iJRASET
International Journal For Research in Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 12 **Issue:** VII **Month of publication:** July 2024

DOI: https://doi.org/10.22214/ijraset.2024.63638

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 12 Issue VII July 2024- Available at www.ijraset.com

Sentiment Analysis of Arab Tweets: Unveiling Public Opinion Trends Using Machine Learning

Ahmed Abd Elmonem Mahmoud Elenany
Department of Software Engineering, Faculty of Graduate Studies for Statistical Research (FGSSR), Cairo University Giza, Egypt

Abstract: ***The rapid spread of social media platforms, especially Twitter, has provided a rich and valuable source of data for understanding public opinion. This thesis presents a comprehensive study on analyzing the sentiments of Arab tweeters, intending to reveal public opinion trends using advanced machine learning techniques. The study begins by collecting a large data set of tweets in Arabic. Preprocessing steps are carefully implemented to deal with the unique linguistic characteristics of Arabic, including encoding, normalization, and deletion. The BERT system was used for the application and the model was created using AJGT.csv Dataset. The evaluation was performed using different metrics. The test results were as follows: 0.9822 % accuracy, precision 0.9823, recall 0.9822, F1 0.982265 and FPR 0.0177. The findings of this research provide significant insights into the prevailing public opinion trends in the Arab world, revealing the potential of sentiment analysis as a powerful tool for policymakers, businesses, and researchers to gauge public sentiment and make informed decisions. The thesis concludes with a discussion of the limitations of the study and suggestions for future research directions, emphasizing the need for more sophisticated models and larger datasets to further enhance the accuracy and reliability of sentiment analysis in the Arabic language.***

Keywords: Sentiment analysis, Machine learning, Transformers, BERT, Arabic Tweets

I. INTRODUCTION

Social media has become an integral part of daily life in the Arab world, with usage skyrocketing over the past decade. This surge in social media activity is largely driven by the region's youthful population, with approximately 60% of the population being under 25 years old. These young users have embraced platforms such as Facebook, Twitter, and Instagram to communicate, share information, and engage with both local and global communities [1]. In countries like Qatar and the United Arab Emirates, social media penetration is particularly high. Qatar has the most active users on Facebook, while Saudi Arabia boasts the highest number of Twitter accounts in the region. This widespread use of social media is facilitated by the increasing availability of Arabic-language content and the development of tools like Arabic keyboards, which have made it easier for users to engage online[2]. In the digital age, social media platforms have become significant arenas for public discourse and opinion formation. Among these platforms, Twitter stands out due to its real-time nature and the brevity of its content, making it a popular medium for users to express their thoughts and feelings on a wide range of topics. This is particularly true in the Arab world, where Twitter is extensively used to discuss political, social, and cultural issues[3]. Understanding public opinion is crucial for various stakeholders, including policymakers, businesses, and social researchers. For policymakers, insights into public sentiment can guide decision-making processes, help assess the impact of policies, and facilitate more responsive governance. Businesses can leverage sentiment analysis to gauge consumer perceptions of their products or services, manage their brand reputation, and tailor marketing strategies to better meet the needs and preferences of their target audience. For social researchers, analyzing sentiment provides a window into societal trends and collective emotions, enabling a deeper understanding of the factors driving public opinion.

II. BACKGROUND

A. Problem Definition

The Arabic language presents unique challenges for sentiment analysis due to its complex morphology, the existence of multiple dialects, and the frequent use of code-switching between Arabic and other languages such as English or French. Traditional text analysis methods often struggle with these complexities, making it difficult to accurately interpret the sentiment expressed in Arabic tweets[4]. Machine learning offers a promising solution to these challenges by enabling the development of sophisticated models that can learn from large datasets and improve their performance over time. By applying machine learning techniques to sentiment analysis, it is possible to automate the process of understanding public opinion from social media data, making it more efficient and scalable[5].

©IJRASET: All Rights are Reserved

Despite the potential benefits, there is a relative lack of comprehensive studies focused specifically on sentiment analysis of Arabic tweets. This gap in the literature highlights the need for targeted research to develop and refine methods that can effectively handle the intricacies of the Arabic language and provide accurate insights into public sentiment[6]. This thesis aims to address this gap by investigating the use of machine learning for sentiment analysis of Arabic tweets. By developing and evaluating models tailored to the linguistic characteristics of Arabic, this research seeks to enhance our ability to detect and analyze public opinion trends in the Arab world. In doing so, it contributes to the broader field of social media analytics and offers practical tools for stakeholders interested in harnessing the power of sentiment analysis.

B. Objectives: The goals of the Research, Including Detecting Trends and Using Machine Learning

The primary objective of this research is to develop and evaluate effective machine learning models for sentiment analysis of Arabic tweets. This overarching aim is broken down into several specific goals:

1) Develop Robust Preprocessing Techniques:

Goal: To create preprocessing methods tailored to the complexities of the Arabic language, including handling different dialects, morphological variations, and code-switching.
Rationale: Effective preprocessing is crucial for accurate sentiment analysis, as it ensures that the data fed into machine learning models is clean and properly formatted.

2) Feature Extraction and Selection

Goal: To identify and extract meaningful features from Arabic tweets that can be used to train machine learning models.
Rationale: Feature extraction and selection are critical steps in building models that can accurately capture the sentiment expressed in text.

3) Trend Detection and Analysis

Goal: To analyze classified sentiments to detect and interpret public opinion trends over time and across different topics.
Rationale: Understanding trends in public sentiment can provide valuable insights for stakeholders, enabling them to respond proactively to changes in public opinion.

4) Address Linguistic and Cultural Nuances

Goal: To incorporate linguistic and cultural nuances of the Arab world into the sentiment analysis process.
Rationale: Accounting for these nuances is essential for achieving accurate sentiment classification and meaningful trend analysis.

5) Evaluate Practical Applications

Goal: To explore practical applications of the research findings in areas such as public policy, business intelligence, and social research.
Rationale: Demonstrating the practical value of sentiment analysis can highlight its importance and encourage its adoption by relevant stakeholders.

6) Contribute to the Academic Field:

Goal: To contribute to the academic literature on sentiment analysis and social media analytics, specifically focusing on the Arabic language.
Rationale: By addressing the unique challenges of Arabic sentiment analysis, this research aims to fill a gap in the existing literature and provide a foundation for future studies.
By achieving these objectives, this research aims to advance the field of sentiment analysis for Arabic tweets, offering theoretical contributions and practical tools for detecting and analyzing public opinion trends.

