ISSN: 2690- 912X

Research Article

Journal of Genetic Engineering and Biotechnology Research

Arabic Opinion Mining System for E-Services Based On Ontology and Machine Learning

Maha Mansour*, Randa Elwakil ,Hoda Wageih and Magdy Aboul-Ela

Sadat Academy for Management Science, AI and Computer Science Department, Cairo, Egypt

***Corresponding Author**
Maha Mansour, Sadat Academy for Management Science, AI and Computer Science Department, Cairo, Egypt.

Submitted: 2023, Nov 13; **Accepted**: 2023, Dec 04; **Published**: 2023, Dec 26

Citation: Mansour, M., Elwakil, R., Wageih, H., Ela, M. A. (2023). Arabic Opinion Mining System for E-Services Based On Ontology and Machine Learning. *J Gene Engg Bio Res, 5*(3), 392-410.

Abstract
With the growing use of E-Services in the Arab world, there is a need for efficient methods to analyze customer feedback and opinions. This research is proposing an Arabic opinion mining system for E-Services based on ontology and machine learning techniques. The system utilizes an Arabic ontology to capture domain-specific vocabulary and relationships between features and a machine learning model to classify customer feedback into positive, negative, or neutral sentiments. By evaluating the performance of the system using several domains datasets for Arabic customer reviews of various e-services, and the results show that the system achieves a high accuracy level. Comparing the performance of the proposed system to traditional polarity and subjectivity approaches.

Keywords: Opinion Mining, Sentiment Analysis, Opinion Mining, Data Mining.

1. Introduction

Opinion mining in natural language processing has been the center point of research lately since e-service owners start to focus on user's needs and requirements. The main idea is where to search for customers opinions through different domains and how, Arabic E-service focus on the customer opinion at different domain levels and it is considered to be a hard task as it requires deep understanding of Arabic sentence structure and domain knowledge as early process of sentiment analysis focus on context with individual features only identifying and extracting subjective information from text data.

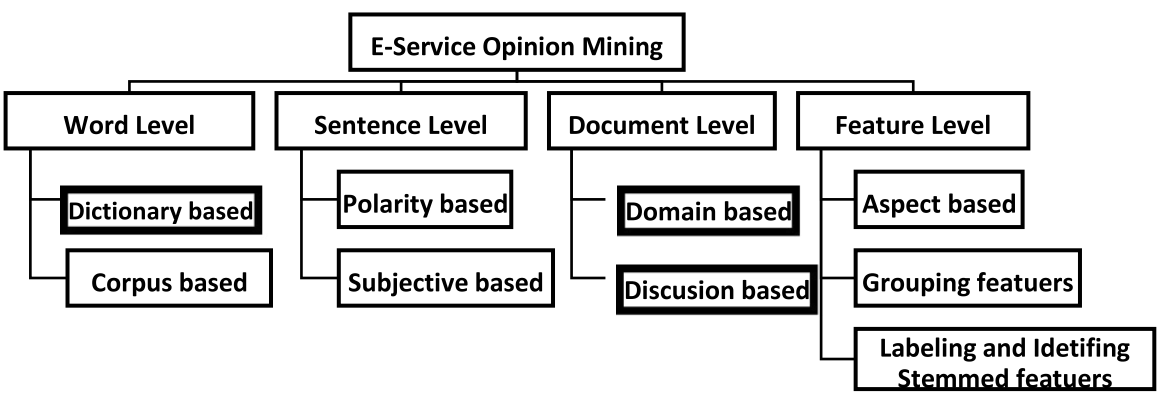

Figure 1: E-Service Opinion Mining Levels

Then machine learning helped focusing on supervised learning techniques that categorize texts like LR , KNN , Naïve Bayes and Support vector machine classifiers. Very few of Arabic sentiment analysis researches focused on more advanced levels of feature selection methods with useful information to the classifiers.

Addressing the different Arabic ontology feature selection methods including the change in feature weighting depending on sentence chunks and domain chunk. The weighting method include TFIDF, BOW, and Chunks All tests were conducted using both SVM and multinomial NB classifiers.

Opinion mining is the process of identifying and extracting subjective information from text data. It is a rapidly growing field of research, with applications in various domains such as marketing, politics, and social media. Arabic sentiment analysis is particularly challenging due to the complexity of the Arabic language and the lack of publicly available resources. In this paper, we will explore the use of machine learning techniques for sentiment analysis in Arabic, using an ontology based approach.

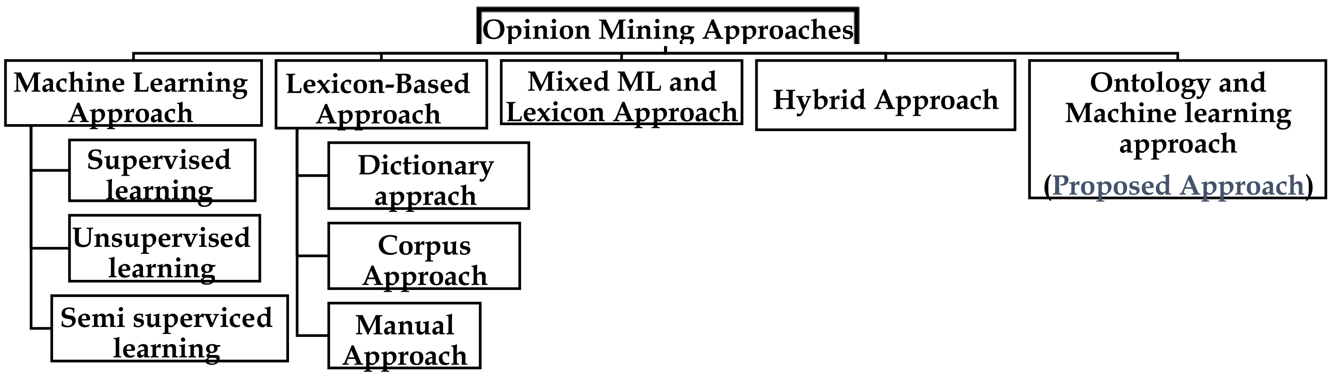

Figure 2: Opinion Mining Approaches

1.1. E-Services and Arabic Ontology

Ontologies play an important role in achieving interoperability across E-services and on the semantic web, because they aim to capture domain knowledge and their role is to create semantics explicitly in a generic way, providing the basis for agreement within a domain. In other words, the current web is transformed from being machine-readable to machine understandable. So, ontology is a key technique with which to annotate semantics and provide a common, comprehensible foundation for resources on the semantic web [1]. Ontology represents the science of existence where it is used to represent a domain knowledge model. Using ontology became a great value for text representation and opinion mining process such as clustering, classification and summarizing of words [2].

1.2. Research Objectives

Developing an Arabic opinion mining system for E-service based on a combination of ontology, which represents a domain knowledge, for mining opinions at the domain feature level and classifying the overall opinion on a multi-point scale. The proposed approach will analyze a collection of customer's reviews datasets at the domain feature level and produce a set of structured information that associates the expressed opinions with specific domain features. When dealing with the limitation of the corpus in Arabic sentiment analysis several features are proposed and investigated. In this research we will try to find the most suitable method to gather features that might work better with Arabic dialects through different types (subjectivity, objectivity or polarity) meaning and reserve its influence on the document level.

1.3 Contribution

• Developing an ontology based Arabic Opinion mining system, based on a set of rich ontology features, domain knowledge and lexicon information.
• By proposing an approach based on ontology BOW and ML classifiers, the system automatically determine the sentiment percentage of the user depending on the review subjectivity, objectivity and polarity.
• Filling the gap between lexicons classification and machine learning classification determining the ontology features and their subjectivity level to increase the process of sentiment analysis accuracy levels.
• Transforming the highest accuracy ratings of binary classification to even more accurate results with trinary classification giving every feature related to the domain its realistic weight
• Handling the Arabic stop words, negation, n-gram and stemming achieving better polarity from the dataset context.

2. Literature Review

Several studies have been conducted on ontology learning from Arabic text. In the authors provide an overview of the research that works on ontology learning for Arabic text and their results. They found that the research focused on extracting the concepts and relations from Arabic text and building the ontology. In the authors propose a model for automatic ontology extraction from Arabic unstructured text. They present a formal Arabic WordNet built based on a corpus of Arabic text. In the authors propose an ontology-based approach to enhance explicit aspect extraction in standard Arabic reviews. They use machine learning and ontology-based approaches to perform ontology-based sentiment analysis on Arabic reviews. Opinion mining is the process of analyzing opinions , sentiment , emotions and evaluating people feelings concerning a product or a business the field is still under a lot of researched and the Arabic market is facing a lot of challengers due to the different slangs , dialects ,expressions and people background [3].

2.1. Arabic Opinion Mining

One approach to Arabic sentiment analysis is to use a hybrid scheme that combines different features, including word-based features such as term frequency (TF), term frequency-inverse document frequency (TF-IDF), and bag of words (BOW) another approach is to use base phrase chunking, which involves identifying the base phrases in a sentence and analyzing their sentiment [4].

2.2 Feature Extraction Methods

Addressed the limitations of the bag-of-words (BOW) model in sentiment analysis (SA) of Arabic datasets. The BOW model treats words as independent features, ignoring semantic associations between them. This leads to synonymous words being represented as different independent features, reducing the accuracy of the model. To overcome this limitation, the research proposed enriching the domain representation with concepts utilizing Arabic WordNet (AWN) as an external knowledge base. Developing and evaluating different concept representation approaches with naïve Bayes (NB) and support vector machine (SVM) ML classifiers on an Arabic domain dataset. The experimental results show that using ontology features improves the performance of the ATSA model compared to the basic BOW representation. With observing an improvement of 4.48% with the SVM classifier and 5.78% with the NB classifier. The findings suggest that incorporating external knowledge bases such as Arabic word net can enhance the accuracy of SA models in Arabic domain observed the increase of the classification model performance by removing space, stop words, null, noise and applying stemming with feature correlation [5,6].

