

Deep Learning for Sentiment Analysis of Tunisian Dialect

Abir Masmoudi, Jamila Hamdi, Lamia Hadrach Belguith

University of Sfax,
ANLP Research group, MIRACL Lab.,
Tunisia

{masmoudiabir, jamila.hamdi90, lamia.belguith}@gmail.com

Abstract. Automatic sentiment analysis has become one of the fastest growing research areas in the Natural Language Processing (NLP) field. Despite its importance, this is the first work towards sentiment analysis at both aspect and sentence levels for the Tunisian Dialect in the field of Tunisian supermarkets. Therefore, we experimentally evaluate, in this paper, three deep learning methods, namely convolution neural networks (CNN), long short-term memory (LSTM), and bi-directional long-short-term-memory (Bi-LSTM). Both LSTM and Bi-LSTM constitute two major types of Recurrent Neural Networks (RNN). Towards this end, we gathered a corpus containing comments posted on the official Facebook pages of Tunisian supermarkets. To conduct our experiments, this corpus was annotated on the basis of five criteria (very positive/positive/neutral/negative/very negative) and other twenty categories of aspects. In this evaluation, we show that the gathered features can lead to very encouraging performances through the use of CNN and Bi-LSTM neural networks.

Keywords. Sentiment Analysis; Tunisian Dialect; Social networks; Aspect-based Sentiment Analysis; Sentence-Based sentiment analysis; Big data; CNN; RNN.

1 Introduction

Sentiment analysis has received special attention in the fields of advertising, marketing and production. Indeed, the emergence of several social media platforms, such as Facebook, Instagram, LinkedIn and Twitter has encouraged individuals to express their opinions and feelings towards various subjects, products, ideas, people, etc. Therefore, automatic sentiment analysis has become one of the most dynamic research areas in the NLP field.

In this respect, our work is part of the automatic analysis of Internet users' comments that are posted on the official pages of supermarkets in Tunisia on Facebook social networks. To do this, We have gathered comments from the official Facebook pages of Tunisian supermarkets.

To conclude, the main objective of this research is to propose an automatic sentiment analysis of Internet users' comments that are posted on the official pages of Tunisian supermarkets and on Facebook social networks. We began by studying the issues related to the term 'opinion' and the existing solutions. At the end of this study, we proposed a method for the analysis of feelings.

The primary contributions of this paper are as follows:

- We gather comments from the official Facebook pages of Tunisian supermarkets, namely Aziza, Carrefour, Magasin Général, Géant and Monoprix. This corpus is made up of comments written in the Tunisian Dialect by taking into account two main scripts; Latin script and Arabic script.
- We annotate our corpus for sentiment analysis based on five sentiment classes (very positive/positive/neutral/negative/very negative) and twenty aspect categories.
- We present our proposed method with the different steps of the automatic sentiment analysis.
- We eventually disclose our elaborate experiments and the obtained outcomes.

The remaining of this paper is structured as follows. Section 2 surveys the literature on sentiment analysis. In Section 3, we deal with the Tunisian Dialect dataset that were collected and used in our experiments. Section 4 is dedicated to a detailed presentation of our proposed method for opinion analysis in the Tunisian dialect. In this context, we propose two different approaches: The first allows to classify the collected comments into five sentiment classes at the sentence level, while the second aims to introduce the aspect-based sentiment analysis of the Tunisian dialect by implementing two models, namely the aspect category model and the sentiment model. Section 5 provides a discussion that demonstrates the efficiency and accuracy of both RNN and CNN-based features. We finally draw some conclusions and future work directions in Section 6.

2 The Main Approaches of Sentiment Analysis

Sentiment analysis is one of the most vigorous research areas in NLP research field that focuses on analyzing people's opinions, sentiments, attitudes, and emotions towards several entities such as products, services, organizations, issues, events, and topics [34]. A large number of research works on sentiment analysis have been recently published in different languages. Hence, to achieve this, we first laid the foundation for research on the sentiment analysis by reviewing relevant literature on past studies conducted in this field. According to this study, sentiment analysis works fall into three major approaches, namely a machine learning based approach, a knowledge based approach and a hybrid approach.

The next subsection highlights the related works of Arabic and Arabic dialects sentiment analysis.

2.1 Machine Learning Based Approach

Machine learning helps data analysts build a model with a large amount of pre-labeled words or sentences in order to tackle the classification problem. [54] examined opinion analysis of the Tunisian dialect. Their corpus was made up of 17k comments.

Three main classifiers were adopted for the classification task, namely Support vector machines, Bernoulli naïve Bayes and Multilayer perceptron. Their obtained error rates were 0.23, 0.22, and 0.42 for support vector machine, multilayer perceptron and Bernoulli naïve Bayes, respectively.

Deep learning is a crucial part of machine learning that refers to the deep neural network suggested by G.E. Hinton [56]. It includes five main networks or architectures: CNN (convolutional neural networks), RNN (recursive neural networks), RNN (recurrent neural networks), DBN (deep belief networks), and DNN (neural networks deep).

2.2 Knowledge-Based Approach

The knowledge-based approach also known as lexicon-based approach consists in building lexicons of classified words. In this respect, [55] relied on a lexicon-based approach to be able to construct and assess a very large sentiment lexicon including about 120k Arabic terms.

They put forward an approach that enables them to utilize an available English sentiment lexicon. To evaluate their lexicon, the authors made use of a pretreated and labeled dataset of 300 tweets and reached an accuracy rate of 87%.

In their attempt to construct a new Arabic lexicon, [59] proposed a subjectivity and sentiment analysis system for Egyptian tweets. They built an Arabic lexicon by merging two modern standard Arabic lexicons (called MPQA and ArabSenti) with two Egyptian Arabic lexicons. The new lexicon is composed of 300 positive tweets, 300 negative tweets and 300 neutral tweets, for a total of 900 tweets achieving an accuracy of 87%.

2.3 Hybrid Approach

The Hybrid approach is a combination of the two approaches already mentioned above. [60] developed a semantic model called ATSA (Arabic Twitter Sentiment Analysis) based on supervised machine learning approaches, namely Naïve Bayes, support vector machine and semantic analysis. They also created a lexicon by relying on available resources, such as Arabic

WordNet. The model's performance has been improved compared to the basic bag-of-words representation with 4.48% for the support vector machine classifier and 5.78% for the classifier NB. In another study conducted by [61], a hybrid approach combining supervised learning and rules-based methods was applied for sentiment intensity prediction. In addition, the authors utilized not only well-defined linear regression models to generate scores for the tweets, but also a set of rules from the pre-existing sentiment lexicons to adjust the resulting scores using Kendall's score of about 53%.

3 Tunisian Dialect Corpus Description

The existence of a corpus is mandatory for a precise analysis of sentiments, because it is used to train and evaluate the models developed. In our case, we need a corpus in Tunisian dialect for the opinions analysis in the field of supermarkets. Due to the lack of available public datasets of this field, in this research project, we built our dataset.

During the data collection phase, two types of corpus are utilized. The comments in the first type are written based on a Latin script (also called Arabizi). The comments in the second type of corpus, however, are written by using an Arabic script. As mentioned above, this is due to the habit and ease of writing in Latin, especially that Tunisians often introduce French words into their writings and conversations. Based on these characteristics, we decided to divide the corpus into two different parts, namely Arabizi corpus and Arabic corpus. This section deals with a breakdown of the datasets used in our work and gives an overview of the corpus statistics.

3.1 Arabizi Corpus

Arabizi in general refers to Arabic written using the Roman script [8]. In particular, Tunisian Arabizi appeared especially in social media, such as Facebook, Instagram, SMS and chat applications. Arabizi corpus comprises numbers instead of some letters.

For example, the sentence
حتى كان بلاش منهزهمش خيت [even if they are

free, I will not take them] can be converted to Arabizi to become: "7ata blech manhezhomch 5iit", here the author used the numbers "7" and "5" which successively replaced the Arabic letters ح and خ.

Moreover, our Arabizi corpus is characterized by the phenomenon of code switching which is defined in [36] as "the mixing, by bilinguals (or multilinguals), of two or more languages in speech, often without changing the speaker or subject". This phenomenon is the passage from one language to another in the same conversation. For example, "solde waktech youfa" / [When will the promotion expire?]. This phenomenon appeared in our Arabizi corpus which includes foreign words of French origin, such as "promotion" [promotion], "bonjour" [hello], "caissière" [cashier], etc.

