ASTES

Advances in Science, Technology and Engineering Systems Journal Vol. 7, No. 3, 12-18 (2022)

www.astesj.com

ASTESJ

ISSN: 2415-6698

Analysis Methods and Classification Algorithms with a Novel Sentiment Classification for Arabic Text using the Lexicon-Based Approach

Bougar Marieme*, Ziyati El Houssaine

Laboratory c3s, higher School of technology, UHC, CASABLANCA, 20100, MAROC

ARTICLE INFO

Article history:
Received: 24 February, 2022
Accepted: 04 May, 2022
Online: 16 May, 2022

Keywords:
Sentiment analysis
Lexicon based Approach
Classification
Social networks

ABSTRACT

Social networks have become a valuable platform for tracking and analyzing Internet users' feelings. This analysis provides crucial information for decision-making in various areas, such as politics and marketing. In addition to this challenge and our interest in the field of big data and sentiment analysis in social networks, we have dedicated this work to combine different aspects of methods or techniques leading to the facilitation of feelings classification in social networks, including text analysis and sentiment analysis. We expose the approaches and the algorithms of supervised machine learning for the classification of feelings. We further our research to concisely present the methods of data representation and the parameters used to evaluate a sentiment analysis method in the context of social networks, with a section presenting our novel lexicon-based approach to give more accurate results in classifying Arabic text. The proposed approach has shown a promising accuracy percentage, especially the precision of the sentiment detected from text with F-Score up to 66%.

1. Introduction

Currently, the Internet allows billions of users to connect to each other, share information, communicate their ideas and opinions, and express their attitudes toward content through social networks. Thus, all actions generate high-volume, varied, high-velocity data called big social data [1].

Typically, this data is a set of opinions that can be processed to evaluate trends, audience preferences, and satisfaction related to a product, service, event, or even people.

Several areas are of interest in social networks, such as politics, health, and marketing. Because the particularity of data is unstructured texts, the data's exploitation makes text analysis an important factor for knowledge extraction and data mining.

In several of our published works, we have been particularly interested in the pre-processing of this data in the Arabic language [2], hence our interest in this research to move to the classification phase. Therefore, in this paper, we treat the analysis of textual data including the polarity of opinions (positive, negative, or neutral) and machine learning.

*Corresponding Author: Bougar Marieme, marieme.bougar7@gmail.com

2. Text mining

Textual analysis entails computer processing that extracts filtered and useful information from unstructured textual data. This analysis describes the content's structure and functions to extract patterns. This area of analysis includes techniques and algorithms, such as data mining and natural language processing. Text analysis occurs in two main actions: analyze and interpret.

2.1. The analysis phases

The analysis phase in Figure 1 consists of recognizing words, sentences, their relationships, and grammatical roles. This phase produces a standardization for the text through several methods or to automatically determine the language of a given content. We are interested in this field, and we have conducted several works in this direction especially for the Arabic language, which is recognized for its complex morphology.

The popular processes used in this phase are as follows:

