A Comparative Analysis for Arabic Sentiment Analysis Models In E-Marketing Using Deep Learning Techniques.

Sara Almutairi [1], Fahad Alotaibi [2]

1. faculty of computer and information technology, King Abdulaziz University, Saudi Arabia, Riyadh, 11564, smateralmutairi@stu.kau.edu.sa

2. faculty of computer and information technology, King Abdulaziz University,Saudi Arabia, Riyadh, 11564, fmmalotaibi@kau.edu.sa

Abstract

The Internet has a huge amount of information when it comes to analysis, much of which is valuable and significant. Arabic Sentiment Analysis (SA) is a method responsible for analyzing people's thoughts, feelings, and responses to a variety of products and services on social networking and commercial sites. Several researchers utilize sentiment analysis to determine the opinions of customers in various areas, including e-marketing, business, and other fields. Deep learning (DL) is a useful technology for developing sentiment analysis models to improve e-marketing operations. There are a few studies targeting Arabic sentiment analysis (ASA) in e-marketing using deep learning algorithms. Due to a number of difficulties in the Arabic language, such as the language's morphological features, the diversity of dialects, and the absence of suitable corpora, sentiment analysis on Arabic material is restricted. In this paper, we will compare several Arabic sentiment analysis models. Also, we discuss the deep learning algorithms that are employed in Arabic sentiment analysis. The domain of the collected papers is Arabic sentiment analysis in e-marketing using deep learning. Our first contribution is to introduce and present deep learning models that are used in ASA. Secondly, investigate and study Arabic datasets utilized for Arabic sentence analysis. We create and develop a new Arabic dataset for Saudi Arabian communication companies, namely Sara-Dataset, to increase the quality and quantity of their services. Third, each collected study is assessed in terms of its methodology, contributions, deep learning techniques, performance, Arabic datasets in emarketing, and potential improvements in developing Arabic sentiment analysis models. Fourth, we analyzed several papers' performance in terms of accuracy, F-measure, recall, pre-procession, and area under the curve (AUC). Also, our comparative analysis includes feature selection (e.g., domain-specific selection) methods that are used in Arabic sentiment analysis. Fifth, we also discuss how to improve Arabic sentiment analysis using preprocessing techniques (e.g., word embedding). Finally, we provide a design model for analyzing Arabic sentiment about communications services provided by Saudi Arabian enterprises.

Keywords: Deep Learning; Comparative; Arabic Sentiment Analysis; E–marketing; Accuracy; Dataset; Feature Selection; Pre–processing CNN; LSTM.

Introduction

Sentiment analysis is an artificial intelligence technique that employs techniques to analyze whether an opinion is positive or negative. It's a powerful tool in the election process and social media to classify people's opinions towards things (e.g., products) [1–3]. Sentiment analysis is recognized as a significant technology for effectively studying customers' opinions. Preparing data, recognizing and identifying respondents, and evaluating findings are the primary components of sentiment analysis[4–10]. There are several studies targeting sentiment analysis in e-Marketing using deep learning algorithms. In this paper, we will conduct and analyze several studies addressing sentiment analysis in e-marketing using deep learning algorithms. First, we discuss and analyze Arabic sentiment analysis studies in e-marketing. Then, we will conduct a comparative analysis of Arabic sentiment analysis models using deep learning. The collected study is evaluated in terms of its methodology, contribution, deep learning techniques, performance, Arabic datasets, Emarketing, and potential improvements in developing Arabic sentiment analysis models. After that, we evaluated the performance of several papers in terms of accuracy, F-measure, recall, pre-procession, and area under the curve (AUC). In this paper, we will introduce a design model for Arabic sentiment analysis. We create a dataset for communications services provided by Saudi Arabian enterprises.

The rest of the paper is structured as follows: In Section 2, we present Arabic sentiment analysis and deep learning techniques, including artificial intelligence, machine learning, Arabic sentiment analysis, and Arabic datasets. Section 3 discusses ASA studies that employ deep learning, and Arabic sentiment analysis in e-marketing. In Section 4, we address comparative analysis for ASA models. Section 5 introduces a proposed design model for analyzing Arabic sentiment about communications services provided by Saudi Arabian enterprises. In Section 6, we conclude the paper and list future works.

Deep learning and sentiment analysis

In this Section, we will address the artificial intelligent, machine learning, deep learning models (CNN and LSTM), sentiment analysis, Arabic sentiment analysis, and our created Arabic dataset.

Artificial intelligence and machine learning

Artificial intelligence (AI) is the simulation of human intelligence processes by computer systems. There are many applications of AI, including speech recognition, natural language processing, expert systems, handwriting recognition, and robotics [11].

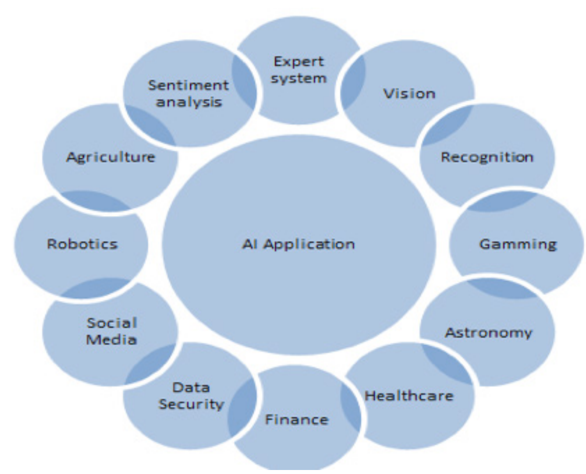

Fig. 1. AI applications

As shown in Fig. 1, AI has different applications such as sentiment analysis, robotics, and gaming. In this paper, we will focus on sentiment analysis. Machine learning and deep learning are parts of AI.

