

MULTILEVEL SENTIMENT ANALYSIS IN ARABIC

ARAPÇA İÇİN ÇOK DÜZEYLİ DUYGU ANALİZİ

Ahmed NASSAR

Prof.Dr. Ebru AKÇAPINAR SEZER
Supervisor

Submitted to Graduate School of Science and Engineering of Hacettepe
University as a Partial Fulfillment of the Requirements
for the Award of the Degree of Doctor of Philosophy in
Computer Engineering

2017

This work named “**MULTILEVEL SENTIMENT ANALYSIS IN ARABIC**” by **AHMED NASSAR** has been approved as a thesis for the degree of **DOCTOR OF PHILOSOPHY IN COMPUTER ENGINEERING** by the below mentioned Examining Committee Members.

Doç. Dr. Hacer KARACAN

Head

Prof. Dr. Ebru AKÇAPINAR SEZER

Supervisor

Doç. Dr. Ahmet Burak CAN

Member

Yrd. Doç. Dr. Mehmet Serdar GÜZEL

Member

Yrd. Doç. Dr. Gönenç ERCAN

Member

This thesis has been approved as a thesis for the Degree of **DOCTOR OF PHILOSOPHY IN COMPUTER ENGINEERING** by Board of Directors of the Institute for Graduate School in Science and Engineering.

Prof. Dr. MENEMŞE GÜMÜŞDERELİOĞLU
Director of the Institute of Graduate School of Science and Engineering

YAYINLAMA VE FİKRİ MÜLKİYET HAKLARI BEYANI

Enstitü tarafından onaylanan lisansüstü tezimin/raporumun tamamını veya herhangi bir kısmını, basılı (kağıt) ve elektronik formatta arşivleme ve aşağıda verilen koşullarla kullanıma açma iznini Hacettepe üniversitesine verdiğimi bildiririm. Bu izinle Üniversiteye verilen kullanım hakları dışındaki tüm fikri mülkiyet haklarım bende kalacak, tezimin tamamının ya da bir bölümünün gelecekteki çalışmalarda (makale, kitap, lisans ve patent vb.) kullanım hakları bana ait olacaktır.

Tezin kendi orijinal çalışmam olduğunu, başkalarının haklarını ihlal etmediğimi ve tezimin tek yetkili sahibi olduğumu beyan ve taahhüt ederim. Tezimde yer alan telif hakkı bulunan ve sahiplerinden yazılı izin alınarak kullanması zorunlu metinlerin yazılı izin alarak kullandığımı ve istenildiğinde suretlerini Üniversiteye teslim etmeyi taahhüt ederim.

- Tezimin/Raporumun tamamı dünya çapında erişime açılabilir ve bir kısmı veya tamamının fotokopisi alınabilir.**
(Bu seçenekle teziniz arama motorlarında indekslenebilecek, daha sonra tezinizin erişim statüsünün değiştirilmesini talep etmeniz ve kütüphane bu talebinizi yerine getirirse bile, tezinin arama motorlarının önbelleklerinde kalmaya devam edebilecektir.)
- Tezimin/Raporumun tarihine kadar erişime açılmasını ve fotokopi alınmasını (İç Kapak, Özet, İçindekiler ve Kaynakça hariç) istemiyorum.**
(Bu sürenin sonunda uzatma için başvuruda bulunmadığım takdirde, tezimin/raporumun tamamı her yerden erişime açılabilir, kaynak gösterilmek şartıyla bir kısmı ve ya tamamının fotokopisi alınabilir)
- Tezimin/Raporumun tarihine kadar erişime açılmasını istemiyorum, ancak kaynak gösterilmek şartıyla bir kısmı veya tamamının fotokopisinin alınmasını onaylıyorum.**
- Serbest Seçenek/Yazarın Seçimi**

07 / 09 / 2017

Ahmed Nassar

(İmza)

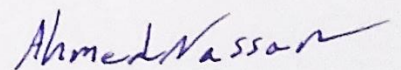
Ahmed NASSAR

ETHICS

In this thesis study, prepared in accordance with the spelling rules of Institute of Graduate School in Science and Engineering of Hacettepe University, I declare that

- all the information and documents have been obtained in the base of the academic rules
- all audio-visual and written information and results have been presented according to the rules of scientific ethics
- in case of using other works, related studies have been cited in accordance with the scientific standards
- all cited studies have been fully referenced
- I did not do any distortion in the data set
- And any part of this thesis has not been presented as another thesis study at this or any other university.

24 August 2017



Ahmed NASSAR

ABSTRACT

MULTILEVEL SENTIMENT ANALYSIS IN ARABIC

Ahmed NASSAR

Doctor of Philosophy, Department of Computer Engineering

Supervisor: Prof.Dr. Ebru AKÇAPINAR SEZER

August 2017, 89 pages

Sentiment analysis has a great necessity to classify sentences like review, news, blog, etc. in order to hold the overall sentiment (i.e. negative, positive or neutral) embedded in them. The vast majority of studies focused on sentiment analysis for English texts, while there is small number of researches has focused on other texts such as Arabic, Turkish, Spanish and Dutch. In this study, we aimed at improving the performance results of Arabic sentiment analysis in the level of document by: firstly, investigating the most successfully Machine Learning (ML) methods to classify sentiments, at the same time rules have been implemented to create new vector formats for representation of inputs with ML based modeling process. Secondly, applying Lexicon Based (LB) approach in both term and document levels by using different formulae based on aggregating functions like maximum, average and subtraction. However, the rules have been applied in the experiments. Performance results of LB approach have been used to identify the best formulae can be used with term level and document level of lexicon based SA at Arabic Language, also the effectiveness of using rules in both levels has been illustrated.

As a final point, employed methods of the two different approaches (i.e. ML and LB) have been tried to create a combined method with considering rules.

The OCA corpus was used in the experiments and a sentiment lexicon for Arabic sentiments (ArSenL) was used to resolve the challenges of Arabic Language. Several experiments have been performed as followed: Firstly, features have been selected for both term and document levels of the OCA corpus independently. Secondly, different linear ML methods such as Decision Tree (D-Tree), Support Vector Machine (SVM), and Artificial Neural Network (ANN) have been applied on both of OCA corpus levels with considering applying and not applying rules on both levels of the corpus. Thirdly, LB approach have been applied on the document level with considering applying rules to each term in a document. And finally comparisons between the results have been done to identify the best way to classify sentiment Arabic documents.

The most successful results in the study are as follows: (i) In ML approach, ANN classifier has been nominated as best classifier in the term level and in the document level of Arabic SA. Furthermore, the average of F-score achieved in the term level for positive testing classes is 0.92, and also in negative classes is 0.92, however, in the document level, the average of F-score for positive testing classes is 0.94, while in negative classes is 0.93. (ii) In the LB approach, it is concluded that the best results have been achieved by applying rules for each term, then computing each sentence score by D_{Max_Sub} formula, and finally, using first *sentence score* formulae for document score computing. In general, the results of the ML approach are better than the results of the LB approach.

Keywords: Arabic sentiment analysis, Opinion mining, Machine learning approach, Lexicon based approach, Rules.

ÖZET

ARAPÇA İÇİN ÇOK DÜZEYLİ DUYGU ANALİZİ

Ahmed NASSAR

Doctor of Philosophy, Department of Computer Engineering

Supervisor: Prof.Dr. Ebru AKÇAPINAR SEZER

August 2017, 89 pages

Duygu analizi yorum, haber, blog gibi türlerde gömülü olan genel duyguyu (yani olumlu, olumsuz veya tarafsız) belirlemek için tümcelerın sınıflandırılması işi olarak tanımlanabilir. Bu alanda yapılan çalışmaların büyük çoğunluğu İngilizce metinler için duygu analizi üzerine yoğunlaşırken; Arapça, Türkçe, İspanyolca ve Hollandaca gibi diller için de yapılmış sınırlı sayıda araştırma bulunmaktadır. Tez çalışmasında duyguları sınıflandırmak için sıklıkla kullanılan makine öğrenme (ML) yöntemlerini araştırarak, Arapça belgeler üzerinde duygu analizinin performans sonuçlarının iyileştirilmesi amaçlanmıştır. Aynı zamanda ML ile modelleme sürecinde girdilerin temsilinde yeni vektör biçimleri oluşturulmuştur. İkinci olarak maksimum, ortalama ve çıkarma gibi toplama işlevlerine dayalı farklı formülleri kullanarak terim ve belge düzeylerinde sözcük tabanlı yaklaşımın uygulanması amaçlanmıştır. Sözcük tabanlı yaklaşımın performans sonuçlarındaki en iyi yaklaşımların, Arapça'da sözcüğe dayalı duygu analizinin terim ve belge seviyesi ile birlikte kullanılabileceğini göstermek için kullanılmıştır. Ayrıca, her iki seviye için kuralların kullanım etkinliği de gösterilmiştir. Son olarak, kuralların ML yöntemlerine entegre edildiği hibrid bir yöntem önerilmiştir.

Deneylerde OCA derlemi ve Arapça duygular için Arapça'nın zorluklarını çözümlenmede geniş kapsamlı bir duygu sözlüğü olan (ArSenL) kullanılmıştır. Deneylerin ilkinde, modellemde kullanılan özellikler OCA derlemininden hem terim hem de doküman düzeyleri için bağımsız olarak seçilmiştir. İkincisinde ise, derlemde Destek Vektör Makinesi (SVM), Karar Ağacı (D-Ağacı) ve Yapay Sinir Ağı (ANN) gibi farklı ML yöntemleri hem tek başlarına ve hem de kurallarla birlikte uygulanmıştır. Üçüncüsünde ise, bir belgedeki her terim için kuralların uygulanabilmesi için belge seviyesinde sözcük temelli yaklaşım uygulanmış ve sonuçlar arasında karşılaştırmalar yapılmıştır.

Bu çalışmada elde edilen en başarılı sonuçlar şöyledir: (i) makine öğrenimi kullanımında, ANN sınıflandırıcısı, pozitif test sınıfları için terim ortalamasında elde edilen F-skorunun ortalaması 0.92 olan Arapça duygu analizinde terim düzeyinde ve belge düzeyinde en iyi sınıflandırıcı olarak adlandırılmıştır ve negatif sınıflarda da 0.92'dir. ANN modelinin belge seviyesinde, pozitif test sınıfları için F-skor ortalaması 0.94, negatif sınıflarda ise 0.93'tür. (ii) Sözcüğe dayalı yaklaşıma göre, en iyi sonucun her terim için kurala dayalı yöntemler uygulayarak, sonra her cümle skorunu D_{Max_Sub} formülüyle hesaplayarak ve son olarak, belge skoru hesaplaması için ilk cümle puanı formülü kullanılarak elde edildiği sonucuna ulaşılmıştır. Genel olarak makine öğrenimi yaklaşımı sonuçları sözlük kullanımı yaklaşımına göre daha iyi sonuçlar vermiştir.

Anahtar Kelimeler: Arapça Duygu Analizi, Düşünce Madenciliği, Makine Öğrenmesi Yaklaşımı, Sözlük Tabanlı yaklaşım, Kurallar.

ACKNOWLEDGEMENT

Firstly, I would like to express my sincere gratitude to my advisor Prof.Dr. Ebru AKÇAPINAR SEZER, for the unlimited guidance and the continuous support throughout my PhD study and research, who always welcomed my questions and advised me. Without her supervision and expertise; this work would not have been possible.

Besides my advisor I would like to express my thanks to my thesis committee members for their support, guidance and helpful feedback, also I thank Dr. Fatih Mehmet GÜLEÇ for his guidance and advices.

My sincere thanks also goes my fellow lab mates in multimedia Information retrieval research group, specially Alaettin UÇAN. I am also very grateful to my friend Tamer ALLOH for his technical support.

Moreover, I would like to thank TiKA and YTB for sponsoring and funding my study, the Turkish people for their friendly accommodation, and IUGAZA for their assistance.

I have no words to express my gratitude to my daer wife Nour AWADALLAH and to my children Majd, Asmaa, Nesma and Mohammed for their unlimited support, patience and encouragement.

Finally, my sincere thanks to the spirit of my father who was my pacemaker and to my beloved family my mother, my sisters and my brothers for their unlimited encouragement.

24 August 2017
Ahmed NASSAR

CONTENTS

	<u>Page</u>
ABSTRACT	I
ÖZET	III
FIGURES	IX
TABLES	X
ABBREVIATIONS AND SYMBOLS.....	XII
1. Introduction	14
1.1 Sentiment Analysis Importance and Definition	14
1.2 Sentiment Analysis General Tasks	14
1.3 Study Tasks	15
1.4 Study Objective and Motivation	17
1.5 Study's Originality.....	18
1.6 Organization of the Study.....	18
2. Literature Review	19
2.1 Sentiment Analysis in English Language	19
2.2 Sentiment Analysis in Arabic Language	23
2.3 Sentiment Analysis in Other Languages	25
2.4 Literature Discussion.....	28
3. Background.....	30
3.1 Machine Learning Approach	31
3.1.1 Support Vector Machine Classifier (SVM)	31
3.1.2 Artificial Neural Network Classifier (ANN).....	32
3.1.3 Decision Tree Classifier (D-Tree)	32
3.2 Lexicon Based Approach	33
3.2.1 Lexicon Based Computing.....	33

3.2.2	Lexicon Building.....	35
3.3	Arabic Language Overview	36
4.	Sentiment Analysis in Arabic Language.....	38
4.1	Study Methodology.....	38
4.1.1.	Preprocessing and Quality Measuring of OCA Corpus.....	42
4.1.2	Stemming Process and Identifying Sentiment Terms	46
4.1.3	Prior Polarity Computing Formulae for Term Level.....	47
4.1.4	Rules	48
4.1.5	Sentence Score Computing Formulae.....	48
4.2	Arabic Sentiment Analysis Using Machine Learning (ML) Approach	49
4.2.1	Features Extraction.....	49
4.2.2	Classification Task.....	51
4.3	Arabic Sentiment Analysis using Lexicon Based (LB) Approach.....	52
4.3.1	Lexicon Based Approach without Rules	52
4.3.2	Lexicon Based Approach with Using Rules	54
4.4	Evaluation Measures.....	55
4.5	Experimental Results and Evaluation.....	55
4.5.1	Experiment A: Term Level Sentiment Analysis using Machine Learning Approach	57
4.5.2	Experiment B: Document Level Sentiment Analysis Using Machine Learning Approach	61
4.5.3	Experiment C: Sentiment Analysis Using Lexicon Based Approach without Rules	65
4.5.4	Experiment D: Sentiment Analysis Using Lexicon Based Approach with Applying Rules in the Document Level.....	71
4.5.5	Results Summary	73
5.	Conclusion.....	78

References.....	80
APPENDEX 1: Samples from the OCA Database of Sentiment Words based on M _{Max_Sub} Prior Polarity Computing Formula.....	85
APPENDEX 2: Samples of Intensification and Negation Words in Arabic Language	89
CURRICULUM VITAE	90



FIGURES

	<u>Page</u>
Figure 1 Sentiment analysis general tasks	15
Figure 2 Statistics of articles with respect to the approach over years, Medhat et al. [1].	29
Figure 3 Applied Sentiment Analysis approaches in this study	30
Figure 4 Proposed methodology of multi-level SA in Arabic Language using ML approach	40
Figure 5 Proposed methodology of multi-level SA in Arabic Language using LB approach	41
Figure 6 (a) OCA frequencies VS Rank, (b) Ideal (Zipf's) Frequency VS Rank	44
Figure 7 Comparison between Actual and Ideal frequencies VS Rank	44
Figure 8 Comparisons of the ML Approach Experiments Results	76
Figure 9 Comparisons of the LB Approach Experiments Results	77

TABLES

	Page
Table 1. Samples of positive and negative OCA reviews.....	42
Table 2. Statistical details of OCA after pre-processing.....	43
Table 3. Comparison between D_{KL} for OCA and D_{KL} for Arabic corpora by Rosso and Benajiba [43]	45
Table 4. Comparison between D_{KL} for OCA and D_{KL} for Arabic corpora by Salhi and Yahya [44]	45
Table 5. Five ArSenL positive and negative scores for the term: "hot" (ساخن).....	46
Table 6 Prior polarity computing equations	47
Table 7 Sentence score computing equations	49
Table 8 Sample of one of the OCA files including the total eight features in term level.....	50
Table 9 Sample of one of the OCA files including the total seven features in the sentence level	51
Table 10 Term level based document score computing formulae.....	53
Table 11 Sentence score computing formulae.....	53
Table 12 Sentence level based document score computing formulae	54
Table 13. Evaluation measures.....	55
Table 14 Average of performance by using total 8 and 6 features with ANN classifier	58
Table 15 Average of performance by using total 8 and 6 features with SVM classifier	59
Table 16 Average of performance by using total 8 and 6 features with D-Tree classifier	59
Table 17. Average of performance for using total 8 features with applying rules with the ML classifiers in the term level	60
Table 18 Average of performance of using total 7, 5 and 4 features based on D_{Max_Max} formulae in the document level without rules	62
Table 19 Average of performance of using total 7, 5 and 4 features based on D_{Max_Sub} formulae in the document level without rules	63

Table 20. Average of performance of using total 7 features with applying rules based on D_{Max_Max} formulae in the document level	64
Table 21. Average of performance of using DT_{MAX_MAX} formula	66
Table 22. Average of performance of using DT_{MAX_SUB} formula	66
Table 23. Average of performance of using DT_{AVG_MAX} formula	66
Table 24. Average of performance of using DT_{AVG_SUB} formula.....	67
Table 25 Average of performance of using DT_{AVG_AVG} formula	67
Table 26 Average of performance of using DS_{MAX_MAX} formula.....	69
Table 27 Average of performance of using DS_{MAX_SUB} formula.....	69
Table 28 Average of performance of using DS_{AVG_MAX} formula.....	69
Table 29 Average of performance of using DS_{AVG_SUB} formula	70
Table 30 Average of performance of using DS_{AVG_AVG} formula.....	70
Table 31 Average of performance of using D_{Max_Sub} formula for calculating each sentence score	70
Table 32 Average of performance of using D_{Max_Max} formula for calculating each sentence score with applying rules on each term.....	72
Table 33 Average of performance of using D_{Max_Sub} formula for calculating each sentence score with applying rules on each term.....	72
Table 34 Average of performance by (i) using total 8 features and ANN classifier in the term level with applying M_{Max_Sub} formula for prior polarity computing, (ii) using total 7 features and ANN classifier in the document level with applying D_{Max_Max} formula for sentence score computing.	74
Table 35 Average of performance (i) without applying rules on both term and document levels. (ii) Applying rules on document level.	75

ABBREVIATIONS AND SYMBOLS

ANN	Artificial Neural Network
ArSenL	Arabic sentiment analysis lexicon
AvgNeg	Average of Negative Scores in document
AvgPos	Average of Positive Scores in document
AWN	Arabic WordNet
BOW	Bag of Words
CountPos	Counts of Positive Scores in document
CountNeg	Counts of Negative Scores in document
D-Tree	Decision Tree
ESWN/ SWN	English SentiWordNet
EWN	English WordNet
F_Subj	First Subjective Score in document
IMDB	Internet Movie Database
KNN	Key Nearest Neighbor
L_Subj	Last Subjective Score in document
LB	Lexicon Based
MaxNeg	Maximum Negative Score in document
MaxPos	Maximum Positive Score in document
ME	Maximum Entropy
ML	Machine Learning
MSA	Modern standard Arabic
NB	Naïve Bayes
NegScore	Negative Score
OCA	Opinion Corpus for Arabic
OM	Opinion Mining
PMI	Pointwise-mutual Information
POS	Part of Speech
PosScore	Positive Score
SA	Sentiment Analysis

SAMA	Standard Arabic Morphological Analyzer
SO	Semantic Orientation
SumNeg	Summation of Negative Scores in document
SumPo	Summation of Positive Scores in document
SVM	Support Vector Machine



1. Introduction

1.1 Sentiment Analysis Importance and Definition

The widespread of using social networks, forums and personal blogs enabled millions of people to post and share their comments or reviews on the web. These reviews can cover several topics including products, films and others. This actuality encouraged many companies, governments, customers and other parties to make analysis of these opinions.