https://dx.doi.org/10.25046/aj070302

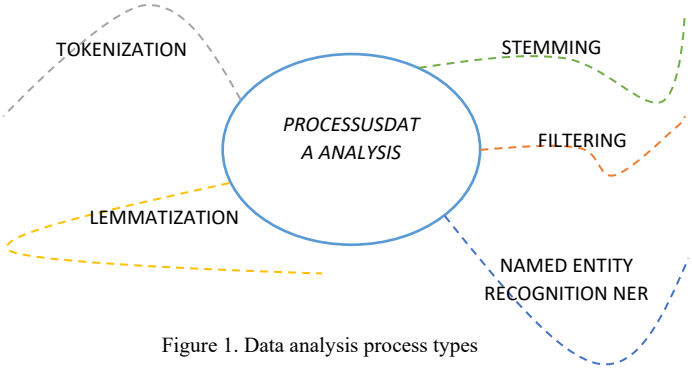

Figure 1. Data analysis process types

• Tokenization is the process of converting character strings into a list of symbols (tokens in English). Tokens are strings with meaning that eliminate spaces, punctuation, etc.
• The lemmatization operation assembles derivational variations (e.g., verbs, adjectives) or inflectional variations (e.g., plural, conjugations)
• Stemming is a process of transformation of bends or derivatives into their roots. The root of a word corresponds to the part of the word remaining once the prefixes and suffixes have been removed. There are specialized procedures for each language, for example, KHOJA Stemmer, which we treated for the stemming of the Arabic language. Stemming is especially advantageous because it is fast, it is based on precise and easy dictionaries and rules of derivation [3], and it allows for the treatment of the peculiarities of certain words.
• Filtering consists of applying filters that remove empty words
• Named entity recognition (NER) is a sub-task that extracts information from textual documents. This sub-task consists of labeling text with tags and searching for text objects that can be categorized into classes. Recognition is based on statistical systems, labeling a corpus that will serve as a learning tool, but these systems are expensive in human time.

2.2. The interpretation phases

Although data analysis is important to begin the interpretation phase to draw results and conclusions, this phase is based on data mining methods that establish reliable prediction models. This phase is a selection based on a lexical property, the presence or absence of a keyword, or other criteria. If it is a new element, we seek relationships that were not explicit between two distant elements in the text. For similarities, we try to discover texts that correspond most to a set of descriptors, such as the text's most frequent nouns and verbs.

As mentioned, the textual analysis process refers to different approaches, allowing data to be filtered, cleaned, and modeled afterward.

2.3. Sentiment analysis

Sentiment analysis appeared in the early 2000s and consists of classifying polarity into two opposing feelings, such as wanting/hating, positive/negative, or black/white.

Opinion analysis is used for different purposes and on a variety of corpora. This method has become essential in several areas, especially in marketing to evaluate a brand, analyze consumer opinions, or retain customers by detecting their emotional state and in politics to predict the results of presidential elections and the will of the public.

Classification begins with the retrieval of textual features in Figure 2:

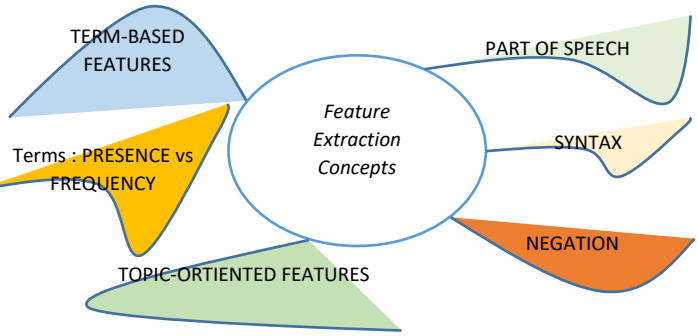

Figure 2. Feature extraction recapitulation

• Presence versus frequency of the term: According to a study by [4], the representation of a text by a vector, in which the elements indicate the existence of a term (1) or not (0) provides a better result than the frequency method (frequency of occurrence of the term) for the classification of polarity.
• Term-based features: In text, the position of a word (for example, in the middle or near the end of a document) can affect overall sentiment or subjectivity. Thus, position information is sometimes encoded in the characteristic vectors used [5].
• Part of speech analysis: This feature explains how a word is used in a sentence. There are several main parts of speech (also called word classes), including nouns, pronouns, adjectives, verbs, adverbs, prepositions, conjunctions, and interjections. Sentiment analysis by machine learning as well as by lexicon has an attraction toward adjectives [6].
• Syntax: This feature refers to the integration of syntactic relationships in the analysis and seems particularly relevant with short texts.
• Negation: In the bag-of-words method, for example, negation is not considered. The phrases "I hate" and "I like" are considered similar, hence the interest in the treatment of negation.
• Topic-oriented features: These features are the interaction between topic and opinion.