One of the major characteristics of Arabizi corpus is the presence of abbreviated of foreign words. Using abbreviated foreign words is one of the major characteristics of our Arabizi corpus. Taking the example of the word "qqles" instead of "quelques" [a few]. Apart from abbreviations, some spelling mistakes can be detected from the internet users' performances of some foreign words. These mistakes are due to the users' low French language proficiency. Take the example of the word "winou el cataloug" instead of "winou el catalogue" [where is the catalog?]. Here, we notice that Facebook users get used to writing words as they listen to them.

3.2 Arabic Corpus

Our Arabic script corpus is made up of words and expressions extracted from the Tunisian Dialect, like يزو من الكذب [stop lying]. We noticed a case in which a specific comment contains foreign words which are written using Arabic alphabets. For example, the french word "climatisseur" becomes كليماتيزور /klymAtyzwr/. Table 1 reports the most salient characteristics of our Tunisian Dialect corpus in terms of the number of sentences and words.

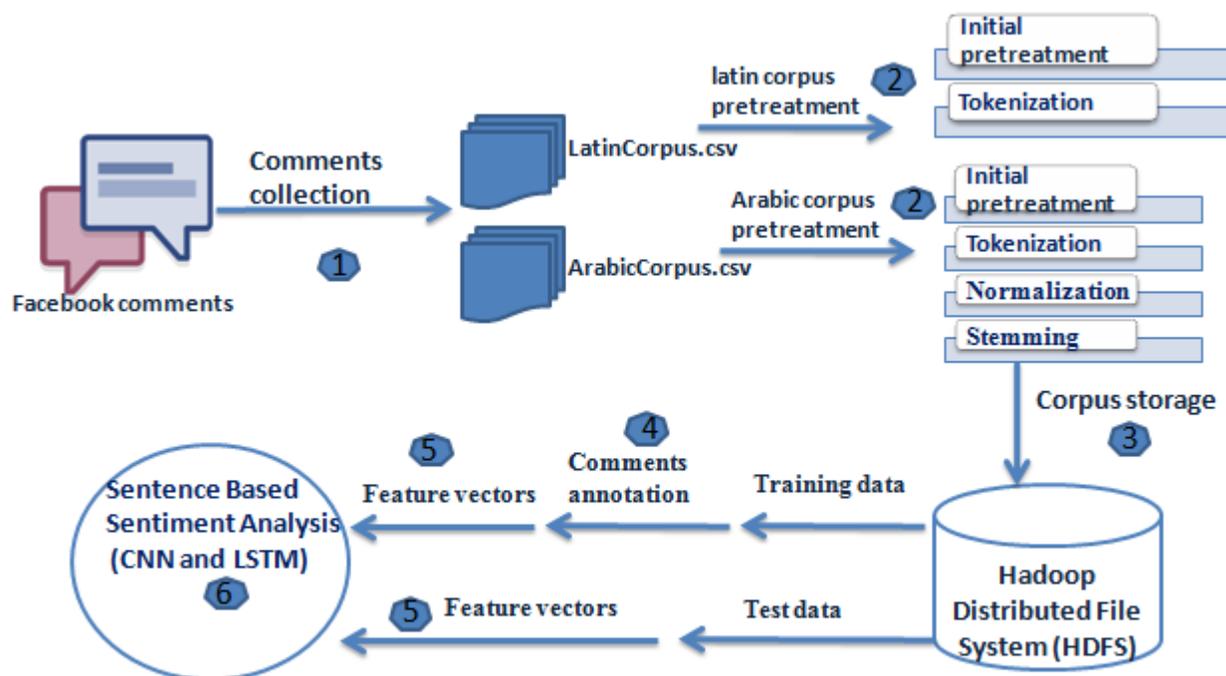


Fig. 1. Sentence-based sentiment analysis method

Table 1. Statistics of the corpus

Script	Number of sentences	Number of words
Arabic	17k	196k
Arabizi	27K	274k

4 Method Overview

Recall that in our work, we are interested in the sentiment analysis of Facebook comments published in the Tunisian Dialect in the field of Tunisian supermarket services. Our sentiment analysis was performed at two levels, sentence level and aspect level. The purpose of sentence based sentiment analysis is to determine the general opinion. For aspect level, the primary goal is to present and discover sentiments on entities and their distinct aspects. Thereafter, we will describe the main tasks of each step of our method.

4.1 Steps of the Proposed Method

As we have conducted two methods for the analysis of opinions, namely, the analysis of opinions at the aspect level and the analysis of opinions at the sentence level, we will follow six steps in our proposed method. We begin by presenting common steps for both levels of analysis. First of all, we collect comments from Facebook. Then we divide our corpus constructed into two parts: Latin corpus, which contains Latin writing and Arabic corpus, which contains Arabic writing. In the following, the corpus used will be stored in a HDFS (Hadoop Distributed File System), which aims to store large volumes of data on the disks of many computers.

4.1.1 Sentiment Analysis at Sentence Level

At this level of analysis, we will exploit the two corpus, corpus Arabic and Arabizi. Thereafter, we will apply pre-treatment steps of these corpus in order to eliminate the noises that include, initial

pre-treatment steps for each corpus, followed by a tokenization step. We will also go through a step of normalization and a step of racinisation for the Arabic corpus. Then we make the manual annotation of two corpus in five classes, very negative, negative, neutral, positive and very positive. Then we move on to a main stage which is the construction of the text representation model, which takes tokens as input, where each token presents a word.

In order to perform the classification task, we will choose the three deep learning algorithms, CNN, LSTM and Bi-LSTM. For example, following this level of analysis, the classification of the comment “hh 7asilou allah la traba7kom or latfar7kom” هه حصيلو الله لا تبارككم و لا تفرحكم highlights the “very negative” result that results in disappointment. Figure 1 shows the method we propose for sentiment analysis at the sentence level.

4.1.2 Sentiment Analysis at Aspect Level

At this level, our analysis of the data was carried out at the aspect level in the Arabic corpus.

4.2 Collection of Dataset

We selected the social network “Facebook” by concentrating on the supermarket field in Tunisia in order to gather suitable comments. We have identified five official pages of supermarkets (Carrefour, Magasin general, Monoprix, Aziza and Géant). This corpus construction is based on two important tools: “Facepager”¹ and “Export Comments”².

4.3 Processing of the Dataset

Once the corpus is collected, the available resources must go through a preprocessing stage in order to create a usable corpus.

¹<https://fr.freedownloadmanager.org/Windows-PC/Facepager-GRATUIT.html>

²<https://exportcomments.com/>

4.3.1 Repeated Comments

A comment sometimes appears several times in the corpus. Repeated comments don't add anything to the sentiment analysis. Therefore, we have kept only one occurrence of each comment.

4.3.2 Normalization

The Tunisian Dialect is characterized by the use of informal writing that does not respect precise orthographic rules. This fact makes their automatic processing difficult. Indeed, some words in the corpus have more than one form. Thus, a special pre-processing step was implemented in order to alleviate these problems and to obtain consistent data. This step consists in normalizing Arabic words by converting the multiple forms of a given word into a single orthographic representation. In our work, we used the CODA-TUN (Conventional Orthography for Tunisian Arabic) tool [51].

4.3.3 Light Stemming

Words written in the Tunisian Dialect are often made up of more than one word; hence the importance of the root word task. Light rooting aims at removing all prefixes and suffixes from the word and keeping its root. For example, the word مغازتنا mgAztnA / [our supermarket] is changed to مغازة [supermarket] by removing the suffix نًا.

4.4 Corpus Annotation

Since the primary goal of supervised learning is to determine the polarity of opinions in advance, we move forward to another crucial step, known as the annotation of the corpus after pretreating the two corpora. Manual annotation is therefore necessary to build a learning corpus for sentiment analysis. Indeed, our corpus was annotated by native Tunisian Dialect speakers who were asked to classify the comments according to already well-defined categories and sentiment classes.

4.4.1 Annotation for Sentiment Analysis at the Aspect Level

the annotation of the aspect-based sentiment analysis depends only on the Arabic script corpus, in which we first labeled our collected comments with five distinct classes. The first class is "very positive". It's used when the comment contains words that express total satisfaction with a service or a product. For example, تفتق [very nice]. The second class "positive" is used if the comment expresses a positive feeling, such as satisfaction, enthusiasm, etc. For example, بنية [delicious].

The third class "neutral" is performed if the comment is informative or expressing no sentiment. For example, the comment باش تخدم في العيد [do you work during Eid] is informative with no word of sentiment. Negative means if the comment expresses a negative feeling, such as dissatisfaction, regret or any other negative feelings. For example, ياسر حايين [they are not beautiful], etc. Finally, very negative means when the comment expresses a very negative feeling, such as annoyance, disappointment or any other very negative feelings. For example, the word ماعجبتيش [very bad].