Whereas, companies follow customer interests and comments in products, governments also track people opinions to make decision about them. However, customers are interested in collecting information from other online people, to know their opinions regarding any product they intend to buy [23]. Manually collecting of people opinions through huge amount of reviews or comments is time consuming and it could be impossible, especially with the speed growth of e-commerce. Therefore, the effective solution for this problem is sentiment analysis.

Sentiment Analysis (SA) is a task of defining polarity (negative or positive) of opinions, emotions or evaluations. However, sentiment analysis tends to extract the stance of a person automatically toward some topic or a document [39]. Processing the huge number of reviews and comments is a challenge that faced researchers in the fields of text mining and information retrieval. This processing task is included inside sentiment analysis or opinion mining field, at the same time, it is sub task of text mining [23]. However, the expression of sentiment analysis and the expression of the opinion mining (OM) are the same [1].

1.2 Sentiment Analysis General Tasks

In general, SA is a field that includes important tasks presented in Fig. 1. Whereas, the reviews of products, movies, tweets or others are normally including subjective (opinioned) words. On the other hand, they may include numbers, punctuations and other unnecessary information. However, the first process of SA is sentiment identi

fication, which tends to extract all opinion words or phrases from a review. Another interesting task is feature selection; this task can be applied by using LB methods or statistical methods in [1]. Another popular task is sentiment classification, whereas ML or LB approaches or a combination of them (i.e. hybrid approach) can be applied on the features from the previous process to decide if the document is positive or negative. However, there are many ML classifiers can be used in text classification including Key Nearest Neighbor (KNN), Support Vector Machine (SVM), Naïve Bayes(NB), Decision Tree (D-Tree) and others, while the LB approach depends on a list of sentiment terms.

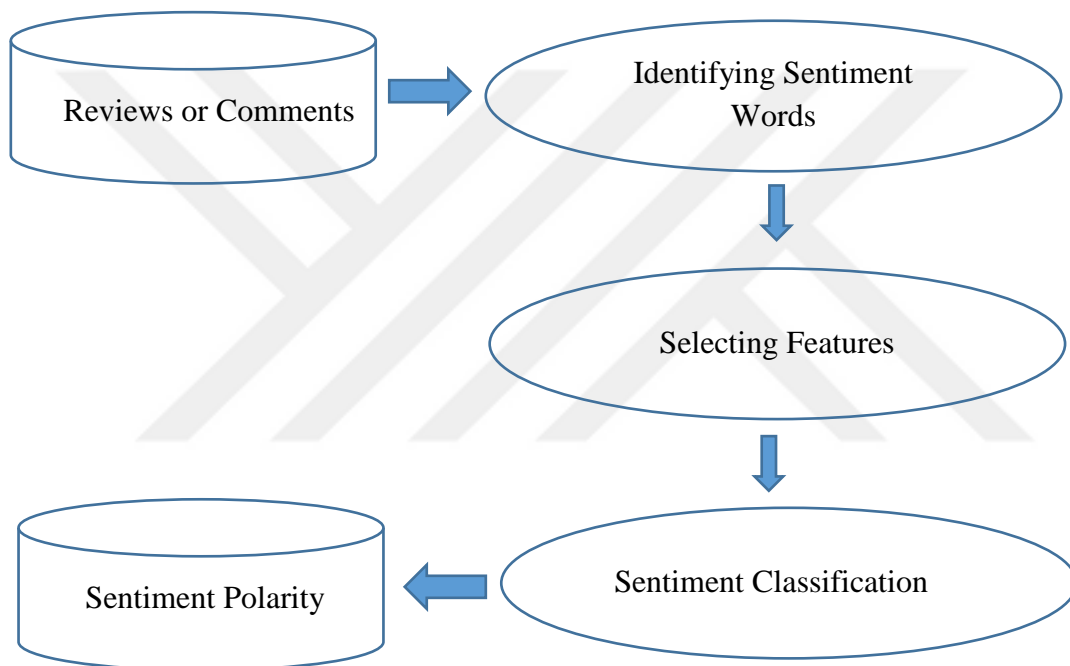


Figure 1 Sentiment analysis general tasks

1.3 Study Tasks

In our study, we are focusing on sentiment analysis (SA) at multi-level (i.e. term and document levels) in Arabic Language. The experiments of Arabic SA have been done in several interesting tasks. In the first task, an Arabic data set called OCA produced by Rushdi-Saleh et al. [23] was selected. It is composed of 500 films reviews; 250 of them are positive and the rest are negative, however, preprocessing of OCA corpus was made by omitting unnecessary terms like numbers and

punctuations, then the high quality of the OCA corpus was ensured by using Zipf's law measure.

The opinion terms have been decided from the corpus by firstly, stemming all terms in OCA by using Standard Arabic Morphological Analyzer (SAMA) [38]. Then sentiment scores have been given for each term root in OCA by comparing term roots placed in OCA with the large scale Arabic sentiment analysis lexicon (ArSenL) created by Badaro et al. [27]. Whereas, each word root (lemma) in the lexicon has many senses or scores which are rated in the interval of [-1...1]. At this stage, one positive or negative score (prior polarity) was computed from many scores (posterior polarities) for each opinion word in OCA by creating different aggregation formulae, then the ML approach was used to nominate the best formula for computing prior polarity for each Arabic word, more details for prior polarity computing are in the later fourth chapter.

The second task is selecting features in two levels of SA, the term level and the document level. In the term level total eight features have been created for each document in OCA, in which the features have been ordered as: *count of positive scores, count of negative scores, summation of positive scores, summation of negative scores, average of positive scores, average of negative scores, first score in the document and last score in the document*. While in the document level total seven features have been created, the order of features is: *count of positive sentences, count of negative sentences, maximum score of positive sentences, maximum score of negative sentences, first sentence score, middle sentence score and last sentence score*. Whereas, two versions of the features have been made for both of the term level and the document level. However, the first version includes all features have been created by applying intensification and negations tools (i.e. rules) on each term placed in OCA, for example, the negation tools “ليس، لم، لن، لا” in Arabic have the same usage of “not” word in English. While in the second version of features the rules have been not applied on each term placed in OCA. In the third task, ML and LB approaches have been used to classify all OCA documents, and the classification results can lead us to nominate the best approach regarding to SA in Arabic Language.

In the case of applying ML approach, three encouraging ML classifiers have been used from SA literature, which are ANN, D-Tree and SVM. The performance was shown in the previous related studies encouraged us to select these classifiers.

In the term level of SA, the three classifiers have been applied on total eight features and total six features (i.e. by omitting last two features) of each document of OCA. However, the results of the three classifiers are used to (i) prove the efficiency of our features by proving that, decreasing the number of features decreases the classification results, and (ii) to nominate the best formula can be used for computing prior polarity from posterior scores in the term level of SA for Arabic Language. On the other hand, LB approach was applied in both term and document levels by using different formulae based on aggregating functions like maximum, average and subtraction. However, the intensification and negations rules have been applied in the experiments. The LB approach results have been used to identify the best formulae can be used with term and document levels of LB SA for Arabic Language, also to illustrate the effectiveness of using rules on both of the two levels.

1.4 Study Objective and Motivation

Further to the previous summary about what we do, our objectives in this study are: firstly, enhancing term level sentiment score by creating new successful formula to compute one prior polarity score for each sentiment term. Secondly, improving the performance results of both term and document levels at Arabic sentiment analysis by investigating the most successfully ML classifier to classify opinion documents. Thirdly, creating new successful LB formula to classify Arabic opinion documents. Finally, implementing rules tools with both of ML approach and LB approach to identify the tools effectiveness on both approaches results.

Several factors motivated us to carry out a research in SA in the Arabic Language. The first factor is the consideration of a large number of Arabic audience, in that the Arabic language is evaluated as one of the most widely used languages on the Internet according to the statistic carried out by the Internet World Stats [49] and it has the fourth rank. Secondly, the Arabic Language has an interesting historical relationship with its people and the region they lived in. Lastly, the large-scale use

of the Internet, social media and social networks plays a great and important role on the Arabic Language.

1.5 Study's Originality

There are many studies conducted in SA in the literature but the vast majority of these studies are dedicated to English Language and not directly applicable to Arabic or any other languages as well. Indeed, critical stage of a typical SA is the specification of the polarity, and according to our best knowledge, there is no research work in Arabic SA that interests the computing of the prior polarity from posterior polarities. Actually proposed SA steps are creating of formula for computing prior polarity, applying ML and LB based classifications and taking into account negations and intensifications. In other words, we consider polarity of term at first and polarity of document at last.

1.6 Organization of the Study

The rest of this dissertation is ordered as the following: Chapter 2 presents and discuss previous studies in SA with respect to this study. Chapter 3 discusses the background including ML approach, LB approach and Arabic Language. Chapter 4 illustrates in details SA in Arabic Language methodology including applying ML approach in term and document levels, applying LB approach in both levels, applying rules in both of the approaches, and the study experimental results and evaluations. Chapter 5 shows the conclusions of the study and its future work.

2. Literature Review

SA task tends to classify the polarity, which can be positive or negative in three common levels: aspect-level, document-level and sentence-level [1] and [55]. There is no difference between sentence and document level classifications because short documents are just sentences [2]. The vast majority of studies are for English Language and not directly applicable for other languages such as Arabic. For this reason, there are other studies interested in other languages. The main objective of any study related to sentiment analysis is to indicate the opinion of reviews or comments in any web content. In this part of the study, we highlight the studies that were applied to English, Arabic and other languages, at the same time are beneficial to our study. Then we generally make discussion about them

2.1 Sentiment Analysis in English Language

In publications, there are many works interested in SA of English Language. Li and Jain [3] used four methods for document classification: decision trees classifier, naive Bayes classifier, nearest neighbor classifier and subspace method. The data set was selected from different groups of yahoo news. Features were represented using standard bag of words (BOW). They indicated that naive Bayes classifier and the subspace method give better results than decision trees and nearest neighbor classifiers on the same data sets. They also studied the combination of the three classifiers that did not always enhance the accuracy of classification compared to the best individual classifier. The best classification accuracy achieved is approximately 83%, which was comparable to the other similar previous studies.

NB classifier was applied by Janyce et al. [4] to make classification for documents collected from the journal of Wall Street. They used linguistic features such as pronouns, adjectives and verbs. The average accuracy of the classifier reaches to 81.5%, which is strong performance.

Pang et al. [5] used ML methods such as ME, NB, and SVM to define the polarity of movie reviews. They downloaded the data from Internet Movie Database (IMDb). The reviews were divided to 700 positive and the other 700 reviews are negative. They applied ML classifiers and standard bag of features (BOW) on documents. They also considered the effect of adding negation “not”. The position of word and the part of speech were also treated. They performed many experiments using n-grams approaches, and according to the results they concluded that using of unigrams is the most effective method, and the SVM classifier outperforms the other classifiers used in their work.

Hong Yu and Vasileios Hatzivassiloglou [6] used NB classifier with features contain words, bigrams, and trigrams, as well as the parts of speech in every sentence to classify documents to actual and sentimental. The documents were collected from wall street journal. The archived results were encouraging with very high precision and recall (F-score of 97%).

Mullen and Collier [7] worked in the same corpus used by Pang et al. [5]. They used the average rating to classify the reviews, were as the reviews above the average rating were classified as positive and those under this average rating were classified as negative. They also treated with many features such as word unigrams. Moreover, they used another movie corpus, in that situation they extracted same features and additional features based on the range of the movie. They applied SVM as ML classifier. Finally, they accomplished that using the collection of word unigrams and their lemmas outperforms the other models that did not use this kind of information.

Chesley et al. [8] applied SVM classifier on linguistic features such as pronouns, adverbs, verbs and adjectives, to discover polarity in reviews from blog spots. The classifying results were encouraging in spite of the noisy nature of the data in blog spots.

Jia et al. [9] built a system that can classify a document as sentimental if there is an opinion sentence in the document at least. They used SVM classifier with both of unigrams and bigrams features, which were chosen using Chi-square test to specify

orientation of opinion in the sentence level. Then a decision tree classifier was used to determine the polarity of opinionated documents.

Prabowo and Thelwall [10] used several classifiers such as SVM, Statistics Based Classifier (SBC), General Inquirer Based Classifier (GIBC) and Rule Based Classifier (RBC) to classify reviews from different corpora. They used hybrid classification, whereby one classifier send the document to the next classifier when it fails in the classification process, and the passing process continue until no classifier remains. They concluded that the effectiveness of SBC and SVM classifiers is improved in the hybrid classification.

SVM was applied by Li and Li [11] as a sentiment polarity classifier. They debated that sentimental subjectivity and expresser credibility should also be taken into account. A framework that gives a compact numerically summarization of opinions on micro-blogs platforms was proposed in their work. The topics mentioned in the opinions shared with the queries of users were extracted, and then SVM was used to classify them. They used twitter posts for their experiment. They found out that aggregating micro-blog opinions essentially needs consideration of user credibility and sentimental subjectivity. They proved that the proposed mechanism could effectively detect market intelligence (MI) for supporting decision-makers.

Moraes and Valiati [12] made experimental comparison between ANN and SVM classifiers in the document level of sentiment analysis because in the literature of sentiment learning ANN classifier has little mentioned, while SVM classifier has widely used by the researchers in successful way. They have used the adoption of a criterion evaluation context with common supervised methods for selection of feature and weighting in a classical bag of words (BOWs) model. Their work indicated that the experiments results of ANN outperform experiments results of SVM for some unbalanced data contexts. Their testing included three types of data sets including Films, GPS, Camera and Books documents from *amazon.com*. In the experiments of film reviews data set, they indicated that ANN outperformed SVM by a statistically significant difference. They confirmed some limitations that have been seldom discussed in the SA literature which are the cost of computation of SVM at running time and the training time of ANN. They concluded that the reducing of

computational effort of both classifiers can be achieved by using Information gain (a computationally inexpensive feature selection Method) without significantly affecting the resulting classification accuracy.

On the other hand, there are many studies interested in LB (semantic orientation) approach. Hatzivassiloglou et al. [13] focused on word orientation in their study. They used 1987 Wall Street Journal corpus in their experiments. They detected the text orientation by using adjectives as an effective indicator. Additionally, they supposed that at the level of phrase, if the adjectives linked with a conjunction such as “but” are likely to express opposing opinions, whereas, if the conjunctive such as “and” connects two adjectives probably indicates the same semantic orientation. They used log-linear regression model to predict whether two conjoined adjectives are of the same orientation or not. In this task 82% accuracy was achieved.

Turney and Peter [14] detected document sentiment based on adjectives or adverb phrases by using a proposed unsupervised learning algorithm. Then, they computed the semantic orientation by using Point Mutual Information (PMI). They assigned a class of “recommended” or “not recommended” to the sentences based on the average semantic orientation of the phrases. Finally, the achieved average accuracy is 74% when the data were taken from different domains (410 reviews of banks, films, automobiles and travel destinations). While the accuracy is 84% for automobile reviews and 66% for film reviews.

Another technique was proposed by Hu et al. [15] for predicting opinion at the sentence level to summarize the customer's reviews of a product. First, they extracted features by investigating frequent words. Then, they realized the sentiment sentences, which include at least one adjective and one feature. The prediction of the orientation of a sentiment sentence depends on the sentiment word in that sentence. The determination of the orientation of the sentence is based on whether most sentiment words tend to be positive or negative. If the number of positive and negative sentiment words is the same, the orientation of the nearest sentiment sentence is taken.

Thet et al. [16] focused on the clause level of sentiment analysis. They used Mean formula to calculate the average of positive and negative scores for all given scores of words lemma to derive polarity from SWN. They have experimented with a dataset of movie review sentences. They made a system that generates dependency and splits sentences into clauses. Then this system can compute the sentiment score of each clause taking into account the word prior polarity score derived from SWN, grammatical dependencies of the word and negations. The result of experiments described the effectiveness of this approach on short documents such as message posts.

Guerini et al. [17] used sentiwordnet (SWN) [18] to derive prior polarity sentiment score from term posterior polarities, thus they tested many formulae that combine posterior polarities in different ways. They compared the previously most used techniques to their proposed ones and incorporated all of them to test whether mixing them can make more improvement to the computing of prior polarity scores. They established motivating bases in computing prior polarity scores.

Moreover, there are many studies applied hybrid approach. Kouloumpis et al. [19] have constructed a model of two classifiers for classifying the sentiment tweets in Twitter of English. One classifier uses n-grams, while the other applies both n-grams and features of lexicon. Three features for each tweet were created according to the existence of any words from the lexicon. They have assigned prior polarity to each word by using MPQA subjectivity lexicon [20], which includes a list of English words that are labeled with their prior polarity: negative, positive, or neutral. According to the outcomes of both classifiers, it was remarkable the accuracy results of second classifier (n-grams and lexicon features) is better than the first one (n-grams only).

2.2 Sentiment Analysis in Arabic Language

Some studies have covered the idea of Arabic sentiment analysis. Almas and Ahmad [21] described a method that can automatically extract specialist terms for different languages such as Arabic, English and Urdu. However, they applied their method on financial news texts. Regarding the method performance, their result was good in the part of precision which is (88.1 %), but the recall result is very low which is (17.2 %).

A. Abbasi et al. [22] applied sentiment classification on forum that posts were posted in English and Arabic languages, their approach based on using stylistic and syntactic features linked to entropy weighted genetic algorithm (EWGA) feature selection method. The achieved result of using EWGA with SVM indicated high efficiency, with accuracy of over 91% on the data set. EWGA has better results than other feature selection methods, indicating the usefulness of these features and methods of document level sentiment classification.

In the case of applying ML approach for Arabic Language, Rushdi-Saleh et al. [23] generated new Arabic corpus called OCA, the data of OCA corpus are reviews of movies were collected from several movie blogs, having a total of 500 reviews (250 positive and the remaining are negative). Some experiments were also executed to evaluate the classifiers used to determine the polarity of a review. The best result is achieved by using SVM classifier. They observed that the best result is 90 % accuracy measure using 10 fold cross validation.

Shoukry and Rafea [24] investigated ML classifiers (SVM and NB) to Arabic sentence level SA using 1000 tweets from twitter. They used unigram and bigram as features. However, they found that the SVM classifier outperforms the NB classifier in all result metrics.

For the LB approach in Arabic, Farra et al. [25] used a list of Arabic words' roots which are selected by using stemmer application. To classify each word in the sentence, they were checked against the list of words' roots. If the root is found in the list, its polarity is selected as positive, negative, or neutral. Else, the user will add the root of the absent word to the list of learned roots. This study also tried to benefit from the advantages of both LB and ML approaches by combining them together (i.e. hybrid approach). They used an approach that combining both syntactic and semantic features to classify sentiment of Arabic sentence level. They used many features such as frequency of negative, positive, and neutral terms in each document, with applying the semantic list (dictionary) they have constructed, frequency of contradiction words, frequency of negation tools like “لا، لن، لم، ليس” which are negation tools in Arabic and have the same usage of “not” word in English and other features are also used. El-Halees [26] also developed a mechanism for Arabic

documents to be classified as positive or negative. His experiments were applied to 1143 posts contains 8793 Arabic Statements. Firstly, the documents were entered to a classification model that is based on lexical resources. Most of the documents were classified at this part. Secondly, Maximum Entropy (ME) method used the classified documents by the previous model as training set, and then it classifies some other documents. Finally, the classified documents from the previous two models were entered to k-nearest method to classify the remaining documents. In average, his system achieved an accuracy reached to 80%, and the f-measure of positive documents has better results than f-measure of negative documents. Badaro et al. [27] created a large sentiment lexicon for Arabic (ArSenL) by using Arabic and English resources: (i) English SentiWordNet (ESWN) [18]; (ii) English WordNet (EWN) [36]; (iii) Arabic WordNet (AWN) [37] ;(iv) Standard Arabic Morphological Analyzer (SAMA) [38]. The English synsets were connected to the lemma entries in the Arabic resources. ArSenL provides number of positive and negative scores for each word. Their results show that ArSenL lexicon outperforms all lexicons in all measures without exceptions.

Regarding the comparison studies, Nawaf et al. [28] addressed two approaches to sentiment analysis in Arabic: corpus based and LB. They built manually annotated data set and then they built a lexicon. They carried out experiments to note the improvements obtained from the accuracy of the system and compare them to the corpus based approach. They observed that the highest accuracy was achieved when SVM classifier was used for classifying a light stemmed data set. Additionally, they perceived that there is a direct proportion between size of the lexicon and the accuracy of the LB tool.