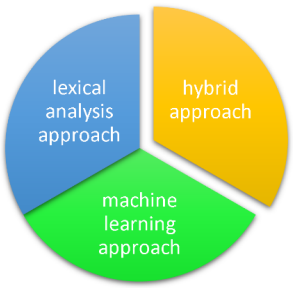

Figure 3. Sentiment analysis categories

There are three main categories of sentiment analysis as shown in Figure 3: a lexical analysis approach, a machine learning approach, and a hybrid approach that combines the first two.

3. Related work

3.1. Lexicon based approaches

Opinion extraction approaches based on lexical analysis consist of extracting the polarity of a sentence using a semantic analysis of words. Thus, a sentence is classified by its instances (words of opinion) for which emotions are already attributed. In the literature, words of opinion are also known as polar words or words carrying opinion. Positive opinion words are used to express certain wanted states, while negatives are used to express unwanted one. Examples of positive opinion words are good, genius, and appreciable. Examples of negative opinion words are scary, horrible, and mediocre.

In lexical analysis, the input text is converted into tokens. If the token has a positive match, its score is added to the total score of the input text. For example, if the word "perfect" has a positive match in the list of opinion words, the total score of the text is incremented by the associated weight. Otherwise, the score is decremented by the same amount when the word is labeled as negative.

To generate the list of opinion words, there are three main approaches resumed in Figure 4: a manual approach, a dictionary-based approach, and a corpus-based approach.

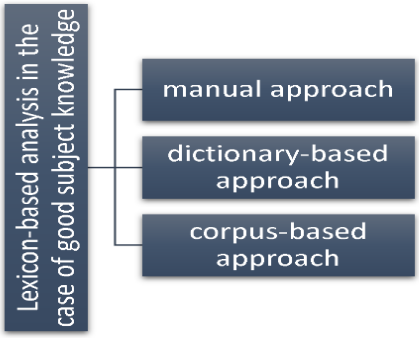

Figure 4. Lexical analysis approaches

• Manual approach: This method is precise, but it is time-consuming so is generally not used alone; instead, it is combined with automated approaches as a final verification.

• Dictionary-based approach: In this approach, a small set of annotated opinion words is collected manually and then developed by searching for their synonyms and antonyms in a dictionary. Newly found words are added to the starting list. This process is iterative and stops when no new words are found. Once the process is complete, a manual inspection is performed to correct any errors. Several researchers [7] have used this approach and generated lists of opinion words.

• Corpus-based approach: This method uses dictionaries to annotate words as well as the context for which polarity is valid. This approach begins with a list of words of opinion that is then expanded based on a large corpus. In [8] authors proposed a "sentiment coherence" technique that begins with a list of opinion adjectives and identifies additional adjective opinion words and their orientations using a set of linguistic constraints or sentence connections (e.g., OR, BUT NI). Another usual constraint is the conjunction "AND" indicates association with two similar orientations. For example, in the sentence "This man is brave and kind", if "brave" is known to be positive, "kind" is also positive because people generally express the same opinion on both sides of a conjuncture. The following sentence is rather unnatural: "This man is brave and authoritarian." If this sentence is changed to "This man is brave but authoritarian," it becomes acceptable. Learning is applied to a large corpus to determine whether two adjectives in the same sentence ("conjoined adjectives") have the same or different orientations. In practice, this is not always consistent. Indeed, in [9] authors demonstrated that the same word could indicate different orientations in different contexts, even in the same field. For example, in the laptop field, the word "long" expresses opposing opinions in these sentences: "Battery life is long" (positive) and "Startup time is long" (negative). Therefore, the generation of opinion words according to the domain becomes insufficient. For this issue, they suggest considering both the possible words of opinion and the aspects: use the couple (aspect, opinion word) as a context of opinion.

To link our contribution with the presentation of the aim methods of sentiment analysis, we present in the following the most relevant work of the SA employing the social networks and which are close to our contribution.

The method of [8] proposed a system based on corpus that retrieve automatically positive and negative semantic information using indirect information from a large corpus of adjectives. The system of prediction is based on linear regression, achieving 92% accuracy for the classification.