Combined approach of Machine learning with ontology information to have better opinion mining classification performance. This study used manual entry for the ontology tree to extract explicit product features from the review, also to determine the important features from the review, and to generate feature-based summary. It requires a domain name and number of levels of the ontology parameter and using Concept Net and WordNet databases to construct domain specific ontology tree. The research was conducted on only two Arabic domains with a total of 2000 reviews with equal number of positive and negative reviews. Subjective evaluation results were taken into consideration never the less due to the limited number of reviews and the pre-defined polarity of this approach with the manual ontology pre-feeding the accuracy of the opinion extraction process decreased neglecting the diversity of social media lexicon and dialects [7].

2.3. Negation Handling

An important aspect of Arabic sentiment analysis, as negation words can change the polarity of a sentence. Negation intensifiers terms are a valence shifter and the main obstacle is that they come in different forms and patterns like tokenized (مش حلوة), fake inverts (هلىن الب), odd negation (بعت شالب), implicit (بعتلا الول انحرف انك) and neutral targets (صقان شم) [8]. Different techniques have been proposed to handle negation in Arabic sentiment analysis, including using different word window sizes, base phrase chunking, and machine learning-based approaches words chunking is another technique that can be used to handle negation in Arabic sentiment analysis. This involves identifying the base phrases in a sentence and analyzing their sentiment. Negation words can be used to identify the scope of negation, and the sentiment of the base phrases within the scope of negation can be reversed [9]. SVM, NB, and K-NN classifiers reported better improvement on the results after applying the exceptional negation algorithm there are three baseline models, the first is baseline in which the simple uni-gram model is used without considering the negation problem. Secondly, a uni-gram model which consider a negation scope of five words that directly follow a negation term, where, each term within the scope will be tagged with the negation mark. The last one, is a uni-gram model for an inclusive negation scope that includes all the words that follow a negation term until the end of the sentence, where, each term within the scope will be tagged with the negation mark [10].

2.4. Research Gap and Hypothesis

All that concerns increase the need for developing an Arabic sentiment analysis system for e-services that leverages both Arabic rich ontology-based approaches and machine learning techniques to address the limitations of existing Arabic opinion mining systems and improve the accuracy and efficiency of sentiment analysis for e-services by leveraging domain specific knowledge represented in the ontology and machine learning techniques for feature extraction and classification. By addressing this research gap, the proposed thesis contributes to addressing large, divers and different types of datasets and domains.

3- Machine learning and Arabic Ontology

3.1. Machine Learning Classification

ML classification is the common approach for supervised learning and is typically used when conducting sentiment analysis. Supervised model is based on machine learning algorithms like Support Vector Machine (SVM), Naive Bayes (NB), Logistic Regression (LR) and K-Nearest Neighbor (KNN), the supervised technique is used to analyze the dataset. To train the machine learning classifiers a substantial amount of labeled data is needed. The aim of the training procedure is to create a classification model that can foretell the polarity of the testing dataset.In this research the following classifiers are used : Support vector machine (SVM), Naïve Bayes (NB), Logistic Regression (LR) and K-Nearest Neighbor (K-NN) to evaluate the performance of our method.

3.1.1. Support Vector Machine (SVM)

SVM is a Set of supervised learning methods used for classification, regression and outline detection methods. Very effective with large datasets and it saves a lot of memory. SVM is Support Vector Machine (SVM) perform a linear classification and also select a line with a maximum-margin. The process of selecting the best decision boundary is based on some points called support vectors. SVM looks at the interactions between features to a certain degree. SVM is good for high-dimensional spaces and in cases where the number of features is greater than the number of observations but in the case of huge number of features the accuracy rate start to decrease.

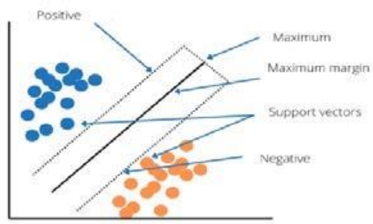

Figure 3: SVM Algorithm

3.1.2. Naïve Bayes (NB)

The Naïve Bayes Classifier belongs to the family of probability classifier, using Bayesian theorem. It is called 'Naïve' because it requires rigid independence assumption between input variables. So, it is more proper to call Simple Bayes or Independence Bayes. It is one of popular methods to solve text categorization problem, the problem of judging documents as belonging to one category or the other. Naive Bayes requires a strong assumption of independent predictors, so when the model has a bad performance, the reason leading to that may be the dependence between predictors. NB work based on Bayes theorem, so it assumes that any feature is independent than the other Thus, NB classifiers can learn easier from small training data sets due to the class independence assumption [11].

The Bayes rule is the main part of the Bayesian model this rule is calculated as follows: Where: P (c|x) is the posterior probability of the hypothesis, P(c) is the prior probability of hypothesis, P(x) is the prior probability of Evidence, and P (x|c) is the conditional probability of Evidence given Hypothesis (likelihood).Bernoulli Naive Bayes: It assumes that all our features are binary such that they take only two values. 0s Means can represent "word does not occur in the document" and 1s as "word occurs in the document" (Equation (3.1)).

$$\underbrace{P(c\mid x)}_{\text{Posterior Probability}} = \frac{\overbrace{P(x\mid c)}^{\text{Likelihood}}\overbrace{P(c)}^{\text{Class Prior Probability}}}{\underbrace{P(x)}_{\text{Predictor Prior Probability}}}$$

$$P(c\mid X) = P(x_1\mid c)\times P(x_2\mid c)\times\cdots\times P(x_n\mid c)\times P(c) \qquad (3.1)$$

3.1.3. K-Nearest Neighbor

KNN as a supervised learning model needs no training. This algorithm classifies data by measuring the distance between training and test sets in order to determine the nearest neighbors.

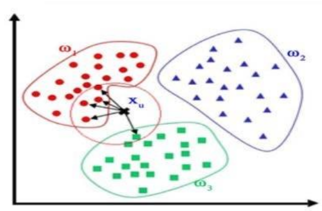

Figure 4: KNN Algorithm

To identify the short answer, two alternative methods are used. The first distance between the points is determined by computing the Euclidean distance, and the second distance is determined by computing the cosine similarity between the training and test data. The first strategy was applied in the context of this work. The documents are shown as points in a Euclidean space. This technique allows us to determine the Euclidean [12].

3.1.4. Logistic Regression

Logistic Regression (LR) is a probabilistic classification model using the sigmoid function. LR is similar to SVM in that they both can divide the feature space with a decision boundary. SVM also takes logistic regression to the next level by allowing non-linear decision boundary effectively thanks to kernel functions. In corresponding to supervised classification as a reliable and well-defined procedure.

The term "logistic" originates from the cost function (logistic function) with a form of Sigmund function with a distinctive S-shaped curve (Equation (3.3)). The LR is a sigmoid function-based transformation of a linear regression. The likelihood for a specific categorization is represented on the vertical axis, while the value of x is represented on the horizontal axis. It is presumed that y | x has a Bernoulli distribution. The formula of LR is as follows Here, β0 + βlx is comparable to the linear model y = ax + b.

$$F(x) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 x)}} \tag{3.2}$$

The logistic function uses a Sigmund function to tie they value from a broad scale to a range of 0, 1. Multinomial logistic regression is a more generalized version of logistic regression that models a categorical variable with more than two values [13].

3.1.5. Random Forest

A supervised machine learning approach based on ensemble learning is known as random forest. It creates a forest of trees by combining various decision trees, hence the name "random forest." Both regression and classification tasks can be performed using the random forest approach.

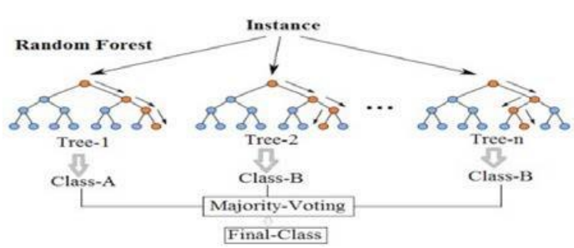

Figure 9: Random Forest Classifier

3.1.6. Voting Classification

A voting classifier is an ensemble classification technique that has the benefit of merging predictions from various machine learning algorithms via majority voting. It makes use of each algorithm's strengths. In this work, a number of single classification method combinations are employed with a majority vote. RF, KNN, SVM, NB, and LR were incorporated in the voting classifier [14].

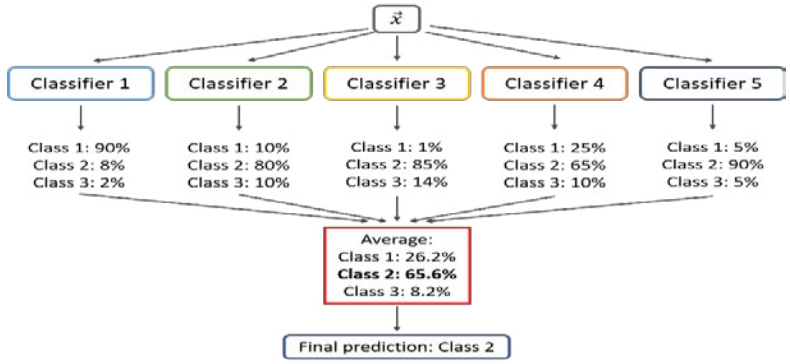

Figure 10: Ensemble Voting Classifier

3.2. Arabic Ontology Building

Ontologies can be used to provide a background knowledge in a similarity based analysis while machine learning is used to build the feature engineering heterogeneous data accessible to machine learning workflow.Ontology based mapping is used in machine learning workflows to generate better features for every learning domain model the effectiveness of ontology bases mapping in machine learning workflow can be measured comparably to every domain richness. Ontologies can be used to compute similarity and incorporate them in machine learning methods [15,16]. Ontological Modeling can help the cognitive machine learning model by broadening its' scope. They can include any data type or variation and set each diver data to a specific task.