Then, we annotated our Arabic corpus with 20 already well-defined categories. Each category comprises two major parts that form the tuple "E#A", namely the aspect entity (E) and the attribute of aspect entity (A). The list of aspect entities is composed of 9 aspects, namely, "drinks", "aliments", "services", "locations", "cleaners", "electronics", "utensils" and "others". However, the list of attributes of aspect entities is, namely, "general", "quality" and "price". These attribute are defined for each entity, except for the two entities; "locations" and "services". Thus, we fixed a single entity attribute "general" for these last two entities. Table 2 reports aspect categories with examples of topics discussed in each category.

Aspect entity attributes corresponding to entity labels are shown as follows:

- **Price:** this attribute includes promotions and payment facilities. For example, the words

رخصولنا [Please reduce the price] , بقدأش سوم [how much?] , بزموسيون [promotion], etc.

- **Quality:** This refers to an advice via a product or a service. For example: تعيف [it's disgusting], تفتق [that's wonderful] , etc.
- **General:** The user does not express his opinion clearly. It is difficult to understand whether his opinion is related to either the price of the product or the quality of the product.

4.4.2 Annotation for Sentiment Analysis at the Sentence Level

As regards the method of analyzing opinions at the sentence level, the two corpus have been labelled into five classes which are shown in table 3.

4.5 Feature Vectors

Our method applied two main feature vectors, namely word embedding vectors and the morpho-syntactic analysis. Although the preprocessing step was performed in the two corpora, the comments are not ready to be used by two neural network algorithms because they require a representation of the words that were considered as feature vectors. Our classifiers, CNN, LSTM and Bi-LSTM deep learning networks, take as input the vector representations of each word.

Indeed, it is necessary to go through the creation of a characteristic vector for each word in a comment. Consequently, we implemented the Gensim³ of the Word2Vec model. This model plays a key role as it uses a deep neural network, manipulates sentences in a given document, and generates output vectors. The Word2Vec model was developed by Google researchers led by [37]. The word2vec algorithms include two different models: skip-gram and the Continuous Bag-of-Words (CBOW) model. In the first model, the distributed representations of the input word are used to predict the context. In the second

³<https://radimrehurek.com/gensim/models/word2vec.html>

Table 2. Aspect categories with the topics discussed in each category

Aspects	Topics discussed
Drinks	Coffee / tea, juice, soda, water.
Aliments	Bakery, canned products, dairy products, dry food, frozen food, vegetables, individual meals, ice cream, meat, etc.
Cleaners	Laundry detergent, dishwashing liquid detergent, etc.
Electronics	Laptops, telephones, televisions, etc.
Services	Complaints and suggestions from the customer.
Locations	Concerning one of the supermarkets location
Utensils	Kitchen utensils
Others	Baby items, pet items, batteries, greeting cards, games, gifts, vouchers, catalogs, recipes, etc.

Table 3. Sentiment classes with examples of comments from the two corpus

	Examples of comments written in Arabic	Examples of comments written in latin
Positive	وَاللهِ باهين و صحاح يهزو قضيا باهية و ما يتقطعوش	7lou 3jebni ha9
Negative	بيعو في الحوت المرابي و يقولو فرشك هه	les caissiere hala yahkyou m3ana bkelet tourbya
very positive	وَأوروعة	a7ssen afar wahsen kadya w service fi mg
very negative	عزيزة الي تببع اغلى من عطار الحومة ما نشريش منها	7asilou allah latraba7kom ou latfar7kom
Neutral	هذي هي الماكينة الي تقضي على الناموس	famech rakadha on promotion svp

model, however, the representations of the context are combined to predict the word in the middle.

For the morpho-syntactic analysis, we employed the tool developed by [52], who proposed a method to remove the ambiguity in the output of the morphological analyzer from the Tunisian Dialect. They disambiguated results of the Al-Khalil-TUN Tunisian Dialect morphological analyzer [53] Their suggested method achieved an accuracy of 87%. For our work, entity names are identified by words which are labeled with "noun" and "prop-noun". about sentiment words, they are recognized by words labeled with the words "adj" and "verb".

4.6 Classification

For the implementation of these methods, we selected three algorithms, namely CNN, LSTM and Bi-LSTM by relaying the deep learning methods in order to implement the two tasks,

namely, sentence-based sentiment analysis and aspect-based sentiment analysis.

We applied the Word2vec algorithm as input for both neural networks. The general architecture of neural networks is illustrated in figure 2.

The neural network is defined by determining the number of input layers, the number of hidden layers and the number of output layers. As illustrated in Figure 2, the neural network architecture can be expressed in this way:

In the input layer, the inputs $(1 \dots n)$ indicate a sequence of words. Then, $i = (i_1 \dots i_n)$ present the vector representations of entries entered. The neural network consists of $h = (h_1 \dots h_n)$ hidden layers that aim to map the vectors i to hidden layers h . The last layer, the output layer, receiving the last hidden layer output to merge and produce the final result o .

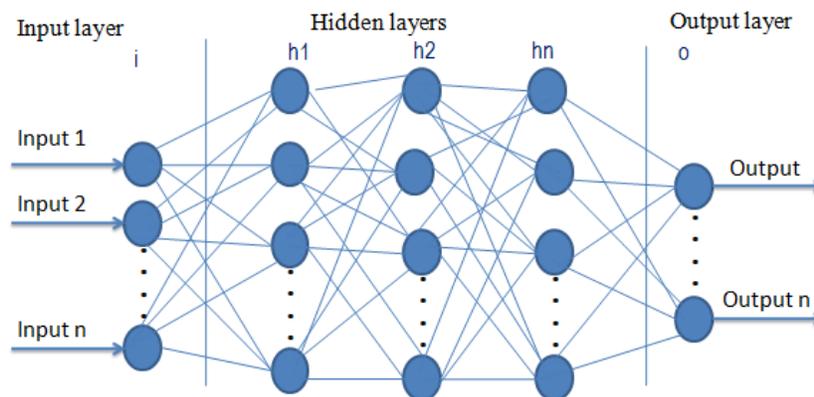


Fig. 2. Neural Network Architecture

4.6.1 Convolution Neural Network model (CNN)

CNN is also called Convnets or Cnns, it is one of the models of deep learning that marks impressive results in the field of TALN in general, and in the analysis of feelings.

Our CNN model consists of an input and an output layer, along with numerous hidden layers. Conventionally, the layers of a neural network are fully connected. This explains the fact that the output of each layer is the input of the next layer. The hidden layers consist of convolution layers, pooling layers, two fully connected layers, and activation functions. The pooling layer is applied to the output of the convolution layer. The fully connected layer aims to concatenate all the vectors into one single vector. The activation function introduces non-linearity into the neural network through a softmax classification layer. The output layer generates the polarity of each input comment.

To summarize, to train our CNN model, we adopted the Word2vec algorithm for the representation of words with a size of 300. This representation presents the input of the convolution layer. The convolution results are grouped or aggregated to a representative number via the pooling layer. This number is sent to a fully connected neural structure in which the classification decision is based on the weights assigned to each entity in the text.

Indeed, the main purpose of the fully connected layer is the reduction or compression of the

dimensional representation of input data generated by the convolution layer. The output from the pooling layer is passed to a fully connected softmax layer. As a multi-class classification was adopted in our work, we used a Softmax classification layer which predicts the probabilities for each class.

The figure 3 summarizes how the CNN model works for sentiment analysis at the two levels. For the aspect level, for the aspect category model, the input is a vector representation of each sentence with its aspect words, while, for the sentiment model, is a vector representation of each sentence with its sentiment words. Concerning the output, for the aspect category model, the output is a list of 20 categories classes, and, for the sentiment model, it's a list of 5 classes. While, for the sentence level, the input of the model is a vector representation of each sentence. The output is a list of 5 classes.

4.6.2 Long Short-Term Memory Model (LSTM)

The RNN Model is characterized by the sequential aspect of an entry in which the word order is of a significant importance. In addition, the RNN gives the possibility of processing variable length entries.