In [10] authors propose model based on the Arabic corpus created by hand where they annotated dataset and give steps to build their lexicon. They noted that more the lexicon is rich the performance is high. Authors propose approaches based on the Arabic corpus and on the lexicon, they created by hand an annotated dataset and they prove that SVM (Sector Vector Machine) method used for classifying text give the highest precision.

In [11], the authors used special approach based on correlation and indication of emotional signals and was tested on sets of twitter data. The results have shown the effectiveness of their approach and the importance of including the different emoticons in their analysis.

In [12], the authors developed an unsupervised approach to automate non-concatenative morphology, which they apply to generate a standard Arabic lexicon of roots. They use a recursive

notation based on hypothesized patterns and root frequencies. Their morphological analysis with the induced lexicon fulfils a root identification accuracy to 95%.

In [13], the authors developed classification model for the Egyptian dialect using Arabic tweets that are analyzed to identify their polarity (positive or negative), using Maching Learning approach that uses supervised classifiers and the semantic approach that requires the construction of a lexicon.

3.2. Machine learning-based approaches

In these approaches, the machine is trained to detect patterns in a corpus by having it learn on a first test corpus. In machine learning, the machine learns from data collected in the past, like human learned from their past experiences in real-world applications.

Machine learning involves five steps Figure 5: data collection, preprocessing, training, classification, and results.

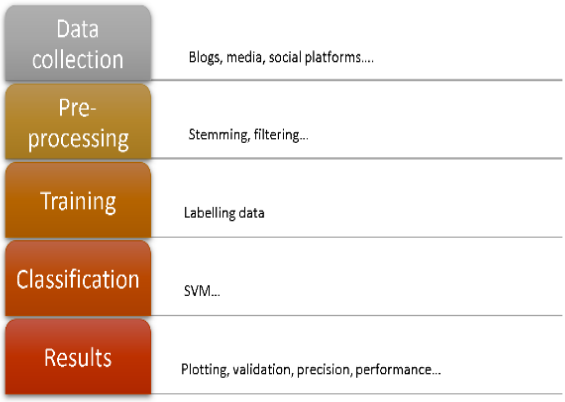

Figure 5. Machine learning steps

• Data collection: The data to be analyzed is collected from various sources, depending on the need of the application, the field, and the context studied.

• Pre-processing: The collected data is cleaned and prepared to be entered into the classifier. Pre-processing is a crucial step and has a direct impact on the quality of the classification operation. Textual data cleansing is completed in several steps and includes tokenization, stemming, or filtering.

• Training: This step consists of labeling a collection of data by hand to generate the training data. The most used method is crowdsourcing. This data is entered into the chosen algorithm for learning purposes.

• Classification: The classifier is trained to detect patterns or patterns in the corpus based on descriptors explained later. After completing the training and building the forecast model, the classifier is deployed on the new data to extract feelings.

• Results: The results are plotted according to the type of representation selected. Then, the classifier's performance is measured according to several methods, including accuracy, recall, F-score, and cross-validation, discussed later.

3.3. The evaluation of the classifier

As mentioned, once a classifier is chosen and built, we must evaluate it to measure its performance and accuracy. This is an important step for any classification project. There are several methods and measures to evaluate a classifier as shown in Figure 6, but the accuracy of the classification remains the main measure. This measure represents the number of documents in the test set that are properly classified, divided by the total number of documents in the test set. Next, we present other metrics and methods commonly used for the evaluation of classifiers.