2.3 Sentiment Analysis in Other Languages

Some studies have concentrated on sentiment analysis for non-English. Denecke [29] used machine translation software to translate a German review into English. She used three classifiers to predict polarity of the English version of the review. The used classifiers are SWN with ML, SWN with classification rule and Ling Pipe. Sentiment classification was also applied by Zhang et al. [30] to Chinese reviews.

Agić et al [31] applied sentiment analysis on financial news in Croatia by using annotated corpus. Ghorbel and Jacot [32] assessed the polarity of classified French film reviews by using ML and SWN.

Martín-Valdivia et al. [33], used the parallel translation of a Spanish movie reviews corpus into English. They proposed new method for polarity classification of these reviews. They joined three models produced by supervised and unsupervised learning methods to prepare a meta-classifier. The first and the second models were created by applying a ML algorithm to Spanish corpus and its parallel translated version. The third model was generated for the English version of the corpus using SWN. Lastly, they have combined various features of the two supervised models and of the third unsupervised model into meta classifier. The outcomes of using combination techniques were hopeful.

Akba et al. [34] used corpus includes Turkish film reviews as data set. They aimed to use small number of features that extracted from each film reviews. For this reason, information gain and chi-square methods have been used for extracting features to decrease costs of calculations and to increase success rate. In their work, they employed support vector machine and naïve bayes classifiers for classification purpose. For performance evaluation, F-score was used. They noted that, in case of classifying film reviews in to two (positive and negative) categories, the SVM classifier have achieved 83.9% performance value, but in case of classifying film reviews in to three categories, the SVM classifier have achieved 63.3% performance value.

Dehkharghani et al. [35] proposed a comprehensive sentiment analysis system that cover multi-level sentiment analysis of Turkish such as aspect, sentence, and document levels. They also considered some linguistic rules such as intensification and conjunction in Turkish sentiment analysis. The data used for evaluation are Turkish movie reviews. The resulted accuracies range from sixty percent to seventy-nine percent in binary and in ternary classification tasks at different levels of analysis.

According to Pai and Chu [51] study, semantics of electronic word of mouth (eWOM) content is used to test eWOM content analysis. Both of positive and negative

evaluations were extracted and their customers were helped in decision-making. Their proposed approach can be used as assisting tool for companies to understand the evaluations of products and services, then converting these sentiments into business intelligence to be applied in the improvement of service or product. The data set used is reviews of Taiwanese Fast food. Their outcomes presented that their approach is effective in producing eWOM evaluations related to products and services.

Yan and Bing [52] used graph-based Approach for Chinese Language. They incorporated the inside sentence feature (intra-document evidence) and outside sentence feature (inter-document evidence) by presenting a propagation approach. They suggested that there is a need to more than features inside the sentence to be classified. Camera reviews have been used as a data set in their experiments, and their method was compared to NB and SVM as supervised approaches and also compared to unsupervised approach. The results of their proposed approach outperforms the results of both supervised and unsupervised approaches in case of not using outside sentence features.

The work presented by Maks and Vossen [53] for Dutch Language is based on using semantic approach to build lexicon model that describes adjectives, nouns and verbs to be used in SA task. The subjectivity relations between the actors in a sentence that express about attitudes for each actor was illustrated in their model. These subjectivity relations among the actors are labeled with information related to both the orientation (positive vs. negative) of the attitude and the identity of the attitude holder. A categorization into semantic categories related to SA was included in their model. It also provided the polarity of the attitude, means for the identification of the attitude holder and the sentiments discription of the different actors contained in the text. Dutch WordNet was used in their work. From their study results, it is concluded that the subjectivity of the speaker can be identified and also the subjectivity of the actor sometimes can be identified.

2.4 Literature Discussion

As can be seen we mostly summarize the studies that are related with Arabic Language and when we compare our approach with these studies, we are contributing them by proposing a general formula to aggregate the posterior polarities of the terms and assigning one polarity for one Arabic term. Additionally, we clearly define employed and proposed vectors while usage of ML methods. Finally, we propose the clear way that is interpretable and repeatable to find out sentiments of documents in Arabic Language.

Moreover, we discuss the main entities of recent SA studies including SA levels, SA algorithmic approaches and text domains of SA experiments. Firstly, there are two main SA levels: term (word) level and document level. Term level SA aims to classify the word to positive, negative or neutral with considering some aspects of the word such as score, location in the document, part of speech, negation and intensification tools. On the other hand of document level, sentences are just short documents, whereas, there is no difference between document and sentence level of SA. However, document level aims to divide the whole document to subjective sentences or subjective words, and then classify the opinion or sentiment of the document to positive or negative. It is obviously that term level of sentiment analysis can be considered as the basis of document level in sentiment analysis task. However, our study characterized by calculating the prior polarity for each sentiment term. Secondly, there are three algorithmic approaches in SA, which are ML approach, LB approach and hybrid approach. The ML approach can be used with many classifiers such as ANN, SVM, D-Tree, NB or ME.

However, these classifiers can be applied on aggregation features or linguistic features of targeted text. The LB approach depends on sentiment word list (i.e. sentiment lexicon), which may include a collection of sentiment words with their posterior scores. Both ML and LB approaches can be used together in one approach which call hybrid approach to enhance sentiment results of each other. However, these approaches can use aggregated features, linguistic features or standard bag of features (BOW) from the targeted text. However, in literature part of our study, we found that the frequent usage of the three approaches by researchers during last

years is the same frequent of which was presented in Fig. 2, that we borrowed from Medhat et al. [1]. Whereas, the researchers are using the LB approach more than using the ML approaches. However, the hybrid approach is little used.

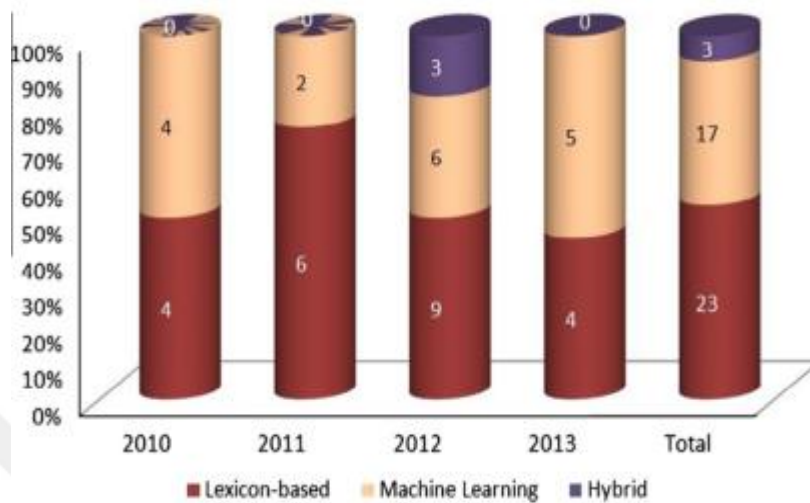


Figure 2 Statistics of articles with respect to the approach over years, Medhat et al. [1].

Thirdly, the mostly used data by researchers in SA task are reviews of products or reviews of movies, whereas, the rest of data such as social media, web Blogs and news texts are frequently used by researchers over recent years. It is noted that most of the studies did not measure the quality of related data sets, in which we did in this study. Finally, most of SA researchers have used evaluation measures including precision, recall, F-Score or Accuracy to present their experiment results. Otherwise, some of them used both of Accuracy and F-Score metrics to satisfy their results.

3. Background

Sentiment analysis can be considered as a classification task which developed by many researchers in recent years to detect sentiments related to a given text of web page. ML and LB approaches are two main directions of SA focused by researchers. The ML approach applies different classifiers and uses linguistic features [1], while, the LB approach depends on known sentiment terms or their roots. The relation between ML, LB approaches and SA is summarized in Fig. 3, whereas the detail methodology of both approaches has been discussed in next chapter.

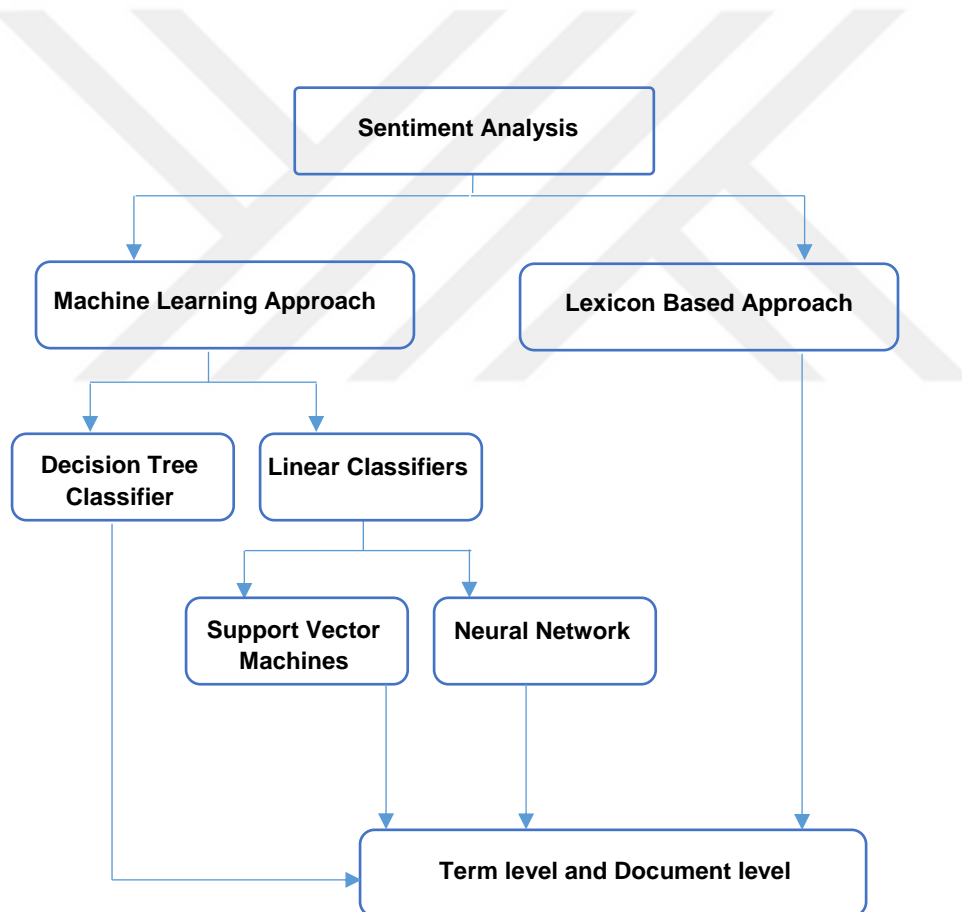


Figure 3 Applied Sentiment Analysis approaches in this study

3.1 Machine Learning Approach

If we have unlabeled data that were needed to be classified, we should build classification model for classification problem, trained with a set of labeled data and their target classes. Two types of classification problems which are binary classification and multiclass classification. If we have two target classes that means binary classification problem, for example, positive or negative target class for text document. Else, if there are more than two target classes that mean multi-class classification. This model can be trained with target vectors; each vector represents a set of features were selected with considering the target class.

In text classification, each vector of features represents document or sentence and includes extracted features values from these documents. For example, the classification model can classify a document to positive if the result value of its included function greater than or equal zero, otherwise the document can be classified as negative. In this section, we theoretically talk about SVM, ANN, and D-Tree, which have been used in our study.

3.1.1 Support Vector Machine Classifier (SVM)

SVM is a supervised learning algorithm with many popular specialties that motivated many researchers to use it. Moreover, it is widely applied in text classification and SA problems due to its superiority over other classifiers [12]. The main idea of SVM is based on the best separation of different classes by using determined linear separator. According to Joachims [40], text data can be classified by using SVM because of the sparse nature of the text, however, some of the text features can be irrelevant but they tend to be organized into categories that can be linearly separated. In SA field, Li and Li [11] used SVM as a sentiment polarity classifier. In this study, we used LIBSVM software package [47] with its default parameters values as implemented in the Matlab software. Moreover, the usual nonlinear kernel Radial Basis Function (RBF) which is used by Moraes et al. [12], is also used to train all the SVM models, as it has better performance than other kernel functions performance in the experiments of our study.

3.1.2 Artificial Neural Network Classifier (ANN)

ANN includes a collection of artificial neurons in which the neuron is the basic unit of ANN. There are input and output for ANN, the inputs are represented by the vector \bar{A}_i . Each neuron has a set of weights that are represented by W . The ANN can be represented by the linear function: $P_i = W * \bar{A}_i$. In binary classification, the sign of predicted function (P_i) yields B_i , which represents the class label of \bar{A}_i .

According to [1], multilayer ANN is more complex and the training process is difficult, because, in back propagation process, errors back-propagated through different layers, though, multilayer ANN is still used for nonlinear boundaries. For the ANN classifier applied in this study, traditional feedforward network with single hidden layer that includes 15 neurons was applied. The training process for each model is repeated more than three times to avoid the problem of convergence to a satisfactory solution [46]. Additionally, gradient descent with momentum and adaptive learning rate back-propagation (traingdx) has been applied as implemented in Matlab software.

3.1.3 Decision Tree Classifier (D-Tree)

D-Tree classifier has a hierarchical structure that includes attributes or nodes which represent training data space, each condition on the attribute value is used to partition the nodes or the data [41]. In text classification field, the condition can be about the absence or the presence of the word. The division of nodes is continued in recursive fashion until certain minimum numbers of records contained in leaf nodes are found. These records can be used for the classification purpose. However, there are D-Tree implementations for text classification task, such as J48 implementation in the Weka Data Mining tool, which is based on the C4.5 D-tree algorithm. Jia et al. [9] built a system at the document level in which the document is classified to be opinionated in the case of finding one sentiment sentence in it. Unigrams and bigrams features were chosen using Chi-square test at the sentence level. Then a D-Tree classifier was used to determine the polarity of opinionated documents. In this study, the default parameters values of D-tree were used as

implemented in the Matlab software. For example, the confidence factor value is 0.25, and the value of the minimum number of instances for the leaf is 2.

3.2 Lexicon Based Approach

The LB approach is based on sentiment lexicon (i.e. list of words) and it is unsupervised approach. There are three ways of creating sentiment lexicon: automatically, semi-automatically or manually whereas each word is represented by numbers indicates its polarity. In order to compute overall document polarity there are two ways: The first way is deriving one score or prior polarity from posterior polarities of each word in the lexicon, then aggregating all prior polarities of the document to indicate its polarity.

The second way is based on the first way by computing the prior polarity for each word in the document, then computing the polarity of each sentence in the document and aggregate the sentences scores to get the polarity of overall document. However, the LB approach applicable for all domains, for this reason, it is domain independent. Section 3.2.1 illustrates two methods for computing LB, one for term level and the other one for document level, while in section 2.3.2, some methods were used in constructing lexicon will be described.

3.2.1 Lexicon Based Computing

LB approach includes two levels of sentiment computing which are term and document levels, for term level sentiment computing, the main strategies for computing prior polarities are based on having lemma with n scores, every formula f is independently applied to all positive scores and negative scores of each sentiment word. This produces two scores, F (positive score) and F (negative score) for each lemma. More details can be found in [16] [17].

Guerini et al. [17] illustrated that obtaining a unique prior polarity from the resulted two scores F (positive score) and F (negative score) for each lemma can be mapped according to different strategies as presented in 3.1 and 3.2 formulae:

$$F_m = \text{MAX} (|\text{PosScore}|, |\text{NegScore}|) \quad (3.1)$$

Where F_m takes the maximum of the two scores.

$$F_d = |\text{PosScore}| - |\text{NegScore}| \quad (3.2)$$

Where F_d takes the difference of the two scores.

It is noticeable that F (negative score) is always positive by construction. In order to obtain final unique prior polarity for a word, the negative sign is added.

Regarding the document level sentiment computing, Turney [14] determined the polarity of the document by applying the average semantic orientation of various phrases in document. However, the semantic orientation linked to every phrase is computed by subtracting similarity of the given phrase to negative reference word “bad” from similarity of the given phrase to positive reference term “nice”:

$$\text{SO (phrase)} = \text{PMI (phrase, “nice”)} - \text{PMI (phrase, “bad”)} \quad (3.3)$$

Whereas, if the semantic orientation of a phrase has strong association with the word “nice” it will be positive, else if the semantic orientation of a phrase has strong association with the word “bad” it will be negative.

The similarity between the pairs of terms or phrases was measured by using the similarity score measure, Pointwise-mutual Information (PMI). The formula of calculating PMI is:

$$\text{PMI}(\text{word}_1, \text{word}_2) = \log_2 \left(\frac{P(\text{word}_1 \& \text{word}_2)}{P(\text{word}_1)P(\text{word}_2)} \right) \quad (3.4)$$

Where $P(\text{word}_1 \& \text{word}_2)$ represents the probability of founding two words together. The ratio between $P(\text{word}_1 \& \text{word}_2)$ and $P(\text{word}_1) p(\text{word}_2)$ is used to measure the degree of statistical dependence between the two terms. The amount of information gained about the presence of one of terms when the other term is detected is given by the log of this ratio.

3.2.2 Lexicon Building

It is known that building lexicons or dictionaries for Arabic SA task is very limited, however, few researchers tried to construct Arabic sentiment lexicons. For example, Badaro et al. [27] created a large sentiment lexicon for Arabic (ArSenL) by using Arabic and English resources: (i) English SentiWordNet (ESWN) [18]; (ii) English WordNet (EWN) [36]; (iii) Arabic WordNet (AWN) [37]; (iv) Standard Arabic Morphological Analyzer (SAMA) [38]. The English synsets were connected to the lemma entries in the Arabic resources.

ArSenL provides number of positive and negative scores for each word. Their results show that ArSenL lexicon outperforms all lexicons in all measures without exceptions. Another method used for building Arabic sentiment lexicon was by Farra et al. [25], they used a list of some Arabic words' roots which are selected by using stemmer application. To classify each word in the sentence, they were checked against the list of words' roots. If the root is found in the list, its polarity is selected as positive, negative, or neutral. Else, the user will add the root of the absent word to the list of learned roots. This study also tried benefits from the advantages of both LB and ML approaches by combining them together (i.e. hybrid approach). They used an approach that combining both semantic and syntactic features to classify the sentiment of Arabic document level. They used many features such as frequency of negative, positive and neutral terms in every document using the semantic list (dictionary) they have constructed, frequency of contradiction words, frequency of negation tools as "لا، لن، لم، ليس" that are negation tools in Arabic that have the same usage of "not" word in English and other features also were used.

Lexical resources for SA obtainable in English like SentiWordNet was also used in the approach of Arabic lexicon building, in which lexical data base of English Language [42] was used to build SentiWordNet. Each word score in SentiWordNet was defined using two step processes. In the first process binary search was used to find words with similar polarities of English WordNet [36]. Then, in the second process an iteration was run on the words to find final polarity of words. According to each word polarity each word is classified as positive, negative or neutral in which each word intensity ranged from 0 to 1 and associated with its part of speech [18].

To get benefit from this lexicon in Arabic Language, the Arabic documents can be translated into English Language using standard translation software as noted in [24]. Then these documents can be classified basing on its sentiment scores into positive or negative. Whereas the Arabic word scores are gained from its translated English words in SentiWordNet.

3.3 Arabic Language Overview

Arabic Language is one of the first ranked 10 languages primarily used on the Internet based on the statistic has been done by the Internet World Stats [49]. It is spoken by millions of people. Making it one of the five most spoken languages in the world, and it is one of the formal languages in United Nations organization [54]. The Arabic language was spoken in Hejaz and surrounding areas before the Islamic age, and because it is the Quran (The Holy Islamic book) language which became the official language in the Islamic regions, and many Muslims from different cultures learned this language in order to understand Islamic religion deeply. The direction of writing is from right to left, with alphabet consists of 28 letters. The alphabet of Arabic can be expanded to 90 elements by writing additional vowels, marks, and shapes [50].

There are two forms of Arabic Language: (i) Modern Standard Arabic (MSA). (ii) Dialectal Arabic. In the Arabic countries, Modern Standard Arabic (MSA) is derived from the language of the Quran. In addition, commonly used in books, newspapers, media, formal speeches, movie reviews, etc. MSA hugely follows the grammatical rules of Quranic Arabic and uses a lot of its vocabulary. It has ignored some grammatical constructions and vocabulary that no longer have any importance in the spoken varieties, and has edited new constructions and vocabulary from the spoken varieties [54]. Dialectal Arabic includes all forms of day life currently spoken Arabic and it is derived from the Standard Arabic [23].