Machine learning: The challenge of text categorization with syntactic or linguistic characteristics It is part of artificial intelligence that is used to classify objects and things Machine learning is classified as supervised, unsupervised, or hybrid [12]:

Supervised learning: supervised learning is a sort of machine learning technique that makes predictions using a data set called the training data set. These data sets include both input and response values. It makes use of a high number of variables in supervised learning methods. Several techniques exist depending on their work on classification, such as classification trees, fuzzy logic, Naïve bayes network, genetic algorithms, neural networks, and support vector machine.

Unsupervised learning: This is a unique form of machine learning that is utilized in most situations to draw varied conclusions from data groups made up of input data with no labeled responses. When labeled training papers are unavailable, this method is utilized. Dividing the graph or data into groups—each group is called a cluster—makes it easier to analyze. Each cluster consists of elements that are similar. Clustering techniques can detect groups without prior knowledge or previous groups (the original data is unclassified, so it is considered unsupervised learning), and there are several cluster approaches used such as k-means, k-medoids, EM clustering, and outlier detection algorithms.

Deep learning algorithms

Deep learning is machine learning that enables computers to learn to perform classification methods directly from images, text, or voice. We will explain the convolutional neural network and long short-term memory architecture.

Convolutional Neural Network (CNN)

The Convolutional Neural Network is one of the most well-known and often utilized deep learning networks. CNN is a type of deep learning architecture or multilayer neural network. CNN is a neural network with several hidden layers, each of which has a number of two-dimensional planes filled with many neurons. Where the feature extraction module integrated into the CNN architecture is concerned, each neuron operates independently [13]. Fig. 2 depicts a CNN architecture.

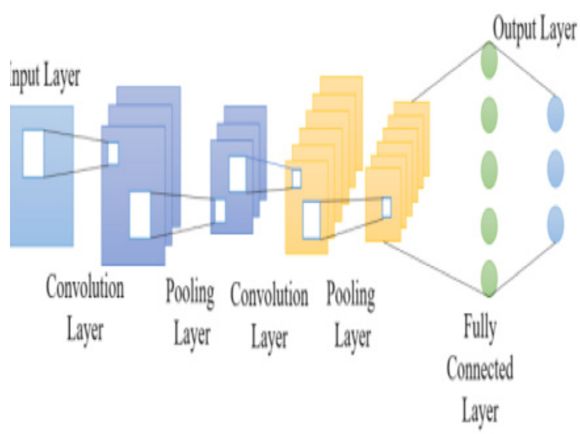

Fig. 2. CNN architecture [13]

CNN architecture includes:

- Input layer: The input layer can accept the input raw data set directly, where, the pixel values of one image are considered the input layer of CNN.
- Convolutional layer: It is composed of several convolutional filters. The output feature map is created by convolving the input picture, which is ex-

pressed as N-dimensional metrics.

- Pooling layer: It is also known as the down-sampling layer. Its primary purpose is to complete the second feature data extraction before moving on to the convolution layer.
- Fully connected layer: All the features maps are connected as inputs in this fully connected layer. The nodes of the neurons in the later layer are connected to the nodes of the neurons in the previous layer in general, but the nodes of the neurons in each layer are unconnected.
- Output layer: The number of neurons required for classification is usually proportional to the number of types to be identified.

Long Short-Term Memory

The LSTM architecture is regarded as a recurrent neural network (RNN) that was developed to overcome the limitations of the conventional RNN in terms of developing long-term dependencies. The parts of the LSTM unit are the gate, memory cell, output, and input gate, as shown in Fig. 3. The extra gates are in charge of regulating the flow of data into and out of the cell, while the memory cell is responsible for retaining values over time. Specifically, Fig 3 reports two architectures of LSTM. First, in Fig 3.a, an LSTM with a memory cell and two gates is shown. Figure 3.b, on the other hand, depicts an LSTM with a memory cell and a forget gate. Besides, one input layer, one output layer, and one self-connected hidden layer make up an LSTM cell. It is possible that the concealed unit contains basic units that can be fed into successive LSTM cells [14].

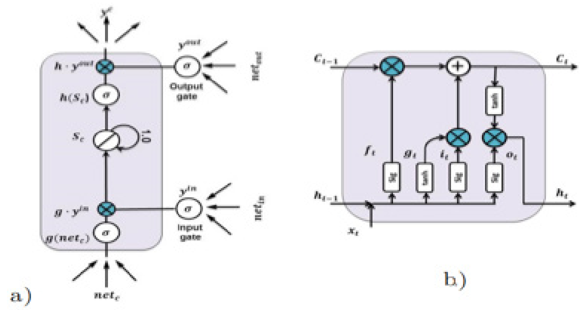

Fig. 3. LSTM architecture [14]

Arabic Sentiment Analysis (ASA)

SA is the automated extraction of expressed concepts from a given text. Using traditional techniques for managing and analyzing huge amounts of data is considered a critical challenge. The researchers developed an effective approach for studying and managing people’s opinions on social media called sentiment analysis

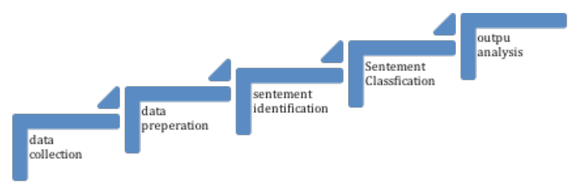

Fig 4. Sentiment analysis model.