In this work, long short-term memory (LSTM) is a particular type of neural network that is used to learn sequence data. As a type of RNN, LSTM reads the input sequence from left to right. It is also capable of learning long-term relationships. That is why the prediction of an input depends on

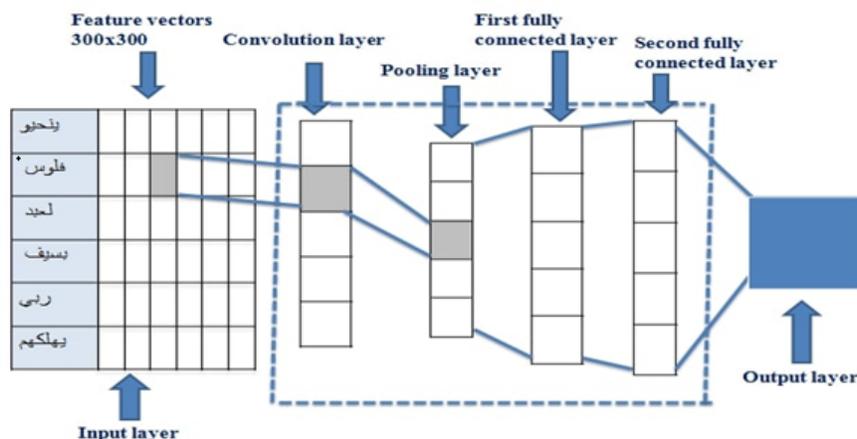


Fig. 3. CNN architecture for sentiment analysis at the two levels

the anteposed or postposed context. So, RNN is designed to capture long distance dependencies.

The entry into the LSTM model is a sequence of word representations using the Word2vec algorithm. Then, those representations are passed to an LSTM layer. The output of this layer is also passed to a softmax activation layer which produces predictions on the whole vocabulary words.

The figure 4 summarizes how the LSTM model works for sentiment analysis at the two levels. As mentioned above in the CNN model, For the aspect level, the input is a vector representation of each sentence with its aspect words and sentiment words. The output of this level is a list of 20 categories classes and other 5 sentiment classes. For the sentence level, the input of the model is a vector representation of each sentence. While, the output is a list of 5 classes.

4.6.3 Bidirectional Long Short-Term Memory Model (Bi-LSTM)

The LSTM network reads the input sequence from left to right, while Bi-LSTM which is variant of LSTM, it relies on the connection of two LSTMs layers of reverse directions (forward and backward) on the input sequence. The first layer is the input sequence and the second layers is the reversed copy of the input sequence The output layer combines the outputs of the forward layer and

backward layer by receiving both status information from the previous sequence (backward) and the next sequence (forward). Bi-LSTM is therefore very useful where the context of the entry is necessary, for example when the word negation appears before a positive term.

In our work, we used a one-layer bi-LSTM where the entry of the model is a sequence of words M represented using the Word2vec algorithm. Then, M is passed to a Bi-LSTM layer, the output of this layer is passed to a softmax activation layer which produces predictions on the whole vocabulary, for each step of the sequence. Each method of CNN and RNNs in general has its characteristics and differs from the other.

RNNs process entries sequentially, while CNNs process information in parallel. In addition, CNN models are limited to a fixed length entry, while RNN models have no such limitation. The major disadvantage of RNNs is that they train slowly compared to CNNs.

The architecture of the Bi-LSTM model for the two levels, sentence and aspect level is shown in Figure 5. As mentioned above in the CNN and LSTM models, the input of the aspect level is a vector representation of each sentence with its aspect words and sentiment words. The output of this level is a list of 20 categories classes and other 5 sentiment classes.

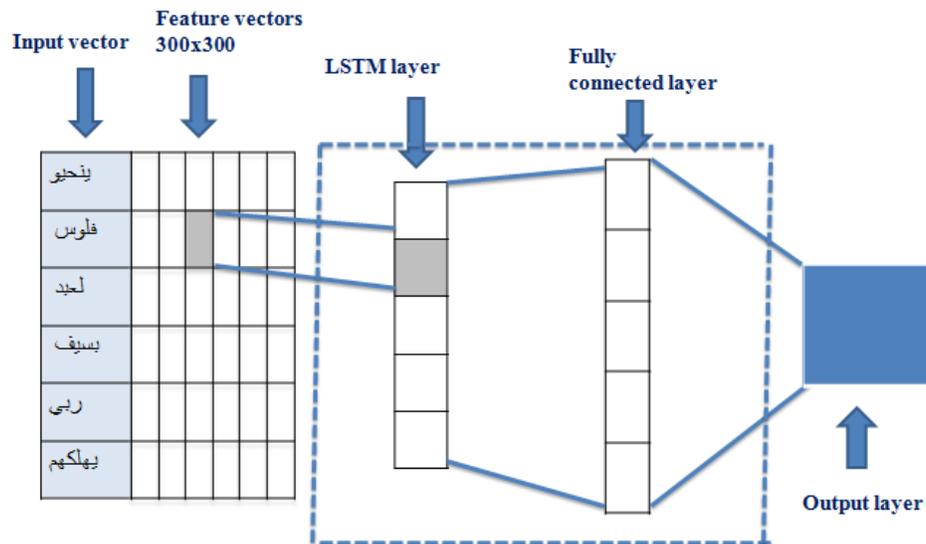


Fig. 4. LSTM architecture for sentiment analysis at the two levels

For the sentence level, the input of the model is a vector representation of each sentence. The output generating is a list of 5 classes.

5 Experiments

In this experiment, three deep Learning methods (CNN, LSTM and Bi-LSTM) were used.

In order to evaluate the two systems. Due to the informal nature of the Tunisian Dialect, we are required to convert all the different forms of a given word into a single orthographic representation. This explains our choice of the CODA-TUN tool (Conventional Orthography for Tunisian Arabic) that is introduced by [51].

One of the major features used in our work is word embedding technique that was described in section 6.6 as input features for our models. In our method, each word of a comment was replaced by a 1D vector representation of dimension d , with d is the length of the vector that equals to 300.

The sentiment analysis at aspect level requires not only the extraction of entity names for the formation of category aspect models, but also the use of sentiment words in order to construct the sentiment model. For this reason,

we made use of Al-Khalil-TUN Tunisian Dialect morphological analyzer developed by [52] to remove the output ambiguity.

5.1 Dataset

In order to present our corpus used for sentiment analysis at the two levels, we have classified our collected data into two parts. The first part is the training corpus which represents around 80% of the total size while the second part constitutes 20% of the corpus used for the test. In what follows, we will present the training corpus size and the test corpus size at each level of analysis in terms of the number of comments and the vocabulary size.

For the sentiment analysis at the sentence level, we used the two corpus. On the contrary, for the analysis at the level of aspect, we used the Arabic corpus only because of the absence of a morphosyntactic analyzer for Latin writing. For the sentiment analysis at sentence level, table 4 reports the corpus size followed by the vocabulary size for the training and test corpus. For sentiment analysis at aspect level, table 5 presents the corpus size followed by the vocabulary size for the category model and the sentiment model for the training and test corpus. The figure 6 presents the

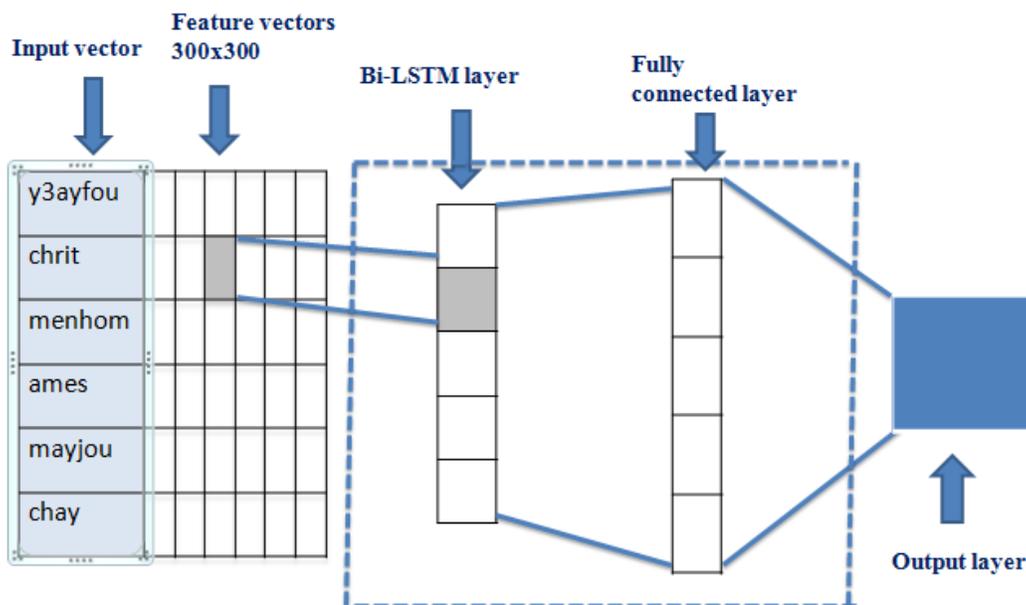


Fig. 5. Bi-LSTM architecture for sentiment analysis at the two levels

Table 4. The corpus size followed by the vocabulary size for sentence-based sentiment analysis

	# of comments	vocabulary size
Training corpus	35k	77k
Test corpus	9k	38k

20 aspect category classes of our Arabic corpus according to the estimated number of occurrences.