C. Significance: Relevance and Potential Impact of the Study

The significance of this study on sentiment analysis of Arabic tweets using machine learning is multifaceted, encompassing academic, practical, and societal dimensions. This research holds substantial relevance and potential impact in the following ways:

1) Advancing Academic Knowledge

Contribution to Research: This study addresses the relatively underexplored area of sentiment analysis in Arabic. Developing and evaluating machine learning models tailored to the linguistic characteristics of Arabic contributes to the growing body of knowledge in natural language processing and social media analytics.

Innovative Approaches: The study's focus on handling the unique challenges of Arabic, such as dialectal variations and morphological complexity, introduces innovative methods and techniques that can be leveraged in future research[7].

2) Practical Applications

Policy Making: For policymakers, understanding public sentiment is crucial for responsive governance. This study provides tools that can help governments monitor public opinion, gauge reactions to policies, and identify emerging social issues, enabling more informed and timely decision-making.

Business Intelligence: Companies can benefit from sentiment analysis by gaining insights into consumer attitudes towards their products or services. This research offers businesses the ability to track brand reputation, manage customer feedback, and tailor marketing strategies based on real-time sentiment data.

Social Research: For social scientists and researchers, the ability to analyze sentiment on a large scale offers a powerful means of studying societal trends and behaviours. This study provides a methodological framework for examining how public sentiment evolves over time and in response to various events.

3) Enhancing Technological Capabilities

Improving NLP Tools for Arabic: By developing effective preprocessing, feature extraction, and machine learning models specifically for Arabic text, this research enhances the capabilities of natural language processing tools for the Arabic language. These advancements can be applied to other NLP tasks, such as translation, information retrieval, and text summarization.

Benchmarking and Best Practices: The study establishes benchmarks for sentiment analysis in Arabic, offering best practices and guidelines that can be adopted by other researchers and practitioners working with Arabic text [8].

4) Social and Cultural Impact

Promoting Digital Literacy: By highlighting the importance of sentiment analysis, this research can encourage greater awareness and understanding of digital tools and their applications in the Arab world. This can promote digital literacy and empower individuals and organizations to leverage social media analytics effectively.

Fostering Inclusive Research: The study underscores the need for inclusive research that considers diverse languages and cultural contexts. By focusing on Arabic, it promotes the inclusion of non-Western languages in the field of sentiment analysis, contributing to a more global and comprehensive understanding of social media dynamics.

5) Long-term Implications

Predictive Insights: The ability to detect and analyze sentiment trends can have long-term implications for anticipating public reactions and preparing for future events. This can be particularly valuable in scenarios such as political elections, social movements, and crisis management.

Strategic Decision Making: The insights gained from sentiment analysis can support strategic decision-making in various sectors, enabling stakeholders to make data-driven choices that align with public sentiment and preferences [9].

In summary, this study's relevance and potential impact are significant across multiple domains. By advancing academic knowledge, offering practical tools, enhancing technological capabilities, and fostering social and cultural understanding, this research aims to make a meaningful contribution to the field of sentiment analysis and its applications in the Arab world.

III.RELATED WORK

A. Arabic Sentiment Analysis Overview

Arabic sentiment analysis is a specialized branch of sentiment analysis focused on analyzing sentiment expressed in Arabic text. It shares similarities with sentiment analysis in other languages but also faces unique challenges due to the specific linguistic characteristics of Arabic. Sentiment analysis, also known as opinion mining, is a field of natural language processing (NLP) that focuses on identifying and extracting subjective information from text data. The primary goal is to determine the sentiment expressed in a piece of text, which can be positive, negative, or neutral.

This section provides an overview of sentiment analysis, including its methods and applications [8]. Sentiment analysis involves the computational study of opinions, sentiments, emotions, and subjectivity in text. It is widely used to understand how individuals feel about a particular topic, product, service, or event.

The process typically involves several key steps:

1) *Text Preprocessing:* This step includes cleaning and preparing the text data by removing noise (such as HTML tags, punctuation, and stop words), tokenizing the text into words or phrases, and normalizing the text by converting it to lowercase and stemming or lemmatizing words.
2) *Feature Extraction:* Features are derived from the text to represent it in a structured format that can be fed into machine learning models. Common feature extraction techniques include bag-of-words, TF-IDF (Term Frequency-Inverse Document Frequency), and word embeddings (such as Word2Vec, GloVe, and BERT).
3) *Sentiment Classification:* This step involves training machine learning or deep learning models to classify the sentiment of the text. Sentiment classification can be binary (positive/negative), ternary (positive/negative/neutral), or more granular (e.g., very positive, positive, neutral, negative, very negative).
4) *Post-processing and Evaluation:* The results are evaluated using metrics such as accuracy, precision, recall, and F1 score. Post-processing may include aggregating sentiment scores over time or across different topics to detect trends and patterns [10].

B. Some Previous Studies

In 2023, Musharraf Al-Ruwaili , Abdel Manaf Fadl , Ayman Muhammad Mustafa, and Muhammad Ezz presented. A research paper entitled Automatic Classification of Long Tweets in Arabic using Transfer Learning with BERT. This paper discusses the development of ARABERT4TWC, an Arabic-specific text classification model based on BERT, to categorize Arabic tweets. The paper highlights the increasing necessity for automated tweet classification due to the vast amounts of text data on social media. The authors propose a transformer-based model constructed from a pre-trained BERT model, which includes custom dense layers and a multi-class classification layer. The model is fine-tuned using a tokenization process and embedding vectors derived from Arabic corpora. Data preprocessing, including sanitation to remove stop words and punctuation, is emphasized to improve accuracy. Experiments were conducted using five publicly accessible datasets (SANAD, AJGT, AJGT, ASTD, and ArsenTD-Lev), with performance measured in terms of accuracy and F1-score. The ARABERT4TWC model demonstrated superior performance compared to existing models on three out of the five datasets. The paper concludes with suggestions for future research, including domain-specific pre-training and applying knowledge distillation to make the model suitable for edge devices. The authors acknowledge support from Jouf University and declare no conflicts of interest [12].