As depicted in Fig. 4, sentiment analysis consists of data collection, data preparation, identification, classification, and output analysis. The collecting data stage involves collecting data from social media about the domain problem. Then the data is cleaned to be suitable for the classifier. Then the classifier will categorize the data as positive or negative.

Effective data science methodologies are used to classify a person's feelings and sentiments on social media. Data science is an interdisciplinary field that makes use of scientific methods, procedures, and algorithms to extract attitudes from tweets on Twitter. The classification of emotions has a key role in many application domains, including marketing, and business sectors can develop appropriately with the help of human emotions. Numerous methods, including deep learning, neuro-fuzzy, and optimization algorithms, are used to extract and classify sentiment [15].

Twitter provides a very important platform for speech and ideas. The users can discuss a wide range of occasions, products, and e-marketing strategies. There are several studies focusing on customer opinions toward products and the public's thoughts and attitudes toward a certain quality or a specific product [10, 16, 17, 18]. Nhan et al.[8] suggest that modern computational linguists have not paid enough attention to the Arabic language. They used sentiment analysis on social networks, which is a critical strategic technique for learning about customer interests. The main challenges, on the other hand, are efficiency, accuracy, and time consumption.

ASA datasets

They mention several limited datasets collected from Twitter, webpages, and blogs. OCA, LABR, and NA are examples of Arabic datasets. Also. In this subsection, we will explain most famous datasets and our created dataset called Sara-dataset

- OCA (Opinion Corpus for Arabic): this Arabic Corpus includes 500 film reviews, 250 good and a lot of bad, gathered from various Arabic online pages and blogs. It's only available in a limited size and for a certain film domain.
-
- LABR (a large scale Arabic book reviews dataset): is Arabic dataset contains over 63,000 book reviews have been written and graded on a scale of one to five stars. It applies only to a certain domain.
- NA: is an Arabic dataset which is used for Arabic sentiment analysis and contains a huge multi-domain dataset (33K annotated reviews for movies, hotels, restaurants, and products).
- Arabic Health Services Dataset (AHSD): This dataset contains over 2,000 posts, but the number of negative and positive records is not equal. This information relates to medical services in the context of health care.
- Arabic Twitter Dataset (ArTwitter): its politics datasets collected from Twitter social media. This dataset has thousands of records with balanced data.
- The Arabic Sentiment Tweets dataset (ASTD) was compiled. Twitter consists of fifty-four thousand Arabic tweets. It is an unbalanced dataset. It has 2479 tweets, distributed into positive and negative posts.
- *Our dataset, "Sara-Dataset."* We develop a new Arabic dataset for three Saudi Arabian communication firms to improve the quality and quantity of their services by collecting customer feedback. We selected three firms, which are Zain, Mobily, and Saudi Telecom (STC). We are targeting Twitter. We applied several preprocessing techniques, such as removing re-tweets, removing diacritics, removing punctu-

ations, removing repeating characters, normalization, removing URLs, creating word tokens and removing them, and removing Arabic stop words. Our dataset consists of 32336 tweets. After preparation, it becomes 20,000 tweets. Fig. 5 shows that Zain company take around 14000 tweets.

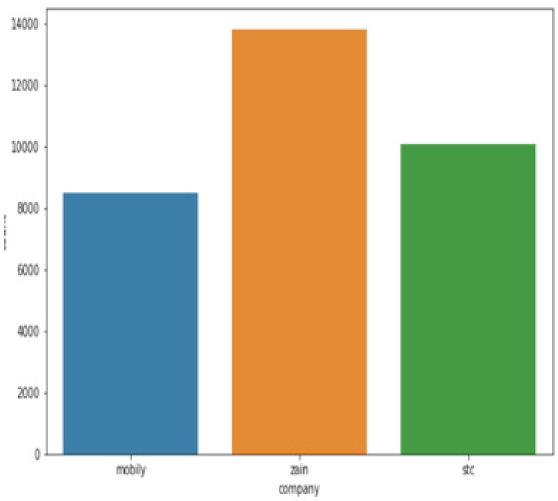

Fig 5. The distribution of tweets amongst companies.

The features of our dataset are (id, text, created_at, author_id, language, city, is_retweet, company, followers_count, tweet_count, listed_count, name, and Label). Table 1 displays a tweet sample from our dataset.

Table 1. Displays a tweet sample features

id	1.55365E+18
text	"السلام عليكم انترنت زين حاليا ضعيف جدا وبفر وثقيل ويعلق وبه عطل الان زين فايبر والبنق جدا سيء"
language	Arabic
city	Riyadh
company	Zain

ASA studies using Deep learning

There are several studies targeting ASA models using deep learning, such as:

Deep learning in ASA

Nassif et al. [1] introduce a full survey about using deep learning techniques in Arabic sentiment analysis. They selected papers from journals, conferences, and workshops in the field of Arabic sentiment analysis using deep learning techniques. The fields of collected datasets are marketing, social media, news, healthcare, education, business, politics, sport, and economics. The sources of the collected datasets are Wikipedia and blogs, news pages, online comments, Facebook, product reviews, and Twitter. Twitter is the most popular source, with 55 papers using it to generate datasets. Convolutional Neural Network (CNN), Long Shot Term Memory (LSTM), Deep Belief Network (DBN), Gated Recurrent Unit (GRU), Recursive Auto Encoder (RAE), Artificial Neural Network (ANN), Hierarchical Bidirectional LSTM (HBiLSTM), Bidirectional LSTM (BiLSTM), Bidirectional GRU (BiGRU), and Deep Neural Network (DNN) are used to perform sentiment analysis in Arabic.