The axis refers to the aspect categories in our dataset. x shows the estimated number of occurrences. The number of comments classified as (others-general) is the largest part of the dataset.

The figures 7 and 8 presents some statistics of sentiment classes for both, Arabic corpus and Arabizi corpus.

5.2 Evaluation Metrics

In order to assess the performance of the two levels, at sentence level and aspect level. Concerning the aspect level, for the aspect category model and the sentiment model, we calculated the F-Measure 1 of each model by combining both

precision and recall measures. The F-Measure also called F-score is computed as the harmonic mean of the precision and recall with values ranging between 0 (representing the worst score) and 1 (representing the best score). Therefore, the F-Measure is calculated as follows:

$$F - Measure = \frac{2 \times (Precision \times Recall)}{(Precision + Recall)}. \quad (1)$$

5.3 Evaluation Results and Discussion

In this part we will present the outcomes of the evaluations carried out for the deep learning models (CNN, LSTM and Bi-LSTM). The performances of those models are represented by using F-Measure measurement. At aspect level, we present the F-Measure of each model

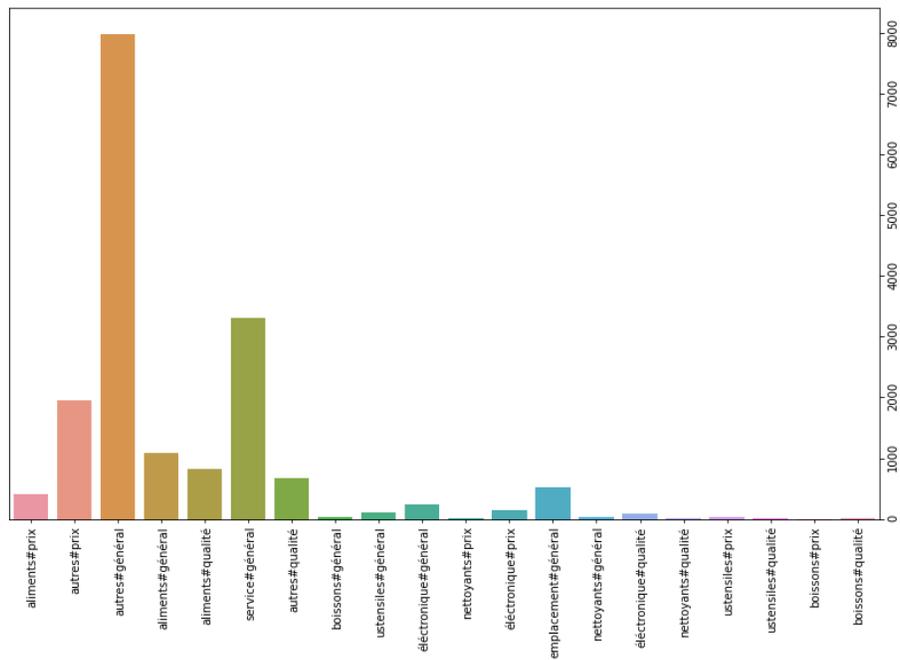


Fig. 6. Statistics of 20 aspect categories

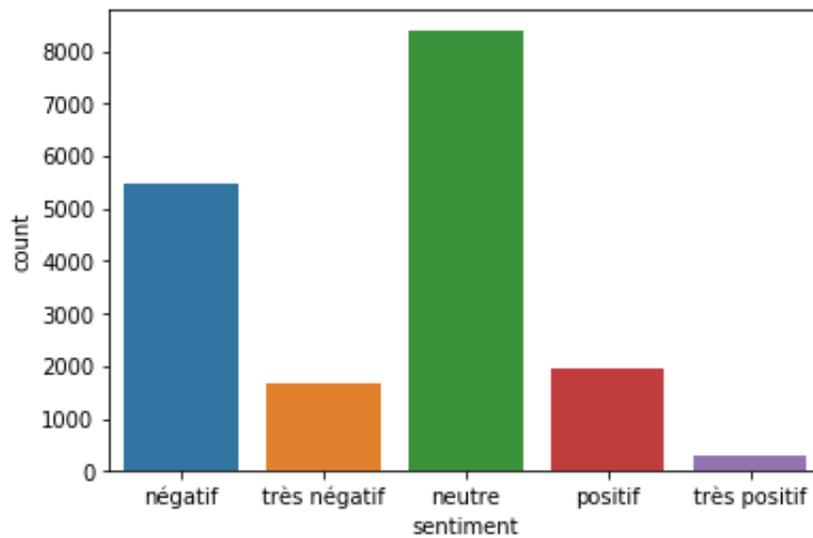


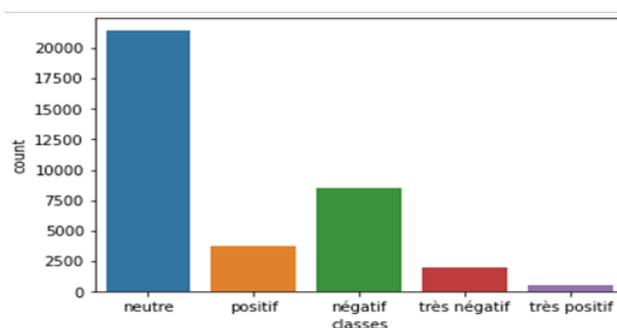
Fig. 7. Statistics of 5 sentiment classes of Arabic corpus

Table 5. The corpus size followed by the vocabulary size for aspect-based sentiment analysis

	# of comments	Vocabulary size for category model	Vocabulary size for sentiment model
Training corpus	14k	4k	4k
Test corpus	3k	2K	2k

Table 6. F-Measures of aspect category model and sentiment model

	Aspect category model			Sentiment model		
	CNN	LSTM	Bi-LSTM	CNN	LSTM	Bi-LSTM
Without stop words and stemming removals	46%	46%	48%	51%	47%	50%
With stop words and stemming removals	47%	46%	49%	51%	47%	50%

**Fig. 8.** Statistics of 5 sentiment classes for Arabizi corpus

depending on the stop words and stemming used in the pretreatment phase.

Table 6 shows F-Measure values of both the aspect category model and the sentiment model. According to table 6, we noticed that the use of stop word removal and stemming had almost no improvements for the three models. For this reason, we decided to carry out the other experiments by ignoring stemming and stop-word removal.

To achieve better results, we will test the performance of our models by classifying our experiences into 8 distinct groups. We will try to modify the number of aspect categories and sentiment classes. To do this, we changed the number of sentiment classes (5, 4, 3 and 2) without modifying the number of aspect categories presented in table 7. For a four-way classification, we ignored the neutral class. For a three-way

classification, we made some modifications. In this vein, the “very positive” class is replaced by the “positive” class, and the “very negative” class is changed to the “negative” class. Finally, for a binary classification, we went through the same steps of the three-way classification by eliminating the neutral class.

According to table 7, the best F-Measure reached 49% with 20 categories and 5 sentiment classes for the aspect category model using Bi-LSTM model. However, the highest F-measure was 78% for the sentiment model using Bi-LSTM model.

In the table 8, we will focus on the modified number of aspect categories. We will also evaluate our models based on each sentiment class (5, 4, 3 and 2). Indeed, we modified the aspect categories by removing the list of attributes of aspect entities (“general”, “quality and “price”).

Table 7. F-Measures of the aspect category model and sentiment model with 20 aspect categories and different sentiment classes

	Aspect category model			Sentiment model		
	CNN	LSTM	Bi-LSTM	CNN	LSTM	Bi-LSTM
with 20 categories and 5 sentiments	47%	46%	49%	51%	47%	50%
with 20 categories and 4 sentiments	42%	33%	40%	65%	58%	60%
With 20 categories and 3 sentiments	48%	46%	47%	58%	47%	55%
with 20 categories and 2 sentiments	42%	28%	41%	77%	75%	78%

Table 8. F-Measure of aspect category model and sentiment model with different sentiment classes and 8 aspect categories

	Aspect category model			Sentiment model		
	CNN	LSTM	Bi-LSTM	CNN	LSTM	Bi-LSTM
with 8 categories and 5 sentiment	61%	61%	62%	48%	46%	50%
with 8 categories and 4 sentiment	54%	49%	55%	60%	58%	60%
With 8 categories and 3 sentiment	62%	61%	62%	54%	47%	49%
with 8 categories and 2 sentiment	56%	49%	55%	78%	75%	77%

Table 9. F-Measure values for the three deep learning models

	CNN	LSTM	Bi-LSTM
With 5 classes	66%	68%	69%
With 4 classes	71%	72%	73%
With 3 classes	69%	87%	72%
With 2 classes	86%	87%	87%

Based on the outcomes displayed in table 8, the CNN-based classification and Bi-LSTM-based classification achieved best F-Measures values with 62% for the aspect category model. For the sentiment model, the best F-Measure value obtained was 78% using CNN model.