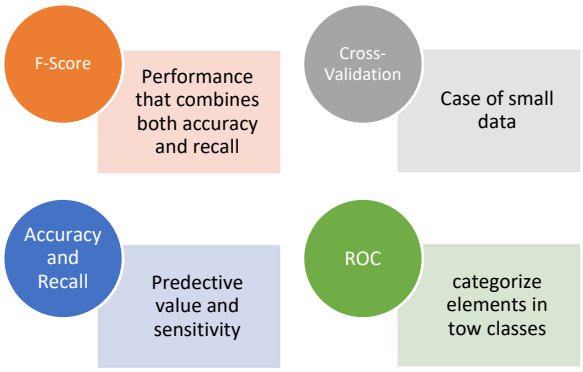

Figure 6. Methods to monitor the performance of a classifier

a) Cross-validation

This method is used especially when the data set is small. The goal of cross-validation is to define a set of data to test the model in the learning phase. In this evaluation method, the available data is partitioned into n disjoint subsets of equal size. Each subset is then used as a test set and the remaining n-1 subsets are combined as a learning set. This procedure is then performed n times, which gives n precision. The estimated final accuracy of learning from this data set is the average of the n precisions. In general, cross-validations 10 and 5 are most used. Cross-validation can also be used for parameter estimation [14].

b) Accuracy and recall

Accuracy and recall are two criteria for statistical measures evaluating classifiers, also known as predictive value and sensitivity. We note VP, which is the number of items correctly labeled positive (true positive), as well as FN, or the number of incorrect classifications of positive examples (false negatives); FP, or the number of items that were incorrectly labeled positive (false positives); and TN, the number of correct classifications of negative examples (true negative)

In an opinion classification task, the precision p of a class is the number of true positives divided by the total number of positive categorized elements:

$$p = VP/VP+FP \quad (1)$$

The recall r in this context is defined as the number of true positives divided by the total number of elements that belong to the positive class:

$$r = VP/VP+FN \quad (2)$$

An accuracy score of 1 for a class C means that each element associated with class C belongs to that class.

However, this score says nothing about the number of Class C items that have not been properly labeled).

A recall score means that each item belonging to Class C has been correctly labeled (but this score says nothing about the number of items that have been incorrectly associated with Class C). There is an inverse relationship between accuracy and recall: one may be increased at the expense of the other.

c) F-score

F-score is a popular measure of test performance that combines both accuracy and recall. This method is often used to compare different classifiers with a single measure.

$$F = 2 \times (p \times r / p + r) \quad (10)$$

F-score, also called F1-score or F-measure, is the weighted harmonic mean of accuracy and recall:

$$F = 2 / ((1/p) + 1/r) \quad (4)$$

d) The efficiency function of the receiver

The receiver efficiency function, more commonly referred to as the ROC curve or sensitivity/specificity curve, is a performance measure of binary classifiers (systems that categorize elements into two distinct classes). Graphically, this measure is represented by the rate of true positives compared to the rate of false positives. The true TPP positive rate refers to the fraction of positives detected and the TFP false positive rate refers to the fraction of negatives incorrectly detected.

$$TVP = VP/(VP+FN) \text{ and } TFP = FP/(VN+FP) \quad (5)$$

DVT is the reminder of the positive class and is also called sensitivity. Another measure, called specificity, represents the rate of true negative (TVN), or the recall of the negative class. TVN is defined as follows:

$$TVN = VN/(VN+FP) \quad (6)$$

An ROC space is defined by the axes x and y referring to TVP and TFP r, respectively, this representation demonstrates a state between true positive and false positive. Each prediction result is represented by a point in the ROC space.

In Figure 7 the points above the diagonal line represent accurate classification results and the points below the line represent poor results. A perfect classification would give the coordinate (0.1) in the ROC space, representing 100% sensitivity (no false negatives) and 100% specificity (no false positives) [15].

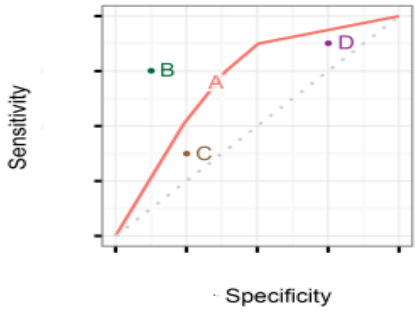

Figure 7. ROC space

4. Novel sentiment classification for Arabic text using the lexicon-based approach

The task of classifying feelings involves labeling the text with a sentiment class. There are two main families of approaches: supervised machine learning and lexicon based. Naturally, supervised methods use machine learning algorithms trained with labeled data (positive, negative, or neutral).