Al-Ayyoub et al. [2] introduce numerous studies focusing on sentiment analysis in Arabic. There are few studies in Arabic SA, as depicted in Table 2.

Table 2. Arabic sentiment analysis published during 2010-2018

Year	# number of papers
2010-2012	Around 28 papers
2013-2016	Around 175 papers
2017-2018	Around 125 papers

Abbasi et al. [4] merge Arabic and English sentiment analysis. They are targeting binary SA problems in Web Forum (WF) postings related to English and Arabic. Each post is classified as supporting or not supporting labeling on a given issue. They use extraction and selection feature methods. They used a root extraction algorithm. They offered pre-processing features (e.g., word n-grams) in English text. They used heuristic data for ranking features, which has better performance than a genetic algorithm (GA). They used the Support Vector Machines (SVM) classifier to test the performance of their technique on a couple of small datasets, each including thousands of postings written in a mixture of languages. The experimental results using an entropy-weighted genetic algorithm with SVM achieve accuracy of over 91% on the 83 benchmark datasets in English and Arabic forums.

Guellil et al. [3] do a survey about sentiment analysis, which is used to classify and identify the sentence as a positive sentence or not. The collected study uses three approaches: a lexicon-based approach, a corpus-based approach, and a hybrid approach. Two of the most serious issues confronting the ASA are the diversity of dialects and the scarcity of Arabic resources. This study focuses on the most recent work, with the period of collected papers being between 2015 and 2019.

Medhaffar et al. [5] used three supervised machine-learning methods: support vector machine (SVM), Naive Bayes (NB), and MultiLayer-Perceptron (MLP) algorithms. They used the Python language, extracting their features using the Doc2Vec tool to classify the text as positive or negative. They evaluate the performance of the three models (SVM, NB, and MLP) in classifying the text as positive or negative in terms of accuracy, precision, and recall. The MLP achieves more than 78% accuracy, which is better than the other two algorithms. They achieved an error rate of 0.23, 0.22, and 0.42 using SVM, MLP, and binary NB, respectively. SVM has a precision of 76%, while BNB has a precision of 60%, and MLP has a precision of 78%.

Alayba et al. [7] used a combination of CNN and LSTM to improve the accuracy of sentiment classification of short messages taken from Twitter as positive or negative. They used many data sets, such as Main-AHS, SubAHS, Ar_Twitter, and ASDT. Then, by applying the deep learning combination algorithms, which are CNN and LSTM, they achieve accuracy as shown in Table 3.

Table 3. ASA models using Arabic dataset

Dataset Description	Algorithms	Accuracy
Arabic Health Services Dataset	CNN+LSTM	94.24%
Arabic Health Services Dataset	CNN+LSTM	95.68%
Arabic Twitter Dataset (ArTwitter)	CNN+LSTM	88.10%
Arabic Sentiment Tweets Dataset (ASTD)	CNN+LSTM	77.62%

Alayba et al. [7] introduce an Arabic-language dataset containing opinions on healthcare services that was gathered via Twitter. The dataset consists of 2026 tweets. where they show how to acquire data from Twitter as well as how to filter, preprocess, and annotate Arabic text to create an Arabic sentiment analysis dataset. They show how to acquire data from Twitter as well as how to filter, preprocess, and annotate Arabic text to create an Arabic sentiment analysis dataset. They used modern, standard Arabic dialects. They used machine learning and deep learning algorithms. They used multinomial NB, Bernoulli NB, LR, linear support vector, stochastic gradient descent, and nu-support vector as machine learning algorithms. Also, they used CNN and DNN as deep learning algorithms. Table 4 shows that multinomial NB, SVM, LSV, stochastic gradient descent, and DNN achieve over 90% accuracy in the classification of health services as positive or negative, as depicted in Table 4.

Table 4. Machine and deep learning algorithm in healthcare ASA.

Algorithm	Type	Accuracy
Multinomial NB	Machine learning	90.14%
Bernoulli NB	Machine learning	89.16%
LR	Machine learning	86.94%
Support Vector machine	Machine learning	90.88%
LSV	Machine learning	91.37%
Stochastic Gradient Descent	Machine learning	91.87%
Nu-Support Vector	Machine learning	87.82%
CNN	Deep learning	85%

According to Maha et al. [9], improving Arabic sentiment accuracy is considered an open problem. They used deep learning to improve Arabic sentiment analysis. Additionally, they explain the difficulties that face Arabic language accuracy. They used convolutional neural networks and long short-term memory models to predict the sentiment of Arabic tweets, but they did not achieve high accuracy. To classify the sentiment of Arabic tweets, they used ensemble models that combined CNN and long-short-term memory models. The CNN has an accuracy of 64.30 and an F1 of 64.09. LSTM has a dropout rate of 0.2, an accuracy of 64.75%, and an F1-score of 62.08%. The accuracy of the ensemble model is 65.05%, and the F1-score is 64.46%. They obtained the dataset from Twitter and prepared it using the following methods: Dates, numbers, and URLs are removed. Emojis are separated by a space to be treated as words, and diacritics and elongation are removed.