Arabic Language is one of the Semitic languages in which the morphology is uncommon and complex (i.e. approach of building terms from a basic root). Arabic Language has a non-concatenative "root-and-pattern" morphology: A root includes of a set of bare consonants (usually three).

In Arabic Language different patterns can be used to form numerous words. For example, from a single root **k-t-b** various words can be formatted: **katabtu** 'I wrote', **aktabtu** 'I dictated', **aktubu** 'I write', **kutiba** 'it was written'...etc.

This study is concerned with MSA. The challenge here is that all approaches of natural language processing that have been applied to most languages are not directly applied to Arabic Language. The text needs additional pre-processing before applying some methods.



4. Sentiment Analysis in Arabic Language

The aim of this study is to enhance Arabic SA by applying both ML and LB approaches on both term and document levels independently and nominating the best SA approach for Arabic. The experiments include implementing rules with both of ML approach and LB approach to identify their effectiveness on both approaches results.

This unit is ordered as follows: section 4.1 presents the methodology of the study including:

- (A) Preprocessing task of OCA corpus, then measuring the corpus quality by comparing it with the other Arabic corpora by using the same quality measures.
- (B) Stemming and identifying each term sentiment scores placed in OCA.
- (C) Computing one sentiment score (prior polarity) for each term.
- (D) Overlooking rules (i.e. negation and intensification).
- (E) Computing sentence score formulae.

Section 4.2 discuss features extraction for term and document levels to be used with ML approach, with applying and without applying rules on the features. Classification task is also illustrated.

In section 4.3, applying ML and LB approaches on both term and document levels, with applying and without applying rules in each approach are discussed.

Section 4.4 summarizes the measures used to evaluate both ML and LB approaches results. Finally, in section 4.5, all experiments and results have been discussed, compared and evaluated.

4.1 Study Methodology

In our dissertation, we are concerned in sentiment analysis at multi-level (i.e. term level and document level) for Arabic Language. The proposed methodology of this

study in case of applying ML and LB approaches has been presented in Fig. 4 and Fig. 5 respectively. However, the methodology steps are:

- (A) Preprocessing of the OCA corpus to remove unnecessary terms then the corpus quality has been measured.
- (B) Computing one sentiment score for each sentiment term of OCA corpus.
- (C) Features have been selected for both term and document levels independently.
- (D) Applying machine (ML) learning approach on the term level and in the document level of OCA corpus with considering applying and not applying rules on both levels.
- (E) Applying LB approach on the document level with considering applying rules to each term in the document.
- (F) Comparisons between the results have been done to identify the best way to classify sentiment Arabic reviews.

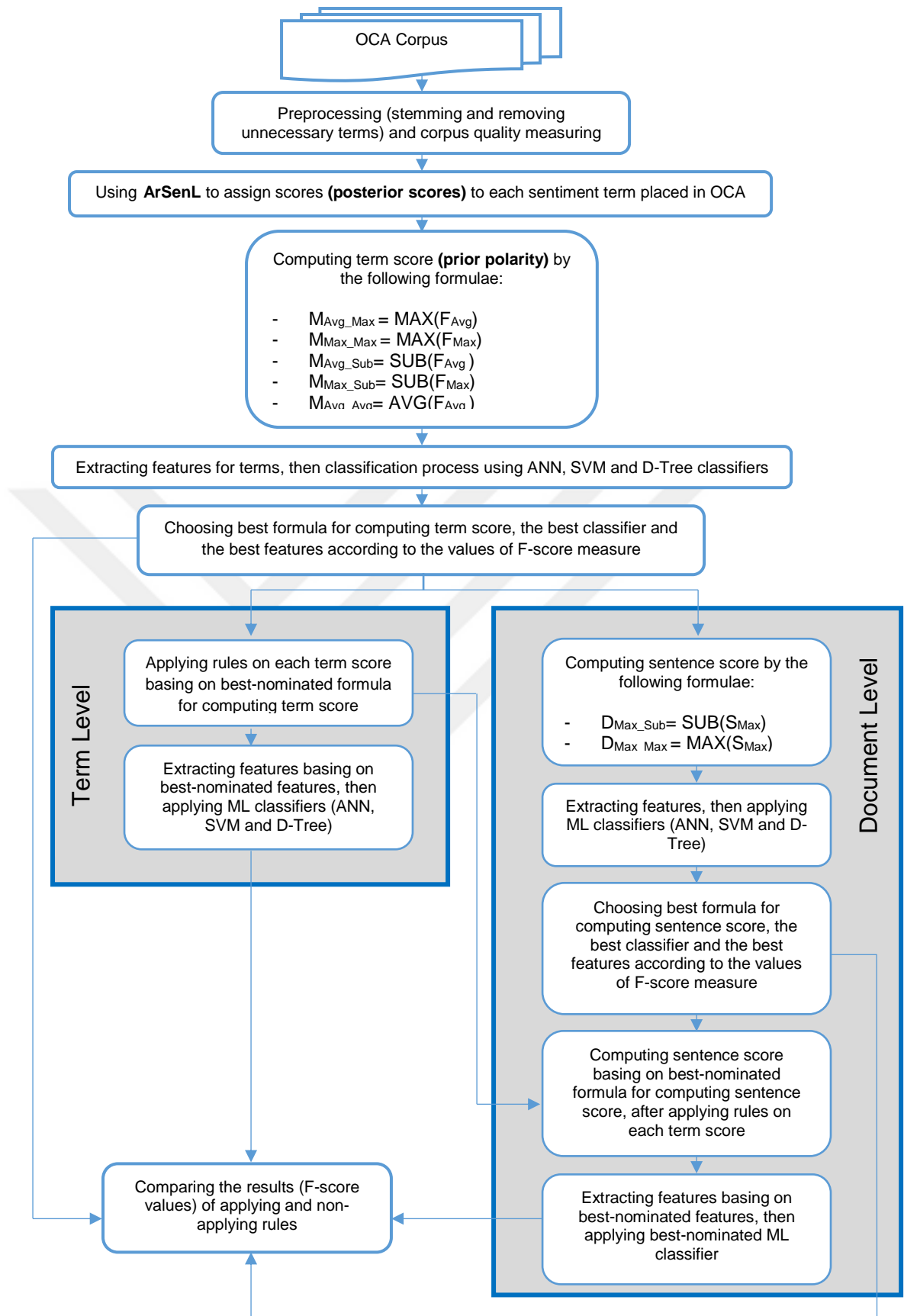


Figure 4 Proposed methodology of multi-level SA in Arabic Language using ML approach

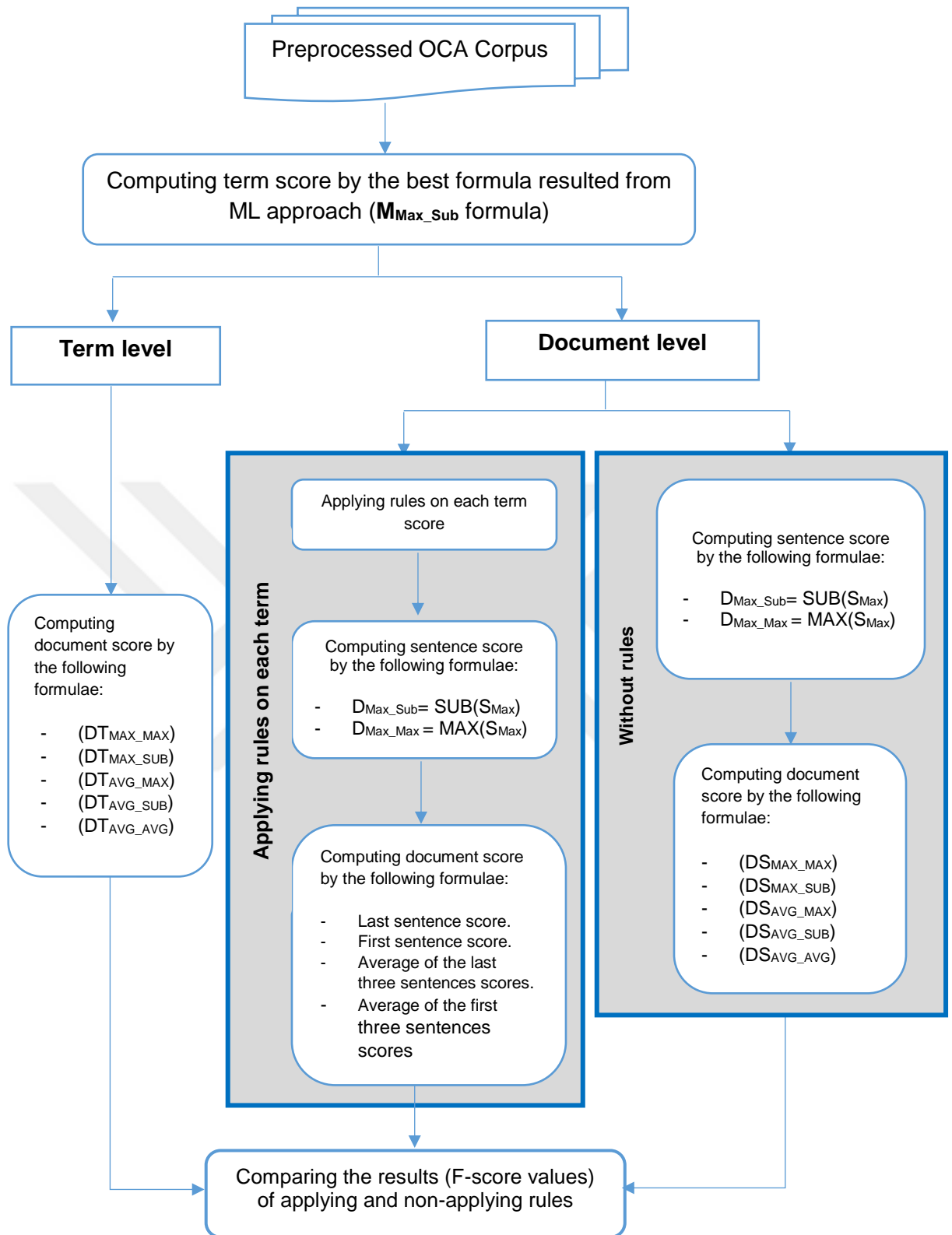


Figure 5 Proposed methodology of multi-level SA in Arabic Language using LB approach

4.1.1. Preprocessing and Quality Measuring of OCA Corpus

OCA corpus is a new Arabic resource that is made obtainable to the scientific purposes to be used in SA task [23]. The corpus contains 500 film reviews collected from different Arabic blogs and web pages, 250 of the reviews are classified as positive, and the rest are classified as negative. Two samples of positive and negative reviews of the OCA corpus are given in Table 1 to be more clear. Preprocessing for OCA dataset is made by removing useless terms (i.e. spaces, punctuation marks, numbers, etc.) in SA task and Table 2 presents statistical definition of the OCA corpus after preprocessing.

Table 1. Samples of positive and negative OCA reviews

<p>Positive review</p>	<p style="text-align: right;">In Her Shoes</p> <p>نعم إنه فيلم "الحسناوات" ولكن موضوعه ذكي ، كما أنه مؤثر وسيثير لديك رغبة شديدة في البكاء ، إلى جانب أنه مرح وعلى درجة عالية من الإخراج والتمثيل. إنه يستحق ثمن التذكرة والساعتين اللتين ستضيعهما من وقتك الثمين على مشاهدته. التقييم العام : 3</p> <p>In Her Shoes Yes, it is the film "Belles" But the theme of intelligent, influential as it is, and will raise you have a strong desire to cry, to the side because it's funny and a high degree of output and representation. It is worth the price of the ticket and two hours of your precious time to watch it. User Rating: 3</p>
<p>Negative review</p>	<p>اسم الفيلم: 7 ورقات كوتشينة إسم الكاتب: د.صلاح الغريب من وجهة نظري الشخصية الفيلم يعتبر ضعيف جداً جداً من جميع النواحي ... ولم يلفت نظري سوى الأداء الجيد للفنان (محمد سليمان) و الفنانة (رانيا شاهين) . التقييم: 10/1</p> <p>Movie Name: 7 papers Kuchinp Author Name: Dr. Salah strange from the point of my personal film is very, very weak in all respects ... Did not draw my eyes only good performance of the artist (Mohammad Suleiman) and the artist (Rania Shaheen). Rating: 1 / 10</p>

Table 2. Statistical details of OCA after pre-processing

	Positive	Negative
Total No. of documents	250	250
Total No. of tokens (words)	77704	60748
Total No. of unique tokens in OCA	5984	5812

First of all, we aimed to measure OCA corpus quality before its usage in the task of sentiment analysis, thus, we used Zipf's law to measure the quality of OCA corpus as it is presented in the previous studies by Rosso and Benajiba [43] and Salhi and Yahya [44]. Zipf's law states that: "in a corpus, the frequency of a word is inversely proportional to its rank"; Zipf's law can be expressed by Eq.4.1

$$\mathbf{F} = \mathbf{C}/\mathbf{r}^{\mathbf{a}} \quad (4.1)$$

Where **F** is the frequency of words, **a** is a constant close to 1, **r** is the rank of the word and **C** is the highest observed frequency.

According to this approach, the most frequent word takes place twice as often as the second most frequent word, and the third most frequent word occurs one-third of the most frequent word and so on.

In our experiments, a relationship has been drawn between actual frequency and the rank using the logarithmic measure as presented in Fig. 6a, and the relationship in Fig. 6b, is also drawn between ideal (Zipf's) frequency and the rank using logarithmic measure. When a comparison was made between actual and ideal frequencies as seen in Fig. 7, it is concluded that few words are often used and many of the words are rarely used.

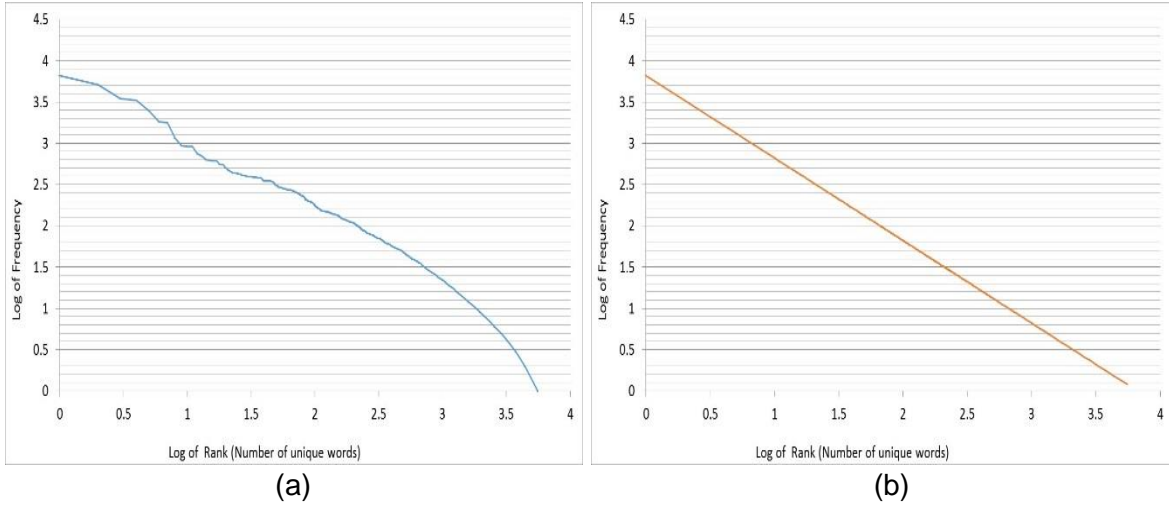


Figure 6 (a) OCA frequencies VS Rank, (b) Ideal (Zipf's) Frequency VS Rank

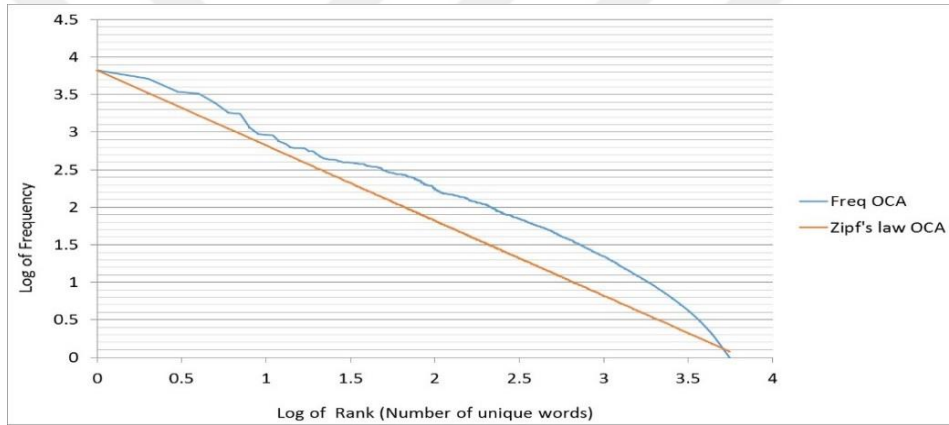


Figure 7 Comparison between Actual and Ideal frequencies VS Rank

To assess the adherence of word frequencies (i.e. the word frequency distribution of text to Zipf's law distribution), we used Kullback-Liebler distance measure (D_{KL}), this distance is asymmetric and measures the distance from a “true” probability distribution \mathbf{P} (Zipf's law distribution based on rank) to an arbitrary probability distribution \mathbf{Q} (word frequency distribution) as discussed by Rosso and Benajiba [43]. Eq.4.2 can express this function:

$$D_{KL}(\mathbf{P}, \mathbf{Q}) = \sum_i P(i) \cdot \log P(i) / Q(i) \quad (4.2)$$

Where \mathbf{P} is Zipf's law distribution and \mathbf{Q} is word frequency distribution. For OCA corpus the D_{KL} obtained has been compared to the D_{KL} resulted from Rosso and Benajiba [43] study as presented in Table 3.

Table 3. Comparison between D_{KL} for OCA and D_{KL} for Arabic corpora by Rosso and Benajiba [43]

Corpus	Kullback-Liebler Distance
OCA	18980.41
Abu-Taïb AlMoutanabbi poetry	22120.32
Newspaper articles	32836.98
A Linux Red Hat installation tutorial book	41983.44
Religious book of the Imam Ibnu Qayyim El Jawziyah	28870.38

Furthermore, it is also compared to Salhi and Yahya [44] study as presented in Table 4. According to the comparisons between our study result and the results of previous studies regarding D_{KL} , we concluded that the OCA corpus has the best D_{KL} . Thus, OCA data set quality outperforms other Arabic data sets quality. For this reason, we can dependably use it in SA task.

Table 4. Comparison between D_{KL} for OCA and D_{KL} for Arabic corpora by Salhi and Yahya [44]

Corpus	Kullback-Liebler Distance
OCA	0.037424
Al-Quds	0.084
Ar. Wikipedia	0.100
Computing	0.143
Economics	0.134
History	0.134
Literatures	0.132
Physics Related	0.139
Medicine Related	0.136
Politics	0.128
Religions	0.139
Sports	0.106

4.1.2 Stemming Process and Identifying Sentiment Terms

After the preprocessing phase of OCA corpus, the opinion terms from the corpus are decided through stemming all terms placed in OCA by using the commonly used Standard Arabic Morphological Analyzer (SAMA) [38]. This analyzer can produce all possible reading out of context for a given word, whereas each word lemma has English gloss and part of speech tag. After stemming process, sentiment scores for each term root placed in OCA is given by comparing them to large-scale Arabic sentiment analysis lexicon (ArSenL). ArSenL has been created by [27]. ArSenL is a big sentiment lexicon for the Arabic Language in which each word seed (lemma) has more than one score (or posterior polarities). The English synsets were connected to the lemma entries in the Arabic resources. Noting that, each word root (lemma) in the lexicon have many senses or scores which are rated in the interval of [-1..1]. For example, Table 5 presents total five ArSenL positive and negative scores for the word: "hot" (ساخن).