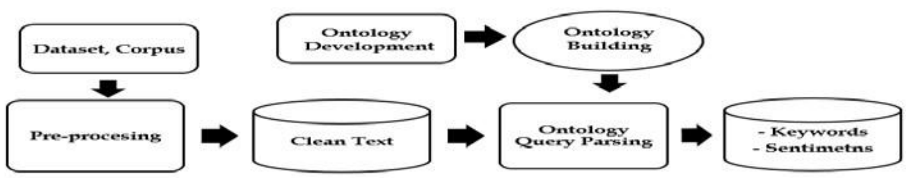

Figure 5: Ontology Building Process

Domain ontology is a formal representation of concepts and their relationships in a specific domain. It consists of four main components: Entity-Class, Entity, Relationship, and Attribute. Entity-Class represents the superclass of concepts and attributes, while Entity represents the subclasses or individuals. Relationship represents the defined object properties between classes in the domain ontology, and Attribute represents the defined data properties of each class.

3.3. Challenges of ML and Arabic Ontology

The ML algorithm can be used in the process on ontology mapping to create a more specific domain for the oncology text heretical setup and preparation also it is used to classify the text data into categories and to compare the performance of the ontologies-based domain algorithms. The ontology-based method will be used to improve the accuracy of the classification by incorporating domain-specific knowledge into the dataset [17].

Secondly, the lack of standardized Arabic datasets for ML training and evaluation can make it challenging to develop accurate and reliable models. Thirdly, the scarcity of Arabic-language resources, such as ontologies, can make it challenging to develop robust models. Additionally, ML can be used to develop an Arabic ontology for social media analysis that can provide insights into the sentiment and opinions of Arabic-speaking communities on various topics.

4- Proposed System for Arabic Opinion Mining

The proposed system is composed of two main components: ontology-based semantic annotation module and machine learning sentiment classification module.

4.1. Overall System Design

The methodology is divided 4 steps, first, gathering datasets with reviews from different websites about a variety of Eservices domains .the research divided the dataset into 30% testing and 70 % training, started with 10% for the test then by keeping on adding to the test size till 40% till it reached that the ration 30-70 provides more useful and meaningful insights from the trigrams generated ,then the research created two data frames for testing and training every data frame concatenated an array of testing and training .To describe the system methodology, a number of resources were used from the internet consists of five different datasets for Book, Hotel, Printer, Mobile and Mobile Application. Then by comparing each extracted domain ontology with other mutli-domain dataset to evaluate the model, each dataset consists of different size and ontology features and generated an ontology lexicon from the datasets. Using OOP to annotate the datasets to primarily polarities.

Datasets	Book	Hotel	Mob.App	Printer	Mobile
No. of reviews	60,055	20,047	11,000	1511	1500
No. of Sentence	348,205	110,543	80,245	11,522	11,309
No. of Positive	20810	14111	6300	945	566
No. of Negative	4844	2100	2550	322	420
No. of Neutral	34401	3836	2150	244	514

Table 1: Tested Domains Datasets

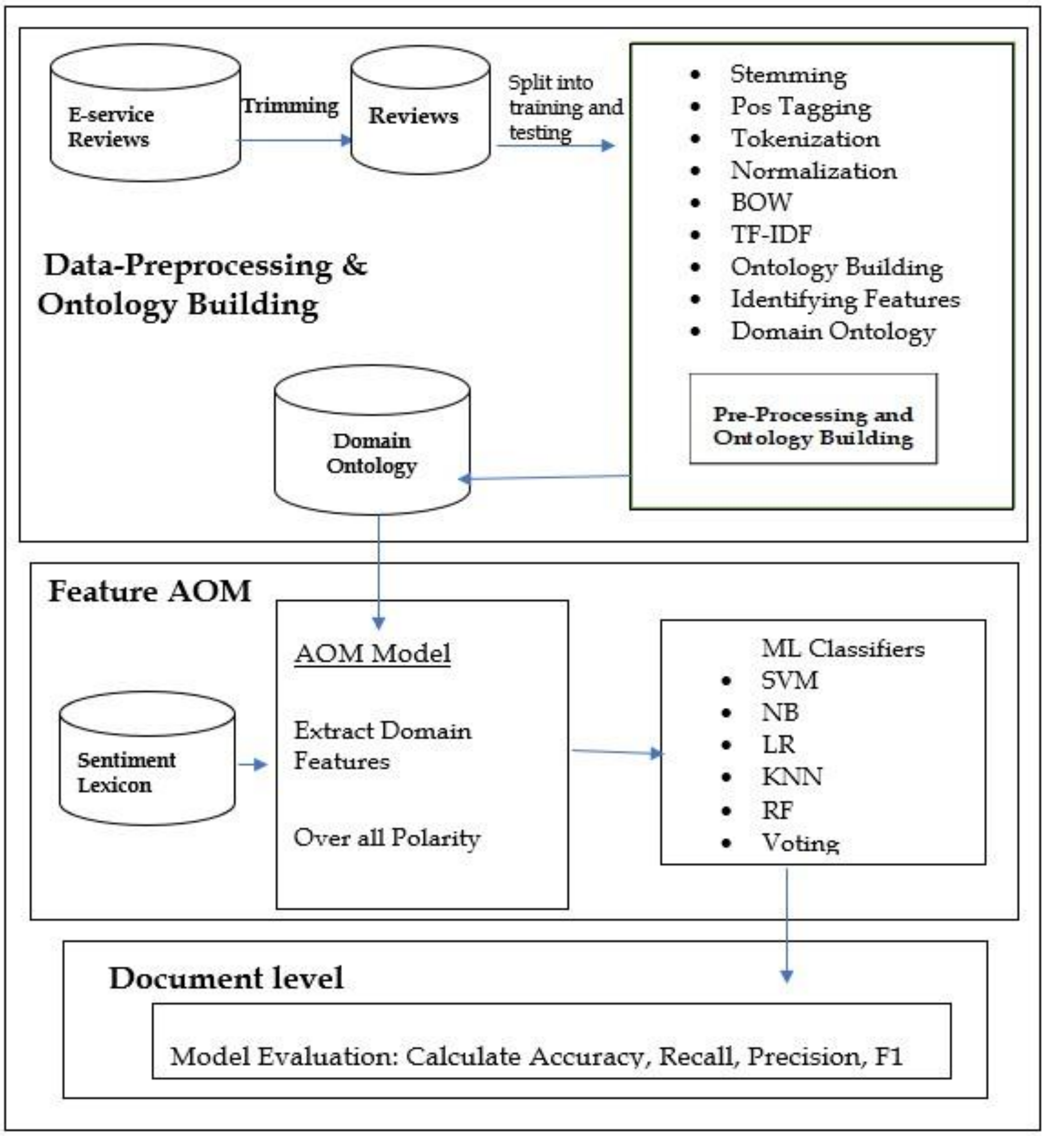

Figure 11: Arabic OM System

4.2. Data Preparation

To describe the system methodology, used a number of resources were used from the internet that consists of five different datasets for Book, Hotel, Printer, Mobile and Mobile Application.

Comparing each extracted domain ontology with other mutli-domain dataset to evaluate our model, each dataset consists of different size and ontology features and generated an ontology lexicon from the datasets. Using OOP to annotating the rest of the datasets to primarily polarities.

By classifying the Arabic sentiment lexicon using ArSenl Arabic SentiWordNet as every domain lemma has a meaning and a combined polarity that indicates weather the text is positive or negative or a mixed text value. As it contains POS tags (noun, verb, adjective and adverbs). The lexicon presents an advanced NLP processing. In the past studies, the sentiment analysis features of a certain domain was described by a lot of means like aspect, conceptual and notions this research focus on ontology studying and representation, by gathering the document text with bag of words and applied TF-IDF to create a feature list from every document-based o that list a label for the document is placed, the other list contains the words surrounding this every label. This labels is a group of keywords, which have a different weight in the domain for each group of documents.

Domain Features	Semantic Synonyms
موبيل	الارسال – الشبكة – التصميم –الجهاز – المحمول –البطارية – الغطاء – كفر – جراب –شاحن- سماعه – شاشه – سوكت

Table 2: Example of semantic synonyms for Mobile Domain Features

The Arabic language as an NLP involves many issues, such as morphological complexities and dialectal varieties. Thus, it requires progressive pre-processing and lexicon-building steps. In The pre-processing steps cleaning the dataset from the irrelevant and unnecessary data and punctuation marks, each review is split into a set of distinct sentences.so number of techniques were applied, normalization, stop word removal, tokenization, POS tagging, and stemming. Table 3.