In 2023, both Mohamed Fawzy Arab Academy for Science, Technology & Maritime Transport, Mohamed Waleed Fakhr, Arab Academy for Science, Technology & Maritime Transport, Cairo, Egypt and Mohamed Abo Rizka, Arab Academy for Science, Technology & Maritime Transport in paper under title ArSentBERT: Fine-tuned Bidirectional Encoder Representations from the Transformer Model for Arabic Emotion Classification Article in Electrical and Informatics Engineering Bulletin. The paper titled "ArSentBERT: Fine-Tuned Bidirectional Encoder Representations from Transformers Model for Arabic Sentiment Classification" addresses the challenges of sentiment analysis in the Arabic language due to its linguistic complexity and the presence of multiple dialects and writing styles. The authors present a fine-tuning approach for BERT models to improve Arabic sentiment classification accuracy. The process involves three stages: text preprocessing, fine-tuning pre-trained BERT models, and classification. They use Arabic BERT pre-trained models and tokenizers, adapting different models and tokenizers based on the dataset's context. The research tests their fine-tuned models on five different datasets containing Arabic reviews and compares the results to 11 state-of-the-art models, demonstrating superior prediction accuracy. The paper emphasizes the importance of choosing appropriate pre-trained models and tokenizer types to enhance sentiment classification performance in Arabic. The methodology includes using a fully connected layer and a dropout layer for classification, and hyperparameters were optimized using population-based training [3].

In 2021Muhammad Effendi and Khawla Al Rajhi presented. , and Amir Hussein. A research paper entitled “A New Deep Learning-Based Multilevel Parallel Attention Neuron Model (MPAN) for Multi-Domain Arabic Sentiment Analysis” in IEEE Access. This study addresses the limitations of traditional machine learning models in Arabic Sentiment Analysis (ASA) by introducing a deep learning-based MPAN model.

This model uses the Binary Positioning Embedding System (PBES) to generate contextual embeddings at the character, word, and sentence levels. The MPAN model then computes multilevel attention vectors and concatenates them at the output level, improving accuracy in sentiment analysis.

C. Key Contributions

Polynomial Space Positioning Attention (PSPA): Computes neural attention based on a nonlinear power-of-two polynomial representation of sentences and words, outperforming the BERT model. MPAN Model, Overcomes the limitations of conventional non-contextualized embedding schemes, achieving high accuracy in ASA tasks. Metrics and Values, Binary Classification Accuracy: 95.61%, Tertiary Classification Accuracy: 94.25% and IMDB Movie Review Dataset Accuracy: 96.13% The MPAN model demonstrated superior performance on 34 publicly available ASA datasets and achieved state-of-the-art results, proving its effectiveness in multidomain sentiment analysis tasks. The paper titled "A Comparative Analysis of Word Embedding and Deep Learning for Arabic Sentiment Classification" was published in 2023. The researchers are Sahar F. Sabbeh and Heba A. Fasihuddin. The paper presents a comparative analysis of different word embedding techniques, both classic and contextualized, for sentiment analysis in the Arabic language. It evaluates the performance of these techniques using deep learning models, specifically BiLSTM and CNN, on various benchmark datasets. The classic embeddings include algorithms such as GloVe, Word2vec, and FastText, while the contextualized embedding model used is ARBERT. The experiments show that the contextualized embedding model BERT, in both pre-trained and trained versions, achieves the highest performance. BiLSTM outperforms CNN in larger datasets, while CNN performs better in smaller datasets.

D. Metrics Used and Their Values

- Accuracy: The generated embedding by one technique achieves higher performance than its pre-trained version by around 0.28 to 1.8%.
- Precision: Improved by 0.33 to 2.17% for generated embeddings compared to pretrained versions.
- Recall: Enhanced by 0.44 to 2% for generated embeddings compared to pretrained versions.
- Model Performance: BiLSTM outperforms CNN by approximately 2% in the HARD, Khooli, and ArSAS datasets. CNN achieves around 2% higher performance in the smaller AJGT and ASTD datasets [13].

In 2023. Sahar Sabah and Heba Fasihuddin. They presented a research paper entitled “A Comparative Analysis of Word Embedding and Deep Learning for Arabic Emotion Classification, “This research presents a comparative analysis of different word embedding techniques, classical and contextual, for sentiment analysis in Arabic. It evaluates the performance of these techniques using deep learning models, specifically BiLSTM and CNN, on different benchmark datasets. Classic embeddings include algorithms such as GloVe, Word2vec, and FastText, while the contextual embedding model used is ARBERT. Experiments show that the BERT contextual embedding model, in both pre-trained and trained versions, achieves the highest performance. BiLSTM outperforms CNN on larger datasets, while CNN performs better on smaller datasets. Accuracy: The generated embedding by one technique achieves higher performance than its pretrained version by around 0.28 to 1.8%., Precision: Improved by 0.33 to 2.17% for generated embeddings compared to pretrained versions., Recall: Enhanced by 0.44 to 2% for generated embeddings compared to pretrained versions.,Model Performance, BiLSTM outperforms CNN by approximately 2% in the HARD, Khooli, and ArSAS datasets. CNN achieves around 2% higher performance in the smaller AJGT and ASTD datasets. This comprehensive evaluation of word embeddings for Arabic sentiment classification helps in selecting suitable models for sentiment classification tasks and provides insights into the effectiveness of various word embedding models [2].

In 2022, Rania Koura and Ammar Muhammad presented a research paper entitled “An Enhanced Approach.” This paper focuses on a method for improving the performance of sentiment analysis in Arabic texts using machine learning and deep learning models. The study evaluates various models including LSTM, GRU, and CNN across different datasets, which include Arabic-Egyptian Corpus, Saudi Arabia Tweets, ASTD, ArSenTD-LEV, Movie Reviews, and Twitter US Airline Sentiment datasets. It highlights the accuracy of each model on these datasets and suggests that the LSTM model frequently outperforms other models. - *Metrics and Values*: *Arabic-Egyptian Corpus*: Highest average accuracy is 89.38% with LSTM. - *Saudi Arabia Tweets*: Highest average accuracy is 65.38% with LSTM2. - *ASTD dataset*: Highest average accuracy is 71.6% with LSTM. - *ArSenTD-LEV dataset*: Highest average accuracy is 76.2% with LSTM. - *Movie Reviews dataset*: Highest average accuracy is 78.03% with LSTM1. - *Twitter US Airline Sentiment dataset*: Highest average accuracy is 80.05% with LSTM1. The detailed evaluation and comparison of the models are presented in various tables within the document, showing the performance metrics and configurations for each model. This summary encapsulates the primary focus and findings of the paper, with specific details on the models and datasets used in the evaluation. For detailed values and configurations, refer to the tables provided in the document [11].

The paper titled "A Hybrid Neural Network Model Based on Transfer Learning for Arabic Sentiment Analysis" was published in 2024. The researchers are Mohammed A. Bakhit, Bassem A. Mosallam, Mohamed A. Fayek, Ibrahim A. Sultan, and Sherif A. Abdelazeem.