In this section, we propose solutions to overcome the difficulties and limitations related to the exploration of opinion in terms of contextual semantic orientation and adaptability. Adopting a lexicon-based methodology, we present a new adaptable approach that allows one to associate a polarity to words according to context through the construction of dictionaries based on instantiation rules. Our aim is to improve the finesse of the classification of feelings of an Arabic text.

4.1. Problematic

The Problematic of approaches based on lexical analysis generally involve the aggregation of polarity scores from generic repositories to classify text. These approaches are more flexible and therefore more suitable for the classification of feelings in the context of big social data, especially for morphologically complete languages such as Arabic.

Nevertheless, they themselves face challenges such as defining the semantic orientation of words that could be strongly influenced by the context, a word can be considered a negative word in a tweet and at the same time considered a positive word if the tweet is related to a different context. As a result, approaches based on lexical analysis do not give very good results if they are not contextualized. In addition, lexicon-based approaches and/or machine learning do not consider the informality of messages published in social networks. Indeed, these messages could contain special words such as those written in a repetitive or extended word. These special words can be used to weight the emotional load of posts. Therefore, they could be considered as intensifiers or diminutives of polarity.

In this paper, we seek to propose solutions to overcome the difficulties and limitations associated with the exploration of opinion facing semantic orientation presenting constraint. Based on lexicon-based approach, we propose a new approach that allows words to be assigned polarity based on context by creating lexicons assigning polarity, with extended rules to analyze a maximum of semantic orientation, improving the finesse of the classification of feelings of a text, and we focus on this work the sentiments expressed in standard Arabic.

To build a sentiment analysis model, we propose a methodology composed of three steps, as illustrated in Figure 8, which is detailed in the following.

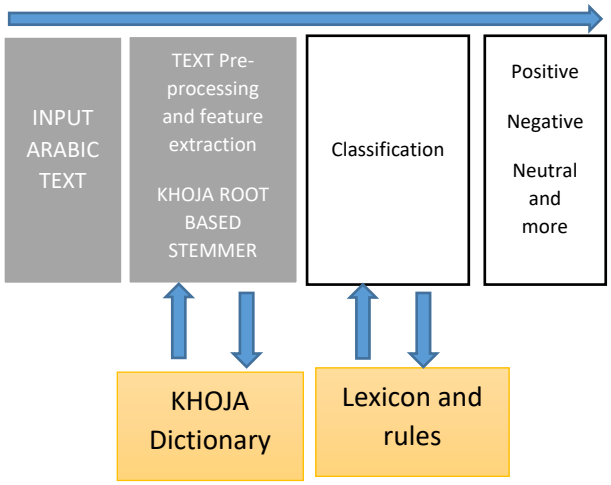

Figure 8. Novel lexicon-based approach

4.2. Preparing DATA

Extraction of data set: We use to scrap specific Tweets replies with Python to handle the verification of our result. The comments are written in Standard Arabic and about given opinion of a specific cosmetic product.

As a tweet could include more than one hashtag, the tweet could be extracted several times. Thus, we deleted duplicate tweets to avoid overweight.

Preprocessing Data: As illustrated in Table 1, this step consists first of translating the different words composing the comments in Arabic text pasted in their roots. To perform this step, we chose KHOJA, an algorithm for preprocessing data based on the roots of the words, we used KHOJA because it is an algorithm that we have perform in many previews work [16], it shows its effectiveness in term of accuracy, available and simple to implement. This algorithm processes the words under the following major steps:

- Tokenization: the process of converting character strings into a list of tokens. Tokens are strings with an assigned and identified meaning. Tokenization consists of eliminating "noises" from the source text, including comments, white space, and punctuation.
- Normalizing
- Stop word removal
- Root stemming: a process of transformation of bends (or sometimes derivatives) into a base or root form.