Tokenization	The Arabic sentence or phrase is partitioned into tokens or words **الخامة جيدة ويأتي بجراب و في شاشة حماية السعر ليس باالجيد والاداء العام متوسط**
Normalization	**الخامة جيدة ويأتي بجراب في شاشة حماية السعر ليس بالجيد والاداء العام متوسط**
Stopwords	Remove stopwords using the list defined in the Isri stemmer **الخامة جيدة يأتي بجراب شاشة حماي السعر ليس بالجيد الاداء العام متوسط**
POS Tagging Stanford parser	Generate parse tree and determine the POS for each token **“الخامه DTNN” - ” جيدة ADJ ” - ” يأتي VB ” - ” جراب NN ” - “ شاشة NN” “السعر NN” - ”ليس ADV ” - ” الجيد DTADJ” - ” الادا DTNN” - “العام DTADJ “ - ”متوسط ADJ”**
Combination of Pos-Tagging used in the research	**1-Adj+Neg, 2-Adv+Adv, 3-VB+Neg, 4-VB+Adv, 5-VB**
Light Stemming ISRI	**متوسط + العام+الاداء + جيد + ليس+حماية+ السعر + شاشة + يأتي +جراب + جيدة+الخامة**
Root Stemming Farasa	**ال+خام+ة جيد+ة و+يأت+ي ب+جراب+في شاش+ة حماي+ة ال+سعر ليس ب+ال+جيد و+ال+اداء ال+عام +متوسط**
Stemming Algorithm used in the research	**Comparing Arabic stemming template, exceptions and morphological rules. Apply Arabic expressions and Arabic word base heuristics.**

Table 3: Dataset Pre-Processing

Stemming or Text Standardization, is converting the word into its root form. For example, “المراجعة ,” المراجعات “ and “المراجع”, all will be stemmed “راج ع” . The Stemming approach finds the three-letter roots for Arabic words without depending on any root or pattern files. Light stemming removes the common suffixes and prefixes from the words [18]. I needed to collect the main features in the reviews to determine which feature is more important to our OM process and generate feature-based ontology. The research used The Information Science Research Institute’s (ISRI) Stemmer for Arabic text browser this stemmer does not use root dictionary. Also, An Arabic Stemmer called Farasa, ISRI stemmer returned normalized form, and Farasa returned the root stem word rather than returning the origina unmodified word.

Normalization, on the Arabic words is performed to make all letters written in the same format which will help alleviate spelling variations which will provide better recall rate.eg “بااااااالجدی” to “بالجدی”.

Assigning a grammatical category to the given word. It’s commonly referred to as POS Tagging. Stanford Arabic part of speech tagger is used to tag the words. This information will help us identify Feature term and opinion words. To determine the Feature term, all the terms with noun category are extracted. As well, to identify the opinion words, all the terms with adjective and verb category are extracted.

4.3. Feature Engineering

Feature extraction: Extract features from the preprocessed Arabic text data. The features selection methods are bag of words, chunking, tied (term frequency-inverse document frequency), and chunking.

4.3.1. POS-Tagging

Sentiment words, using POS –tagger, where the nous, adjectives, adverbs are considered a domain feature that illustrate the building of domain ontology. The domain dictionary contains mainly certain nous the other verbs and adjectives are supporting sentiment word in the corpus. The tags of the POS-tagger are.

POS	Notation	POS	Notation
NN	Noun	DT	Determiner
VB	Verb	DTNN	Noun with DT ال
ADJ	Adjective	DTADV	Adjective with DT ال
ADV	Adverb	NEG	Negation

Table 3 : Pos-Tagging tags

With using I used Stanford part of speech tagger tool to collect each review and to get the Arabic POS tag for each word divide them to noun, verb and adjectives. POS tagging can detect noun, verb, present tense and adjective, applying this to the domain ontology and excluding all the not related words like “اوي” and “بشدة” [19]. The research matched the sentiment words with the Arabic corpus our matching method was built on matching any of the word from the original word, the stemmed word or the word root then ignored the unmatched sentiments. Used ISRI and Farasa stemmers for matching the sentiment word which is supported in python.

4.3.2. Handling Negation

Negations and reflective are collected during the identification of our domain feature sentiment and handled with the Farasa stemmer as the Arabic language dialects have several expressions for the negation forms adding a negation expression before a verb, noun or adjective will enhance the negative polarity and remove the sentiment from the positive list to the negative affecting the weight of the feature in the evaluation measure, With using The ArSenL lexicon to detect the negation words it was much easier to enhance the system the ArsenL for better results .

Intensifiers are considered a power word of the sentiment like (جدا) they are added after the sentiment to exaggerate the meaning and give it high strength levels so we doubled the positive sentiment of the words with intensifiers to increase the weight of the positive polarity in the evaluation measures. At the Lexicology level Most of the work use adjectives only for sentiment analysis, and some of them use nouns, verbs, adverbs or a combination of them.

4.3.3. N-gram

Sentiments of the domain features are extracted using the N-gram model in which the basic most informative words will be preserved and used to build our feature model. In order to collect the sentiment word which is related to the domain feature the research tested several models where Unigram model: the BOW feature will contain only one word of the corpus, bigram: the BOW contains a combination of two words, tri-gram: the BOW contain a combination of three words.

4.3.4. Bag of Words (BOW)

Using bag of words as a feature to the model depending on a distinct word in the corpus meaning. Bag of words level where the main morphological aspects are used to build the feature model and a distinct word in the Arabic corpus are used for getting the best results.

4.3.5. TF-IDF

Term frequency is the number of times a particular word appeared in the comment. TF-IDF measure is a main important feature extractor for all the past systems, also inverse document frequency implies the common terms in each document the research used TF-IDF Vector for Feature Selection and Extraction so the text could be converted into an understandable form.

To examine the influence of weighing scheme, the research computed the weight of each feature using two different methods: TF-IDF (term frequency–inverse document frequency) and BOW (Bag of Words) provides different models of sentence representing.

4.4. Arabic Ontology Selection

Building an Arabic semantic ontology of the domain and measure each word hierarchical tree compared with the Arabic corpus retrieving all the facts in the ontology level using python , OOP and SQL query to select similar words from the Arabic dictionary to the domain all the facts of the domain ontology retrieved with their types (class, entity, object ,property, inheritance).

The gathered data are organized to features and semantic groups with an indicator of their class hierarchy level and type which is either an entity or an attribute. After organizing the data retrieved from the corpus to keywords (features/concepts) and semantics, by gathering all of this entries the building of a dictionary of semantics represents the domain ontology.

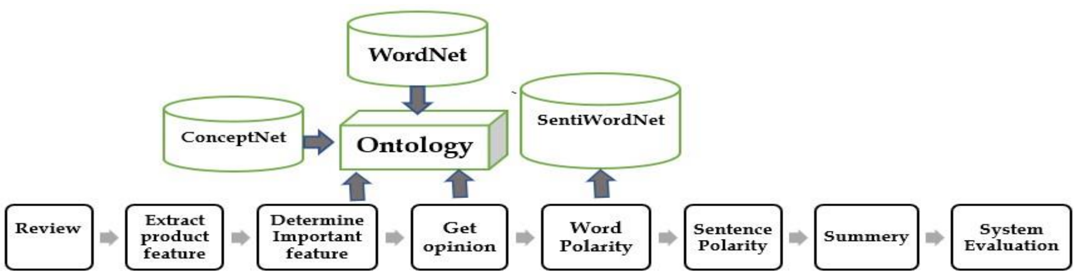

Figure 12: Ontology Based Arabic OM System Diagram

Gathering all the Arabic dialect in the dataset using SentiWordNet and BOW chunks as an associating semantic for extracting the text features, as the data is collected in unigram , bigram and trigram models depending on the classification of the text vectors , collected a group of useful features with close meanings in chunks with different weight in the dataset wand with measuring e measure the TF-IDF of the features to calculate which feature is considered a domain ontology , the frequency of the feature determine the importance of the word in the domain ontology tree.

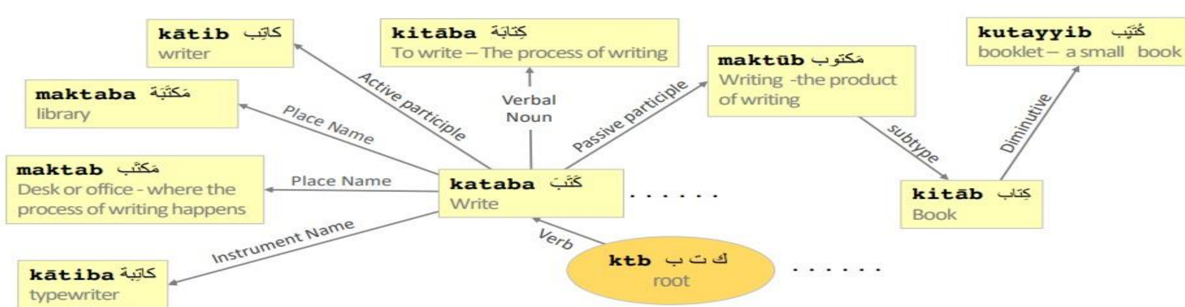

Figure 13: Ontology Cleaning Process from Arabic Corpus

4.4.1. The bag of words approach: involves representing text as a collection of words, ignoring grammar and word order. Each word is treated as a separate feature, and the frequency of each word is used to represent the text. This approach is simple and effective but does not capture the context of the words [20].Use the bag of words representation model to represent the preprocessed text data as vectors of word frequencies.

4.4.2. Chunking approach identifies and groups together related words in a sentence. This technique can help capture the context of words and improve the accuracy of sentiment analysis Chunking is a technique used in natural language processing (NLP) to group words together based on their part of speech (POS) tags. We used it to extract meaningful phrases from gathered texts [21].

4.4.3. Parts of speech (POS) tagging is the process of labeling each word in a sentence with its corresponding part of speech, such as noun, verb, adjective, etc. This technique can help identify the role of each word in a sentence and improve the accuracy of sentiment analysis. It also highlight each domain keywords. To identify the role and context of words in the text data, and filter out irrelevant or noisy words.