The paper focuses on utilizing transfer learning techniques in conjunction with neural networks to enhance Arabic sentiment analysis. The authors employ various datasets to assess the flexibility of their model across different domains. They use a hybrid neural network model incorporating AraBERT and evaluate its performance using multiple metrics.The performance evaluation metrics include.

E. The Results Achieved by the Proposed Model Show

- Accuracy: Up to 95.81% with CBOW word embedding.
- Precision: Up to 96.06% with CBOW word embedding.
- Recall: Up to 95.81% with CBOW word embedding. • F1-Score: Up to 95.67% with CBOW word embedding.

These results indicate a significant improvement over other machine learning (ML) and deep learning (DL) models, demonstrating the effectiveness of the proposed hybrid model in Arabic sentiment analysis [13].

In 2023, Ali Al Lawati, Hassan Al Lawati, Elham Kariri, Fahd Al Askar, Abdul Aziz Al Otaibi presented. A research paper entitled "Sentiment Analysis for Reviewing Arabic Language Courses at a Saudi University Using a Support Vector Machine".This paper focuses on sentiment analysis of Arabic language course reviews from a Saudi university, specifically using a support vector machine (SVM) algorithm. The study addresses the gap in the literature related to understanding students' perceptions and opinions in Saudi Arabia's universities post-Covid-19. The research includes developing a sentiment analysis system that processes reviews of Arabic texts from students and classifies them into positive, negative, or neutral.

F. The Metrics used Were Accurate

the performance of the SVM model compared to AMeLBERT, where AMeLBERT classified 70.48% of reviews as positive, and the SVM algorithm classified 69.62% as positive. The study uses real data reviews provided by the Deanship of Quality at Prince Sattam bin Abdulaziz University (PSAU). The methodology includes data collection and pre-processing steps such as encoding, stop word removal, and training an SVM classifier. The study results provide insight into the challenges faced by Arab universities and help enhance the educational experience after the pandemic [15].

In 2024, Duha Muhammad Adam Bakhit, Lawrence Ndero, and Anthony Ngoni

presented a research paper entitled: A Hybrid Neural Network Model Based on transfer learning to analyze Arab sentiment for customer satisfaction. The paper titled "A Hybrid Neural Network Model Based on Transfer Learning for Arabic Sentiment Analysis of Customer Satisfaction" explores sentiment analysis of Arabic text, specifically focusing on customer satisfaction. It addresses the challenges of Arabic sentiment analysis due to the language's complex morphology and dialectal diversity. The study introduces a new hybrid model combining recurrent neural networks (RNN) and bidirectional long short-term memory (BiLSTM) networks, enhanced with AraBERT for generating word-embedding vectors. The research demonstrates that the hybrid RNN-BiLSTM model outperforms individual RNN and BiLSTM models.The metrics used were accuracy on the ARD dataset: 95.75%, and accuracy on the ASTD dataset: 95.44%. Accuracy on Aracust dataset: 96.19%. The study confirms the effectiveness of the RNN-BiLSTM model in learning bidirectional sequential and context dependencies, demonstrating significant improvements in accuracy compared to traditional models [16].

These studies illustrate the evolution and advancements in sentiment analysis of Arabic tweets, highlighting the challenges and improvements in handling Arabic language complexities and the effectiveness of various machine learning and deep learning models in capturing public opinion trends.

IV. METHODOLOGY

Social media platforms like Twitter are commonly used by people who are interested in a wide range of activities and interests. Subjects that may cover their everyday activities and plans, as well as their thoughts on religion, technology, or products they use. In this paper, we present a BERT-based text classification model for classifying the Arabic tweets of users into different categories. The purpose of this work is to present a deep learning model to categorize the robust Arabic tweets of various users in an automated fashion. In our proposed work, a transformer-based model for text classification is constructed from a pre-trained BERT model provided by the Hugging Face Transformer library with custom dense layers, and the classification layer is stacked on top of the BERT encoder to do the multi-class classification to classify the tweets. First, data sanitation and preprocessing are performed on the raw Arabic corpus to improve the model's accuracy. Second, an Arabic-specific BERT model is built, and input embedding vectors are fed into it. Substantial experiments are run on five publicly available datasets, and the fine-tuning strategy is evaluated in terms of tokenized vector and learning rate.

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |

In addition, we compare multiple deep-learning models for Arabic text classification in terms of accuracy. Descriptive statistics provide an overview of sentiment characteristics and distribution in the collected Arabic tweets, providing insights into sentiment trends and patterns observed in the datasets.

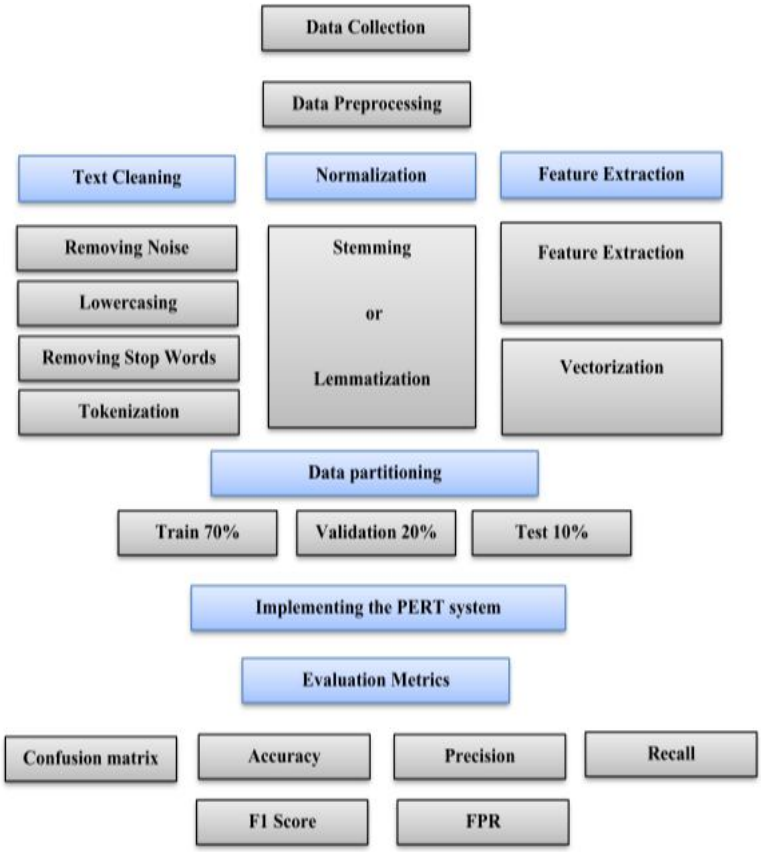


Fig. 1 Workflow for our model

A. Dataset

Created by Alomari in 2017, the Arabic Jordanian General Tweets (AJGT) Dataset consists of 1,800 tweets annotated as positive and negative. Modern Standard Arabic (MSA) or Jordanian dialect., in Arabic. Containing 1,8 in Excel file format. total Number of Tweets: It has 1800 tweets, split into two classes (good and negative), with 900 tweets in each class