Table 1. Data preprocessing

Preprocessing step	Output
Basic text	من أحسن المنتوجات, أكثر من رااائع (One of the best products, the most amazing)
Tokenization	من-أحسن-المنتوجات-أكثر-من- رااائع
Normalizing	من- أحسن -المنتوجات-أكثر- من- رائع
Remove stop word	أحسن -المنتوجات -أكثر- رائع
Stemming	احسن -منتوج - اكثر- رائع

4.3. Lexicon dictionary Construction

Lexicon-based sentiment analysis involves extracting sentiment scores from a dictionary. The polarity of a position is therefore calculated by adding the score values {+1, -1} of the words.

To construct our lexicon dictionary, we associate each word to a weight (polarity). If the word does not exist in the lexicon its polarity is null. For feature construction, and to calculate the global polarity of the text, it suffices to aggregate the polarity score of each word constituting the text from the lexicon.

Our method is based on an ordinary polarity, which is founded on rules considering the intensity of the word appearing in the comment in its weight (positive, negative, or neutral), for example:

- Negation: a negative value is assigned to words that inverse the polarity of the word for example 1 turn to -1.
- Intensification: we define specific word that give unadded accent we add +1 to a positive word and −1 to a negative one. For example, the sentence (This cream is very smooth), has a polarity equals to 2 1 refers to the positive word smooth and +1 to the presence of the word very.
- Conflicting phrases: when two words with opposite polarity in the same comment the negative polarity reign.

The objective of this rule is to enhance performance to decide on our classification. Then, we present the results of our rankings according to several classes as follows.

We classified the tweets rated in seven classes according to their degree of polarity, limiting ourselves to this distinction according to the high weight result of polarity detected throw an empirical test on a sample of 1200 tweets to determine the limits of each classroom:

- Highly satisfied: polarity ≥ 7.
- Moderately satisfied: $4 \leq$ polarity ≤ 6.
- Slightly satisfied: $0 <$ polarity ≤ 3.
- Slightly unsatisfied: $-3 \leq$ polarity < 0.
- Moderately unsatisfied: $-6 \leq$ polarity ≤ 4.
- Strong negative: polarity ≤ -7
- If 0, the comment is considered neutral.

4.4. Results

In this stage, we compared results of the same data set using simple polarity and polarity with rules that give the ordinal scale with rules application. In the total text, we used 5000 words in Arabic text: 2000 positives and 3000 negatives. These are opinions of customer satisfaction on a cosmetic product.

Table 2. special words impacting polarity.

Rules	Word
Negation	لا ، لم ، ليس
Intensity	كثيرا ،جدا، قوي
Contradictory	غامض، عجيب

We present in Table 2 some specific words figuring in our training data set and make impact to word polarity.

Table 3. Examples of global polarity

Comment	Binary scale	Ordinal scale and rules
هدا منتوج رائع جدا (This is a great product)	+2	+6
صراحة لم يعجبني مطلقا (I don't like it at all)	-1	-3

Finally, we compared results of the same data set using simple polarity and polarity with rules that give ordinal scale with rules application.

Table 4. Comparing result of binary and ordinal scale with rules: accuracy and F-Score percentage

Method	Accuracy	F-Score
Binary scale	67.30%	61.11%
Ordinal scale and rules	69.08%	**66.30%**

Regarding to our related work [13] implemented sentiment classification for Arabic tweets and obtained the F-score about 65.4% and [14] their F-score obtained was equal to 59.6%.

Even if our F-score result is more interesting which exceeds 66% as shown in Table 4, the result of our training requires improvements, to make more exceptional and efficient output.

5. Conclusion

Currently, the main goal of applications using big social data is to make a machine able to identify emotions and feelings in various areas in real time. The analysis of big social data therefore creates many challenges, such as adaptability and processing of big data in real time.

The first objective of this work was to study and compare the different big data tools available and choose the appropriate tools to study and classify the nature of the data.

The second objective was to establish a methodology based on the lexicon. We present a new adaptable approach that allows assigning polarity to words based on context taking semantic constraint into account, through the construction of dictionaries based on rules. This method has remarkably improved the finesse of the results obtained.