Altaher et al. [19], make shifting from traditional techniques to deep learning algorithms. They used stop-word and stemming as pre-processing methods and they used feature and information gain as feature weighting. Then, applying a deep learning algorithm to effectively and accurately classifies Arabic tweets either as positive or negative tweets. We collected a dataset consisting of 500 Arabic tweets, and the tweets mainly discuss general topics about education. the feed-forward architecture as deep learning. The accuracy when using deep learning is 90%. Using SVM, DT, neural network they accuracy are 85%, 67%, and 80%, respectively.

Khalil et al. [20], introduce semantic classification for multi-label Arabic emotion analysis (e.g., annoyed, happy and angry). They presented an optimized Bidirectional LSTM network. Where they used four optimizer techniques, which are Adam, Adamax, Nadam, and RMSpro. They used word embedding models for pre-processing. They used Task E-c: Detecting Emotions (multi-label classification) dataset. Where this task contains 2278 tweets for training, 585 tweets for development, and 1518 tweets for test data.

Manshu et al. [21], they used the Amazon dataset for books, electronics, kitchen domain. They apply CNN deep learning algorithm to measure the sentiment analysis via products. They achieve accuracy 81.98%. They use a new method of feature selection method where they combine between domain feature specification and domain independent features.

ASA in E-marketing

Several studies are targeting marketing and business analysis [15] [22-25]. In [15] Stock market investment is a significant aspect of every country's economy. Market research is an important part of making an investment in that filed. The SA is applied into market data, where these data are collected and per-processed before training and testing. This will assist investors in predicting where their money should be placed in the stock market, as well as in preserving the market's economic equilibrium.

The goal of Rosool et al. [10] is to determine what the public thinks about the top two worldwide apparel companies and compare the positive and negative sentiments of everyday consumers toward each one. It was found that positive reviews of Adidas are higher than those of Nike. While the neutral values record the satisfaction level among the online Twitter users for both brands, which is more than 60% of total reviews as depicted in Table 5,

Table 5. People opinions about Nike and Adidas products

Nike			Adidas		
Positive	Negative	Neutral	Positive	Negative	Neutral
24.5%	11.9%	63.6%	27.2%	11.7%	61.1%

Elzayady et al. [26], using traditional machine learning and deep learning algorithm in sentiment analysis. They used machine learning algorithm in sentiment analysis, which are KNN and DT. They used two Arabic datasets, which are HTL (hotel review) and LBAR (book review). The experimental results showed that there is no correlation between classifier and N-gram for feature representation. The best result in accuracy is 76.6% using the KNN algorithm with the HTL dataset. Using the NB algorithm with the LABR dataset and bigram features achieves an accuracy of 81.1%. On the other hand, using deep learning Using the HTL dataset, the accuracy is 85.38%; using the LABA dataset, the accuracy is 86.88%.

Al-Bayati et al. [27] mention that deep learning models have been provided to address and solve the Arabic sentiment analysis problem. They develop a sentiment analysis model using the LSTM deep learning algorithm. This model may be used to forecast Arab reactions to current political events, making this a crucial sector in decision-making. In commercial and marketing, where, Arabic enterprises and others sell their products to Arab consumers, they would gain from such a project since it would allow them to automatically collect feedback on their products and services to improve them.

They used Large-Scale Arabic Book Reviews (LABR), then selected a deep neural network, which is the LSTM. Converting texts to sequence numbers using word embedding method. Finally, the outputs of the LSTM layer are given to the SoftMax layer, which normalizes them and classifies the emotion of the input text as positive or negative. The experimental results show that the best accuracy and F-score are achieved when the hyperparameter LSTM out is set to 50 and the batch size is 256, where the accuracy is 82% and the F-score is 81.6%.

Yadav et al. [28], mention that social media is a strong tool for individuals to communicate their feelings in the form of thoughts and viewpoints. Marketing should research and analyze people's emotions and reactions to products.

Al-Bayati et al. [27], introduce a comprehensive overview of the most widely used deep learning models in sentiment analysis. They present a sentiment classification taxonomy and explore the consequences of common deep learning architectures, which are convolutional neural networks, recursive neural network, recurrent neural network (LSTM and gated recurrent units), and deep belief networks. They also highlight the most Arabic datasets used by deep learning to develop Arabic sentiment analysis models, which are the Stanford large movie review (IMDB), Yelp dataset, Stanford sentiment treebank (SSTb), Amazon review dataset, CMUMOSI, MOUD, Getty Images, Twitter Dataset, and Twitter Image Dataset. They mention the applications where sentiment analysis is performed, which are crime prediction, politics, business review analysis, business review, and financial market prediction. Also, they introduce the drawbacks of using deep learning for Arabic sentiment analysis. Whereas the main disadvantage is ensuring that the ASA provides the expected results. DL algorithms need a large amount of labeled data for training. In addition, some deep learning algorithms need to restart their parameters from the starting point, which incurs overhead in terms of time.

Comparative analysis for ASA models

In this section, we introduce a comparative analysis for Arabic sentiment analysis using a deep learning algorithm. We will analyze other researchers' reflections on our research. Then we are going to keep listing all the technologies used in developing ASA models.