4.4.4. Noun parsing is important NLP technique used to join important keywords into a hierarchical model in order to improve the ontology performance and accuracy. The parsing process starts with the parse tree textual unit which represent a data structure of sentences.

The corpus entries comprise of two arguments: the semantic category (type) and the list of concepts belongs to it as shown in the following structure:

Dictionary = { < Y1, C1, C2,..,Cj,..Cm > < Y2, C1, C2,..,Cj,..Cm > < Yi, C1, C2,..,Cj,..Cm >
Y is the semantic groups; Y = {Y1, Y2,..,Yi,..Yn} And C is the concepts C = {C1, C2,..,Cj,..Cm}
An ontology of predefined utilities is built based on examining the most frequently mentioned feature in the domain dataset. Matching based on the ArSenl noun phrases to their opponent semantic categories.

Then using a Professional Arabic corpus matcher to find all matches of phrases against the domain features. The exacted corpus chunkier based on Exponential search for matching phrases to their semantic categories, this created bag of words chunks from the corpus to create Arabic ontology based on bow chunks with the domain feature and their semantic category. The semantic categories were constructed based on the features retrieved from the domain ontology knowledge represents the concepts in the domain ontology which is the following types:

- Entity-Class: the super class of concepts and its attributes (such as ق دنف or باتك class).
- Entity: the concepts that can be subclasses or individuals (such as ه حفص).
- Relationship: the defined object properties between classes that is related to the domain ontology such as (هحفص هب باتكلا).
- Attribute: the properties of the entity class in the domain ontology such as (تاملك).
- Attribute-Value: data properties values for each class in the domain ontology such as (ة ىؤورقم).
- A group of predefined values of attributes which is built for class attributes based on the most frequently mentioned features in the domain during the process of creating domain ontology, all the domain data retrieved and organized to features and sets of semantic groups with the indication of their level and class hierarchy and type.

4.5. Methods

4.5.1. Polarity Level

The research considers the sentiment word to emphasize a meaning and to indicate the meaning strength and weaknesses some intensifiers are added after a word to exaggerate the word or a negation to reflect to another polarity and change the test polarity levels. Building a feature model just based on the test level polarity is an excellent place to start in sentiment analysis but the feature model has to include more details about the domain and the ontology of the text this is called semantic orientation. Domain ontology and features is the main useful judgement on the sentiment lexicon so the research have made a model to calculate the text polarity level of sentiment hoping to prove that domain ontology research will achieve better results.

```
Algorithm : Polarity level Feature Extraction

INPUT: Dataset, Arabic _Lexicon dictionary {T, BOW}
OUTPUT: Sentiment Labels δ = {Pos, Ng, Nu} :

3: PosCount= Number of words Positive
4: NegCount= Number of words Negative
5: PosScore = The accumulated Positive
6: NegScore = The accumulated Negative
7: for all τi ∈T do
8: if PosScore > NegScore then
9: τi= Positive 10: else if NegScore > PosScore then
11: τi = Negative
12: else [PosScore = NegScore]
 13: if PosCount > NegCount then
14: τi = Positive
15: else if NegCount > PosCount then
16: τi = Negative
Else τi = Neutral
17: end if
```

As a result, when this feature is included, the ML methodology is employed as the primary classifier and is supported by some of the semantic technologies. Although relying solely on the word itself when creating a feature model is a solid place to start in sentiment analysis, the feature model needs to include more details about the text. When the ML methodology is employed as a primary classifier and is supported by some of the semantic technologies, the addition of this feature results in a hybrid method.

4.5.2. Subjectivity Level

In order to obtain a better value, the research included subjectivity level analysis used SentiWordNet.

We created a vector model to include all the features and their representation this model preserves essential domain knowledge each phase in this model displays the feature representation and values using BOW as a feature and SentiWordNet calculating the whole sentiment of the document we reached the subjective level classification results.

Algorithm : Subjectivity_level Feature Extraction

```
INPUT: Dataset, Arabic _Lexicon dictionary {T, BOW}
OUTPUT: Sentiment Labels δ = {Pos, Ng, Nu}

3: PosCount= Number of words having Positive Sentiment Intensity
4: NegCount= Number of words having Negative Sentiment Intensity
5: PosScore = The accumulated Positive Sentiment Intensities for each review
6: NegScore = The accumulated Positive Sentiment Intensities for each review
7: for all τi ∈T do
8: if PosScore > NegScore then
9: τi= Positive 10: else if NegScore > PosScore then
1: τi = Negative
12: else [PosScore = NegScore]
13: if PosCount > NegCount then
14: τi = Positive
15: else if NegCount > PosCount then
16: τi = Negative
Else τi = Neutral
17: end if
18: end for
```

Depending only on the post-tagging identification of the text weather it a noun verb or an adjective the model start classifying the sentiment passed on their feature. The sentiment in the phrase has a grammatical structure that consists of subjective and objective meaning. The phrase has the domain features represented in sentiments and an opinion object that carry less sentiment information but most of the opinionated information. Most of the studies ignore the noun (الكتاب) and focus on the verb (ادج دىج) to get the overall polarity. Incorporating semantic concepts to the review features space. In this level the model is deployed based only on the subjectivity level of the domain to register the change in domain before adding the ontology level sentiment analysis. The model collects all the verbs, adjectives and adverbs concerning the domain and extract the result.

4.5.3. Ontology Level

In this strategy, reviews are represented by their ontology tree concept including objective and subjective words. The extracted domain features are included in the sentiment analysis process, the implicit and explicit value of the sentiment are taking into consideration and reviewed.

the research used the polarity level and the subjectivity level sentiment analysis to make this collective level sentiment analysis process adding to them domain knowledge and ontology representation as the research believe that it might increase the accuracy levels of the sentiment and gives a deeper better results , users express their feelings with placing actual meaning about the domain and combining the different expressions to form an ontology importance value is the main concept of this research therefore the research investigated the Arabic domain ontology by gathering the most subjective features of the domain in BOW chunks depending on their logistic meaning and value in the phrase simple this chucks gather words in the text that represent the same meaning I applied a weighting process.

Considering review sentience segmentation, sentiment pattern of the review and the novel in the classification. The following ML classifiers were used (SVM, K-NN, LR and Naive Bayes) to classify the collected data into negative-, neutral or positive class. For the SVM classifier, a set of experiments were carried out with different values in order to choose the value that allows us to achieve the best results.

```
Algorithm : Ontology_level Feature Extraction

INPUT: Dataset, Arabic _Lexicon dictionary {T, BOW}
OUTPUT: polarity and ontology Labels δ = {Pos, Ng, Nu}

3: PosCount= Number of domain words having Positive Sentiment Intensity
4: NegCount= Number of domain words having Negative Sentiment Intensity
5: PosScore = The accumulated Positive Sentiment Intensities for each domain review
6: NegScore = The accumulated Positive Sentiment Intensities for each domain review
7: for all τi ∈T do
8: if PosScore > NegScore then
9: τi= Positive 10: else if NegScore > PosScore then
11: τi = Negative
12: else [PosScore = NegScore]
13: if PosCount > NegCount then
14: τi = Positive
15: else if NegCount > PosCount then
16: τi = Negative
Else τi = Neutral
17: end if
18: end for
```

The experiments in this section used different ML classifiers to carry out our approach in Arabic sentiment analysis using a set of features. Working on the Arabic language and Egyptian colloquial and most internet users who write in these languages are expressing their opinions and feelings with sarcastic way, old wisdom, old saying and idioms, which exceeded more than 20% of the total opinions.

First, all data are grouped into the algorithm and the process of the grouping started based on the features of each group then the research have a database of all the words of our corpus, to investigate the model in different domains the tags were gathered by the meaning and by similarities then by applying the BOW model the word is attached to its feature in the cluster as it follows the post then by combining the clusters with the BOW the best ontology feature possible for the model will be achieved.

4.6. Evaluation

Sentiment analysis metrics are used to evaluate the performance of the sentiment analysis system in identifying the intensity of sentiment in the text. The commonly used sentiment analysis metrics include:

- Polarity: It measures the degree of positivity or negativity in the text.
- Subjectivity: It measures the degree of personal opinion or emotion in the text.
- Confidence: It measures the level of confidence in the sentiment analysis result.
- Agreement: It measures the degree of agreement between the sentiment analysis result and human annotators.

4.6.1. Polarity and Subjectivity

Two main measures were used Polarity and Subjectivity then by adding the ontology features measure the weight of the sentiment is calculated. Polarity ranges from - 1 to 1 (1 is more Positive, 0 is Neutral, - 1 is more Negative, Subjectivity goes from 0 to 1(0 being exceptionally objective and 1 being extremely subjective).

