Unsupervised Fuzzy-Multi-Core Aspect Sentiment Analysis for Big Data of Online News Users' Persian Opinions

Abstract:

An online news article can cover various topics or contain different aspects of a subject, encouraging readers to express their opinions on specific topics or aspects. Sentiment analysis evaluates the overall sentiment of the audience towards the entire news article, whether it is positive, negative, or neutral. However, in aspect-based sentiment analysis, the focus is on determining which aspect of the news article the opinion is referring to. Extracting the relevant aspect in sentiment analysis involves identifying the part of the article that the reader has expressed an opinion about. This task can lead to a more precise analysis of audience reactions to future news and events. To accomplish this, the news text is segmented into constituent sentences and transformed into a vector space. Then, an unsupervised clustering method is applied to extract various aspects of the news. Fuzzy multi-core clustering is employed as the clustering technique, which has lower computational overhead and can handle uncertain, noisy, and outlier data easily. The implemented approach is based on the concept of feasibility and utilizes multi-core learning to detect clusters in complex data structures. This method remains robust against issues such as ineffective cores or unrelated features by automatically adjusting the core weights within an optimized framework. Furthermore, support vector machines are employed to establish the relationship between opinions and relevant aspects. The transition to the vector space, the mapping process, clustering operations, and aspect extraction are performed in the reducer.

Keywords: Aspect-based Sentiment Analysis, Fuzzy Multi-Core Clustering, Opinion Mining, Persian Online News, Big Data.

1. Introduction

The exponential growth in the number of online news articles has captured extensive research attention towards their appropriate utilization. Each news piece not only imparts valuable information to the audience but also encompasses multiple challenging aspects that provoke readers to express their opinions on various relevant, and sometimes even unrelated, dimensions of the news. To elaborate on the matter, it can be stated that establishing and maintaining sustainable interaction with the audience is a vital aspect for any media outlet striving to exert influence on society. On the other hand, the high volume of online news production and the feedback received from readers in the form of comments have practically made it infeasible for humans to analyze and summarize all these opinions. Additionally, media outlets have a social responsibility to regulate their news content in a manner that minimizes negative impacts on the minds and psyche of their audience. Hence, one of the contemporary needs of online media is the analysis of audience sentiments and emotions towards published news on one hand, and the anticipation of audience reactions to tailored news content intended for dissemination on the other. The analysis of the semantic meaning and sentiment mining of a news article reflects whether the audience's opinion about it is positive, negative, or neutral. However, this article endeavors to present a more innovative and distinct approach to analyzing the audience's opinions towards online news, which has not been previously addressed in research on Persian-language news websites. This approach focuses on uncovering hidden aspects within a news article. In this regard, after extracting the various aspects of a news article, a correspondence is established between the extensive data of opinions and these aspects. This approach not only enables unsupervised classification of news but also provides us with the knowledge that lays the groundwork for predicting the behavior and response of the audience to newsworthy topics and challenges. In addition to scientifically revealing the current inclinations and positions of individuals in society, such predictions can significantly assist policymakers in making decisions and directing future policies [1]. To further illustrate the subject, consider the following example: Among today's hot topics, let's assume that a news article has been published regarding the president's provincial visit to one of the oil-rich provinces in the country. As a result, this news article will encompass various aspects. The main aspect of this news is the provincial visit and the underlying aspects involve reporting on the visits and the topics discussed during them. These topics may include the cost of petroleum products, the economy of the region's people, and the challenges of remote schooling. By reading this news, the audience becomes stimulated and engages in providing opinions on different aspects of the news. Now, by extracting the various aspects of the news and analyzing the opinions expressed about them, we can first understand what topics the audience is more inclined towards. Secondly, we can make predictions. For example, if the president continues his provincial visits, what kind of reactions can be expected from the people? Or if gasoline prices increase, how will society react? Or if schools switch to remote learning, what kind of response will be received from the people? As we have seen, access to this level of knowledge can play a significant role in formulating and implementing executive policies in society. The field of health and hygiene [2], recommendation systems [3], and business intelligence are among the other notable applications of sentiment analysis of audience opinions.