Table 5. Five ArSenL positive and negative scores for the term: "hot" (ساخن)

Lemma by SAMA	Positive scores	Negative scores
sAxin	0.375	0.25
sAxin	0.75	0.125
sAxin	0.5	0.375
sAxin	0.25	0.25
sAxin	0.125	0

In our work, we utilized ArSenL to derive one numerical score from positive scores and one numerical score from negative scores. Then, we formulated aggregation formulae to derive one score (or prior polarity) from the posterior scores for each term placed in OCA as illustrated in the next section 4.1.3.

4.1.3 Prior Polarity Computing Formulae for Term Level

In this phase, one prior sentiment score is extracted from posterior positive or negative scores as shown in Fig 4. However, total 5 formulae (or equations) are constructed to derive one prior sentiment score from posterior scores in two steps. In the first step, two scores which are F (Positive Score) and F (Negative Score) are derived from posterior scores for each word lemma using Eq.4.3 and Eq.4.4 in Table 6. Secondly, we derived unique prior polarity score from the resulted two scores of F_{Avg} and F_{Max} separately by using the Equations 4.5 – 4.9 in Table 6.

Table 6 Prior polarity computing equations

The equation	Discription	Eq.#
$F_{Avg} = (Avg (Positive\ Score), Avg (Negative\ Score))$	F_{Avg} takes the average of the absolute value of positive and negative scores independently for each term.	(4.3)
$F_{Max} = (Max (Positive\ Score), Max (Negative\ Score))$	F_{Max} takes the maximum of the absolute value of positive and negative scores independently for each term.	(4.4)
$M_{Avg_Max} = MAX(F_{Avg})$	M_{Avg_Max} takes the maximum score of the resultant two scores from F_{Avg} .	(4.5)
$M_{Max_Max} = MAX(F_{Max})$	M_{Max_Max} takes the maximum score of the resultant two scores from F_{Max} .	(4.6)
$M_{Avg_Sub} = SUB(F_{Avg})$	M_{Avg_Sub} takes the difference of the resultant two scores from F_{Avg} . That is, the subtraction of the negative score from the positive score.	(4.7)
$M_{Max_Sub} = SUB(F_{Max})$	M_{Max_Sub} takes the difference of the resultant two scores from F_{Max} . That is the subtraction of the negative score from the positive score.	(4.8)
$M_{Avg_Avg} = AVG(F_{Avg})$	M_{Avg_Avg} takes the average of the resultant two scores from F_{Avg} .	(4.9)

In Eq.4.9, the negative sign is added to the negative score before taking the average. In addition, in the Eq.4.5 and Eq.4.6, it is important to note that, a negative sign is added to the final unique prior score if it was negative before taking the absolute value. To nominate the best of the five equations (or formulae) of deriving prior polarity, five OCA files based on the five formulae (M_{Avg_Max} , M_{Max_Max} , M_{Avg_Sub} , M_{Max_Sub} and M_{Avg_Avg}) were created, after that, we tested each OCA file independently using ML methods, that is, the ANN, D-Tree and SVM classifiers. Furthermore, the M_{Max_Sub} formula has the best results amongst all other formulae, in which the average F-score achieved in the term level for both positive and negative testing classes is 0.92. Thus, the M_{Max_Sub} formula is certified for giving one score for each term placed in OCA corpus.

4.1.4 Rules

Rules designed for Arabic Language include two main situations that may change the meaning of a word or a sentence; these are “intensification” and “negation” tools. For example, the intensification tools in Arabic including the words “إفراط، جداً، كبيراً،” “مطلق” can have the same usage of the word “very” in the English language. On the other hand, the negation tools “ليس، لم، لن، لا” in Arabic have the same usage of the word “not” in the English language. In our study, it is decided that for each term: (i) in case of positive terms, if the term has intensification word before or after it, its score will be increased to 1, while in case of negative terms, the term score is decreased to -1. (ii) If the term has negation word before it, its score will be negated, in which the positive term becomes a negative term and the negative term becomes a positive term.

4.1.5 Sentence Score Computing Formulae

One score for each sentence in each of the OCA documents has been computed considering that M_{Max_Sub} formula has been certified for giving one score for each term of OCA corpus. However, two versions of the corpus based on the sentence score have been made. Whereas, in the first version each sentence score is computed applying rules, and the second version was computed without applying rules to make comparisons and to identify the efficiency of the rules on the experiments result. We first of all derived two scores S (Positive term score) and S (Negative term score) using Eq. 4.10 in Table 7. Secondly, one sentence score is derived from the result of two scores of S_{Max} formula by using Eq. 4.11 and Eq. 4.12 in Table 7. To nominate the best of the two equations (Eq. 4.11 and eq. 4.12) of deriving one sentence score, we tested each OCA file independently using ML methods, that is, the ANN, D-Tree and SVM classifiers, therefore, the D_{Max_Max} formula has results better than the results of D_{Max_Sub} formulae.

Table 7 Sentence score computing equations

The equation	Discription	Eq.#
$S_{Max} = (\text{Max} (\text{Positive term Score}), \text{Max} (\text{Negative term Score}))$	S_{Max} takes the maximum absolute value of the positive term score in the sentence and the maximum absolute value of the negative term score in the sentence.	(4.10)
$D_{Max_Sub} = \text{SUB}(S_{Max})$	D_{Max_Sub} takes the difference of the resultant two scores from S_{Max} . That is the subtraction of the negative score from the positive score.	(4.11)
$D_{Max_Max} = \text{MAX}(S_{Max})$	D_{Max_Max} takes the maximum score of the resultant two scores from S_{Max} , and the negative sign is added to the final sentence score in case of the negative term is the maximum.	(4.12)

4.2 Arabic Sentiment Analysis Using Machine Learning (ML) Approach

Features were selected for both term and document levels independently, considering the application and non-application of rules on both levels. Afterward, we applied ML approach on the term and the document level of the OCA corpus. Finally, we used three encouraging ML classifiers from SA literature, which are SVM, ANN and D-Tree to classify OCA documents.

4.2.1 Features Extraction

Basing on the M_{Max_Sub} formula for computing prior polarity for each sentiment term in OCA, features have been extracted in two levels of SA, the term level and the document level. Whereas, we made two versions of features for each of term level and document level, however, the first version includes all features have been created with applying intensification and negations tools (i.e. rules) on each term placed in OCA.

4.2.1.1 Term Level Features

Depending on the former section 4.1.3, five OCA files have been built. Each file words assigned prior score according to the M_{Max_Sub} formula.

Total eight features have been selected for each OCA file independently. However, each document in a file has been represented by a vector of total eight features, and the total eight features are ordered as: *count of positive scores, count of negative scores, summation of positive scores, summation of negative scores, average of positive scores, average of negative scores, first subjective score and last subjective score*. For example, Table 8 is a sample for one of the OCA files including the total eight features. To ensure the effectiveness of our features, we omitted the last two features (i.e. *first subjective score and last subjective score*), thus we have total six features. Then, the same experiments have been repeated for them, whereas comparisons between the testing results of both total six features and total eight features also have been illustrated in section 4.5.1.

Table 8 Sample of one of the OCA files including the total eight features in term level

Class	CountPos	CountNeg	SumPos	SumNeg	AvgPos	AvgNeg	F_Subj	L_Subj
0	166	141	20.19	-18.76	0.12	-0.13	0.04	0.21
0	197	147	25.51	-21.94	0.13	-0.15	0.04	-0.25
0	110	66	16.24	-11.89	0.15	-0.18	-0.08	0.03
0	19	21	2.61	-4.04	0.14	-0.19	-0.56	0.04
1	61	21	5.56	-3.70	0.09	-0.18	0.04	0.02
1	37	32	5.16	-6.00	0.14	-0.19	0.04	0.25
1	45	27	8.29	-3.70	0.18	-0.14	0.04	0.26

4.2.1.2 Document Level Features

With respect to the former section 4.1.5, we built two OCA files, each file sentences assigned one score according to the D_{Max_Max} and D_{Max_Sub} formulae.

Total seven features have been selected for each OCA file independently, the total seven features are ordered as: *count of positive sentences, count of negative sentences, maximum score of positive sentences, maximum score of negative sentences, first sentence score, middle sentence score and last sentence score*. For example, Table 9 is a sample for one of the two OCA files including the total seven features. To guarantee the validation of the features, we firstly omitted the third and

the fourth features (i.e. *maximum score of positive sentences* and *maximum score of negative sentences*), thus we have total five features. Then, the same experiments have been repeated for them. Secondly, we omitted the last three features (i.e. *first sentence score*, *middle sentence score* and *last sentence score*), thus we have total four features. Then, the same experiments also have been repeated for them, whereas comparisons between the testing results of the total seven features, the total five features and the total four features are illustrated in section 4.5.2

Table 9 Sample of one of the OCA files including the total seven features in the sentence level

Class	CountPos	CountNeg	MaxPos	MaxNeg	FirstScore	MiddleScore	LastScore
0	8	2	0.75	-0.75	-0.25	-0.75	0.125
0	38	22	1	-0.875	0.125	0	0.125
0	6	1	0.875	-0.5	0.5	0.75	0.125
1	1	0	0.75	0	0.75	0.75	0.75
1	1	0	0.75	0	0.75	0.75	0.75
1	1	0	0.875	0	0.875	0.875	0.875
1	1	0	0.75	0	0.75	0.75	0.75

4.2.2 Classification Task

In the classification task, we put the features vector into a format that can be processable by each of SVM, ANN and D-Tree classifiers. According to this study experiments, each classifier settings are presented in the previous Sections 3.1.1 – 3.1.3 respectively. Some classifiers need parameter settings, which are still a research issue [12]. The aim of using three deferent classifiers is to prove that decreasing the number of features decreases the value of classification results. We applied 5-fold cross validation for testing and validation purpose to ensure that every sample of OCA corpus has the same chance of appearing in training and validation set [48].

However, the preprocessed OCA corpus is divided into 5 testing files in which each file includes 100 documents in which 50 documents are positive and 50 are negative. The rest corpus is divided into 5 training files, so that each file contains 400 documents (200 positive and 200 negative documents).

4.3 Arabic Sentiment Analysis using Lexicon Based (LB) Approach

In our work, the LB methodology is based on the (M_{Max_Sub}) formula to get one score for each term, however, we applied LB approach on both of term level and document level as illustrated in the following two subsections. However, Fig. 5 describes the methodology of the LB approach in this study in the case of application and non-application of rules.

4.3.1 Lexicon Based Approach without Rules

We firstly used LB approach in the term level, however, we applied five aggregation formulae for each OCA document. The aggregation formulae and their descriptions have been presented in the Table 10.

Secondly, in the document level we also used LB approach, whereas, each sentence score has been obtained by one of two formulae: D_{MAX_MAX} and D_{MAX_SUB} formulae which have been illustrated in the Table 11. Then these formulae have been applied on two versions of the OCA corpus, one version is based on D_{MAX_MAX} formula and the other version is based on D_{MAX_SUB} formula. Each document score is calculated by using different aggregation formulae, which are presented in Table 12.

More details for this section and related experiments including comparing term level resultant scores and sentence (document) level resultant scores with the OCA labels have been discussed in section 4.5.3.

Table 10 Term level based document score computing formulae

The formula	Description	Eq.#
DT_{MAX_MAX}	Takes the maximum positive term score and the absolute value of the maximum negative term score, then the document score takes the maximum score of them. The negative sign is added to the document score if the absolute value of the maximum negative term score is bigger than maximum positive term score.	(4.12)
DT_{MAX_SUB}	Takes the maximum positive term score and the absolute value of the maximum negative term score, then the document score takes the subtraction of the absolute value of the maximum negative term score from the maximum value of the positive term score	(4.13)
DT_{AVG_MAX}	Takes the average of the positive terms scores and the average of the negative terms scores, then the document score will be the absolute value of the maximum of them. Negative sign is added to the document score if (the absolute value for the average of negative terms scores) is bigger than (the average of the positive terms scores).	(4.14)
DT_{AVG_SUB}	Takes the average of the positive terms scores and the average of the negative terms scores, then the document score will be the subtraction of (the absolute value for the average of negative terms scores) from (the average value of positive terms scores).	(4.15)
DT_{AVG_AVG}	Takes the average of the positive terms scores and the average of the negative terms scores, then the document score takes the average of them.	(4.16)

Table 11 Sentence score computing formulae

The formula	Description	Eq.#
D_{MAX_MAX}	Takes the maximum positive term score and the absolute value of the maximum negative term score, then the sentence score takes the maximum score of them with considering the sign of the sentence score in case of the absolute value of the maximum negative term score is bigger than the maximum positive term score.	(4.17)
D_{MAX_SUB}	Takes the maximum positive term score and the absolute value of the maximum negative term score, then the document score takes the subtraction of the absolute value of maximum negative term score from the maximum value of the positive term score.	(4.18)

Table 12 Sentence level based document score computing formulae

The formula	Description	Eq.#
DS_{MAX_MAX}	Takes the maximum positive sentence score and the absolute value of the maximum negative sentence score, then the document score takes the maximum score of them. The negative sign is added to the document score if the absolute value of the maximum negative sentence score is bigger than maximum positive sentence score.	(4.19)
DS_{MAX_SUB}	Takes the maximum positive sentence score and the absolute value of the maximum negative sentence score, then the document score takes the subtraction of the absolute value of the maximum negative sentence score from the maximum value of the positive sentence score	(4.20)
DS_{AVG_MAX}	Takes the average of the positive sentences scores and the average of the negative sentences scores, then the document score will be the absolute value of the maximum of them. Negative sign is added to the document score if (the absolute value for the average of negative sentences scores) is bigger than (the average of the positive sentences scores).	(4.21)
DS_{AVG_SUB}	Takes the average of the positive sentences scores and the average of the negative sentences scores, then the document score will be the subtraction of (the absolute value of the average of negative sentences scores) from (the average value of positive sentences scores).	(4.22)
DS_{AVG_AVG}	Takes the average of the positive sentences scores and the average of the negative sentences scores, then the document score takes the average of them.	(4.23)

4.3.2 Lexicon Based Approach with Using Rules

In this section, intensification and negation methods have been applied on each term as previously presented in Fig. 5. Then we got each sentence score according the D_{MAX_MAX} and D_{MAX_SUB} formulae. Finally, each document score was computed according to four techniques, which are: last sentence score, first sentence score, average of the last three sentences scores and average of the first three sentences scores. All the results of the LB approach with and without using rules have been compared to identify the best LB SA formulae can be used for SA in Arabic

Language, and to illustrate the effectiveness of using rules in both term and document levels of Arabic SA.

4.4 Evaluation Measures

The objective of classification task is to find the best formula for Arabic SA task in the Arabic Language and to prove the efficiency of the features by showing how classification results are decreased when the number of features is decreased with consideration of application and non-application of the rules.

We used three commonly used metrics in the literature by El-Halees [26], Badaro et al. [27] and Al-Azani and El-Alfy [45] which are Precision, Recall, and F-score. The equations representing these metrics are listed in Table 13. The average of F-scores is used in comparing all the results of the classification in our experiments.

Table 13. Evaluation measures

Evaluation measures	The equations	Eq.#
Precision for positive classes	$\text{Precision} = \frac{\text{true_positive}}{\text{true_positive} + \text{false_positive}}$	(4.24)
Precision for negative classes	$\text{Precision} = \frac{\text{true_negative}}{\text{true_negative} + \text{false_negative}}$	(4.25)
Recall for positive classes	$\text{Recall} = \frac{\text{true_positive}}{\text{true_positive} + \text{false_negative}}$	(4.26)
Recall for negative classes	$\text{Recall} = \frac{\text{true_negative}}{\text{true_negative} + \text{false_positive}}$	(4.27)
F-score	$\text{F-score} = \frac{2 * \text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}}$	(4.28)

4.5 Experimental Results and Evaluation

In this study, we focused on sentiment analysis at term level and document level for Arabic Language. However, two main approaches used in the literature have been applied which are ML approach and LB approach.

Moreover, the rules also have been applied to see how they can effect on the classification experiments in both SA levels.

To achieve our goal from this study, our experiments have been done in the following steps: (i) Using Zipf's law for measuring the OCA corpus quality. Then, preprocessing of OCA corpus by omitting unnecessary terms like numbers and punctuations, moreover deciding the opinion terms from the corpus by using the large scale Arabic sentiment analysis lexicon (ArSenL). Whereas, each word root (lemma) in the lexicon have many senses or scores which are rated in the interval of [-1...1]. (ii) Computing one positive or negative score (prior polarity) from many scores (posterior polarities) for each opinion word in OCA corpus by creating different aggregation formulae. (iii) Extracting features based on the term level and features based on the document level with applying and without applying rules in the term level, whereas the document level is based on the term level. (iv) Applying ML approach by using ML classifiers, which are ANN, D-Tree, and SVM. However, the results of the three classifiers are used to: Firstly, prove the efficiency of our features by proving that, decreasing the number of features decreases the classification results. Secondly, to nominate the best formula can be used for computing prior polarity from posterior scores in the term level, and best formula can be used for computing document score in the document level of SA at Arabic Language. (v) Applying LB approach in both term and document levels by using different formulae based on aggregating functions like maximum, average and subtraction. However, the intensification and negations rules have been used in the LB approach experiments. The results of the LB approach have been used to identify the best formulae can be used with term level and document level of LB SA at Arabic Language, and to illustrate the effectiveness of using rules in both term and document levels of Arabic SA. (vi) Comparing the results of each approach and also comparing final results of both approaches by using F-score metric.

This section is ordered as follows: section 4.5.1 discuss computing prior polarity for each term in the OCA corpus by using different features and formulae. However, applying and not applying rules on each term is considered, then using the results of the ML classifiers to nominate the best formula, the best classifier and the best number of features in the term level.

Section 4.5.2 discuss computing each sentence score in each document of the OCA corpus by using different features and formulae with applying and without applying rules for each term, then, using the results of the ML classifiers to nominate the best formula, the best classifier and the best number of features in the document level. In section 4.5.3, firstly, LB approach has been applied on the term level in which rules have not been used, then different aggregation formulae have been used to compute each document sentiment score, and finally, the results of applying each formula have been compared to nominate the best of them. Secondly, LB approach has been applied on the document level, and also without using rules, then different aggregation formulae have been applied to compute document sentiment score, finally, the results of applying each formula have been compared to nominate the best formula. While in section 4.5.4, the LB approach has been applied by using rules for each term. Then each sentence score has been computed by using the D_{MAX_MAX} and D_{MAX_SUB} formulae. Finally, each document score has been computed according to four techniques that have been compared to nominate the most effective of them. In section 4.5.5, the best results summary of both ML approach and LB approach have been discussed.

4.5.1 Experiment A: Term Level Sentiment Analysis using Machine Learning Approach

We applied this experiment firstly without using rules, then with applying them on the term level.

4.5.1.1 Without Rules

- **Objective:** Firstly, we aim at this experiment to nominate the best formula of computing term level score by comparing the results of using ML classifiers (ANN, D-Tree, and SVM) to classify each OCA document with considering the number of used features (8 or 6 features) in each document. However, each OCA document is based on different formula. Secondly, proving the efficiency of our features by decreasing their number.

- **Method:** (i) Each word root (lemma) in the OCA corpus have many senses or scores which are rated in the interval of [-1...1]. At this stage, one positive or

negative score (prior polarity) has been computed from many scores (posterior polarities) for each opinion word in the OCA corpus by using the 5 aggregation formulae that have been illustrated in section 4.1.3. The aggregation formulae are M_{Avg_Max} , M_{Max_Max} , M_{Avg_Max} , M_{Avg_Sub} , M_{Max_Sub} , M_{Avg_Avg} . (ii) Total eight features have been selected for each OCA file independently. The total eight features are ordered as *count of positive scores, count of negative scores, summation of positive scores, summation of negative scores, average of positive scores, average of negative scores, first subjective score and last subjective score*. To ensure the effectiveness of our features and the testing methods, we omitted the last two features (i.e. *first subjective score and last subjective score*), thus we have six features. Now we have 10 OCA files (i.e. 5 files of total 8 features and 5 files of total 6 features) in which each file is based on one of the five formulae. (iii) We applied 5 fold cross validation for training and testing purpose on each of the 10 files. (iv) Then the ML approach (ANN, D-Tree, and SVM classifiers) has been applied on each of the 10 files to nominate the best formula for computing prior polarity for each Arabic word. (v) Precision and recall metrics have been used to calculate F-Score measure, then we calculated the average of F-Score measure for results of five folds of positive training classes, positive testing classes, negative training classes, and negative testing classes to present the numbers shown in the next tables.