Researches reflection on our research

In our research, we will focus on sentiment analysis to analyze people’s opinions about e-marketing products. Previous techniques suffered from low performance and accuracy when dealing with large datasets. Additionally, researchers do not compute the time complexity when the dataset becomes large. These problems will be solved in our proposed model, in addition to improving the company’s sales through the analysis of online product reviews. Tables 6.a and 6.b show machine learning and deep learning algorithms, accuracy, and enhancements that can be made in developing ASA models. Also, Table 6b shows research reflections on our work.

Table 6.a Deep learning algorithms and its reflection for our research

Paper Number	Year	Dataset source	Algorithms
[29]	2016	From Twitter. consisting of 1103 tweets (576 as positive and 527 labeled as negative	Lexical+ SVM.
[26]	2016	HTL and LABR which belongs to book rating	CNN+LSTM
[19]	2017	Arabic tweets are collected that consisting of 500 tweets related for education area	DT and SVM as traditional machine learning and the feed forward architecture as deep learning
[9]	2018	Arabic Sentiment Tweets Dataset (ASTD).	CNN is coupled with LSTM
[30]	2019	Twitter Arabic Hotels reviews	(LSTM) and Bidirectional LSTM
[20]	2021	SemEval 2018) contains 4372 tweets, which are organized into three categories: training, development, and testing	Bidirectional LSTM

Table 6.b Deep learning algorithms and its reflection for our research

Paper Number	Accuracy	Reflection	Enhancements
[29]	84.01%	we will usedeep learning with large dataset	Hybrid algorithm enhance the accuracy of semantic analysis The data set is small
[26]	Using HTl dataset, the accuracy is 85.38%. Using LABA dataset, the accuracy is 86.88%	Several papers using combination of CNN and LSTM	The accuracy is increased when using deep learning algorithm. And we must increase the datasets
[19]	Using deep learning accuracy 90%. using SVM 85'%	Deep learning algorithm outperform machine learning algorithm	Deep learning with weighting characteristic need time consuming
[9]	65.05% using ensemble learning	Ensemble learning can enhance the accuracy of ASA	accuracy has to be improved
[30]	82.6%	When dealing with huge dataset, the RNN has the overfitting problem	Solving overfitting problem
[20]	75.5% accuracy for validation and 49.8% for testing	SemEval is considered as large scale	accuracy has to be improved

Studies analysis

We can conclude that the researchers develop Arabic sentiment analysis using several types of Arabic datasets, and machine-learning, deep learning algorithms. combination between deep learning algorithms.

- Arabic datasets: Most studies are collecting datasets from Twitter web pages, newspapers, and blogs to form Arabic corpora, then making pre-processing to enhance the optimization. Also, some researchers develop a sentiment analysis model using datasets such as HTL for hotel reviews and LABR for book ratings. Main-AHS, Sub-AHS, Ar-Twitter ASDT (Arabic Sentiment Tweets Dataset), and Human-Annotated Arabic Dataset (HAAD) are also available in ASA. Stanford large movie review (IMDB), Yelp dataset, Stanford sentiment treebank (SSTb), Amazon review dataset, CMU-MOSI, MOUD, Getty images, Twitter dataset, and Twitter image dataset.

- Machine learning: Traditional machine learning includes supervised algorithms, unsupervised algorithms, and hybrid algorithms (supervised and unsupervised).
- Deep learning. Researchers use deep learning with different architectures.
- Combinations between deep learning algorithms such as CNN and LSTM increase the accuracy of ASA.
- Ensemble learning merges several learners to enhance the detection rate and performance accuracy. We will use ensemble-learning techniques such as bagging (randomly selected training data) or boosting (focusing on enhancing incorrect classification instances).
- Also, researchers are using several preprocessing techniques, such as
- Word Embedding
- Normalization, stop-word removal, negation terms and stemming
- Tokenization and segmentation of tweets.
- Spelling Correction.

There are several Arabic sentiment analysis levels such as:

- Sentence level
- Document level
- Phrases level
- Multidimensional level
- Multimodal level

Proposed Model

Our proposed model's key contribution is the development of Arabic sentiment analysis to gauge consumer perception of Saudi Arabian communication companies in order to increase the quality and quantity of their services through gathering client feedback. After we created our Arabic dataset, Sara-Dataset, our proposed work is to develop Arabic sentiment analysis for Saudi Arabia's opinions toward Saudi communication enterprises. As mentioned in the dataset section, the size of the datasets is small. Our model will use a large dataset, as mentioned in Section 2. Our model will select the best preprocessing, feature selection, and deep learning methods. We will enhance model accuracy using ensemble learning. As depicted in Fig. 6, our design model consists of the following steps:

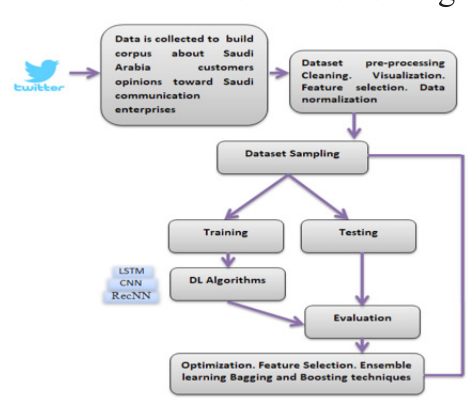

Fig. 6. Our proposed ASA model

1. Collecting dataset. We collected Sara-database from Twitter to build a corpus about Saudi Arabia's customers' opinions toward Saudi communication enterprises.
2. Pre-processing. This step is responsible for cleaning data, normalization, and identifying features. In this step we used several techniques such word embedding, filtering technique, normalization, stop-word removal, negation terms and stemming, tokenization and segmentation of tweets and spelling correction
3. Sampling. To generate a random sample for training and testing, the dataset is divided into K-fold cross validation.
4. Evaluation. Our model will be evaluated in terms of accuracy, precision, recall, f-measure, and area under the curve. Also, improving the quality and quantity of a company's services through customer feedback
5. Optimization. We will enhance our model in terms of feature selection methods and using bagging and boosting ensemble learning techniques. There are several issues related to our proposed ASA models. First, the main limitations of our model are the scarcity of Arabic resources and the fact that we do not cover all Arabic dialects. Second, we will combine several deep learning algorithms to enhance accuracy. Third, we will use bagging and boosting ensemble-learning techniques. Also, we will try using SMOTEDNN (Synthetic Minority Oversampling Technique with Deep Neural Network) to address air pollution classification [31]. SMOTEDNN was created to classify air pollution, and its accuracy is 99.90%. Finally, we'll try to apply transformer deep learning, which has a higher accuracy rate than deep learning [32].

Conclusion and Future Works

Several studies pertain to English sentiment analysis. Due to the difficult structure of Arabic language, Arabic sentiment analysis has several limitations and restrictions. This paper conducts a comparative analysis of Arabic sentiment analysis using deep learning. The domain of the collected papers is Arabic sentiment analysis in e-marketing using deep learning. The comparative metrics are: contribution of papers, deep learning techniques, performance, Arabic datasets, and potential improvements in developing Arabic sentiment analysis models. There is a shortage of Arabic sentiment datasets compared to English. One of the primary results of this analysis is that there is still a huge need for substantial research to acquire a better knowledge of Arabic dialects. There is no Arabic sentiment analysis model that can accurately handle all Arabic dialects. This has created a large gap in our understanding of ASA that researchers will try to fill in future research. Due to its complicated structure, numerous dialects, and scarcity of resources, the Arabic language has a number of constraints. Although the accuracy of Arabic sentiment analysis has increased using a deep learning model, there is still potential for improving accuracy.

A large Arabic dataset that we created belonged to a Saudi Arabia n communication corporation. Consequently, the aim of our future work is to develop a sentiment analysis model by incorporating deep learning techniques into our dataset. To improve the accuracy of our proposed Arabic sentiment analysis model, we must select the best preprocessing, feature selection, and deep learning methods for our dataset.

Conflict of Interest

The authors declare no conflict of interest

Acknowledgements

I thank all who in one way or another contributed in the completion of this thesis. First, I give thanks to God for protection and ability to do work. My special and heartily thanks to my supervisor, Professor Fahad Alotaibi who encouraged and directed me. His challenges brought this work towards a completion. It is with his supervision that this work came into existence. For any faults I take full responsibility. I also thank my family who encouraged me and prayed for me throughout the time of my paper.

References

[1] Nassif, Ali Bou, Ashraf Elnagar, Ismail Shahin, and Safaa Henno “Deep learning for Arabic subjective sentiment analysis: Challenges and research opportunities.” Journal of Applied Soft Computing 98 (2020): 106836.

[2] Al−Ayyoub, M., Khamaiseh, A. A., Jararweh, Y., & Al−Kabi, M. N. (2018). “A comprehensive survey of arabic sentiment analysis.” Journal of Information processing management, 56 (2019): 320-342.

[3] Guellil, Imane, Faical Azouaou, and Marcelo Mendoza. "Arabic sentiment analysis: studies, resources, and tools." Social Network Analysis and Mining 9.1 (2019): 1-17.

[4] Abbasi, Ahmed, Hsinchun Chen, and Arab Salem. "Sentiment analysis in multiple languages: Feature selection for opinion classification in web forums." ACM transactions on information systems (TOIS) 26.3 (2008): 1-34.

[5] Mdhaffar, S., Bougares, F., Esteve, Y., & Hadrich-Belguith, L. (2017, April). Sentiment analysis of tunisian dialects: Linguistic ressources and experiments. In Third Arabic Natural Language Processing Workshop (WANLP) (pp. 55-61).

[6] Alayba, Abdulaziz M., Vasile Palade, Matthew England, and Rahat Iqbal. "A combined CNN and LSTM model for arabic sentiment analysis." In International cross-domain conference for machine learning and knowledge extraction, pp. 179-191. Springer, Cham, 2018.

[7] Alayba, Abdulaziz M., Vasile Palade, Matthew England, and Rahat Iqbal. "Arabic language sentiment analysis on health services." In 2017 1st international workshop on arabic script analysis and recognition (asar), pp. 114-118. IEEE, 2017.

[8] Abdullah, Malak, and Mirsad Hadzikadic. "Sentiment analysis on arabic tweets: Challenges to dissecting the language." International Conference on So-

cial Computing and Social Media. Springer, Cham, 2017.

[9] Heikal, Maha, Marwan Torki, and Nagwa El-Makky. "Sentiment analysis of Arabic tweets using deep learning." Procedia Computer Science 142 (2018): 114-122..

[10] Rasool, Abdur, et al. "Twitter sentiment analysis: a case study for apparel brands." Journal of Physics: Conference Series. Vol. 1176. No. 2. IOP Publishing, 2019.