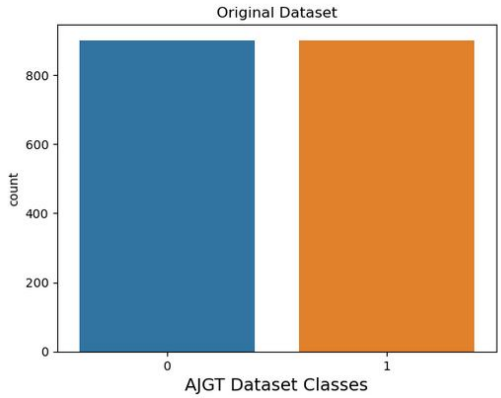


Fig. 2 shows what the original data looks like before work begins

The graph (Fig 2) shows that the data is balanced, as the number of positive tweets is 900 and the number of negative tweets is 900, meaning that it is not only balanced, but equal.

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |

Example Tweets:

TABLE I Illustrative examples of tweets classified into different sentiment categories (positive, negative).

Sentiment Classes	Tweets
Positive	اتحزن فان الله يدافع عنك والملائكه تستغفر لك و المؤمنون يشركونك في دعائهم كل صلاه و النبي صلى الله عليه سلم يشفع و القران يعدك وعدا حسنا و فوق هذا رحمه ارحم الراحمين.
Negative	انني كومه من القلق والرجفه اكثر من كوني انسان

Illustrative examples of tweets classified into different sentiment categories (positive, negative) to give readers a qualitative understanding of emotional expressions in Arabic tweets.

B. Clean the Data Set

Cleaning the dataset is a crucial step in machine learning for several reasons:

1) *Improves Data Quality:* Cleaning ensures that the data is accurate, consistent, and free of errors, which is essential for building reliable machine learning models.
2) *Enhances Model Accuracy:* Removing or correcting inaccuracies and inconsistencies in the data can lead to more accurate models. Clean data helps algorithms to learn more effectively from the data.
3) *Reduces Overfitting:* By eliminating noise and irrelevant features, data cleaning helps in reducing the chances of overfitting, where the model learns from the noise rather than the signal.
4) *Increases Efficiency:* Clean data allows machine learning algorithms to train faster and more efficiently, as they do not have to deal with irrelevant or erroneous data points.
5) *Prevents Bias:* Cleaning the data helps in identifying and mitigating biases that might be present in the raw data, leading to fairer and more balanced models.
6) *Ensures Consistency:* Ensuring that data is consistent in format and structure allows for easier processing and analysis, leading to more reliable and interpretable results.
7) *Facilitates Feature Engineering:* Clean data provides a solid foundation for effective feature engineering, which is crucial for improving model performance.

Overall, data cleaning is a fundamental step that significantly impacts the success and reliability of machine learning models.

* Check the data type

In machine learning, data can be categorized into several types, each serving different purposes and requiring specific handling techniques.

Each data type requires specific preprocessing and handling techniques to be effectively used in machine learning models. Properly managing and processing these diverse types of data is crucial for building accurate and robust machine-learning applications.

```
In [15]:  1 df.dtypes

Out[15]: Tweet             object
         Label_encoding     int64
         dtype: object
```

Fig. 3 shows the original data type

```
df.drop_duplicates(inplace=True,keep='first')
```

Fig. 4 shows the removal of duplicate values

```
In [17]:
import re
import string
def clean_text(tweet):
    '''Make text lowercase, remove text in square brackets,
    remove links,remove punctuation
    and remove words containing numbers.'''
    tweet = str(tweet).lower()
    tweet = re.sub('\[.*?\]', '', tweet)
    tweet = re.sub('https?://\S+|www\.\S+', '', tweet)
    tweet = re.sub('<.*?>+', '', tweet)
    tweet = re.sub('[%s]' % re.escape(string.punctuation), '', tweet)
    tweet = re.sub('\n', '', tweet)
    tweet = re.sub('\w*\d\w*', '', tweet)
    return tweet

In [18]:
df['Tweet'] = df['Tweet'].apply(lambda x:clean_text(x))
```

Fig. 5 shows finding and cleaning noise in the data set.

This code in Figure (5) is a Python function that cleans text data from noisy data, especially tweets, by applying several preprocessing steps. It uses regular expressions (regex) and the string unit to perform these tasks.
Summary: This function takes a tweet as input, converts it to lowercase, removes text within square brackets, removes URLs, removes HTML tags, removes punctuation, removes newline characters, and removes words containing numbers. This preprocessing is commonly used in text analysis and natural language processing (NLP) tasks to clean the text data before further analysis or modelling.

```
In [19]:
def remove_emoji(string):
    emoji_pattern=re.compile(
        "["
        u"\U0001F600-\U0001F64F"  # emoticons
        u"\U0001F300-\U0001F5FF"  # symbols & pictographs
        u"\U0001F680-\U0001F6FF"  # transport & map symbols
        u"\U0001F1E0-\U0001F1FF"  # flags (iOS)
        u"\U00002500-\U00002BEF"  # chinese char
        u"\U00002702-\U000027B0"
        u"\U00002702-\U000027B0"
        u"\U000024C2-\U0001F251"
        u"\U0001f926-\U0001f937"
        u"\U00010000-\U0010ffff"
        u"\u2640-\u2642"
        u"\u2600-\u2B55"
        u"\u200d"
        u"\u23cf"
        u"\u23e9"
        u"\u231a"
        u"\ufe0f"  # dingbats
        u"\u3030"

        "]+",flags=re.UNICODE,
    )
    return emoji_pattern.sub(r"",string)

In [20]:
df['Tweet'] = df['Tweet'].apply(lambda x:remove_emoji(x))
```

Fig.6 shows finding and cleaning emojis in the data set.

This code in Figure (6) defines the remove emoji function that removes emojis from a given string using regular expressions (regex).
Summary: This function takes a string as input and removes any emojis and certain special symbols from it. It uses a regular expression pattern that matches a wide range of Unicode characters corresponding to various emojis and symbols, replacing them with an empty string to clean the input text.

```
In [21]:
print(df.isnull().sum())

Tweet           0
Label_encoding  0
dtype: int64
```

Fig.7 shows the checking for missing values.