- **Results:** The results are presented in Tables 14 -16.

Table 14 Average of performance by using total 8 and 6 features with ANN classifier

	Formulae	Positive classes		Negative classes	
		F-score for training data	F-score for testing data	F-score for training data	F-score for testing data
8 Features	M_{Max_Max}	0.78	0.89	0.76	0.88
	M_{Avg_Max}	0.79	0.87	0.76	0.87
	M_{Avg_Sub}	0.77	0.84	0.73	0.84
	M_{Max_Sub}	0.81	0.92	0.79	0.92
	M_{Avg_Avg}	0.77	0.86	0.75	0.85
6 Features	M_{Max_Max}	0.73	0.80	0.72	0.79
	M_{Avg_Max}	0.75	0.84	0.71	0.81
	M_{Avg_Sub}	0.72	0.80	0.70	0.80
	M_{Max_Sub}	0.74	0.78	0.71	0.77
	M_{Avg_Avg}	0.74	0.82	0.70	0.81

Table 15 Average of performance by using total 8 and 6 features with SVM classifier

	Formulae	Positive classes		Negative classes	
		F-score for training data	F-score for testing data	F-score for training data	F-score for testing data
8 Features	M_{Max_Max}	0.70	0.58	0.48	0.48
	M_{Avg_Max}	0.68	0.61	0.47	0.49
	M_{Avg_Sub}	0.67	0.63	0.53	0.39
	M_{Max_Sub}	0.74	0.60	0.60	0.56
	M_{Avg_Avg}	0.66	0.68	0.51	0.53
6 Features	M_{Max_Max}	0.67	0.55	0.45	0.47
	M_{Avg_Max}	0.67	0.60	0.45	0.43
	M_{Avg_Sub}	0.67	0.63	0.51	0.41
	M_{Max_Sub}	0.74	0.57	0.57	0.55
	M_{Avg_Avg}	0.66	0.66	0.49	0.57

Table 16 Average of performance by using total 8 and 6 features with D-Tree classifier

	Formulae	Positive classes		Negative classes	
		F-score for training data	F-score for testing data	F-score for training data	F-score for testing data
8 Features	M_{Max_Max}	0.75	0.75	0.69	0.76
	M_{Avg_Max}	0.76	0.77	0.72	0.76
	M_{Avg_Sub}	0.72	0.75	0.69	0.77
	M_{Max_Sub}	0.72	0.77	0.72	0.78
	M_{Avg_Avg}	0.75	0.76	0.68	0.77
6 Features	M_{Max_Max}	0.74	0.67	0.58	0.61
	M_{Avg_Max}	0.74	0.69	0.58	0.62
	M_{Avg_Sub}	0.69	0.68	0.56	0.60
	M_{Max_Sub}	0.71	0.70	0.58	0.60
	M_{Avg_Avg}	0.73	0.72	0.56	0.61

- **Discussion:** By comparing the results of the three ML classifiers in the tables, three conclusions were reached. Firstly using 8-features performs better than using 6-features in the three tables. The second conclusion is that the results of ANN classifier in Table 14 outperforms both the results of SVM classifier in Table 15 and that of D-Tree classifier in Table 16. Lastly, M_{Max_Sub} formula has the best results amongst the employed formulae in Table 14 in which the average F-score achieved in the term level for positive testing classes was **0.92**, and same value was obtained

for the negative classes as well. So the M_{Max_Sub} formula is credited to be used in all future experiments in order to get better one score for each Arabic sentiment term.

4.5.1.2 With Rules

- **Objective:** To identify the effectiveness of applying rules in term level of SA by using the ML approach.
- **Method:** From the previous experiment the M_{Max_Sub} formula is credited to be used for calculating one sentiment score for each term and the total 8 features are also credited to be used in the term level experiment of ML approach. To do this experiment, (i) The score for each term has been calculated by using M_{Max_Sub} formula. (ii) Then rules have been applied for each term. (iii) 5-fold cross validation has been applied for training and testing purpose on the file that has been based on M_{Max_Sub} formula. (iv) Then the ML approach (ANN, D-Tree, and SVM classifiers) has been applied on the same file to see the effectiveness of the rules on the classification results compared to the previous results of not applying rules. (v) The average of the F-Score measure has been calculated for the results of the five folds of positive training classes, positive testing classes, negative training classes, and negative testing classes to present the numbers shown in the next table.
- **Results:** Results are presented in the following Table 17.

Table 17. Average of performance for using total 8 features with applying rules with the ML classifiers in the term level

Classifier	Positive classes		Negative classes	
	F-score for training data	F-score for testing data	F-score for training data	F-score for testing data
ANN	0.78	0.91	0.77	0.91
D-Tree	0.71	0.80	0.65	0.80
SVM	0.66	0.58	0.54	0.37

- **Discussion:** The results of the ANN classifier outperform the results of the SVM and D-Tree classifiers, and the results of non-application of rules in Table 14 outperform the results of application of rules in Table 17 specifically in the case of applying M_{Max_Sub} formula and a total of 8 features in both experiments. For example, the F-score average in positive and negative classes for testing data in Table 14 is **0.92**. While the F-score average in positive and negative classes for testing data in Table 17 is **0.91**.

4.5.2 Experiment B: Document Level Sentiment Analysis Using Machine Learning Approach

We applied this experiment firstly without using rules, then with using rules on the document level.

4.5.2.1 Without Rules

- **Objective:** We aim at this experiment to identify the best formula of computing one score for each sentence in the document. In addition, to see the efficiency of the features by decreasing their number in the classification task of the document level.
- **Method:** (i) It is proved in the first experiment of section 4.5.1.1 that the M_{Max_Sub} formula is credited to compute one sentiment score (prior polarity) for each sentiment term. However, one positive or negative score (sentence polarity) has been computed from many terms scores for each sentence in the OCA corpus by using the 2 aggregation formulae in section 4.1.5 which are D_{Max_Sub} and D_{Max_Max} formulae. (ii) For each file we extracted total seven features which are in order: *count of positive sentences, count of negative sentences, maximum score of positive sentences, maximum score of negative sentences, first sentence score, middle sentence score and last sentence score*. To ensure the effectiveness of the features and the testing methods, we firstly omitted the third and the fourth features (i.e. *maximum score of positive sentences and maximum score of negative sentences*), thus we have five features. secondly, the last three features (i.e. *first sentence score, middle sentence score and last sentence score*) have been omitted, thus we have four features. Now we have 6 OCA files in which 3 files are based on D_{Max_Sub}

formula and the other 3 files are based on D_{Max_Max} formula. (iii) 5-fold cross validation has been applied for training and testing purpose on each of the 6 files. (iv) Then the ML approach (ANN, D-Tree, and SVM classifiers) have been applied on each of the 6 files to nominate the best formula for computing one score for each Arabic sentence in a document. (v) We calculated the average of the F-Score measure for results of the five folds of positive training classes, positive testing classes, negative training classes, and negative testing classes to show the numbers shown in the next tables.

- **Results:** Results are presented in Tables 18 - 19.

Table 18 Average of performance of using total 7, 5 and 4 features based on D_{Max_Max} formulae in the document level without rules

	Classifier	Positive classes		Negative classes	
		F-score for training data	F-score for testing data	F-score for training data	F-score for testing data
7 Features	ANN	0.78	0.93	0.78	0.92
	D-Tree	0.71	0.87	0.68	0.88
	SVM	0.60	0.79	0.63	0.82
5 Features	ANN	0.77	0.93	0.77	0.92
	D-Tree	0.76	0.87	0.71	0.88
	SVM	0.61	0.82	0.63	0.85
4 Features	ANN	0.67	0.87	0.60	0.84
	D-Tree	0.65	0.75	0.58	0.76
	SVM	0.58	0.79	0.57	0.83

Table 19 Average of performance of using total 7, 5 and 4 features based on D_{Max_Sub} formulae in the document level without rules

	Classifier	Positive classes		Negative classes	
		F-score for training data	F-score for testing data	F-score for training data	F-score for testing data
7 Features	ANN	0.73	0.90	0.70	0.89
	D-Tree	0.64	0.81	0.64	0.88
	SVM	0.59	0.79	0.62	0.82
5 Features	ANN	0.71	0.86	0.71	0.87
	D-Tree	0.64	0.80	0.65	0.86
	SVM	0.60	0.78	0.60	0.81
4 Features	ANN	0.68	0.86	0.61	0.85
	D-Tree	0.52	0.78	0.63	0.82
	SVM	0.57	0.79	0.57	0.82

Discussion: Comparisons between the results of the three ML classifiers in the two tables can make three conclusive result statements. The first conclusion is that using total 7 features is better than using a total of 5 or 4 features in both tables. The second conclusion is that ANN gives better results than SVM and D-Tree in both tables, which was based on a total of 7 features. The last conclusive statement is that the ANN classifier results based on the D_{Max_Max} formula in Table 18 have better result output than the ANN classifier results based on D_{Max_Sub} formula in Table 19. Therefore, the D_{Max_Max} formula is recommended to be used when the need to getting one score for each Arabic sentiment sentence is needed.

4.5.2.2 With Rules

- **Objective:** Identifying the effectiveness of applying rules in document level of SA by using the ML approach.
- **Method:** From the previous experiment, the D_{Max_Max} formula is credited to be used in getting one score for each Arabic sentiment sentence, and the total 7-features are also credited to be used in this document level experiment of ML approach. To do this experiment, (i) Rules have been applied for each term in the document. (ii) The score for each sentence has been calculated by using D_{Max_Max}

formula. (iii) Total seven features have been extracted which are in order: *count of positive sentences, count of negative sentences, maximum score of positive sentences, maximum score of negative sentences, first sentence score, middle sentence score and last sentence score*. (iv) 5-fold cross validation has been applied for training and testing purpose on the file which is based on D_{Max_Max} formula. (v) The ML approach (ANN) is applied on the file to see the effectiveness of the rules on the classification results compared to the previous results of not applying rules. (vi) We calculated the average of the F-Score measure for the results of the five folds of positive training classes, positive testing classes, negative training classes, and negative testing classes to present the numbers shown in the next table.

- **Results:** Results are presented in Table 20.

Table 20. Average of performance of using total 7 features with applying rules based on D_{Max_Max} formulae in the document level

	Positive classes		Negative classes	
Classifier	F-score for training data	F-score for testing data	F-score for training data	F-score for testing data
ANN	0.81	0.94	0.78	0.93

Discussion: The results of applying rules in the document level are compared to the results of non-application of rules in the Table 18 of the previous section, especially when D_{Max_Max} formula for sentence score computing and total 7 features are used. For example, the average F-score in positive classes for testing data in Table 18 is **0.93**. While the average F-score in positive classes for testing data in case of applying rules is **0.94** in Table 20, and the average F-score in negative classes for testing data in Table 18 is **0.92**. While the average F-score in negative classes for testing data when rules are applied is **0.93** in Table 20. From the results in section 4.5.1 and section 4.5.2, we conclude that the effectiveness of the rules appears clearly in the document level rather than in the term level of Arabic SA.

4.5.3 Experiment C: Sentiment Analysis Using Lexicon Based Approach without Rules

We applied this experiment without using rules firstly in the term level, then in the document level.

4.5.3.1 Term Level Experiment without Rules

- **Objective:** Nominating the most effective formula for computing document sentiment score by comparing the performance of different formulae.
- **Method:** It is proved in the first experiment that the M_{Max_Sub} formula is nominated to calculate one sentiment score for each sentiment term, however we did our experiment in the following steps: (i) One sentiment score is calculated for each term in the OCA corpus using the M_{Max_Sub} formula. (ii) 5-fold cross validation has been applied to ensure that every sample of the OCA has the same chance of appearing in the calculations. (iii) Each document score has been calculated by using different aggregation formulae that have been presented in Table 10. The aggregation formulae are DT_{MAX_MAX} , DT_{MAX_SUB} , DT_{AVG_MAX} , DT_{AVG_SUB} , and DT_{AVG_AVG} . (iv) We tried to enhance results by calculating one sentiment score for every term in the OCA using the M_{Max_Sub} formula in which each $|term\ score| \geq 0.3$. Then we repeated step 3. (v) We also tried to enhance results by calculating one sentiment score for every term in the OCA using the M_{Max_Sub} formula in which each $|term\ score| \geq 0.5$. Then we repeated step 3. (vi) We used precision and recall metrics to calculate F-Score metric, then we calculated the average of the F-Score measure for the results of the five folds of positive training classes, positive testing classes, negative training classes, and negative testing classes to present the numbers shown in the next tables.
- **Results:** Results are presented in Tables 21 - 25.

Table 21. Average of performance of using DT_{MAX_MAX} formula

	Positive classes	Negative classes
Term score based on M_{Max_Sub} formula	F-score for testing data	F-score for testing data
term score	0.67	0.27
term score \geq 0.3	0.67	0.27
term score \geq 0.5	0.66	0.27

Table 22. Average of performance of using DT_{MAX_SUB} formula

	Positive classes	Negative classes
Term score based on M_{Max_Sub} formula	F-score for testing data	F-score for testing data
term score	0.67	0.27
term score \geq 0.3	0.67	0.27
term score \geq 0.5	0.66	0.27

Table 23. Average of performance of using DT_{AVG_MAX} formula

	Positive classes	Negative classes
Term score based on M_{Max_Sub} formula	F-score for testing data	F-score for testing data
term score	0.59	0.33
term score \geq 0.3	0.52	0.51
term score \geq 0.5	0.55	0.49

Table 24. Average of performance of using DT_{AVG_SUB} formula

	Positive classes	Negative classes
Term score based on M_{Max_Sub} formula	F-score for testing data	F-score for testing data
term score	0.59	0.33
term score \geq 0.3	0.52	0.51
term score \geq 0.5	0.55	0.49

Table 25 Average of performance of using DT_{AVG_AVG} formula

	Positive classes	Negative classes
Term score based on M_{Max_Sub} formula	F-score for testing data	F-score for testing data
term score	0.59	0.33
term score \geq 0.3	0.52	0.51
term score \geq 0.5	0.55	0.49

- **Discussion:** It is notable that: (i) The results of positive classes in the Table 21 and in the Table 22 are acceptable, but the results of negative classes in both tables are not acceptable. Although we tried to enhance results by calculating one sentiment score for every term in the OCA by using the M_{Max_Sub} formula in which the absolute value of each term is \geq 0.3 or the absolute value of each term \geq 0.5, there is no notable change in the results of both tables. It is concluded from the Table 21 and from the Table 22 that in the term level of Arabic SA, the DT_{MAX_MAX} formula and the DT_{MAX_SUB} formula are not dependable in the computing of the document score. (ii) The results of positive and negative classes in the Table 23, Table 24 and Table 25 are acceptable. Although we tried to enhance results by calculating one sentiment score for every term in the OCA by using the M_{Max_Sub}

formula in which the absolute value of each term is ≥ 0.3 or the absolute value of each term ≥ 0.5 , it is notable there is a small enhancement of the results in the three tables. It is concluded from the results in Tables 21 - 25 that the applying of DT_{MAX_MAX} , DT_{MAX_SUB} , DT_{AVG_MAX} , DT_{AVG_SUB} , and DT_{AVG_AVG} formulae is not enough, for this reason, in the next section our experiment has been done in the document level rather than term level.

4.5.3.2 Document Level Experiment without Rules

- **Objective:** Extracting the most effective formulae for computing document sentiment score between five aggregation formulae by comparing their performance in the document level, and without applying rules.
- **Method:** (i) M_{Max_Sub} formula is used to compute one score (prior polarity) for each sentiment term. (ii) For each sentence in the document, one positive or negative score (i.e. sentence polarity) from many terms scores in each sentence in the OCA corpus has been computed by using the 2 aggregation formulae in section 4.1.5 which are D_{Max_Sub} and D_{Max_Max} formulae. (iii) 5-fold cross validation is applied to ensure that every sample of the OCA corpus has the same chance of appearing in the calculations. (iv) Each document score is calculated by using different aggregation formulae which are presented previously in Table 12. The aggregation formulae are DS_{MAX_MAX} , DS_{MAX_SUB} , DS_{AVG_MAX} , DS_{AVG_SUB} , and DS_{AVG_AVG} . (v) Precision and recall metrics are used to calculate F-Score metric. Then we calculated the average F-Score measure for the results of the five folds of positive training classes, positive testing classes, negative training classes, and negative testing classes to present the numbers shown in the next tables. (vi) The D_{Max_Max} formula has better results than D_{Max_Sub} in the ML approach experiments in this study. However, we tried to enhance results by calculating one sentiment score for each sentence in the OCA corpus by using D_{Max_Max} formula, in which each $|\text{sentence score}| \geq 0.3$, then we repeated step 4 and step 5. (vii) We also tried to enhance results by calculating one sentiment score for each sentence in the OCA corpus using the D_{Max_Max} formula in which each $|\text{sentence score}| \geq 0.5$, then we repeated step 4 and step 5.
- **Results:** The results are presented in Tables 26 - 31.

Table 26 Average of performance of using DS_{MAX_MAX} formula

	Positive classes	Negative classes
Sentence score based on D_{Max_Max} formula	F-score for testing data	F-score for testing data
sentence score	0.67	0.27
sentence score \geq 0.3	0.67	0.27
sentence score \geq 0.5	0.66	0.27

Table 27 Average of performance of using DS_{MAX_SUB} formula

	Positive classes	Negative classes
Sentence score based on D_{Max_Max} formula	F-score for testing data	F-score for testing data
sentence score	0.67	0.27
sentence score \geq 0.3	0.67	0.27
sentence score \geq 0.5	0.66	0.27

Table 28 Average of performance of using DS_{AVG_MAX} formula

	Positive classes	Negative classes
Sentence score based on D_{Max_Max} formula	F-score for testing data	F-score for testing data
sentence score	0.60	0.51
sentence score \geq 0.3	0.65	0.42
sentence score \geq 0.5	0.66	0.40

Table 29 Average of performance of using DS_{AVG_SUB} formula

	Positive classes	Negative classes
Sentence score based on D_{Max_Max} formula	F-score for testing data	F-score for testing data
sentence score	0.60	0.51
sentence score ≥ 0.3	0.65	0.42
sentence score ≥ 0.5	0.66	0.40

Table 30 Average of performance of using DS_{AVG_AVG} formula

	Positive classes	Negative classes
Sentence score based on D_{Max_Max} formula	F-score for testing data	F-score for testing data
sentence score	0.60	0.51
sentence score ≥ 0.3	0.65	0.42
sentence score ≥ 0.5	0.66	0.40

Table 31 Average of performance of using D_{Max_Sub} formula for calculating each sentence score

	Positive classes	Negative classes
Formulae	F-score for testing data	F-score for testing data
DS_{MAX_MAX}	0.60	0.34
DS_{MAX_SUB}	0.60	0.34
DS_{AVG_MAX}	0.52	0.48
DS_{AVG_SUB}	0.52	0.48
DS_{AVG_AVG}	0.52	0.48

- **Discussion:** (i) The results in Table 26 and the results in Table 27 of positive classes are acceptable, but it is not acceptable in the negative classes. Therefore, it is concluded that the DS_{MAX_MAX} and DS_{MAX_SUB} formula are not dependable in the computing of document score of Arabic SA. (ii) In the case of not applying $|sentence\ score| \geq 0.3$ and $|sentence\ score| \geq 0.5$ on the document score result, the results in Tables 28 - 30 are acceptable for positive and negative classes. However, it is concluded that the DS_{AVG_MAX} , DS_{AVG_SUB} and DS_{AVG_AVG} formulae can be used for computing of document score of Arabic SA but there is a need to enhance their performance, for this reason, rules and new formulae will be applied in the next experiment. (iii) Moreover, it is notable that the results of Table 31 are not dependable, so it is concluded that using D_{Max_Max} formula is better than using D_{Max_Sub} formula in the computing of sentence score in the document level of Arabic SA without applying rules.