[11] Zhang, Yudong, Saeed Balochian, Praveen Agarwal, Vishal Bhatnagar, and Orwa Jaber Housheya. "Artificial intelligence and its applications." Mathematical problems in Engineering 2014 (2014).

[12] Ang, Jun Chin, Andri Mirzal, Habibollah Haron, and Haza Nuzly Abdull Hamed. "Supervised, unsupervised, and semi-supervised feature selection: a review on gene selection." IEEE/ACM transactions on computational biology and bioinformatics 13, no. 5 (2015): 971-989.

[13] A. Khan, Sohail, A., Zahoora, A. Qureshi," A survey of the recent architectures of deep convolutional neural networks" Artificial Intelligence Review, 53(8), 2020, pp. 5455–5516. doi:10.1007/s10462-020-09825-6.

[14] F. Kratzert, D. Klotz, C. Brenner, K. Schulz, M. Herrnegger,"Rainfall–runoff modelling using long short-term memory (LSTM) networks. Hydrology and Earth System Sciences", 22(11), pp. 6005-6022, 2019.

[15] Bhardwaj, Aditya, Yogendra Narayan, and Maitreyee Dutta. "Sentiment analysis for Indian stock market prediction using Sensex and nifty." Procedia computer science 70 (2015): 85-91.

[16] Berthon, Pierre R., et al. "Marketing meets Web 2.0, social media, and creative consumers: Implications for international marketing strategy." Business horizons 55.3 (2012): 261-271.

[17] Chen, Liang-Chu, Chia-Meng Lee, and Mu-Yen Chen. "Exploration of social media for sentiment analysis using deep learning." Soft Computing 24.11 (2020): 8187-8197.

[18] Revathy, G., Saleh A. Alghamdi, Sultan M. Alahmari, Saud R. Yonbawi, Anil Kumar, and Mohd Anul Haq. "Sentiment analysis using machine learning: Progress in the machine intelligence for data science." Sustainable Energy Technologies and Assessments 53 (2022): 102557.

[19] Altaher, Altyeb. "Hybrid approach for sentiment analysis of Arabic tweets based on deep learning model and features weighting." Int. J. Adv. Appl. Sci 4.8 (2017): 43-49.

[20] Khalil, Enas A. Hakim, Enas MF El Houby, and Hoda Korashy Mohamed. "Deep learning for emotion analysis in Arabic tweets." Journal of Big Data 8.1 (2021): 1-15.

[21] Manshu, Tu, and Wang Bing. "Adding prior knowledge in hierarchical attention neural network for cross domain sentiment classification." IEEE Access 7

(2019): 32578-32588.

[22] Napitu, F., Bijaksana, M. A., Trisetyarso, A., & Heryadi, Y. (2017, November). Twitter opinion mining predicts broadband internet's customer churn rate. In 2017 IEEE International Conference on Cybernetics and Computational Intelligence (CyberneticsCom) (pp. 141-146). IEEE.

[23] Xu, Feifei, and Vlado Keelj. "Collective sentiment mining of microblogs in 24-hour stock price movement prediction." 2014 IEEE 16th conference on business informatics. Vol. 2. IEEE, 2014.

[24] Zvarevashe, Kudakwashe, and Oludayo O. Olugbara. "A framework for sentiment analysis with opinion mining of hotel reviews." 2018 Conference on information communications technology and society (ICTAS). IEEE, pp1-4, 2018.

[25] Singla, Zeenia, Sukhchandan Randhawa, and Sushma Jain. "Statistical and sentiment analysis of consumer product reviews." 2017 8th International Conference on Computing, Communication and Networking Technologies (ICCCNT). IEEE, 2017. Pp1-6.

[26] Elzayady, Hossam, Khaled M. Badran, and Gouda I. Salama. "Arabic Opinion Mining Using Combined CNN-LSTM Models." International Journal of Intelligent Systems & Applications 12.4 (2020).

[27] Al-Bayati, Abdulhakeem Qusay, Ahmed S. Al-Araji, and Saman Hameed Ameen. "Arabic sentiment analysis (ASA) using deep learning approach." Journal of Engineering 26.6 (2020): 85-93.

[28] Yadav, Ashima, and Dinesh Kumar Vishwakarma. "Sentiment analysis using deep learning architectures: a review." Artificial Intelligence Review 53.6 (2020): 4335-4385.

[29] Aldayel, Haifa K., and Aqil M. Azmi. "Arabic tweets sentiment analysis–a hybrid scheme." Journal of Information Science 42.6 (2016): 782-797.

[30] Al-Smadi, Mohammad, Bashar Talafha, Mahmoud Al-Ayyoub, and Yaser Jararweh. "Using long short-term memory deep neural networks for aspect-based sentiment analysis of Arabic reviews." International Journal of Machine Learning and Cybernetics 10, no. 8 (2019): 2163-2175.

[31] Haq, Mohd Anul. "Smotednn: A novel model for air pollution forecasting and aqi classification." Computers, Materials and Continua 71 (2022): 1.

[32] [32] Adoma, Acheampong Francisca, Nunoo-Mensah Henry, and Wenyu Chen. "Comparative analyses of bert, roberta, distilbert, and xlnet for text-based emotion recognition." In 2020 17th International Computer Conference on Wavelet Active Media Technology and Information Processing (ICCWAMTIP), pp. 117-121. IEEE, 2020.