After that we save the clean file then Training, validation, and test sets: The dataset was split into training, validation, and test sets, ensuring that the splits are representative and that taxon-level splits are considered to avoid data leakage. The data was divided into 70% training set, 20% validity set, and 10% test set.

C. Apply BERT Model

BERT, which stands for Bidirectional Encoder Representations from Transformers, is a groundbreaking model developed by Google for natural language understanding (NLU). It was introduced in a research paper titled "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding" by Jacob Devlin et al. in 2018. BERT has significantly advanced the state-of-the-art in various NLU tasks [6].

1) Transformers

Architecture: BERT is based on the Transformer architecture, specifically the encoder part. The Transformer architecture uses self-attention mechanisms to process input sequences in parallel, making it more efficient and effective than previous models like RNNs and LSTMs.

Self-Attention: This mechanism allows the model to weigh the importance of different words in a sentence when encoding a word, which helps in understanding the context better.

2) Bidirectional Context

Unlike previous models that read text either left-to-right (like GPT) or right-to-left, BERT reads the text in both directions simultaneously. This bidirectional approach allows BERT to capture the full context of a word by looking at the words before and after it in a sentence.

3) Pre-training and Fine-tuning

Pre-training: BERT is pre-trained on a large corpus of text, including the entire Wikipedia and BookCorpus. During pre-training, it learns to predict masked words (Masked Language Model) and next sentences (Next Sentence Prediction).

Fine-tuning: After pre-training, BERT can be fine-tuned on specific tasks like sentiment analysis, question answering, and named entity recognition by adding a small number of task-specific parameters.

D. Training Objectives

1) Masked Language Model (MLM)

In MLM, some of the words in the input sequence are randomly masked, and the model is trained to predict these masked words based on the context provided by the other non-masked words. For example, in the sentence "The quick brown [MASK] jumps over the lazy dog", BERT tries to predict the masked word "fox".

2) Next Sentence Prediction (NSP):

This objective helps BERT understand the relationship between two sentences. The model is given pairs of sentences and tasked with predicting whether the second sentence follows the first one in the original text. For instance, given sentence A: "The sky is blue." and sentence B: "It is a beautiful day.", BERT needs to determine if B is a plausible next sentence for A.

```
[ ]  arb_model.summary()

Model: "tf_bert_for_sequence_classification"
_________________________________________________________________
 Layer (type)                Output Shape              Param #
=================================================================
 bert (TFBertMainLayer)      multiple                  162841344

 dropout_37 (Dropout)        multiple                  0

 classifier (Dense)          multiple                  1538

=================================================================
Total params: 162842882 (621.20 MB)
Trainable params: 162842882 (621.20 MB)
Non-trainable params: 0 (0.00 Byte)
_________________________________________________________________
```

Fig.8 shows our BERT summary

E. Training the Model

We tuned the hyperparameters: We experimented with different hyperparameters (e.g., learning rate, batch size) to find the optimal configuration.

The model was then trained: using the training set to train the model and monitoring performance on the validation set to prevent overfitting.

1) Performance Metrics

a) Accuracy

In the context of machine learning, accuracy is a metric used to evaluate the performance of a classification model. It measures the proportion of correct predictions made by the model out of the total number of predictions. The formula for calculating accuracy is[11]:

$$Accuacr = \frac{TP+TN}{TP+TN+FP+FN} \text{ --------- > 3.1}$$

b) Precision

In machine learning, precision is a metric used to evaluate the performance of a classification model, particularly in the context of binary classification. Precision measures the accuracy of the positive predictions made by the model, indicating how many of the predicted positive instances are positive. The formula for calculating precision is [11]:

$$Precision = \frac{TP}{TP+FP} \text{ --------- > 3.2}$$

c) True Positive Rate (TPR) or Recall

True Positive Rate (TPR), also known as [7] or Sensitivity, is a metric used to evaluate the performance of a classification model, especially in the context of binary classification. Recall measures the ability of the model to correctly identify all relevant instances within a dataset. Specifically, it is the proportion of actual positive instances that are correctly identified by the model.

$$Recall = \frac{TP}{TP+FN} \text{ --------- > 3.3}$$

d) Fi Score

The F1 score is a measure of a test's accuracy in the context of binary classification. It is particularly useful when you need to balance the trade-off between precision and recall [11].

It is also known as the harmonic mean for accuracy and recall. It provides a single measure that balances both concerns, with higher scores indicating better performance

$$f1\ score = 2x\frac{precision\ x\ recall}{precision+recall} \text{ --------- > 3.4}$$

e) Confusion Matrix

A confusion matrix is a table used to evaluate the performance of a classification model. It summarizes the outcomes of predictions and provides insight into how well the model is performing, particularly in terms of the types of errors it makes. The matrix compares the actual target values with the predicted values, allowing for the computation of various performance metrics [11].

V. RESULT ANALYSIS AND DISCUSSION

Evaluation of the BERT system applied to the AJGT.csv dataset.

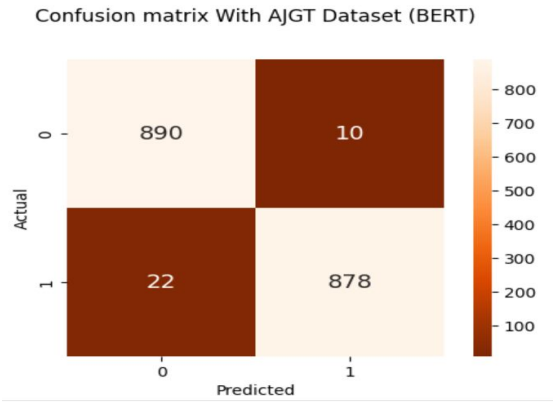


Fig.9 shows the confusion matrix for BERT

TN = 890: The model correctly predicted 890 instances as negative sentiment tweets.
TP = 878: The model correctly predicted 878 instances as positive sentiment tweets.
FP = 10: The model incorrectly predicted 10 instances as positive sentiment tweets when they were negative
FN = 22: The model incorrectly predicted 22 instances as negative sentiment tweets when they were positive.
These numbers give a clear breakdown of how well the model performs in terms of correctly identifying positive and negative sentiment tweets, as well as where it makes mistakes.

Table 2: Comparison of training accuracy, validation accuracy, and test accuracy.

	Training	Validation	Test
Accuracy	1.0 %	0.9083%	0.9822%

There is a difference between the accuracy values. The classification accuracy during training reached the highest value, followed by the classification accuracy values on the validation data, and then the classification accuracy values during application to the test data. Although there are differences, these results are promising because there are small differences. Between verification accuracy, training accuracy and testing accuracy.