4.5.4 Experiment D: Sentiment Analysis Using Lexicon Based Approach with Applying Rules in the Document Level

- **Objective:** Enhancing the results of computing each document score by applying rules on each sentiment term in the OCA corpus and nominating the most effective formulae for computing document sentiment score between new four formulae, by comparing their performance.
- **Method:** (i) M_{Max_Sub} formula has been used to compute one score (prior polarity) for each sentiment term. (ii) Rules have been applied on each term of the preprocessed OCA corpus. (iii) One positive or negative score (i.e. sentence polarity) from many terms scores in each sentence in the OCA corpus has been computed by using the 2 aggregation formulae in section 4.1.5 which are D_{Max_Sub} and D_{Max_Max} formulae. (iv) 5-fold cross validation is applied to ensure that every sample of the OCA corpus has the same chance of appearing in the calculations. (v) Each document score is calculated by using different formulae which are *last sentence score*, *first sentence score*, *average of the last three sentences*, and *average of the first three sentences*. (vi) Precision and recall metrics are used to calculate F-Score metric, then we calculated the average of the F-Score measure for the results of the five folds of positive training classes, positive testing classes,

negative training classes, and negative testing classes to present the numbers shown in next tables.

- **Results:** Results are presented in Tables 32 and 33.

Table 32 Average of performance of using D_{Max_Max} formula for calculating each sentence score with applying rules on each term

	Positive classes	Negative classes
Formulae	F-score for testing data	F-score for testing data
Last sentence score	0.68	0.47
First sentence score	0.68	0.47
Average of last 3 sentences	0.61	0.27
Average of first 3 sentences	0.68	0.36

Table 33 Average of performance of using D_{Max_Sub} formula for calculating each sentence score with applying rules on each term

	Positive classes	Negative classes
Formulae	F-score for testing data	F-score for testing data
Last sentence score	0.43	0.28
First sentence score	0.64	0.55
Average of last 3 sentences	0.59	0.44
Average of first 3 sentences	0.63	0.53

- **Discussion:** The results in Table 32 of positive classes are acceptable, but it is not acceptable for the negative classes. Therefore, it is concluded that it is not

dependable that using D_{Max_Max} formula for calculating each sentence score with applying rules on each term, and then applying the four formulae for calculating document score. In Table 33 the results positive classes and negative classes are acceptable just in case of using D_{Max_Sub} formula for calculating each sentence score with applying rules on each term, and then applying the *First sentence score* formulae for calculating document score. Finally, in the LB approach the best method result is of using D_{Max_Sub} formula for calculating each sentence score with applying rules on each term, and then applying the *First sentence score* formulae for calculating document score.

4.5.5 Results Summary

The best results summary of both ML and LB approaches are discussed in the next two subsections.

4.5.5.1 Summary of ML Approach Results

It is seen that the average performance results are presented in Table 34. In general, we can observe that the average of F-score for testing data is better than that of training data due to the size of testing data, which is smaller than that of training data (i.e. 400 documents for training and 100 documents for testing) and this is noticeable in literature, especially in text classification task.

In the term level of Arabic SA, the best results are obtained by using M_{Max_Sub} formula to compute each term prior polarity score, then, applying total of 8 features with ANN classifier.

While, in the document level, the best results are obtained by applying D_{Max_Max} formula for computing sentence score then, using total 7 features with ANN classifier. In other words, the best results are achieved by firstly, the application of rules for each term, then computing each sentence score using the D_{Max_Max} formula, and lastly, using a total of 7 features and an ANN classifier in the document level.

Table 34 Average of performance by (i) using total 8 features and ANN classifier in the term level with applying M_{Max_Sub} formula for prior polarity computing, (ii) using total 7 features and ANN classifier in the document level with applying D_{Max_Max} formula for sentence score computing.

SA levels	Applying / without applying rules	Positive classes		Negative classes	
		F-score for training data	F-score for testing data	F-score for training data	F-score for testing data
Term level	Without applying rules	0.81	0.92	0.79	0.92
	Applying rules	0.78	0.91	0.77	0.91
Document level	Without applying rules	0.78	0.93	0.78	0.92
	Applying rules	0.81	0.94	0.78	0.93

To evaluate the results of our study with respect to studies in literature, the best results obtained by Rushdi-Saleh et al. [23] are compared with our best results . It is important noting that OCA corpus and ML approach are used in both studies, and the best results are related to Recall, Precision and F-score evaluation metrics. Although Recall of Rushdi-Saleh' study (0.95) is better than the Recall of our study (**0.93**), the Precision of our study (0.94) is better than the Precision of of Rushdi-Saleh' study (0.87), and that for F-score of our study (**0.94**) outperforms that of Rushdi-Saleh's study (0.91). In general, the results of this study are better than the results of Rushdi-Saleh's study.

4.5.5.2 Summary of Lexicon Based Approach Results

In the LB approach, the average of performance results (F-Score average) has been presented in the Table 35 and illustrated as follows: (i) The best results are gained without applying rules on both term and document levels of SA. However, in the term level DT_{AVG_MAX} , DT_{AVG_SUB} , or DT_{AVG_AVG} formulae (that have the same results)

have been used for document score computing. While in the document level D_{Max_Max} formula has been used for sentence score computing, then DS_{AVG_MAX} , DS_{AVG_SUB} , or DS_{AVG_AVG} formulae (that have the same results) have been used for document score computing. (ii) With applying rules on each term, then using D_{Max_Sub} formulae for sentence score computing and the *first sentence score* formulae has been used for document score computing.

Table 35 Average of performance (i) without applying rules on both term and document levels. (ii) Applying rules on document level.

Applying / without applying rules	SA levels	Positive classes	Negative classes
		F-score for testing data	F-score for testing data
Without applying rules	Term Level	0.55	0.49
	Document Level	0.60	0.51
Applying rules	Document Level	0.64	0.55

Discussion: When the Table 34 and the Table 35 results are observed, it is concluded that: (i) In the ML approach the best results are achieved by applying rules for each term, then computing each sentence score by D_{Max_Max} formula, and finally using total 7 features with and ANN classifier in the document level. (ii) In the LB approach the best results are achieved by applying rules for each term, then computing each sentence score by D_{Max_Sub} formula, finally, using first *sentence score* formulae for document score computing. However, the ML approach comparisons of the best results are presented in Fig. 8. And the LB comparisons of the best results are presented in Fig. 9

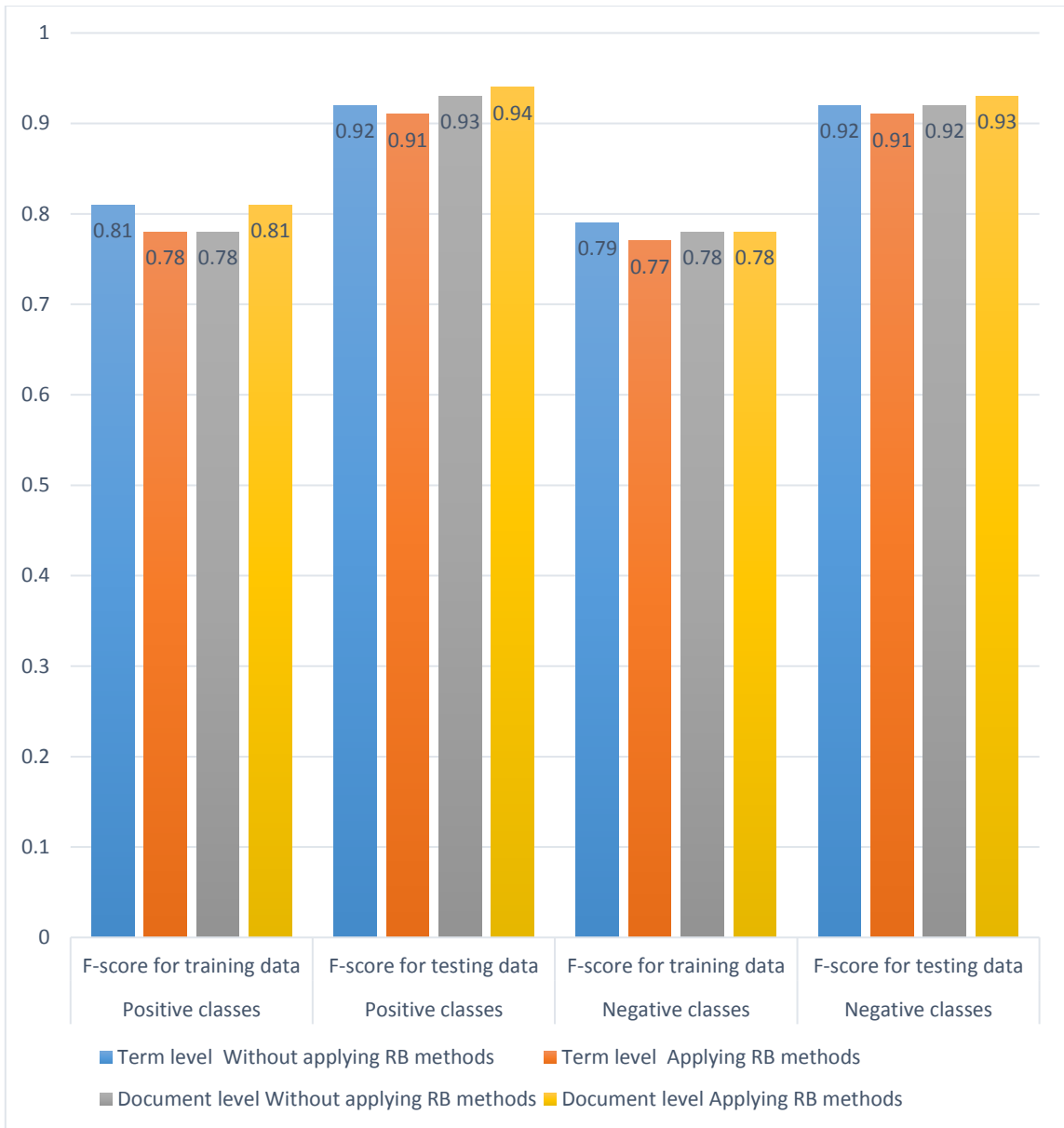


Figure 8 Comparisons of the ML Approach Experiments Results

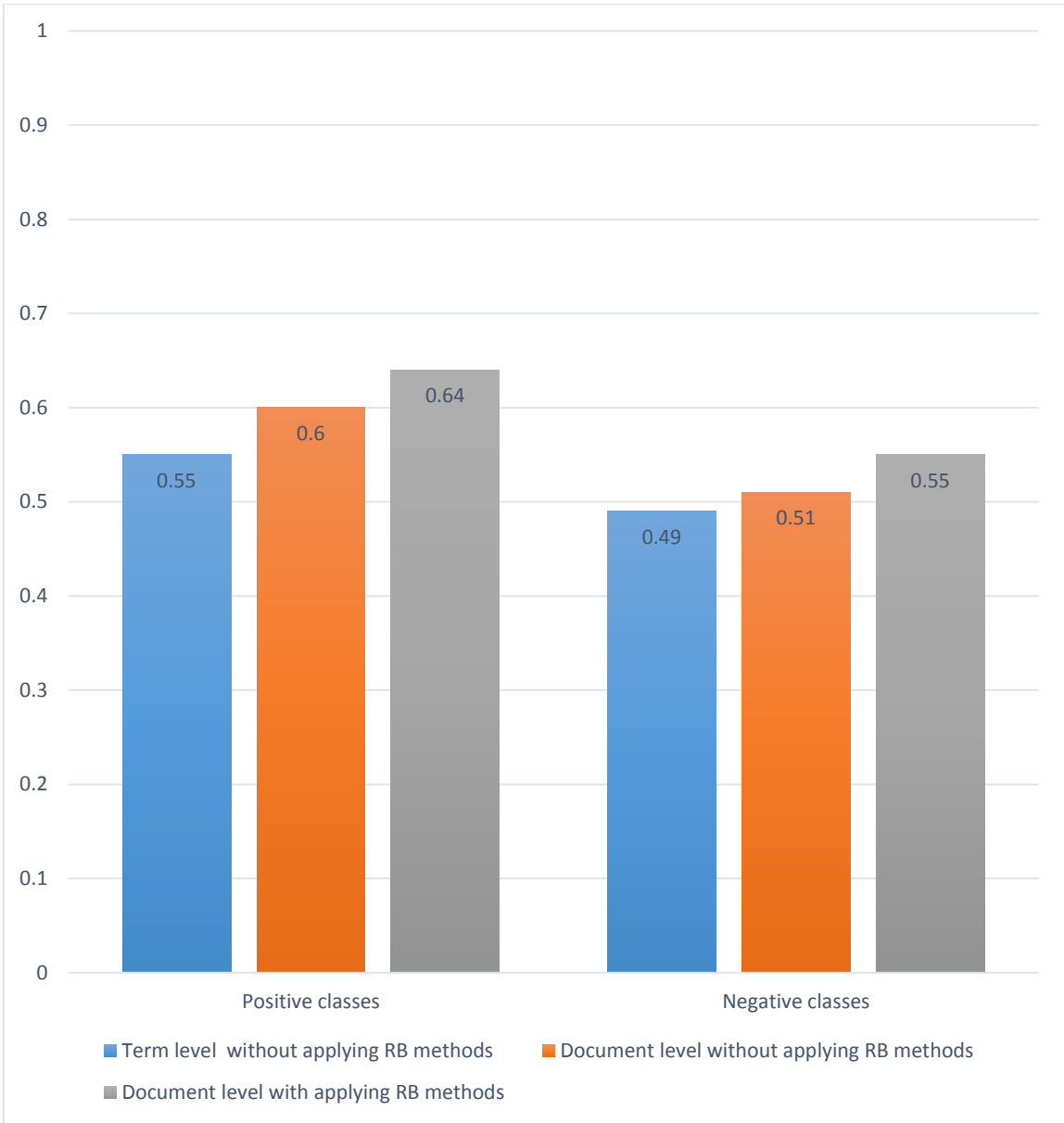


Figure 9 Comparisons of the LB Approach Experiments Results

5. Conclusion

We aimed to enhance term level sentiment score by creating a new success formula to compute one prior polarity score for each sentiment term in this research study. Secondly, to enhance the performance results of both term and document levels at Arabic sentiment analysis by investigating the most successfully ML classifier to prove the efficiency of the features and to classify opinion documents. Finally, implementation of ML and LB approaches with rules together to improve obtained SA performance values. To achieve our aims of this study, the following steps are followed:

- (i) The opinion terms were decided from the corpus by using ArSenL, where, each word root (lemma) in the lexicon have many senses or scores that are rated in the interval of [-1...1]. One positive or negative score (prior polarity) from many scores (posterior polarities) was computed for each opinion word placed in the OCA corpus by creating different aggregation formulae. However, the M_{Max_Sub} formula has the best results compared to the other formulae results. Therefore, the M_{Max_Sub} formula is recommended for use when the need to getting only one score for each Arabic sentiment term arises.
- (ii) Features have been extracted based on the term level and document level with application and non-application of rules in the term level, where the document level is based on the term level. However, it is concluded that increasing number of features gives better results in both term level and document levels.

The main findings achieved by our study are that, (i) ANN classifier is nominated as the best classifier in proving the efficiency of the features in both term and document levels of Arabic SA. In which the average F-score achieved in the term level for both of positive and negative testing classes is 0.92, however, in the document level, the average F-score for positive testing classes is 0.94, and 0.93 in the negative classes. Moreover, (ii) by comparing our study with studies in literature, and using

F-score as evaluation measure, it is observed that, the best result of our study (0.94) outperforms the best result (0.91) of Rushdi-Saleh et al. [23], whereas, OCA corpus and ML approach are used in both studies. On the other hand of applying LB approach, (iii) the best results are achieved by applying rules for each term, then computing each sentence score by D_{Max_Sub} formula, and finally, using first *sentence score* formulae for document score computing.

In general, the integration of rules with both ML and LB approaches has enhanced their experiment results; however, the ML approach results are better than LB results.

When all findings obtained from the experiments are considered, the main contributions of this work may be that one prior polarity is computed for each Arabic sentiment term; moreover, it is found that there is no previous study in Arabic SA that is interested in computing prior polarity from posterior polarities. Lastly, rules were integrated with both ML and LB approaches in the SA for the Arabic Language, and a comprehensive comparison between the results has been made.

For future work, we aim at enhancing the results of the LB approach by trying different formulae and different rules. Finally, integrating the LB approach to the ML approach with the application of rules in both approaches.

References

- [1] Medhat, Walaa, Ahmed Hassan, and Hoda Korashy. "Sentiment analysis algorithms and applications: A survey." *Ain Shams Engineering Journal* 5.4 (2014): 1093-1113.
- [2] Liu, Bing. "Sentiment analysis and opinion mining." *Synthesis lectures on human language technologies* 5.1 (2012): 1-167.
- [3] Li, Yong H., and Anil K. Jain. "Classification of text documents." *The Computer Journal* 41.8 (1998): 537-546.
- [4] Wiebe, Janyce M., Rebecca F. Bruce, and Thomas P. O'Hara. "Development and use of a gold-standard data set for subjectivity classifications." *Proceedings of the 37th annual meeting of the Association for Computational Linguistics on Computational Linguistics*. Association for Computational Linguistics, 1999.
- [5] B. Pang, L. Lee, S. Vaithyanathan, Thumbs up? Sentiment classification using machine learning techniques. *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 2002, pp. 79–86. Association for Computational Linguistics.
- [6] Yu, Hong, and Vasileios Hatzivassiloglou. "Towards answering opinion questions: Separating facts from opinions and identifying the polarity of opinion sentences." *Proceedings of the 2003 conference on Empirical methods in natural language processing*. Association for Computational Linguistics, 2003.
- [7] [Mullen and Collier, 2004] Mullen, T., & Collier, N. (2004). Sentiment analysis using support vector machines with diverse information sources. In: *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP)*. pp. 412–418.
- [8] Chesley, Paula, et al. "Using verbs and adjectives to automatically classify blog sentiment." *Training* 580.263 (2006): 233.
- [9] Jia, Lifeng, Clement Yu, and Wei Zhang. "UIC at TREC 2008 blog track". ILLINOIS UNIV AT CHICAGO DEPT OF COMPUTER SCIENCE, 2008.
- [10] Prabowo, R., & Thelwall, M. (2009). Sentiment analysis: A combined approach. *Journal of Informetrics*, 3(2), 143–157.

- [11] Li, Yung-Ming, and Tsung-Ying Li. "Deriving market intelligence from microblogs." *Decision Support Systems* 55.1 (2013): 206-217.
- [12] Moraes, Rodrigo, João Francisco Valiati, and Wilson P. Gavião Neto. "Document-level sentiment classification: An empirical comparison between SVM and ANN." *Expert Systems with Applications* 40.2 (2013): 621-633.
- [13] Hatzivassiloglou, Vasileios, and Kathleen R. McKeown. "Predicting the semantic orientation of adjectives." *Proceedings of the eighth conference on European chapter of the Association for Computational Linguistics*. Association for Computational Linguistics, 1997.
- [14] Turney, Peter D. "Thumbs up or thumbs down?: semantic orientation applied to unsupervised classification of reviews." *Proceedings of the 40th annual meeting on association for computational linguistics*. Association for Computational Linguistics, 2002.
- [15] Hu, Minqing, and Bing Liu. "Mining and summarizing customer reviews." *Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 2004.
- [16] Thet, Tun Thura, et al. "Sentiment analysis of movie reviews on discussion boards using a linguistic approach." *Proceedings of the 1st international CIKM workshop on Topic-sentiment analysis for mass opinion*. ACM, 2009.
- [17] M. Guerini, L.Gatti, and M.Turchi. 2013. Sentiment Analysis: How to Derive Prior Polarities from SentiWordNet. In *Proceedings of EMNLP*, 2013.
- [18] Baccianella, Stefano, Andrea Esuli, and Fabrizio Sebastiani. "SentiWordNet 3.0: An Enhanced Lexical Resource for Sentiment Analysis and Opinion Mining." *LREC*. Vol. 10. 2010.
- [19] Kouloumpis, Efthymios, Theresa Wilson, and Johanna D. Moore. "Twitter sentiment analysis: The good the bad and the omg!" *Icwsm* 11.538-541 (2011): 164.
- [20] Wilson, Theresa, Janyce Wiebe, and Paul Hoffmann. "Recognizing contextual polarity in phrase-level sentiment analysis." *Proceedings of the conference on human language technology and empirical methods in natural language processing*. Association for Computational Linguistics, 2005.
- [21] Almas, Yousif, and Khurshid Ahmad. "A note on extracting 'sentiments' in financial news in English, Arabic & Urdu." *The Second Workshop on Computational Approaches to Arabic Script-based Languages*. 2007.