The result obtained by the two models was not good for several reasons. First, deep learning methods need a large dataset for best performance. However, this is not the case for our work as our corpus size is only 17k. Second, morpho-syntactic analyzer that produces many empty lines was not able to extract all sentiment words and aspects.

Moreover, the training corpus utilized to train the morpho-syntactic analyzer [52] was a spoken a corpus that consists of not only radio and TV broadcasts but also conversations recorded in

railway stations. Our constructed corpus, however, was gathered from the official supermarket pages. Hence, all these causes have a negative impact on the performance of our developed classifiers.

At sentence level, table 8 shows F-Measure values for the three deep learning models.

In order to achieve better results, we will test the performance of our models by classifying our experiences into 4 distinct groups by changed the number of sentiment classes (5, 4, 3 and 2). For a four-way classification, we ignored the neutral class. For a three-way classification, we made some modifications. In this vein, the “very positive” class is replaced by the “positive” class, and the “very negative” class is changed to the “negative” class. Finally, for a binary classification, we went through the same steps of the three-way classification by eliminating the neutral class.

According to table 9, the highest F-Measure reached was 87% with LSTM and Bi-LSTM with three-way and two-way classification using 5 sentiment classes.

5.4 Discussion

In this section, for the sentence level, we will compare the results obtained by our classifiers with the work [69] and of [70].

Table 10. Comparison of CNN, LSTM and Bi-LSTM results for sentiment analysis at sentence level

Corpus Analyzer	Training corpus	Test corpus	Language	Accuracy
Sentiment analysis at sentence level	35k	7k	Tunisian dialect	For the binary classification: LSTM=87% CNN=86% Bi-LSTM =87% For the three-way classification: LSTM=87% CNN=69% Bi-LSTM =72%
[70]	1k	359	MSA and dialect tweets	For the binary classification: LSTM=84%, CNN=77 %
[69]	8k	2k	MSA and dialects	For the three-way classification: CNN=64% LSTM=64%
[58]	13k	4k	Tunisian dialect	For the binary classification: LSTM =67% Bi-LSTM =70%

Table 10 presents, in detail, a comparison between our classifiers and those of others. We cannot compare the two levels of analysis, at the level of sentence and at the level of aspects, because they do not have the same size of corpus nor the same characteristics.

According to table 10, our classifiers compared to that of [70] surpass the performance of the latter with F-Measure value using CNN model. [70] obtained an accuracy of 85% using CNN, while our classifier recorded an F-Measure of 68% using CNN for a binary-classification.

On the other hand, by comparing our system with that of [69], our classifiers were able to obtain better F-Measure with 87% for LSTM and 68% for CNN with respect to 64% and 64% respectively for the classifiers from [69].

Regarding aspect-based sentiment analysis, the main challenge is the lack of tools available to deal with Arabic content and especially dialects. In this fact, there is not much work that has focused on this level of analysis for the Arabic language using deep learning.

Indeed, based on the outcomes displayed in table 10, the performance of [73] classifier has surpassed our classifier in terms of precision, which obtained an accuracy of 82%, while our classifier recorded an accuracy of 61% for the category appearance model and 75% for the sentiment model.

6 Conclusion and Future Work

This research focuses on the domain of sentiment analysis at two levels for the Tunisian Dialect, the sentence level and aspect level. We chose to work in the field of supermarkets. We selected the comments from the official Facebook pages of Tunisian supermarkets, namely Aziza, Carrefour, Magasin General, Geant and Monoprix. To the best of our knowledge.

At the aspect level, this is the first work that deals with the sentiment analysis problem in the Tunisian Dialect. In order to perform the sentiment analysis task, we utilized a machine learning approach, called deep learning. Indeed, we have developed three deep learning algorithms called, CNN, LSTM and Bi-LSTM. At sentence level, our system achieved a best performance with an F-Measure value of 87% using LSTM and Bi-LSTM. At aspect level, our system achieved a best performance with an F-Measure value of 62% for the aspect category model and 78% for sentiment model using Bi-LSTM and CNN.

Five universal problems for processing Tunisian Dialect affect the performance of the sentiment analysis task. First, the Tunisian Dialect, representing a mosaic of languages, is strongly influenced by other languages, like Turkish, Italian and French. Second, Tunisian people write and comment based on two type of scripts, namely Latin script and Arabic script. Third, due to the

absence of a standard orthography, people use a spontaneous orthography based on phonological criteria. Fourth, the Tunisian Dialect is usually written without diacritical signs. One of the major functions of these signs is to determine and facilitate the meaning of words, phrases or sentences.

Our future works will focus, first, on improve and develop our models to be more precise in detecting negation and to deal with both the sarcasm problem and spam detection. In addition, we will try to increase the dataset size by transliterating the Arabizi corpus into Arabic.

Finally, in order to ameliorate the outcomes of our models (CNN, LSTM and Bi-LSTM), we will try to test other deep learning techniques, like DNN (deep neural networks) and DBN (deep belief networks).

References

1. **Liu, B., Zhang, L. (2012).** A Survey of Opinion Mining and Sentiment Analysis. *Mining Text Data*, pp. 415–463. DOI: 10.1007/978-1-4614-3223-4_13.
2. **Liu, B. (2006).** *Web data mining: exploring hyperlinks, contents, and usage data.* Springer.
3. **Priyadarshana, Y.H.P.P., Gunathunga, K.I.H., Perera, K.K.A.N.N., Ranathunga, L., Karunaratne, P.M., Thanthriwatta, T.M. (2015).** Sentiment analysis: Measuring sentiment strength of call centre conversations. *Proceeding IEEE ICECCT*, pp. 1–9.
4. **Nitin, J., Liu, B. (2006).** Mining comparative sentences and relations. *Proceedings of National Conf. on Artificial Intelligence (AAAI-2006)*.
5. **Noforesti, S., Shamsfard, M. (2015).** Resource construction and evaluation for indirect opinion mining of drug reviews. *PLoS One*, Vol. 10, No. 5, e0124993.
6. **Cambria, E., Poria, S., Bisio, F., Bajpai, R., Chaturvedi, I. (2015).** The CLSA model: A novel framework for concept-level sentiment analysis. *LNCS*, Vol. 9042, pp. 3–22. Springer.
7. **Manna, S. (2018).** Sentiment analysis based on different machine learning algorithms. *International Journal of Computer Sciences and Engineering*, Vol. 6, No. 6, pp. 1116–1120.
8. **Darwish, K. (2013).** Arabizi detection and conversion to Arabic. *ANLP-EMNLP*.
9. **Moussa, M.E., Mohamed, E.H., Haggag, M.H. (2018).** A survey on opinion summarization techniques for social media. *Future Computing and Informatics Journal*.
10. **Hu Ya-Han, Chen Yen-Liang, Chou Hui-Ling (2017).** Opinion mining from online hotel reviews: A text summarization approach. *Information Processing Management*, Vol. 53, pp. 436–449. DOI:10.1016/j.ipm.2016.12.002.
11. **Qwaider, C., Saad, M., Chatzikiyriakidis, S., Dobnik, S. (2018).** Shami: A corpus of Levantine Arabic dialects. *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC-2018)*.
12. **Jiménez-Zafra, S.M., Martín-Valdivia, M.T., Martínez-Cámara, E., Ureña-López, L.A. (2016).** Combining resources to improve unsupervised sentiment analysis at aspect-level.
13. **Noforesti, S., Shamsfard, M. (2015).** Resource construction and evaluation for indirect opinion mining of drug reviews. *PLoS One*, Vol. 10, No. 5. DOI: 10.1371/journal.pone.0124993.
14. **Dua, M., Nanda, Ch., Nanda, G. (2018).** Sentiment analysis of movie reviews in Hindi language using machine learning. *Conference: International Conference on Communication and Signal Processing (ICCSP)*, India. DOI:10.1109/ICCSP.2018.8524223.
15. **Varathan, K.D., Giachanou, A., Crestani, F. (2015).** Comparative opinion mining: A review. *J. Assoc. for Inf. Sci. Technol*, Vol. 68, No. 4, pp. 811–829.
16. **Nabil, M., Aly, M., Atiya, A. (2015).** ASTD: Arabic sentiment tweets dataset. *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, pp. 2515–2519.
17. **Van de Kauter, M., Breesch, D., Hoste, V. (2015).** Fine-grained analysis of explicit and implicit sentiment in financial news articles. *Expert Systems with Applications*, Vol. 42, No. 11. DOI:10.1016/j.eswa.2015.02.007.
18. **Tang, D., Qin, B., Liu, T. (2015).** Document modeling with gated recurrent neural network for sentiment classification. *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, pp. 1422–1432.