The methodology employed in this article can be summarized as follows: Firstly, sentences are transformed into vector space. Then, fuzzy clustering is utilized to segment the document of opinions. Next, opinions are transferred to vector space. Finally, support vector machines are employed to determine the main aspect of the given opinions. The article proceeds with the second section, which introduces the research background. The third section presents the proposed method and relevant technical considerations. Implementation and evaluation are conducted in the fourth section.

Ultimately, the article concludes in the fifth section by providing a summary and outlining future work.

2. Literature Review

Knowing people's opinions in a specific domain can play a significant role in making major decisions within that domain. Sentiment analysis refers to the use of text analysis and computational linguistics to extract the desired objective from textual sources. Opinion mining, while having a similar meaning to sentiment analysis, typically focuses on extracting ideas and perspectives on topics and trends, rather than specifically extracting user opinions about a product [4]. Nonetheless, these two approaches share a high degree of similarity. As shown in Figure 1, text sentiment analysis can be conducted at three different levels: aspect, sentence, and document [5]. The first two levels, document and sentence analysis, are quite engaging and challenging. However, the third level, aspect-level sentiment analysis, is even more difficult due to the need for a detailed examination of the given document. This is because it requires analyzing the meaning related to a specific aspect within the text. Therefore, at this level of sentiment analysis, the focus is on different aspects rather than the entire sentence or document.

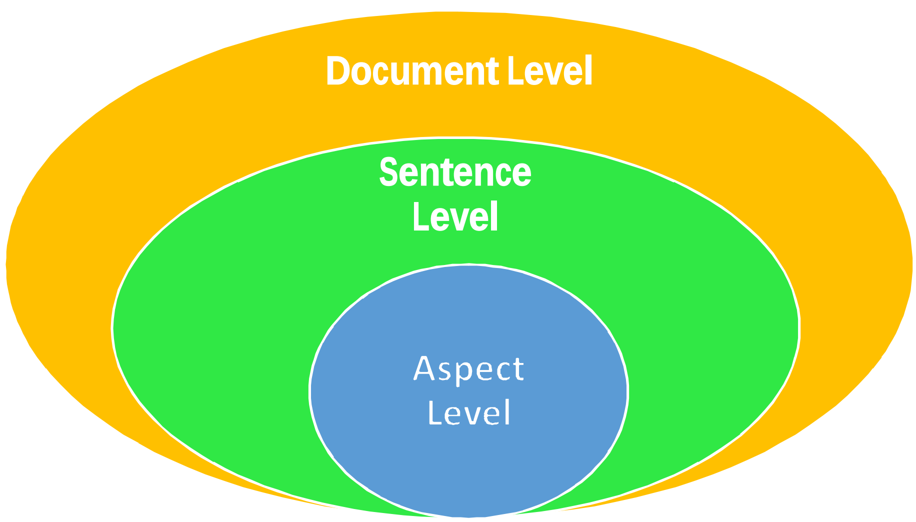

Figure 1: Hierarchical Levels of Semantic Analysis

It is worth mentioning that the combination of aspect-level semantic analysis with sentence-level semantic analysis can provide better results, as the outcome of the sentence-level analysis is dependent on the aspect-level analyses conducted. In sentence-level semantic analysis, the ultimate goal is to categorize the sentence into one of three categories: positive, negative, or neutral [6]. To achieve this goal, the sentences need to be classified as either a factual statement providing information or an opinion-oriented sentence expressing ideas and perspectives [7]. On the other hand, document-level semantic analysis aims to determine whether a text as a whole conveys a positive or negative sentiment. The process of semantic analysis is not a single problem [6]. As depicted in Figure 2, it is a collection of issues that require various natural language processing tasks to be performed.