- [22] Abbasi, Ahmed, Hsinchun Chen, and Arab Salem. "Sentiment analysis in multiple languages: Feature selection for opinion classification in web forums." *ACM Transactions on Information Systems (TOIS)* 26.3 (2008): 12.
- [23] Rushdi-Saleh, Mohammed, et al. "OCA: Opinion corpus for Arabic." *Journal of the American Society for Information Science and Technology* 62.10 (2011): 2045-2054.
- [24] Shoukry, Amira, and Ahmed Rafea. "Sentence-level Arabic sentiment analysis." *Collaboration Technologies and Systems (CTS), 2012 International Conference on*. IEEE, 2012.
- [25] Farra, Noura, et al. "Sentence-level and document-level sentiment mining for Arabic texts." *Data Mining Workshops (ICDMW), 2010 IEEE International Conference on*. IEEE, 2010.
- [26] El-Halees, Alaa. "Arabic opinion mining using combined classification approach." *International Arab Conference on Information Technology (ACIT), 2011*.
- [27] Badaro, Gilbert, et al. "A large scale Arabic sentiment lexicon for Arabic opinion mining." *ANLP 2014* 165 (2014).
- [28] Abdulla, Nawaf A., et al. "Arabic sentiment analysis: Lexicon-based and corpus-based." *Applied Electrical Engineering and Computing Technologies (AEECT), 2013 IEEE Jordan Conference on*. IEEE, 2013.
- [29] Denecke, Kerstin. "Using sentiwordnet for multilingual sentiment analysis." *Data Engineering Workshop, 2008. ICDEW 2008. IEEE 24th International Conference on*. IEEE, 2008.
- [30] Zhang, Changli, et al. "Sentiment analysis of Chinese documents: From sentence to document level." *Journal of the American Society for Information Science and Technology* 60.12 (2009): 2474-2487.
- [31] Agić, Željko, Nikola Ljubešić, and Marko Tadić. "Towards sentiment analysis of financial texts in Croatian." *Proceedings of the Seventh International Conference on Language Resources and Evaluation*. 2010.
- [32] Ghorbel, Hatem, and David Jacot. "Sentiment analysis of French movie reviews." *Advances in Distributed Agent-Based Retrieval Tools*. Springer Berlin Heidelberg, 2011. 97-108.

- [33] MartíN-Valdivia, MaríA-Teresa, et al. "Sentiment polarity detection in Spanish reviews combining supervised and unsupervised approaches." *Expert Systems with Applications* 40.10 (2013): 3934-3942.
- [34] Akba, F., Uçan, A., Sezer, E., & Sever, H. (2014). Assessment of feature selection metrics for sentiment analyses: Turkish movie reviews. In 8th European Conference on Data Mining .2014.
- [35] Dehkharghani, Rahim, et al. "Sentiment analysis in Turkish at different granularity levels." *Natural Language Engineering* (2016): 1-25.
- [36] Miller, George A., et al. "Introduction to WordNet: An on-line lexical database." *International journal of lexicography* 3.4 (1990): 235-244.
- [37] Black, William, et al. "Introducing the Arabic wordnet project." *Proceedings of the third international WordNet conference*. 2006.
- [38] Graff, David, et al. "Standard Arabic morphological analyzer (SAMA) version 3.1." *Linguistic Data Consortium LDC2009E73* (2009).
- [39] Wikipedia, Sentiment analysis.
https://en.wikipedia.org/wiki/Sentiment_analysis (accessed March, 2017).
- [40] Joachims, Thorsten. *A Probabilistic Analysis of the Rocchio Algorithm with TFIDF for Text Categorization*. No. CMU-CS-96-118. Carnegie-mellon univ pittsburgh pa dept of computer science, 1996.
- [41] Quinlan, J. Ross. "Induction of decision trees." *Machine learning* 1.1 (1986): 81-106.
- [42] Miller, George A. "WordNet: a lexical database for English." *Communications of the ACM* 38.11 (1995): 39-41.
- [43] Benajiba, Yassine, and Paolo Rosso. "Towards a measure for Arabic corpora quality." *Proc. Int. Colloquium on Arabic Language Processing, CITALA-2007, Rabat, Morroco*. 2007.
- [44] Salhi, Ali and Adnan Yahya. 2014. Quality Assessment of General and Categorized Arabic Corpora. *Proceedings of 5th International Conference on Arabic Language Processing-CITALA 2014; November 26--27, 2014, Woujda, Morocco*.
- [45] Al-Azani, Sadam, and El-Sayed M. El-Alfy. "Using Word Embedding and Ensemble Learning for Highly Imbalanced Data Sentiment Analysis in Short Arabic Text." *Procedia Computer Science* 109 (2017): 359-366.

- [46] Atakulreka, Akarachai, and Daricha Sutivong. "Avoiding local minima in feedforward neural networks by simultaneous learning." Australasian Joint Conference on Artificial Intelligence. Springer Berlin Heidelberg, 2007.
- [47] Chang, Chih-Chung, and Chih-Jen Lin. "LIBSVM: a library for support vector machines." *ACM Transactions on Intelligent Systems and Technology (TIST)* 2.3 (2011): 27.
- [48] Kohavi, Ron. "A study of cross-validation and bootstrap for accuracy estimation and model selection." *Ijcai*. Vol. 14. No. 2. 1995.
- [49] Internet World Stats, Internet world users by language.
<http://www.internetworldstats.com/stats7.htm> (accessed February 2017).
- [50] Abdelali, Ahmed, Jim Cowie, and Hamdy S. Soliman. "Arabic information retrieval perspectives." *Proceedings of the 11th Conference on Natural Language Processing, Journes d'Etude sur la Parole-Traitement Automatique des Langues Naturelles (JEP-TALN)*. 2004.
- [51] Pai, Mao-Yuan, et al. "Electronic word of mouth analysis for service experience." *Expert Systems with Applications* 40.6 (2013): 1993-2006.
- [52] Yan-Yan, Zhao, Qin Bing, and Liu Ting. "Integrating intra-and inter-document evidences for improving sentence sentiment classification." *Acta Automatica Sinica* 36.10 (2010): 1417-1425.
- [53] Maks, Isa, and Piek Vossen. "A lexicon model for deep sentiment analysis and opinion mining applications." *Decision Support Systems* 53.4 (2012): 680-688.
- [54] Wikipedia, Arabic. <https://en.wikipedia.org/wiki/Arabic> (accessed May 2017).
- [55] Boudad, Naaima, et al. "Sentiment analysis in Arabic: A review of the literature." *Ain Shams Engineering Journal* (2017).

APPENDEX 1: Samples from the OCA Database of Sentiment

Words based on M_{Max_Sub} Prior Polarity Computing Formula

Positive Words with Applying Rules

أولا	ذلك	بالمتعة	كلمات	كبير	ومصطفى	وأضاف	بعقل
مشارك	النجمين	فكر	تولستوي	تعمر	شخصيا	أول	عمل
بنجاح	بينهما	قطعية	كامل	قدمتم	وينقذها	يتضمن	منطقي
نجمان	حقيقة	يرقى	أجل	التقييم	الإعلام	بودي	يتجراً
الرومانسية	جيدة	الرجل	قارئاً	دور	إنناس	العرض	المراد
أثناء	مهما	والفكرية	يعيش	بعقل	دعاوى	حوله	الرسالة
زوجها	سيء	أشاهد	بالإمكان	العربي	دور	كبيراً	توصيلها
القضاء	كلية	أجد	وأبدع	العربي	العرض	فعل	خصوصاً
تريد	ولكنه	أريد	أبطال	مشاهدو	وصفته	ليس	العربي
علاقة	تشاهدها	الأرثوذكسية	تعيشون	توصيفات	حول	بطولة	حسن
منطقة	مشاهدة	الهائلة	المشاهدين	المتفرج	الكلام	أثير	جيد
بتكوين	مرة	مشاهدتي	والمنتجين	صناع	مشاعر	بعيدا	موضحا
الحب	التقييم	فرعي	موجهه	قصة	مسرحية	العربية	قدمته
أول	إعجابي	برسم	الرواية	مؤكدين	مشيرا	الإطلاق	فرض
عائليه	الأدبية	الأدبية	السجود	تشبه	يتيمة	بطل	وجارح
وتجد	دقائق	أجد	طبع	تقديم	سكنت	العمل	مؤكدا
الشاب	يقدم	أجبرت	ليس	حول	جاء	العربية	مكتبة
الخارجة	قدمها	منتجي	حياة	نورة	مصطفى	بين	متسائلا
يعطي	الأصدقاء	الأملاك	القراءة	تتخصص	نجاحا	علاقة	توفيق
المتكررة	تولستوي	أفكاره	رئيسي	بطولة	كوميديا	الإيحاء	يشاهد
تصدق	يريد	تولستوي	وفي	تغرم	مثل	أعمال	مشيرا
والحبكة	لحظات	والحقوق	يمكن	شاب	قبلة	ورجال	حتى
المنطق	محيي	وحتى	محيي	بالشاب	وبالتالي	المشاهدون	ومفيدا
والتسامح	التأكيد	حول	أذهان	بالفعل	دون	مشاهدون	مهدي
وصف	تولستوي	الكنيسة	تولستوي	مشاهدين	دور	بين	القصة

Negative Words with Applying Rules

و دينية	مجنون	صفر	فيما	وتقرر	السجدة	غياهب	وهما
عاما	درجة	وسبئ	به	شيء	شيء	النسيان	بعدها
مجنون	ولا	لا	رد	البقاء	شيء	غير	هوليود
صفرا	المشاهد	المشاهد	ودافع	فيصاب	مشهد	لا	مرور
عاما	مفتعلة	سخيف	عرض	جدران	رحلة	سيئ	مطاردين
سيئ	سلطان	مجنون	مطلقة	الجنسية	أماكن	حياته	القتلة
زايد	عاما	بمرور	نفسه	بالمشاهد	أخرى	كما	تواجه
ولا	ضعيف	سخيف	طرح	محاولة	لا	طوال	زواجهما
مجنون	سيئة	عاما	مشهد	مشاهده	الحزينة	وسنخرج	وبعد
رأي	مجنون	ويجب	الجنسية	قرار	مشهد	الجواب	الدامي
أخرى	تعد	جرح	لا	جنسية	شيء	الماضية	نفسها
درجة	عاما	المشهد	القرار	آخر	طويلا	سوى	نظرة
عاما	جانبها	صفرا	خال	مجنون	وهم	بسيطة	سرا
يكن	وسطحي	مجنون	مشاهد	مجنون	سيئ	لا	صادما
قررت	بعد	ضعيف	سواء	الجماهير	مجنون	التأثير	لتوها
يصنعه	درجة	والدليل	درجات	ولا	ضعيفة	وخلافاته	قتلة
يخجل	عدم	كتاريخ	وغيرها	جنسية	يعرض	يكن	لا
حدة	سيئ	درجة	باللفظ	كما	مجنون	والأراضي	لا
وغير	ومضمونا	عام	صفر	مشاهد	حاليا	شيئا	المشاهد
نمطية	مجنون	تعد	ومشاهد	الإثارة	مجنون	آخر	سوى
بطحها	سخيفة	تعبيره	جنسية	الدهشة	جديد	طرد	عزل
مجنون	درجة	عاما	مجنون	عرضه	فقير	طوال	مشاهد
الرغم	الرأي	وسخافة	مسألة	عرضه	رجال	ولو	خلقت
عاما	تيسير	متماسكا	عاما	فقد	الفقير	المشهد	بعدها
ويجب	ضعيف	صدمت	ولا	سبق	مصادفة	بتلك	أيضا

Positive Words without Applying Rules

موضحا	دون	ومصطفى	موجهه	أجل	إعجابي	بينهما	الرومانسية
الرسالة	العربية	إيناس	القراءة	يعيش	الأدبية	النجمين	بنجاح
بعقل	قبلة	دعاوى	التقييم	تولستوي	أجد	ذلك	نجمان
العربي	بودي	وصفته	مشاهدو	وأبدع	فكر	حقيقة	أولا
منطقي	الإطلاق	الإعلام	المتفرج	قارنا	والفكرية	جيدة	مشترك
يتجرأ	العمل	العرض	قصة	بالإمكان	لحظات	كلية	وتجد
خصوصا	فعل	دور	توصيفات	كامل	محيي	ليس	عائليه
عمل	حول	وأضاف	دور	ليس	الأرثوذكسية	مرة	القضاء
حسن	كبيراً	ليس	صناع	محيي	والحقوق	تشاهدها	زوجها
المراد	مثل	أثير	بعقل	تولستوي	أجد	مهما	تريد
فرض	مشيرا	أول	العربي	وفي	أجبرت	ولكنه	أول
مكتبة	مصطفى	بعيدا	بالشباب	تعمر	أريد	مشاهدة	علاقة
مؤكداً	سكنت	بطل	شخصياً	يمكن	تولستوي	التقييم	منطقة
وجارح	وبالتالي	بطولة	مؤكدين	قدمتم	فرعي	دقائق	أثناء
توفيق	الإيحاء	العرض	نورة	رئيسي	أفكاره	يرقى	الشاب
متسائلاً	علاقة	حوله	تتخصص	الرواية	وحتى	الرجل	الحب
يشاهد	مشاهدون	الكلام	شاب	أذهان	حول	الأصدقاء	بتكوين
مشيرا	ورجال	كوميديّة	بالفعل	حياة	كلمات	يريد	الخارجة
القصة	بين	مشاعر	بطولة	المشاهدين	الكنيسة	قدمها	يعطي
ثقة	أعمال	يتضمن	حول	والمنتجين	الأملاك	يقدم	المنطق
ومفيداً	المشاهدون	دور	مشاهدين	السجود	برسم	التأكيد	المتكررة
حتى	بين	جاء	تشبه	طبع	الأدبية	تولستوي	وصف
مهدي	قدمته	مسرحية	تقديم	أبطال	الهائلة	بالمتعة	تصدق
العاطفية	توصيلها	نجاحاً	تغرم	تعيشون	مشاهدتي	قطعية	والتسامح
لوجودها	جيد	يتيمة	وينقذها	كبير	منتجي	أشاهد	والحبكة

Negative Words without Applying Rules

مجنون	درجة	لا	فيما	جدران	رحلة	النسيان	هوليود
عاما	عاما	المشاهد	الجماهير	الجنسية	شيء	سيئ	وهما
ولا	ولا	صفر	كما	مجنون	المشهد	أيضا	بعدها
زايد	المشاهد	مجنون	لا	رحال	السجدة	بعدها	سرا
صفرا	مفتعلة	عاما	ولا	فيصاب	أخرى	غير	وبعد
سيئ	ضعيف	المشهد	مشاهد	فقير	مشهد	حياته	قتلة
مجنون	سيئة	جرح	نفسه	محاولة	لا	الجواب	زواجهما
درجة	مجنون	ويجب	سبق	مشاهده	طويلا	كما	الدامي
عاما	جانبها	بمرور	جنسية	مجنون	وهم	لا	لتوها
أخرى	ومضمونا	صفرا	مطلقة	قرار	شيء	طوال	القتلة
يكن	درجة	سخيف	القرار	جنسية	الحرينة	بسيطة	مرور
نمطية	بعد	مجنون	وغيرها	آخر	أماكن	الماضية	مطاردين
يصنعه	سيئ	ضعيف	سواء	به	سيئ	سوى	نفسها
قررت	تعد	كتاريخ	جنسية	رد	مجنون	وسنخرج	نظرة
وغير	عدم	والدليل	صفر	خال	يعرض	يكن	صادما
حدة	عاما	درجة	درجات	الدهشة	ضعيفة	طوال	تواجه
رأي	وسطحي	عاما	باللفظ	عرض	حاليا	شيئا	لا
بطحها	مجنون	متماسكا	ومشاهد	الجنسية	مجنون	والأراضي	سوى
يخجل	ضعيف	وسخافة	مشاهد	مشهد	مصادفة	وخلافاته	لا
مجنون	درجة	تعبيره	مجنون	مجنون	وتقرر	آخر	المشاهد
درجات	ودينية	عام	ولا	طرح	شيء	التأثير	عزل
الرغم	سخيفة	صدمت	مسألة	الإثارة	البقاء	طرد	مشاهد
ويجب	تيسير	تعد	سخيف	ودافع	الفقير	بتلك	خلقت
عاما	عاما	مجنون	وسيئ	عرضه	بالمشاهد	ولو	غياهب
لقاءات	الرأي	سلطان	عاما	فقد	جديد	مشهد	لا

APPENDEX 2: Samples of Intensification and Negation Words in Arabic Language

Intensification Words

افراط

فعلا

تماما

جدا

للغاية

كبيرا

شديد

منتهى

مطلقا

Negation Words

ليس

ما

لن

لم

لما

لا

غير

CURRICULUM VITAE

Credentials

Name, Surname : Ahmed NASSAR
Place of Birth : Libya
Marital Status : Married
E-mails : atn.20036@gmail.com
ahnassar@hacettepe.edu.tr
Address : Department of Computer Engineering,
Hacettepe- University, Beytepe, Ankara /
TURKEY

Education

High School : EIKARMAL Secondary School, Gaza /
Palestine
BSc. : Department of Educational Computer, Islamic
University of Gaza, Gaza / Palestine
MSc. : Department of Business Information-
Technology, Middlesex University-Riti, Cairo /
Egypt
PhD. : Department of Computer Engineering,
Hacettepe University, Ankara / TURKEY

Foreign Languages

Arabic (Native), Turkish (Advanced), English (Advanced)

Work Experience

[September.2014 – to Now]: Junior IT professional and teacher at 17 February school of Libya embassy, Ankara / Turkey.

[November.2010 – to 2012]: Lecturer at Information Technology department, College of Science and Technology, Gaza / Palestine.

[2009 – to 2011]: Part time lecturer at journalism and information department– Islamic University, Gaza / Palestine.

[2008 – to 2009]: Senior IT professional (Head of operating Systems department), Ministry of Transport, Gaza / Palestine.

[2007-2008]: Senior IT professional (Head of networks and maintenance department), Ministry of Transport, Gaza / Palestine.

[March. 2004 – October. 2004]: Junior IT professional (Programmer, systems analyst, Computer maintenance and Network manager in Fadda for Technology and Information systems), Gaza / Palestine.

Area of Experience

Sentiment Analysis, Natural Language Processing, Machine Learning, Information Retrieval, Data Base Systems and Operating Systems.

Projects and Budgets

-

Publications

-

Oral and Poster Presentations

-



HACETTEPE UNIVERSITY
GRADUATE SCHOOL OF SCIENCE AND ENGINEERING
THESIS/DISSERTATION ORIGINALITY REPORT

HACETTEPE UNIVERSITY
GRADUATE SCHOOL OF SCIENCE AND ENGINEERING
TO THE DEPARTMENT OF Computer Engineering

Date: 05/09/2017

Thesis Title / Topic: **MULTILEVEL SENTIMENT ANALYSIS IN ARABIC**

According to the originality report obtained by myself/my thesis advisor by using the *Turnitin* plagiarism detection software and by applying the filtering options stated below on **28/08/2017** for the total of **94** pages including the a) Title Page, b) Introduction, c) Main Chapters, d) Conclusion sections of my thesis entitled as above, the similarity index of my thesis is **8 %**.

Filtering options applied:

1. Bibliography/Works Cited excluded
2. Quotes excluded/ included
3. Match size up to 5 words excluded

I declare that I have carefully read Hacettepe University Graduate School of Science and Engineering Guidelines for Obtaining and Using Thesis Originality Reports; that according to the maximum similarity index values specified in the Guidelines, my thesis does not include any form of plagiarism; that in any future detection of possible infringement of the regulations I accept all legal responsibility; and that all the information I have provided is correct to the best of my knowledge.

I respectfully submit this for approval.

Date and Signature

Name Surname: AHMED NASSAR
Student No: N12146216
Department: Computer Engineering
Program: Computer Engineering-Doctor of Philosophy (Ph.D.)
Status: Masters Ph.D. Integrated Ph.D.

05-09-2017
Ahmed Nassar

ADVISOR APPROVAL

APPROVED.

Supervisor

Prof. Dr. Ebru AKÇAPINAR SEZER

(Title, Name Surname, Signature)