhi*.

- [10] Alyami, S. N., & Olatunji, S. O. (2020). *Application of Support Vector Machine for Arabic Sentiment Classification Using Twitter-Based Dataset*.
- [11] Cardoso-Fernandes, J., Teodoro, A. C., Lima, A., & Roda-Robles, E. (2020). *Semi-Automatization of support vector machines to map lithium (Li) bearing pegmatites*.
- [12] Alghamdi, H. M. (2024). *Unveiling Sentiments: A Comprehensive Analysis of Arabic Hajj-Related Tweets from 2017–2022 Utilizing Advanced AI Models*.
- [13] Chawla, A. (2024, August 14). *Formulating and implementing XGBoost from scratch*.
- [14] Al-Gaphari, G., et al. (2025). *Enhancing Arabic sentiment analysis during the Qatar World Cup: A study of generative pre-trained transformers and machine learning techniques*.
- [15] Mhamed, M., Sutcliffe, R., & Feng, J. (2025). *Benchmark Arabic news posts and analyzes Arabic sentiment through RMuBERT and SSL with AMCFFL technique*. Egyptian Informatics Journal, 29, 100601.
- [16] Elhassan, N., Varone, G., Ahmed, R., et al. (2023). *Arabic Sentiment Analysis Based on Word Embeddings and Deep Learning*.
- [17] Khamket, T., & Polpinij, J. (2022). *Automatically correcting noisy labels for improving quality of training set in domain-specific sentiment classification*.
- [18] Basiri, M. E., Nemati, S., Abdar, M., Asadi, S., & Acharya, U. R. (2021). *ABCDM: An attention-based bidirectional CNN-RNN deep model for sentiment analysis*.
- [19] Xu, Z. (2023). *Sentiment analysis of hotel comments based on LSTM and GRU*. In *Proceedings of the 2023 International Conference on Machine Learning and Automation*.
- [20] Alhumoud, S. O. (2024). *CNN-BiGRU-Focus: A Hybrid Deep Learning Classifier for Sentiment and Hate Speech Analysis of Ashura-Arabic Content for Policy Makers*.
- [21] Harrou, F., Dairi, A., Dorbane, A., & Sun, Y. (2023). *Energy consumption prediction in water treatment plants using deep learning with data augmentation*.
- [22] Minaee, S., Kalchbrenner, N., Cambria, E., Nikzad, N., Chenaghlu, M., & Gao, J. (2021). *Deep learning-based text classification: A comprehensive review*.
- [23] Mutawa, A. M., et al. (2025). *A Comparative Evaluation of Transformers and Deep Learning Models for Arabic Text Classification*. Applied Sciences, 15(9), 4941.
- [24] Antoun, W., Baly, F., & Hajj, H. (2020). *AraBERT: Transformer-based model for Arabic language understanding*.

- [25] Abd Elaziz, M. E., Dahou, A., Orabi, D. A., Awadallah, M. A., Lu, S., Al-Qaness, M. A. A., & Ewees, A. A. (2023). *A hybrid multitask learning framework with a Fire Hawk optimizer for Arabic fake news detection*.
- [26] Abdul-Mageed, M., Elmadany, A., Nagoudi, E. M. B., Pabbi, D., Verma, K., & Lin, R. (2021). *ARBERT & MARBERT: Deep bidirectional transformers for Arabic*.
- [27] Gabr, M. A.-E. I. I., Badr, A. Z., & Mahdi, H. M. K. (2023). Exploring strategies for measuring semantic similarity in short Arabic texts. *Indian Journal of Computer Science and Engineering*, 14(3), 480–489.
- [28] Inoue, G., Alhafni, B., Baimukan, N., et al. (2021). *The Interplay of Variant, Size, and Task Type in Arabic Pre-trained Language Models*.
- [29] Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2019). *BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding*.
- [30] Jurafsky, D., & Martin, J. H. (2025). *Speech and language processing* (3rd ed., draft). Stanford University & University of Colorado Boulder.
- [31] Terven, J., Cordova-Esparza, D.-M., Romero-González, J.-A., Ramírez-Pedraza, A., & Chávez-Urbiola, E. A. (2025). *A comprehensive survey of loss functions and metrics in deep learning*. *Artificial Intelligence Review*, 58(7).
- [32] Abdul-Mageed, M., Zhang, C., Elmadany, A., Bouamor, H., & Habash, N. (2022). *NADI 2022: The third nuanced Arabic dialect identification shared task*.
- [33] Abdel-Salam, R. (2022). *Dialect & sentiment identification in nuanced Arabic tweets using an ensemble of prompt-based, fine-tuned, and multitask BERT-based models*.
- [34] Bayrak, G., & Issifu, A. M. (2022). *Domain-adapted BERT-based models for nuanced Arabic dialect identification and tweet sentiment analysis*.
- [35] Qaddoumi, A. (2022). *Arabic sentiment ensemble NADI shared task 2*.
- [36] Belbachir, F. (2023). *Foul at SemEval-2023 Task 12: MARBERT language model and lexical filtering for sentiment analysis of tweets in Algerian Arabic*.
- [37] 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*.
- [38] Baniata, L. H., & Kang, S. (2024). *Switch-transformer sentiment analysis model for Arabic dialects that utilizes a mixture of experts mechanism*.

- [39] Almaqtari, H., Zeng, F., & Mohammed, A. (2024). *Enhancing Arabic sentiment analysis of consumer reviews: Machine learning and deep learning methods based on NLP.*
- [40] Alawneh, H., Hasasneh, A., & Maree, M. (2024). *On the utilization of emoji encoding and data preprocessing with a combined CNN-LSTM framework for Arabic sentiment analysis.*
- [41] Mansour, O., Aboelela, E. M., Talaat, R., & Bustami, M. (2025). *Transformer-based ensemble model for dialectal Arabic sentiment classification.*
- [42] Aljomah, F., Aldhafeeri, L., Alfadel, M., Alshahrani, S., Abbas, Q., & Alhumoud, S. (2025). *Enhancing Arabic sentiment analysis with pre-trained CAMELBERT: A case study on noisy texts.*
- [43] Jaber, A., Bahati, I., & Martínez, P. (2025). *Leveraging pre-trained embeddings in an ensemble machine learning approach for Arabic sentiment analysis.*
- [44] Alharbi, M. I., Chafik, S., Ezzini, S., Mitkov, R., Ranasinghe, T., & Hettiarachchi, H. (2025). *AHaSIS: Shared task on sentiment analysis for Arabic dialects.*
- [45] Alghamdi, S., Rayson, P., & Alotibi, R. (2026). *Improving on state-of-the-art models for sentiment analysis on Saudi-English code-switching text.*
- [46] Alsahafi, S., Alshutayri, A., & Alahdal, S. (2026). *Sentiment analysis of tweets during the Riyadh Season using deep learning.*