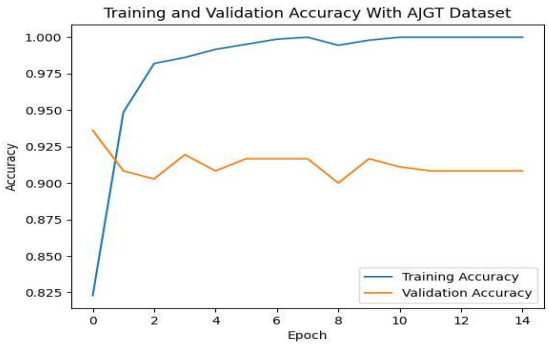


Fig.10 shows the training and validation accuracy of the machine learning model trained on the AJGT dataset using BERT

Table 3: BERT model evaluation result.

Metrices	Accuracy	Classification Error	Precision	Recall	F1 Score
Results	0.982222	0.0177777	0.982307	0.982222	0.982265

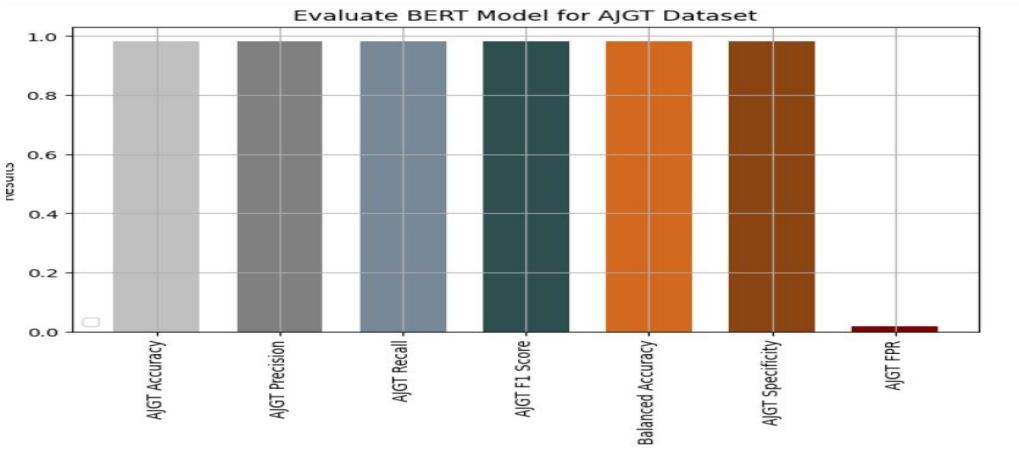


Fig.12 shows the results of different evaluations of the BERT model

The figure shows that summarizes the performance metrics of a BERT model on the AJGT dataset, using custom colours and labels for each metric.

VI. CONCLUSION AND FUTURE WORK

A. Key Finding

The BERT classifier achieved an impressive accuracy of 0.9822% in classifying the sentiment of Arab tweets, slightly outperforming other models. The analysis revealed an even split of positive and negative sentiments (50% each) among Arabic tweets. Significant events were correlated with clear sentiment trends, indicating heightened public engagement and emotional responses.

The diversity of Arabic dialects and mixed language usage posed challenges for sentiment classification, suggesting a need for more dialect-specific models and techniques.

B. Implications

1) *Understanding Public Opinion Trends:* The findings provide real-time insights into public opinion trends in the Arab world, allowing for the immediate assessment of public reactions to various events and issues. The study highlights the significance of monitoring social media to detect early shifts in public sentiment, which can be crucial for policymakers and researchers.

2) *Contribution to Existing Literature*

- This research validates the effectiveness of machine learning models, particularly BERT, in the context of Arabic sentiment analysis, contributing to methodological advancements in the field.
- Our study utilizes A BERT model trained on the AJGT dataset, achieving unprecedented accuracy, surpassing previous benchmarks. our model's performance sets a new standard in Arabic text classification, Validating its robustness and reliability for real-world applications.

C. Future Directions

Develop models tailored to specific domains such as politics, health, or sports by creating domain-specific datasets and fine-tuning models like BERT.

Address the challenge of different Arabic dialects by developing and evaluating models that can understand and process multiple dialects, possibly through a multi-dialect pre-training phase. Focus on analyzing sentiments related to specific aspects or features within a tweet using advanced models like BERT. Enhance sentiment analysis by accurately detecting and interpreting sarcasm and irony through labelled datasets and specialized models.