P is the polarity, R is the reviews, V is the -ve , +ve and neutral, Tags (t) and frequency (F) of each review. First, we calculated the polarity level without the Subjectivity factor with the following equation:

R= (t1, t2 …tn), Polarity (R) = ((V1, F1), , ……, (Vn, Fn))

Then we calculate the subjectivity of the review with (S) Srr=S(R)= (V1, V2, …..Vn)

After extracting the review subjectivity, the polarity could be calculated again differently from the classification of polar (V) and frequency (F) of each review subjectivity of the review towards +v function e or -ve by this polarity:Polarity (Srr)= {(V1, F1), (V2, F2), ……, (Vn, Fn)}

Then, the polarity of the review is Polarity (R)= Max ($SSrr$) ((V1, F1), (V2, F2), ……, (Vn, Fn))

By classifying the review according to the major sentiment in the phrase every sentence will be giving a weight as mentioned before 1, 0, -1 this weight is calculated as follows: If review (R) consists of (n) phrases (P) then the weight (W) of a (j) review (Rj) starts with the weight of phrase (i) which is calculated by equation of i (Pii) is W(Pii)={1, 0, -1} Then the weight (W) of the whole (j) review(R) is:W(Rjj)=$\sum^{nn}_{ii}$=1$WW(PPii)$

And the polarity of the review

Polarity (Rj) = {Neutral if W(Rj)=0 , Positive if W(Rj) > 0, Negative if W(Rj) < 0}

4.6.2. TF-IDF

There are several weighting schemes, unigram features and SentiWordNet groups one of the unigram features is the Term Frequency Inverse Document Frequency (TF-IDF) which is the most used scheme for the task of information retrieval and text classification and due to previous experiments it have proven to have the most accurate results [22].The domain features of every review have an initial polarity value depending on the semantic word in the corpus.
TF (t) = (Number of times term t appears in a document) / (Total number of terms in the document).
IDF (t) = log_e(Total number of documents / Number of documents with term t in it).
W (t) = TF (t) X IDF(t)

4.6.3. F-Measure

Using Precision, Recall, and Accuracy, which are well-known evaluation criteria, to compare the findings in the current evaluation. To evaluate the dataset, some matrices will be used with the system [23]. Accuracy The percentage of the total sample that predicts the correct outcome. To calculate the accuracy of the system the research has compared the ML classifiers performance using the accuracy meter of the test set with the system outcome.

- **Precision**

Calculating precision is done by dividing the number of correctly classified as positive results by the number of overall results that were classified as positive whether the classification was correct or not. The precision is then calculated as follows.

- **Recall**

Recall. This measurement indicates the percentage of the completeness of the result. Low recall would mean that there are many left-out results that were not classified. Recall is calculated by dividing the number of correctly classified results as positive by the number of all positive correct results that should be classified as positive. Recall will be measured as follows.

- **F1 Measure**

The F1 value represents the weighted average of Accuracy and Recall. When the numerical difference between Accuracy and Recall rate is large, the F1 value can effectively combine the two indicators. The formula for calculation of the F1 value is as follows:

$$\mathrm{F1} = \frac{2 \times \mathrm{Precision} \times \mathrm{Recall}}{\mathrm{Precision} + \mathrm{Recall}} \tag{4.1}$$

The next section presents the evaluation metrics used to assess the obtained results. During every model, the F1 metric is calculated that measures the accuracy of the model depending on the subjectivity and polarity. Finally, a weighted average of F1 is calculated, resulting in a single value. The weighted average is calculated as:

$$F1_{weighted\ average} = \frac{\sum_{i=1}^{n} W_i \cdot f_i}{\sum_{i=1}^{n} W_i}, \tag{4.2}$$

Where f is the F1 score for each class, and W is the numbers of documents or sentences that are used in the testing data in each class. This algorithm is programed to calculate the result of the experiment depending on each sentiment analysis level.

5. Implementation and Results

5.1. Experimental setup

The research have implemented the proposed techniques in the Python language with NLTK, Pandas, NumPy, MatPlotLib , Seaborn, Joblib , sklearn and Scikit Learn Python libraries. Furthermore, the experiments ran on Windows 11 Pro with a 12th Gen Intel(R) Core(TM) i9-12900KF 3.20 GHz processor and 32.0 GB RAM

5.2. Data set

This research in opinion mining for Arabic sentiment analysis main purpose is helping e-services to provide the best quality for online users so the most applicable kind of data would have to be found on websites specialized in selling e-service products. The research collected five different domain datasets with different sizes and shapes.

Dividing the dataset into several parts with equal proportions of samples in each class. Most of them are used to train the model while the rest will be used to test the model that is generated during the training process. That means the will be trained on 70 of the data and used on 30% for testing. This process will be repeated through every dataset.

5.3. Text Preprocessing

During abducting this experiment, I found it is difficult for to work with direct Arabic language sentences so the dataset needed a phase of preprocessing before running as I used to work with English Lexicons. I had to write a code for cleaning the Arabic texts by removing stop words, hashtags, and emoji's punctuations which is provided by the NLTK corpus.

Then with the help an assistant parser tool RegexpParser class from the NLTK library to perform chunking of the related tokens using a regular expression grammar, and joins the chunks back into a string parts of speech tagging. The chunking is made based on the feature bag of words and on the TF-IDF features. The data-set is divided into training and testing sets 30-70%. Finally, a machine learning algorithm is used in the classification process, such as Linear Regression, Naive Bayes, KNN and SVM, to classify the sentiment of the text based on the gathered features.

(ISRI) Arabic Stemmer tool is used for stemming reducing words to their roots in the ontology. Also removing some stop words like ("م ن ","الذي ","اذا", etc.) and using a professional root Arabic stemmer Farasa to keep only valence shifter (e.g.," لم" ,"ابدا" ,"ليس", etc.) words that is important part in the domain and for better handling of the negation negation expressions.

An n-gram allows us to see which words tend to occur together. It is helpful in capturing negated words. The research tested unit-gram, bi-gram and tri-gram representation.

5.4. Experiments

5.4.1. Polarity Approach (baseline)

In this experiment establishing a result based on word sentiment of the domain provide basic knowledge about the domain and the domain text, then by using a primarily feature of bag of words depending on the corpus tested the experiment on different ML classifiers. Table .6 shows the result of the polarity sentiment level.

In regard to the datasets, the Book Domain was the hardest domain to classify, duo to its subjectivity sentence level of the classification and due to the huge number of neutral reviews that the system took forever to analyze, also there was a great different between domains as some of them describe the sentiment explicitly and other implicitly so some reviews were not meaningful or clear.

		SVM		**GNB**		**KNN**		**LR**		**RF**		**Voting**	
Datasets		Acc.	F1	Acc.	F1	Acc.	F1	Acc.	F1	Acc.	F1	Acc.	F1
Book	60,055	66%	63%	62%	61%	57%	55%	69%	62%	61%	59%	69%	62%
Hotel	20,047	80%	77%	68%	66%	78%	75%	84%	81%	75%	72%	84%	81%
MobApp	11,000	70%	68%	77%	78%	83%	80%	86%	83%	71%	68%	86%	83%
Printer	1,511	79%	77%	76%	76%	79%	78%	68%	65%	75%	73%	79%	77%
Mobile	1,500	74%	74%	66%	65%	58%	55%	77%	71%	70%	68%	77%	71%

Table 4: Polarity Level Ml Classifiers Accuracy and F1 measure

Although usually the MNB gets the best results for polarity classification in one domain but in this case but it didn't happen indicating that maybe MNB work better with larger reviews sizes and with sentiment containing bi-polarity of only positive and negative. The research further investigated that classifier and compared the all-NB types to find out that in the case of polarity level sentiment Gaussian NB achieved better results and higher accuracy although it was not the height but at least this accuracy results is comparable to the other classifiers

		GNB		CANB		CNB		MNB	
Datasets		Acc.	F1	Acc.	F1	Acc.	F1	Acc.	F1
Book	60,055	62%	61%	62%	61%	61%	60%	56%	58%
Hotel	20,047	68%	66%	65%	66%	68%	66%	68%	63%
MobApp	11,000	77%	76%	75%	75%	76%	76%	74%	73%
Printer	1,511	76%	76%	66%	65%	75%	76%	70%	59%
Mobile	1,500	66%	65%	66%	65%	66%	65%	63%	61%

Table 5: Comparing NB Classifiers with Baseline Polarity Level

5.4.2 Subjectivity Approach

In this experiment the research used the BOW subjectivity to classify the domains but during the process classifying the domain text to subjective and objective took the model long time to process. In sentiment classification for the Arabic text, we tried to capture the effect of BOW on the domain analysis by applying N-gram and SentiWordNet to the model trying to achieve better results.

The research tested the n-gram model with a loop that identify weather the feature of the word is a unigram feature or bigram or a trigram the research combined the results to investigate the subjective meaning of the model on the document level The research notice that the results started to show a good change. Noticing that after using trigram model the system become much slower than it should be the shaded numbers shows the best results the research notice a huge difference in the classification results this is due to the sentiment domain balance as SVM achieves the best results as SVM learned a new knowledge from adding the N-grams while NB didn't have the same effect.Adding subjectivity to the analysis it was noticed that MNB started to achieve higher accuracy scores than GNM as the data started to have real sentiment.

		SVM		**NB**		**KNN**		**LR**		**RF**		**Voting**	
Datasets		Acc	F1	Acc	F1	Acc	F1	Acc	F1	Acc	F1	Acc	F1
Book	60,055	80%	77%	62%	61%	74%	74%	69%	62%	67%	66%	79%	75%
Hotel	20,047	85%	84%	72%	70%	75%	71%	81%	79%	80%	79%	82%	82%
MobApp	11,000	88%	86%	76%	69%	79%	75%	87%	84%	86%	84%	86%	85%
Printer	1,511	85%	83%	70%	66%	71%	69%	84%	82%	80%	77%	82%	80%
Mobile	1,500	70%	68%	60%	55%	58%	57%	74%	71%	72%	72%	66%	60%

Table 6: Subjectivity level Classifiers Accuracy and F1 measure

That shoes that MNB is very sensitive towards adding more SentiWordNet lexicons and N-gram models to the feature selection, however for other classifiers the research noticed a better performance and an increase on the overall accuracy of the system, therefore the research suggest that adding subjectively to the overall classification help increase the results with a noticeable value.