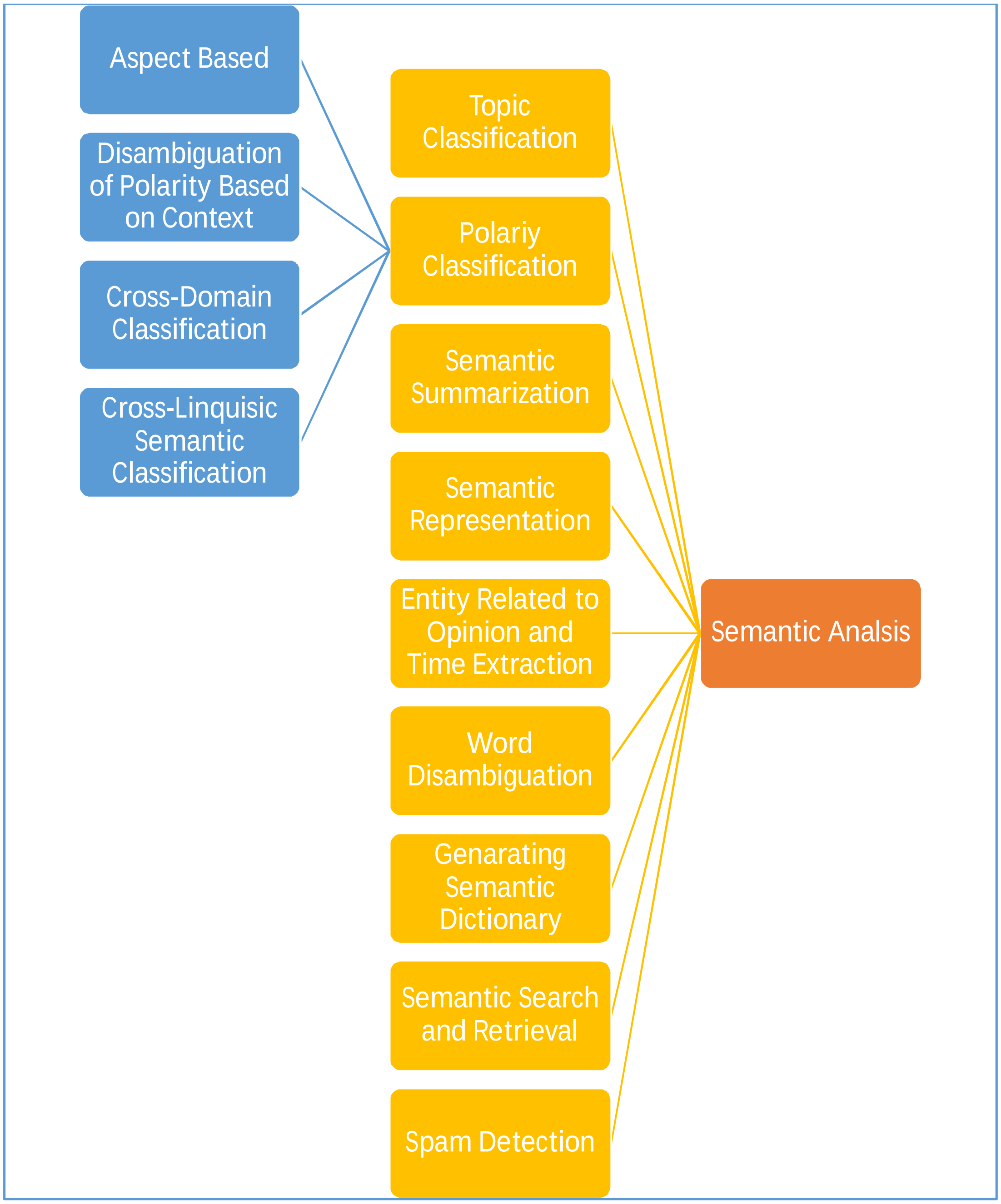

Figure 2: Subtasks of Semantic Analysis

Some of the challenges in textual semantic analysis and opinion mining include detecting fake and spam opinions, identifying figurative expressions, managing negating expressions, resolving lexical ambiguities, dealing with low-resource languages, and domain independence [8].

In this regard, unsupervised approaches are often proposed due to the high human and financial costs associated with generating labeled data. Typically, clustering methods are employed in these approaches, where data points within each cluster have the minimum distance from each other, while data points from different clusters have the maximum distance. Fuzzy logic, as a reflection of human thinking, is an attempt to model human sentiment through words. Consequently, fuzzy logic can lead to more advanced intelligent systems that think and make decisions like humans. Fuzzy clustering is considered a more advanced type of clustering method. In this approach, each data point can belong to multiple clusters with different membership degrees. In each iteration of these algorithms, the membership degree of each data point in each cluster is calculated. Furthermore, the closer a data point is to the center of a cluster, the higher its membership degree in that cluster. Recently, we have witnessed diverse applications of clustering algorithms based on their specific use cases in real-world environments. One of the most effective methods in this regard is multi-cluster-based fuzzy clustering [9]. Among the approaches based on kernel learning, some focus on learning the kernel matrix [10], while others aim to learn the parameters of a specific form of the kernel by defining a suitable objective function [11]. For example, the PCK-Means method defines the objective function by assigning weights to variables and considering a penalty term. This method not only allows for variable weighting along with a penalty term but also attempts to learn an appropriate distance function during the clustering process [12].