19. **Mankar, S.A., Ingle, M. (2015).** Implicit sentiment identification using aspect based opinion mining. DOI:10.17762/ijritcc2321-8169.150491.
20. **Wiebe, J., Wilson, T., Bruce, R.F., Bell, M., Martin, M. (2004).** Learning Subjective Language. *Computational Linguistics*, Vol. 30, pp. 277–308.
21. **Wiebe, J., Riloff, E. (2005).** Creating subjective and objective sentence classifiers from unannotated texts. *Computational Linguistics and Intelligent Text Processing*, 6th International Conference, CILing, pp. 486. DOI: 10.1007/978-3-540-30586-6_53.
22. **Shubham, D., Mithil, P., Shobharani, M., Subramanain, S. (2017).** Aspect level sentiment analysis using machine learning. *IOP Conference Series Materials Science and Engineering*, Vol. 263, No. 4. DOI: 10.1088/1757-899X/263/4/042009.
23. **Oueslati, O., Cambria, E., HajHmida, M.B., Ounelli, H. (2020).** A review of sentiment analysis research in Arabic language. *Future Generation Computer Systems*, Vol. 112, pp. 408–430. DOI: 10.1016/j.future.2020.05.034.
24. **Kim, S.M., Pantel, P., Chklovski, T., Pennacchiotti, M. (2006).** Automatically assessing review helpfulness. *EMNLP'06: Proceedings of the Conference on Empirical Methods in Natural Language Processing*, pp. 423–430. Association for Computational Linguistics.
25. **Bisio, F., Gastaldo, P., Peretti, Ch., Zunino, R., Cambria, E. (2013).** Data intensive review mining for sentiment classification across heterogeneous domains. *Advances in Social Networks Analysis and Mining (ASONAM)*, pp. 1061–1067. DOI: 10.1145/2492517.2500280.
26. **Ofek, N., Poria, S., Rokach, L., Cambria, E., Hussain, A., Shabtai, A. (2016).** Unsupervised commonsense knowledge enrichment for domain-specific sentiment analysis. *Cognitive Computation*, Vol. 8, No. 3, pp. 467–477.
27. **Hammad, A.S.A., El-Halees, A. (2013).** An approach for detecting spam in Arabic opinion reviews. *The International Arab Journal of Information Technology*, Vol. 12.
28. **Ghose, A., Panagiotis I. (2007).** Designing novel review ranking systems: predicting the usefulness and impact of reviews. *Proceedings of the Ninth International Conference on Electronic Commerce (ICEC '07)*, Association for Computing Machinery, pp. 303–310. DOI:10.1145/1282100.1282158
29. **Fsih, E., Boujelbane, R., Belguith, L. (2018).** Tunisian dialect resources for opinion analysis on social media. *JCCO Joint International Conference on ICT in Education and Training, International Conference on Computing in Arabic, and International Conference on Geocomputing (JCCO: TICET-ICCA-GECO)*. DOI:10.1109/ICCA-TICET.2018.8726218.
30. **Chaturvedi, I., Cambria, E., Welsch, R., Herrera, F. (2018).** Distinguishing between facts and opinions for sentiment analysis: Survey and challenges. *Information Fusion*, Vol. 44, pp. 65–77. DOI:10.1016/j.inffus.2017.12.006.
31. **Alharabi, A., Taileb, M., Kalkatawi, M. (2019).** Deep learning in Arabic sentiment analysis: An overview, *Journal of information science. Journal of Information Science*. DOI:10.1177/0165551519865488.
32. **Refaee, E., Rieser, V. (2014).** An Arabic twitter corpus for subjectivity and sentiment analysis. *LREC*, pp. 2268–2273.
33. **Oueslati, O., Ben Hajhmida, M., Ounelli, H., Cambria, E. (2019).** Sentiment analysis of influential messages for political election forecasting. *International Conference on Computational Linguistics and Intelligent Text Processing*.
34. **Oueslati O., Ahmed, I.S.K., Habib O. (2018).** Sentiment Analysis for Helpful Reviews Prediction. *International Journal of Advanced Trends in Computer Science and Engineering*, Vol. 7, No. 3, pp. 34–40. DOI:10.30534/ijatcse/2018/02732018.
35. **Liu, B., Zhang, L. (2012).** A Survey of Opinion Mining and Sentiment Analysis. *Mining Text Data*, pp. 415–463. DOI:10.1007/978-1-4614-3223-4-13.
36. **Poplack, S. (2001).** Code-switching (linguistic).
37. **Mikolov, T., Chen, K., Corrado, G., Dean, J. (2013).** Efficient estimation of word representations in vector space. *arXiv:1301.3781*.
38. **Masmoudi, A., Habash, N., Khmekhem, M., Estève, Y., Belguith, L. (2015).** Arabic transliteration of romanized Tunisian dialect text: A preliminary investigation. *Computational Linguistics and Intelligent Text Processing*, 16th International Conference, CILing, pp. 608–619. DOI:10.1007/978-3-319-18111-0_46.
39. **Abir Masmoudi (2016).** Approche hybride pour la reconnaissance automatique de la parole pour la langue arabe. Thèse de doctorat en informatique, Université de sfax.

40. **Masmoudi, A., Mdhaftar, S., Sellami, R., Belguith, L. (2019).** Automatic diacritics restoration for Tunisian dialect. *ACM Transactions on Asian and Low-Resource Language Information Processing*, Vol. 18, No. 3, pp. 1–18. DOI:10.1145/3297278.
41. **Masmoudi, A., Khmekhem, M.E., Khrouf, M., Belguith, L.H. (2020).** Transliteration of Arabizi into Arabic script for Tunisian dialect. *ACM Transactions on Asian and Low-Resource Language Information Processing*, Vol. 19, No. 32. DOI:10.1145/3364319.
42. **Xing, F., Pallucchini, F., Cambria, E. (2019).** Cognitive-inspired domain adaptation of sentiment lexicons. *Information Processing and Management*, Vol. 56, No. 3, pp. 554–564. DOI:10.1016/j.ipm.2018.11.002.
43. **Younes, J., Achour, H., Souissi, E., Ferchichi, A. (2018).** Survey on corpora availability for the Tunisian dialect automatic processing. *JCCO Joint International Conference on ICT in Education and Training, International Conference on Computing in Arabic, and International Conference on Geocomputing (JCCO: TICET-ICCA-GECO)*. DOI:10.1109/icca-ticet.2018.8726213.
44. **Pang, L. et Vaithyanathan (2002).** Thumbs up? Sentiment Classification using Machine Learning Techniques. *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pp. 79–86. DOI:10.3115/1118693.1118704.
45. **Boukadida, N. (2008).** Connaissances phonologiques et morphologiques dvationnelles et apprentissage de la lecture en arabe (Etude longitudinale). Thèse de doctorat, Université Rennes.
46. **Peng, H., Ma, Y., Li, Y., Cambria, E. (2018).** Learning multi-grained aspect target sequence for chinese sentiment analysis. *Knowledge-Based Systems*, Vol. 148, pp. 167–176.
47. **Wahsheh, H., Al-Kabi, M., Alsmadi, I. (2013).** Spar: A system to detect spam in Arabic opinions. *IEEE Jordan Conference on Applied Electrical Engineering and Computing Technologies (AEECT)*, pp. 1–6. DOI:10.1109/AEECT.2013.6716442.
48. **Gautami T., Naganna, S. (2015).** Feature selection and classification approach for sentiment analysis. *Machine Learning and Applications: An International Journal (MLAIJ)*. DOI: 10.5121/mlaij.2015.2201.
49. **Preeti, Dahiya, S. (2015).** Sentiment analysis using SVM and naive Bayes algorithm. *Journal of Computer Science and Information Technology, IJCSMC*, Vol. 4, No. 9, pp. 212–219.
50. **Xing, F., Pallucchini, F., Cambria, E. (2019).** Cognitive-inspired domain adaptation of sentiment lexicons. *Information Processing and Management*, Vol. 56, No. 3, pp. 554–564. DOI:10.1016/j.ipm.2018.11.002.
51. **Zribi, I., Boujelbane, R., Masmoudi, A., El-louze, M., Belguith, L., Habash, N. (2014).** A Conventional Orthography for Tunisian Arabic. *Proceedings of the Ninth International Conference on Language Resources and Evaluation: LREC'14*, pp. 2355–2361.
52. **Zribi, I., Ellouze, M., Belguith, L., Blache, Ph., (2017).** Morphological disambiguation of Tunisian dialect. *Journal of King Saud University-Computer and Information Sciences*, Vol. 29, No. 2, pp. 147–155. DOI:10.1016/j.jksuci.2017.01.004.
53. **Zribi, I., Ellouze, M., Belguith, L. (2013).** Morphological Analysis of Tunisian Dialect. *International Joint Conference on Natural Language Processing*, pp. 992–996.
54. **Medhaftar, S., Bougares, F., Estève, Y., Hadrich-Belguith, L. (2017).** Sentiment analysis of Tunisian dialects: Linguistic resources and experiments. *Proceedings of the Third Arabic Natural Language Processing Workshop*, pp. 55–61. DOI:10.18653/v1/W17-1307.
55. **Al-Ayyoub, M., Bani-Essa, S., Alsmadi, I. (2015).** Lexicon-based sentiment analysis of Arabic tweets. *Int. J. Social Network Mining*, Vol. 2, No. 2, pp. 101–114. DOI:10.1504/IJSNM.2015.072280.
56. **Gupta, A., Pruthi, J., Sahu, N. (2017).** Sentiment analysis of tweets using machine learning approach. *International Journal of Computer Science and Mobile Computing, IJCSMC*, Vol. 6, No. 4, pp. 444–458.
57. **Kwaik, K.A., Saad, M., Chatzikyriakidis, S., Dobnik, S. (2019).** LSTM-CNN Deep learning model for sentiment analysis of dialectal Arabic. *Arabic Language Processing: From Theory to Practice*, pp. 108–121.
58. **Mohamed Amine Jerbi, Hadhemi Achour, Emna Souissi (2019).** Sentiment analysis of code-switched Tunisian dialect: Exploring RNN-based techniques. *Arabic Language Processing: From Theory to Practice*, pp. 122–131. DOI:10.1007/978-3-030-32959-4.9.