5.4.3 Ontology Approach

On the domain level the research tried to collect the relation between features in Arabic text. Capturing the context of the text that preserves meaning and relations specially that BOW that carry the same segmentation and orientation.Noticing that the SVM classifier achieve the best results in the datasets with the LR in some other smaller datasets and the research suggest that this is due to the vectorization of the review which most of the review tend to have closer features and datasets , the SVM performance increased by 3 % in the First dataset , the performance exceeded that of the subjectivity by 3% as the F1 measure and on most of the datasets results the accuracy and F1 score increased by 2% to 3% with no interference with the data size. Most of the classifiers increased their performance after adding the Arabic ontology feature domain method to the sentiment analysis.

		SVM		**NB**		**KNN**		**LR**		**RF**		**Voting**	
Datasets		Acc	F1	Acc	F1	Acc	F1	Acc	F1	Acc	F1	Acc	F1
Book	60,055	80%	77%	62%	61%	74%	74%	69%	62%	67%	66%	79%	75%
Hotel	20,047	85%	84%	72%	70%	75%	71%	81%	79%	80%	79%	82%	82%
MobApp	11,000	88%	86%	76%	69%	79%	75%	87%	84%	86%	84%	86%	85%
Printer	1,511	85%	83%	70%	66%	71%	69%	84%	82%	80%	77%	82%	80%
Mobile	1,500	70%	68%	60%	55%	58%	57%	74%	71%	72%	72%	66%	60%

Table 7 Ontology level Classifiers Accuracy and F1 measure

5.5. Results Discussion

This research is about building a learning model, to train the model for the system needs a training stages and a dataset to train the model for best performance then applying the model on the training dataset for achieving the best results.After the model is trained, the testing dataset start to calculate its results and depending on this results and the performance the best technique is used to achieve best results. The ontology construction process began by identifying the domain and its important features throughout a set of queries and methods After inserting the parameters and functions for weighting the results, the domain variables are placed in its related function to determine the vectorization process

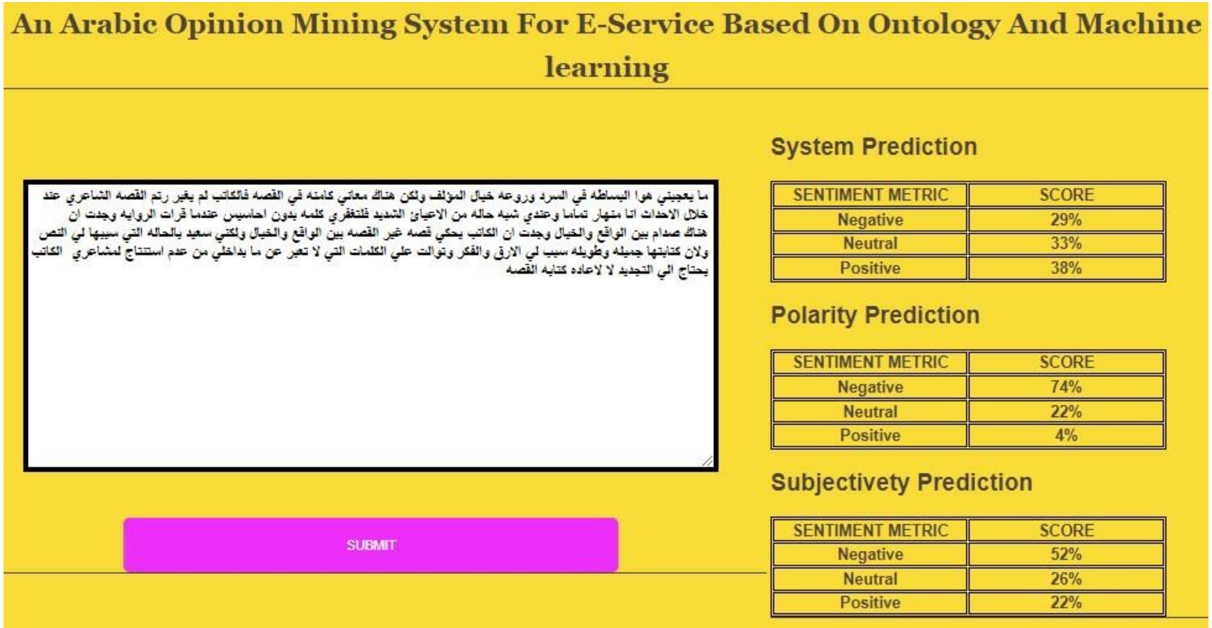

Figure 14: Arabic Opinion Mining System

by using several classifiers the comparing the result to the 0 factor by bigger than or less than the separable vectors are created depending on the function results and weights the right vector is selected for the module, through applying different weights and functions and comparing the outputs the module is trained for best results. By adding different weights to the function and calculating the error percentage with every change till the error percentage is almost none .This training process took a long time for evaluation through feeding the model with the different testing parameters and attributes after reaching the point of the best results the model is converted into feeding the system with the conclusion.

The SVM, NB, LR and KNN classifies were used to calculate the model results, the accuracy, recall and precision of the classifiers were analyzed using different techniques. The SVM classifier after constructing the ontology and applying the N-gram, TF-IDF, chunking features and BOW weighing achieved the best accuracy of 90% with an average dataset size of 11,000 records in the Mobile application dataset. LR came in second with up to 87% sentiment, while MNB and KNN were in third place with up to 79%. Overall, SVM performed the best in almost all settings, except for polarity level analysis where LR exceeded the SVM classifier and this is due to the calibrated probabilities of the polarity process at interpreted from a confidence decision that the word is either positive or negative as the LR is a straight forward classifier it works better with polar classification of extreme positive and extreme negative.

The SVM classifier showed a better performance as the feature number increase and with the smaller size datasets so despite of its good theoretic foundations and high classification accuracy, SVM is not suitable for classification of large data sets, because the training complexity of SVM is highly dependent on the size of the data set, it is also noticed that the SVM classifier does not perform well when the datasets contains more unknown variables or noises the targeted features have an overlapping due to the redundancy of the feature level indicator.

Gaussian naïve Bayes classification showed better distribution in the polarity level of opinion mining as it doesn't requires much training data and it easily handles noise as it deals with every feature as an independent capacity for prediction so it fits the polarity level perfectly. By adding the subjectivity level the feature number increased so the MNB started to achieve better results as it uses a multinomial distribution for each feature it achieved better results than GNB at the document level of opinion mining analysis.

The system conducted on the impact of the representation model on classification showed that using a professional Arabic stemmer combined with light stemming resulted in a significant increase in accuracy. This is due to the fact that reducing words to their three-letter root affects both semantics and sentiment orientation. It was also discovered that using bi-grams and tri-grams improved performance for all classifiers. This is because bi-grams and tri-grams allow for the handling of negation, with the former attaching negation particles to the word before or after it, and the latter enhancing sentence polarity. The study analyzed the impact of weighting schemes on various classifiers. SVM and LR classifiers performed exceptionally well using BOW weighting. However, MNB and KNN classifiers showed poor results with this weighting on all representation models. On the other hand, TFIDF weighting produced favorable outcomes for all classifiers.

ML Classifier	SVM Classifier With Ontology				SVM Classifier Without Ontology			
	Precision	Recall	Accuracy	F1	Precision	Recall	Accuracy	F1
Mob App, Domain Dataset	90%	89%	90%	89%	76%	76%	77%	76%

Table 8: Comparing ML (SVM) Classifier after adding the Ontology

The results show that the proposed sentiment analysis system outperforms existing systems in terms of accuracy, precision, recall, F1 score. The accuracy of our system was 90%, compared to 85% for the best-performing existing system by indicating an accurate domain knowledge representation of the customer review, the system can provide valuable insights into customer sentiment and help businesses and organizations improve their services and productivity.

6. Conclusions and Future Work

6.1. Introduction

This chapter is an overview of the findings of the study on An Arabic Opinion Mining system outcome, it anticipates the study limitations and recommended future investigations. The proposed system for sentiment analysis of e-services based on ontology and machine learning can provide valuable insights for companies to improve customer satisfaction. The system can be applied to various e-services, including e-commerce, food delivery services, and other online services [24-46]. The system can also be used to analyze customer reviews and feedback, and provide valuable insights for companies to improve their products and services.

6.2. Limitation of Research

The research is conducted to help e-services based on ontology domain knowledge provided mainly by online e-service Arabic user's reviews facing the unavailability of unstructured data to build the Arabic ontology frame of the researched domains datasets.

Unavailability of Arabic lexicon for every Arab world dialect reduced the accuracy of the sentiment process. As the research is on Arabic language opinion mining process, the research mainly chooses only the methods that is compatible with Arabic language metaphors and methods. The system serves only three polarity types positive, negative and neutral.

6.3. Future Works

Despite the high predictive accuracy obtained with the proposed system with the selected features the system used, I believe that the proposed approaches can still be extended with other features that may influence the relevance and that the research did not use in this work. For example, it would be interesting to integrate the new reviews and the product updates and to test the proposed method on different domains.

Working with more complex data types such as images and voices, handling Arabic web written in English letters, signs, sarcasm, hints, slangs, punctuations, emoji's, expressive photos and Arabic proverbs.

Analyzing reviews that includes positive and negative opinions at the same time (complex opinions) that is not considered only neutral or only positive or only negative.

Increasing the polarity of the sentiment to include 10 levels of customer satisfaction depending on the customer satisfaction subjectivity of the features rates and the most important domain features hierarchy from very-negative (1), negative (2), weak negative (3), neutral towards negative (4), neutral towards slightly negative (5), neutral (6), neutral towards positive (7), weak positive (8) positive (9), very positive (10).