In this context, the FCM (Fuzzy C-Means) clustering method is another popular fuzzy clustering approach that is more flexible than its crisp counterpart in many cases. This method considers the sum constraint equal to one for the sum of membership degrees of a data point to all clusters. Although this constraint is useful in creating fuzzy clusters, the FCM method is sensitive to noise and outliers [13]. In [14], the PCM (Possibilistic C-Means) method is introduced by removing this constraint. This method provides a possibilistic partitioning of the data, where the degree of possibility of each data point indicates the extent to which that data inherits the properties of the clusters. Since noisy and outlier data are assigned to clusters with lower degrees of possibility, this method is robust against noise and outliers, preventing the data from significantly influencing the clustering result. However, the efficiency of this method heavily relies on the appropriate initialization of cluster representatives. One drawback of this method is the generation of overlapping clusters. By combining the FCM and PCM methods, good results can be achieved [15]. Although the combination of FCM and PCM yields an effective method for clustering data with overlapping and noisy clusters, it is limited to detecting spherical clusters and is not capable of clustering data with complex, voluminous structures in semantic analysis. In this regard, reference [16] proposes a method for clustering non-linear, separable, and overlapping large-scale data that is resilient to noise and outliers, which has attracted the attention of the authors of the current paper. To achieve this, the method combines two approaches: Fuzzy C-Means (FCM) and Possibilistic C-Means (PCM), while improving the objective function by incorporating a penalty term. Furthermore, to detect linearly inseparable clusters in complex high-dimensional data structures, a multi-kernel learning architecture is employed to define the objective function. Since finding an appropriate combination of kernels is a challenging task, this method ensures robustness by automatically adjusting the weights of the kernels, thereby avoiding ineffective kernel choices or irrelevant features. This reduces the sensitivity of the proposed method to inefficient kernel selection. Consider Equation 1, which represents a linear combination of M base kernels in φ for mapping the data into the feature space [16].

$$\varphi(x)=w_1\varphi_1(x)+w_2\varphi_2(x)+\ldots+w_M\varphi_M(x) \qquad \text{(Equation 1)}$$

This equation consists of weighted sums of individual base kernels, where each base kernel is denoted as φi(x) and the corresponding weight is represented as wi. By multiplying each base kernel with its respective weight and summing them up, the mapping φ(x) is obtained. The purpose of this equation is to transform the data into a higher-dimensional feature space, where non-linear relationships can be captured more effectively. Each base kernel represents a different transformation of the input data, and the weights determine the contribution of each transformation to the final mapping. By adjusting the weights, the importance of different transformations can be controlled, allowing for flexibility in modeling complex patterns in the data. It is worth noting that Equation 1 is widely used in various machine learning and pattern recognition tasks, as it provides a flexible and expressive framework for capturing non-linear relationships in the data. The choice of base kernels and their corresponding weights can be optimized through various techniques, such as kernel methods or optimization algorithms, to achieve the best performance in a given task. In summary, Equation 1 represents a linear combination of base kernels in the feature space φ, allowing for the mapping of data into a higher-dimensional space to capture non-linear relationships. The weights associated with each base kernel determine their contribution to the final mapping, offering flexibility and adaptability in modeling complex patterns. In this method, the minimization of the objective function, as stated in Equation 2, is proposed to achieve the mentioned objectives [16]:

$$J_{FPM}(W,T,U,V) = \sum_{c=1}^{C}\sum_{i=1}^{N} u_{ci}^{m} t_{ci}^{p} (\varphi(x_i) - v_c)^T (\varphi(x_i) - v_c) + \sum_{c=1}^{C} \mu_c \sum_{i=1}^{N} u_{ci}^{m} (1 - t_{ci})^p + \alpha\left(\sum_{(i,j)\in M} \sum_{c=1}^{c} \sum_{\substack{l=1 \\ l\neq c}}^{C} u_{ci}^{m} u_{lj}^{m} t_{ci}^{p} t_{lj}^{p} + \sum_{(i,j)\in C} \sum_{c=1}^{C} u_{ci}^{m} u_{cj}^{m} t_{ci}^{p} t_{cj}^{p} \right) \qquad \text{(Equation 2)}$$