59. **El-Makky, N., Nagi, K., El-Ebshihy, A., Apady, E., Hafez, O., Mostafa, S., Ibrahim, S. (2015).** Sentiment Analysis of Colloquial Arabic Tweets. The 3rd ASE International Conference on Social Informatics (SocialInformatics).
60. **Alowaidi, S., Saleh, M., Abulnaja, O.A. (2017).** Semantic sentiment analysis of Arabic texts. *International Journal of Advanced Computer Science and Applications*, Vol. 8, No. 2. DOI:10.14569/IJACSA.2017.080234.
61. **Refaee, E., Rieser, V. (2016).** iLab-Edinburgh at SemEval-2016 Task 7: A Hybrid Approach for Determining Sentiment Intensity of Arabic Twitter Phrases. *Proceedings of the 10th International Workshop on Semantic Evaluation (SemEval)*, pp. 474–480. DOI:10.18653/v1/S16-1077.
62. **Younes, J., Achour, H., Souissi, E. (2015).** Constructing linguistic resources for the Tunisian dialect using textual user-generated contents on the social web. *International Conference on Web Engineering*, pp. 3–14.
63. **McNeil, K. (2011).** Tunisian Arabic corpus: Creating a written corpus of an unwritten language. *International Symposium on Tunisian and Libyan Arabic Dialects*, University of Vienna.
64. **Habash, N.Y. (2010).** Introduction to Arabic natural language processing. *Synthesis Lectures on Human Language Technologies*, Vol. 3, No. 1, pp. 1–187. DOI:10.2200/S00277ED1V01Y201008HLT010.
65. **Hamdi, A., Gala, N., Nasr, A. (2014).** Automatically building a Tunisian lexicon for deverbal nouns. *Proceedings of the First Workshop on Applying NLP Tools to Similar Languages, Varieties and Dialects*, pp. 95–102. DOI:10.3115/v1/W14-5311.
66. **Graja, M., Jaoua, M., Belguith, L.H. (2010).** Lexical study of a spoken dialogue corpus in Tunisian dialect. *Proceedings of the International Arab Conference on Information Technology*, Benghazi-Libya.
67. **Abir Masmoudi (2016).** Approche hybride pour la reconnaissance automatique de la parole pour la langue arabe. Thèse de doctorat en informatique, Université de sfax.
68. **Al-Azani, S., El-Alfy, E. (2018).** Emojis-based sentiment classification of Arabic microblogs using deep recurrent neural networks. *International Conference on Computing Sciences and Engineering (ICCSE)*, pp. 1–6, DOI:10.1109/ICCSE1.2018.8374211.
69. **Heikal Maha, Torki Marwan, El-Makky Nagwa (2018).** Sentiment analysis of Arabic tweets using deep learning. *Procedia Computer Science*, Vol. 142, pp. 114–122. DOI:10.1016/j.procs.2018.10.466.
70. **Al-Azani, S., El-Alfy, E.S.M. (2017).** Hybrid deep learning for sentiment polarity determination of Arabic microblogs. **Liu, D., Xie, S., Li, Y., Zhao, D., El-Alfy, E.S. (eds)** *Neural Information Processing ICONIP. Lecture Notes in Computer Science*, Vol. 10635, pp. 491–500.
71. **Hossam, S.I., Mervat, G., Sherif, M.A. (2015).** Sentiment Analysis for Modern Standard Arabic and Colloquial. *International Journal on Natural Language Computing IJNLC*, Vol. 4, No. 2.
72. **Barhoumi, A., Estève, Y., Aloulou, Ch., Belguith, L.H. (2017).** Document Embeddings for Arabic Sentiment Analysis. *LPKM*.
73. **Al-Smadi, M., Al-Ayyoub, M., Jararweh, Y., Qawasmeh, O. (2019).** Enhancing aspect-based sentiment analysis of Arabic hotels' reviews using morphological, syntactic and semantic features. *Information Processing and Management*, Vol. 56, No. 2, pp. 308–319. DOI: 10.1016/j.ipm.2018.01.006.
74. **Wagh, B., Shinde, Kale, P.A. (2018).** A twitter sentiment analysis using NLTK and machine learning techniques. *International Journal of Emerging Research in Management and Technology*, Vol. 6, No. 37. DOI: 10.23956/ijermt.v6i12.32.
75. **Al-Shalabi, R., Kanan, G.G., Gharaibeh, M.H. (2006).** Arabic text categorization using kNN algorithm.
76. **Masmoudi, A., Khmekhem, M.E., Estève, Y., Belguith, L.H., Habash, N. (2014).** A corpus and phonetic dictionary for Tunisian Arabic speech recognition. *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14)*, pp. 306–310.
77. **Masmoudi, A., Estève, Y., Khmekhem, M.E., Bougares, F., Belguith, L.H. (2014).** Phonetic tool for the Tunisian Arabic. *SLTU*, pp. 253–256.
78. **Zribi, I., Boujelbane, R., Masmoudi, A., Ellouze, M., Belguith, L.H., Habash, N. (2014).** A conventional orthography for Tunisian Arabic. *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14)*, pp. 2355–2361.

- 79. Elsahar, H., El-Beltagy, S. (2015).** Building large Arabic multi-domain resources for sentiment analysis. *Lecture Notes in Computer Science*, Vol. 9042, pp. 23–34. DOI:10.1007/978-3-319-18117-2_2.
- 80. Duwairi, R.M. (2015).** Sentiment analysis for dialectical Arabic. *Proceedings 6th ICICS International Conference on Information and Communication Systems*, pp. 166–170. DOI:10.1109/IACS.2015.7103221.
- 81. Al-Obaidi, A.Y., Samawi, V. (2016).** Opinion mining: Analysis of comments written in Arabic colloquial. *Proceeding of the Conference of World Congress on Engineering and Computer Science WCECS*, Vol. 1, pp. 470–475.
- 82. Bayoudhi, A., Ghorbel, H., Koubaa, H., Hadrich-Belguith, L. (2015).** Sentiment classification at discourse segment level: Experiments on multi-domain Arabic corpus. *Journal for Language Technology and Computational Linguistics JLCL*, Vol. 30 No. 1, pp. 1–25.
- 83. Huang, H.H., Wang, J.J., Chen, H.H. (2017).** Implicit opinion analysis: Extraction and polarity labelling. *Journal of the Association for Information Science and Technology*, Vol. 68. DOI:10.1002/asi.23835.
- 84. Al-Twairesh, N., Al-Khalifa, H., Al-Salman, A. (2014).** Subjectivity and sentiment analysis of Arabic: Trends and challenges. *Computer Systems and Applications AICCSA, IEEE/ACS 11th International Conference on*, pp. 148–155. DOI:10.1109/AICCSA.2014.7073192.

*Article received on 07/09/2020; accepted on 29/09/2020.
Corresponding author is Abir Masmoudi.*