The research would recommend Appling the system methodologies to other complicated languages like (e.g. Chinses) and observer the methodology effect on increasing the accuracy of the system results.

References

1. Gontier, E. M. (2015). Web Semantic and Ontology. Advances in Internet of Things, 5(02), 15.
2. Al Zamil, M. G., & Al-Radaideh, Q. (2014). Automatic extraction of ontological relations from Arabic text. Journal of King Saud University-Computer and Information Sciences, 26(4), 462-472.
3. Pang, B., & Lee, L. (2008). Opinion mining and sentiment analysis. Foundations and Trends® in information retrieval, 2(1–2), 1-135.
4. Aljarah, I., Habib, M., Hijazi, N., Faris, H., Qaddoura, R., Hammo, B., & Alfawareh, M. (2021). Intelligent detection of hate speech in Arabic social network: A machine learning approach. Journal of Information Science, 47(4), 483-501.
5. Alowaidi, S., Saleh, M., & Abulnaja, O. (2017). Semantic sentiment analysis of Arabic texts. International Journal of Advanced Computer Science and Applications, 8(2).
6. Duwairi, R., & El-Orfali, M. (2014). A study of the effects of preprocessing strategies on sentiment analysis for Arabic text. Journal of Information Science, 40(4), 501-513.
7. AlAsmar, A. M. (2016). Feature based approach in arabic opinion mining using ontology. Department of Information Technology, The Islamic University–Gaza, Msc Thesis.
8. Kaddoura, S., Itani, M., & Roast, C. (2021). Analyzing the effect of negation in sentiment polarity of facebook dialectal arabic text. Applied Sciences, 11(11), 4768.
9. Saud S. Alotaibi, "Capturing Negation scope using Base

Phrase Chunk in Arabic Sentiment Classification," Umm Al-Qura University, Information Systems Department, Makkah, Saudi Arabia, vol. 4, 2022.
10. Alharbi, O. (2020). Negation handling in machine learning-based sentiment classification for colloquial Arabic. International Journal of Operations Research and Information Systems (IJORIS), 11(4), 33-45.
11. Amra, I. A. A., & Maghari, A. Y. (2017, May). Student's performance prediction using KNN and Naïve Bayesian. In 2017 8th international conference on information technology (ICIT) (pp. 909-913). IEEE.
12. Zhang, S., Li, X., Zong, M., Zhu, X., & Wang, R. (2017). Efficient kNN classification with different numbers of nearest neighbors. IEEE transactions on neural networks and learning systems, 29(5), 1774-1785.
13. Cessie, S. L., & Houwelingen, J. V. (1992). Ridge estimators in logistic regression. Journal of the Royal Statistical Society Series C: Applied Statistics, 41(1), 191-201.
14. Awajan, I., & Mohamad, M. (2018). A review on sentiment analysis in arabic using document level. Int. J. Eng. Technol., 7(3.13), 128-132.
15. Ichise, R. (2008). Machine learning approach for ontology mapping using multiple concept similarity measures. In Seventh IEEE/ACIS International Conference on Computer and Information Science (icis 2008) (pp. 340-346). IEEE.
16. Kulmanov, M., Smaili, F. Z., Gao, X., & Hoehndorf, R. (2022). Semantic similarity and machine learning with ontologies. Briefings in bioinformatics, 22(4), bbaa199.
17. Wang, X., & Fu, G. H. (2010). Chinese subjectivity detection using a sentiment density-based naive Bayesian classifier. In 2010 International Conference on Machine Learning and Cybernetics (Vol. 6, pp. 3299-3304). IEEE.
18. Duwairi, R., Al-Refai, M., & Khasawneh, N. (2007). Stemming versus light stemming as feature selection techniques for Arabic text categorization. In 2007 Innovations in Information Technologies (IIT) (pp. 446-450). IEEE.
19. Green, S., & Manning, C. D. (2010). Better Arabic parsing: Baselines, evaluations, and analysis. In Proceedings of the 23rd International Conference on Computational Linguistics (Coling 2010) (pp. 394-402).
20. Bordoloi, M., & Biswas, S. K. (2023). Sentiment analysis: A survey on design framework, applications and future scopes. Artificial Intelligence Review, 1-56.
21. Alotaibi, S. S. (2015). Sentiment analysis in the Arabic language using machine learning (Doctoral dissertation, Colorado State University).
22. O'Keefe, T., & Koprinska, I. (2009). Feature selection and weighting methods in sentiment analysis. In Proceedings of the 14th Australasian document computing symposium, Sydney (pp. 67-74).
23. Siddiqui, S., Monem, A. A., & Shaalan, K. (2016). Sentiment analysis in Arabic. In Natural Language Processing and Information Systems: 21st International Conference on Applications of Natural Language to Information Systems, NLDB 2016, Salford, UK, June 22-24, 2016, Proceedings 21 (pp. 409-414). Springer International Publishing.
24. Nandwani, P., & Verma, R. (2021). A review on sentiment analysis and emotion detection from text. Social Network Analysis and Mining, 11(1), 81.
25. Saleh, H., Mostafa, S., Alharbi, A., El-Sappagh, S., & Alkhalifah, T. (2022). Heterogeneous ensemble deep learning model for enhanced Arabic sentiment analysis. Sensors, 22(10), 3707.
26. Siregar, A., Faisal, S., Tukino, T., Puspabhuana, A., & Simarangkir, M. (2019, November). Comparison Study of Term Weighting Optimally With SVM in Sentiment Analysis. In Proceedings of the 2nd International Conference on Advance and Scientific Innovation, ICASI 2019, 18 July, Banda Aceh, Indonesia.
27. Alhamad, G., & Kurdy, M. B. (2020). Feature-based sentiment analysis for Arabic language. International Journal of Advanced Computer Science and Applications, 11(11).
28. Alotaibi, S. S. (2015). Sentiment analysis in the Arabic language using machine learning (Doctoral dissertation, Colorado State University).
29. Khabour, S. M., Al-Radaideh, Q. A., & Mustafa, D. (2022). A new ontology-based method for Arabic sentiment analysis. Big Data and Cognitive Computing, 6(2), 48.
30. Wątróbski, J. (2018). An ontology-based approach to opinion mining tools selection. Studies & Proceedings of Polish Association for Knowledge Management, 89, 42-51.
31. Thakor, P., & Sasi, S. (2015). Ontology-based sentiment analysis process for social media content. Procedia Computer Science, 53, 199-207.
32. El Bakly, A. H., Darwish, N. R., & Hefny, H. A. (2020). Using ontology for revealing authorship attribution of Arabic text. Int. J. Eng. Adv. Technol. (IJEAT), 4, 143-151.
33. Aboelela, E. M., Gad, W., & Ismail, R. (2021). Ontology-Based Approach for Feature Level Sentiment Analysis. International Journal of Intelligent Computing and Information Sciences, 21(3), 1-12.
34. Gautam, G., & Yadav, D. (2014). Sentiment analysis of twitter data using machine learning approaches and semantic analysis. In 2014 Seventh international conference on contemporary computing (IC3) (pp. 437-442). IEEE.
35. Haddi, E., Liu, X., & Shi, Y. (2013). The role of text pre-processing in sentiment analysis. Procedia computer science, 17, 26-32.
36. Potdar, G. S., & Phursule, R. N. (2015). Twitter Blogs Mining using Supervised Algorithm. International Journal of Computer Applications, 126(5).
37. Khalil, A., Jarrah, M., Aldwairi, M., & Jararweh, Y. (2021, November). Detecting Arabic fake news using machine learning. In 2021 Second International Conference on Intelligent Data Science Technologies and Applications (IDSTA) (pp. 171-177). IEEE.
38. Al-Twairesh, N., & Al-Negheimish, H. (2019). Surface and deep features ensemble for sentiment analysis of arabic

tweets. IEEE Access, 7, 84122-84131.
39. Hamouda, A. E. D. A., & El-taher, F. E. Z. (2013). Sentiment analyzer for arabic comments system. Int. J. Adv. Comput. Sci. Appl, 4(3).
40. Ibrahim, M. F., Alhakeem, M. A., & Fadhil, N. A. (2021). Evaluation of Naïve Bayes classification in Arabic short text classification. Al-Mustansiriyah J. Sci., 32(4), 42-50.
41. Brahimi, B., Touahria, M., & Tari, A. (2015). Data and Text Mining Techniques for Classifying Arabic Tweet Polarity. Journal of Digital Information Management, 14(1).
42. Robinson, P. N., & Haendel, M. A. (2020). Ontologies, knowledge representation, and machine learning for translational research: recent contributions. Yearbook of medical informatics, 29(01), 159-162.
43. Badaro, G., Baly, R., Hajj, H., Habash, N., & El-Hajj, W. (2014). A large scale Arabic sentiment lexicon for Arabic opinion mining. In Proceedings of the EMNLP 2014 workshop on arabic natural language processing (ANLP) (pp. 165-173).
44. Brahimi, B., Touahria, M., & Tari, A. (2021). Improving sentiment analysis in Arabic: A combined approach. Journal of King Saud University-Computer and Information Sciences, 33(10), 1242-1250.
45. Abdelminaam, D. S., Neggaz, N., Gomaa, I. A. E., Ismail, F. H., & Elsawy, A. A. (2021). Arabicdialects: An efficient framework for arabic dialects opinion mining on twitter using optimized deep neural networks. Ieee Access, 9, 97079-97099.
46. Alsayat, A., & Elmitwally, N. (2020). A comprehensive study for Arabic sentiment analysis (challenges and applications). Egyptian Informatics Journal, 21(1), 7-12.