6. Perspectives

The challenge that we face in our approach is to detect a ironic comment that can impact the performance of our classification process, so that we have to enrich our lexicon to consider Arabic words with both polarities (positive or negative).

References

[1] J. Ishwarappa, A.Anuradha, "Brief Introduction On Big Data 5vs Characteristics And Hadoop Technology" Procedia Computer Science, **48**, 319-324, 2015.

[2] M. Bougar, E. Ziyati, "Stemming Algorithm For Arabic Text Using Parallel Data Processing" Revue Méditerranéenne Des Télécommunications , **7**(2), 2017.

[3] S. Khoja, R.Garside, "Stemming Arabic Text. Computing Department," In 1999 Lancaster University, Lancaster, http://Zeus.Cs.Pacificu.Edu/Shereen/Research.Htm.

[4] B. Pang, L.Lee, Vaithyanathan, Shivakumar, "Thumbs Up? Sentiment Classification Using Machine Learning Techniques," Proceedings Of Emnlp, 2002.

[5] S. Kim, E.Hovy, "Extracting Opinions, Opinion Holders, And Topics Expressed In Online News Media Text," In Proceedings of the Workshop on Sentiment and Subjectivity in Text, 1-8. 2006. DOI: 10.3115/1654641.1654642

[6] V. Hatzivassiloglou, W. Janyce, "Effects of Adjective Orientation And Gradability On Sentence Subjectivity," In COLING 2000 Volume 1: The 18th International Conference on Computational Linguistics. 2000.

[7] A. Esuli, F.Sebastiani, "Sentiwordnet: A Publicly Available Lexical Resource For Opinion Mining, Proceedings Of The Fifth International Conference On Language Resources And Evaluation" (Lrec'06) May, 2006.

[8] V. Hatzivassiloglou, K.R. Mckeown, "Predicting The Semantic Orientation Of Adjectives," Proceedings Of The 8th Conference On European Chapter Of The Association For Computational Linguistics Madrid, Spain, 174-181, 1997.

[9] X. Ding, Xiaowen, B.Liu, P.Yu, "A Holistic Lexicon-Based Approach To Opinion Mining. Wsdm'08," Proceedings Of The 2008 International Conference On Web Search And Data Mining, 231-240, 2008, 10.1145/1341531.1341561.

[10] N. Bdulla, A.Ahmed, M.Shehab, M.Al-Ayyoub, "Arabic Sentiment Analysis: Lexicon-Based And Corpus-Based," In 2013 Ieee Jordan Conference On Applied Electrical Engineering And Computing Technologies (Aeect), Pp 1–6. Ieee, 2013.

[11] X. Hu, J.Tang, H.Gao, H.Liuunsupervised, "Sentiment Analysis With Emotional Signals,". Proceedings Of The 22nd International Conference On World Wide Web, 607–618, 2013.

[12] B. Khaliq, J.Carroll, "Induction Of Root And Pattern Lexicon For Unsupervised Morphological Analysis Of Arabic," 2013.

[13] A. Shoukry, A. Rafea, Sentence-Level Arabic Sentiment Analysis. In: International Conference On Collaboration Technologies And Systems (Cts), 546–550. Ieee. 2012

[14] L. Bing; " Data-Centric Systems And Applications Series Editors M.J. Carey S. Ceri Editorial Board P. Bernstein U. Dayal C. Falout." Https://Epdf.Tips/Web-Data-Mining-2nd-Edition-Exploring-Hyperlinks-Contents-And-Usage-Data.Html

[15] H. Mohsen, "Knowledge Discovery Considering Domain Literature And Ontologies : Application To Rare Diseases;" Computation And Language [Cs.Cl]. Université De Lorraine, English, Nnt: 2017lorr0092 Tel-01678860v2f, 2017.

[16] M. Bougar, E.Ziyati, (2019). Stemming Algorithm For Arabic Text Using A Parallel Data Processing: Icict 2018, London. 10.1007/978-981-13-1165-9_23.