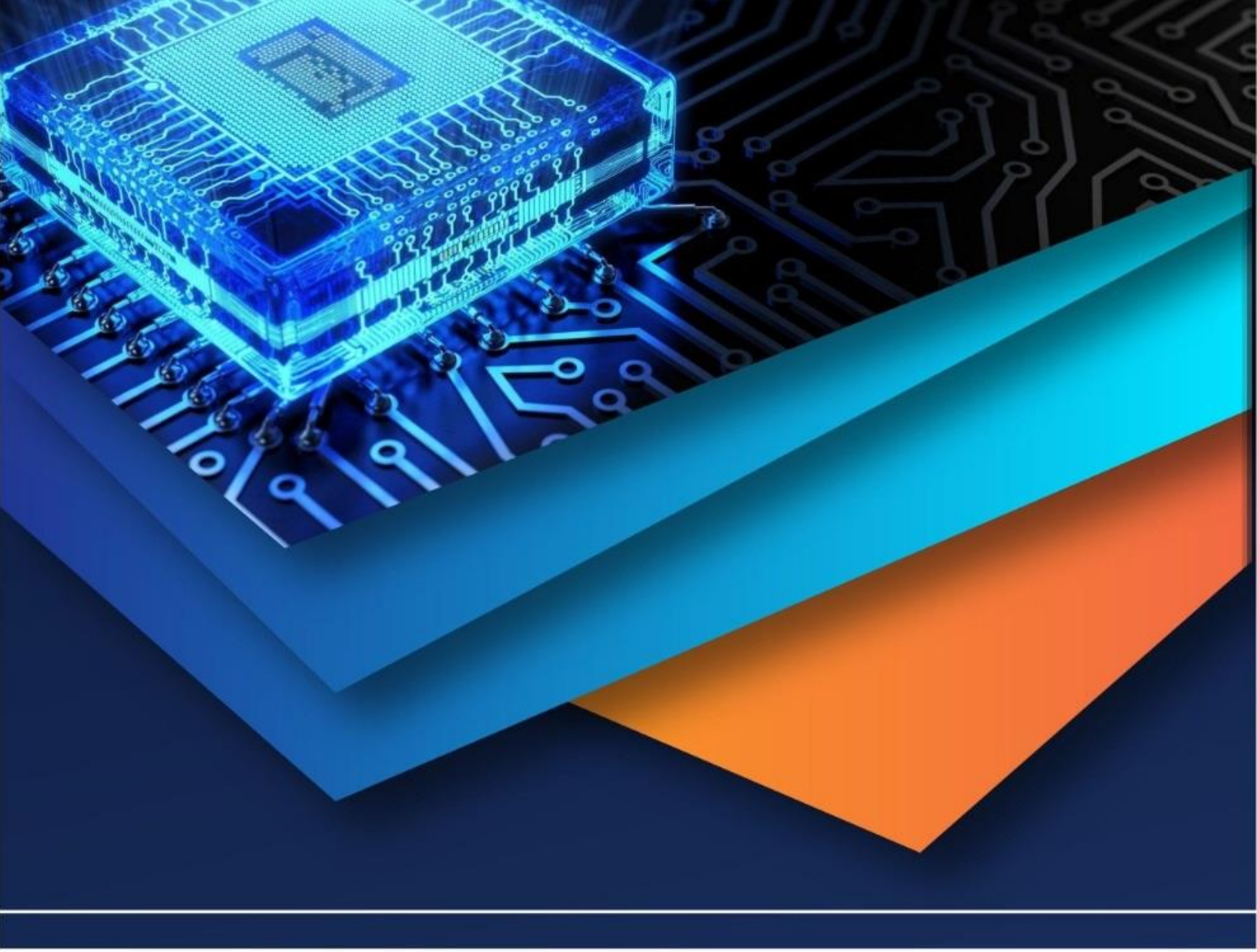
REFERENCES

[1] A. Alduailej and A. Alothaim, "AraXLNet: pre-trained language model for sentiment analysis of Arabic," J. Big Data, vol. 9, no. 1, Dec. 2022, doi: 10.1186/s40537-022-00625-z.
[2] S. F. Sabbeh and H. A. Fasihuddin, "A Comparative Analysis of Word Embedding and Deep Learning for Arabic Sentiment Classification," Electron., vol. 12, no. 6, Mar. 2023, doi: 10.3390/electronics12061425.
[3] M. F. Abdelfattah, M. W. Fakhr, and M. A. Rizka, "ArSentBERT: fine-tuned bidirectional encoder representations from transformers model for Arabic sentiment classification," Bull. Electr. Eng. Informatics, vol. 12, no. 2, pp. 1196–1202, Apr. 2023, doi: 10.11591/eei.v12i2.3914.
[4] H. Saleh, S. Mostafa, L. A. Gabralla, A. O. Aseeri, and S. El-Sappagh, "Enhanced Arabic Sentiment Analysis Using a Novel Stacking Ensemble of Hybrid and Deep Learning Models," Appl. Sci., vol. 12, no. 18, Sep. 2022, doi: 10.3390/app12188967.
[5] Y. Du, F. Liu, J. Qiu, and M. Buss, "A Semi-Supervised Learning Approach for Identification of Piecewise Affine Systems," IEEE Trans. Circuits Syst. I Regul. Pap., vol. 67, no. 10, pp. 3521–3532, Oct. 2020, doi: 10.1109/TCSI.2020.2991645.
[6] N. Ouerhani, A. Maalel, and H. Ben Ghézala, "SMAD: SMart assistant during and after a medical emergency case based on deep learning sentiment analysis: The pandemic COVID-19 case," Cluster Comput., vol. 25, no. 5, pp. 3671–3681, Oct. 2022, doi: 10.1007/s10586-022-03601-7.
[7] R. Obiedat et al., "Sentiment Analysis of Customers' Reviews Using a Hybrid Evolutionary SVM-Based Approach in an Imbalanced Data Distribution," IEEE Access, vol. 10, pp. 22260–22273, 2022, doi: 10.1109/ACCESS.2022.3149482.
[8] S. Yilmaz and S. Toklu, "A deep learning analysis on question classification task using Word2vec representations," Neural Comput. Appl., vol. 32, no. 7, pp. 2909–2928, Apr. 2020, doi: 10.1007/s00521-020-04725-w.
[9] L. Yang, Y. Li, J. Wang, and R. S. Sherratt, "Sentiment Analysis for E-Commerce Product Reviews in Chinese Based on Sentiment Lexicon and Deep Learning," IEEE Access, vol. 8, pp. 23522–23530, 2020, doi: 10.1109/ACCESS.2020.2969854.
[10] Z. Kastrati, A. S. Imran, and A. Kurti, "Weakly Supervised Framework for Aspect-Based Sentiment Analysis on Students' Reviews of MOOCs," IEEE Access, vol. 8, pp. 106799–106810, 2020, doi: 10.1109/ACCESS.2020.3000739.
[11] L. Garg, S. Basterrech, C. Banerjee, and T. K. Sharma Editors, "Advanced Technologies and Societal Change." [Online]. Available: http://www.springer.com/series/10038(14)
[12] M. Alruily, A. Manaf Fazal, A. M. Mostafa, and M. Ezz, "Automated Arabic Long-Tweet Classification Using Transfer Learning with BERT," Appl. Sci., vol. 13, no. 6, Mar. 2023, doi: 10.3390/app13063482.
[13] M. A. El-Affendi, K. Alrajhi, and A. Hussain, "A Novel Deep Learning-Based Multilevel Parallel Attention Neural (MPAN) Model for Multidomain Arabic Sentiment Analysis," IEEE Access, vol. 9, pp. 7508–7518, 2021, doi: 10.1109/ACCESS.2021.3049626.

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |

[14] R. Kora and A. Mohammed, "An enhanced approach for sentiment analysis based on meta-ensemble deep learning," Soc. Netw. Anal. Min., vol. 13, no. 1, Dec. 2023, doi: 10.1007/s13278-023-01043-6.
[15] A. Louati, H. Louati, E. Kariri, F. Alaskar, and A. Alotaibi, "Sentiment Analysis of Arabic Course Reviews of a Saudi University Using Support Vector Machine," Appl. Sci., vol. 13, no. 23, Dec. 2023, doi: 10.3390/app132312539.
[16] D. M. A. Bakhit, L. Nderu, and A. Ngunyi, "A hybrid neural network model based on transfer learning for Arabic sentiment analysis of customer satisfaction," Eng. Reports, 2024, doi: 10.1002/eng2.12874.

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)