Equation 2 plays a crucial role in improving the multi-kernel model by ensuring cluster compactness in the feature space. Let's break down the components of the equation to provide a more detailed explanation: The first term of Equation 2 calculates the weighted sum of squared distances between data points and cluster centroids. It takes into account fuzzy membership degrees and possibility memberships. This term encourages data points to be close to their assigned cluster centroids, contributing to cluster compactness.The second term aims to maximize the possibility of membership degrees. By maximizing the membership degrees, the equation avoids trivial solutions and ensures that each data point has a significant degree of association with its assigned cluster. The third term in the equation introduces a penalty controlled by α, which represents the relative importance of the penalty. This term consists of two penalty components. The first part penalizes different clusters based on their corresponding membership degrees, encouraging distinct clusters to be formed. The second part penalizes identical clusters based on their membership degrees, discouraging redundant or overlapping clusters. Minimizing Equation 2 leads to minimizing the total intra-cluster distances in the feature space, thereby achieving an optimal partitioning of the data. It has been established that minimizing the objective function in Equation 2 guides the algorithm towards one of its local minima, resulting in a meaningful clustering of the data. For a more comprehensive understanding of the objective function, particularly from a multi-cluster perspective, we recommend referring to the relevant reference. The mentioned reference provides further examination and insights into the objective function, shedding light on its behavior and implications in the context of multi-cluster analysis.

Here, M and C represent sets of constraints. $V_c \in R^L$ represents the centroid of cluster c, and $V \equiv [v_c]_{l\times c}$ is an L×C matrix where each column represents a cluster. The vector $W=(w_1, w_2, \ldots, w_M)^T$ represents the weights of the kernels, satisfying the condition $\sum_{k=1}^{M} wk = 1$. $U \equiv [u_{ci}]_{c\times n}$ is a fuzzy membership matrix, where the element uci represents the fuzzy membership degree of data point x_i to cluster c, considering the condition $\sum_{c=1}^{C} u_{ci} = 1$. m is the fuzzifier controlling the degree of fuzziness in cluster assignment. $T \equiv [t_{ci}]_{c\times n}$ is the possibility membership matrix, where each element t_{ci} represents the possibility membership degree of data point xi to cluster c. p represents the weight factor of possibility membership degree. The expression $(\varphi(x_i) - v_c)^T(\varphi(x_i) - v_c$ denotes the distance between data point x_i and the centroid v_c in the feature space. μ_c is the scaling parameter, and an appropriate value for it is suggested in Equation 3 [16].

$$\mu_c = \frac{\sum_{i=1}^{N} u_{ci}^{m} t_{ci}^{p} (\varphi(x_i) - v_c)^T (\varphi(x_i) - v_c)}{\sum_{i=1}^{N} u_{ci}^{m} t_{ci}^{p}} \qquad \text{(Equation 3)}$$

As we mentioned before, the first two terms of Equation 2 contribute to the improvement of the multi-kernel model by ensuring cluster compactness in the feature space. The first term represents the weighted sum of squared distances between data points and cluster centroids, considering fuzzy membership degrees and possibility memberships. The second term aims to maximize the possibility of membership degrees to avoid trivial solutions. The third term in the equation controls the penalty and is weighted by α as a measure of relative importance. This term consists of two penalty terms. The first part penalizes different clusters based on their corresponding membership degrees. In contrast, the second part penalizes identical clusters based on their corresponding membership degrees. Minimizing this equation minimizes the total intra-cluster distances in the feature space to achieve optimal partitioning. It has been proven that by minimizing the objective function in Equation 2, a meaningful clustering of the data converges to one of its local minima [16]. Reference [17] has provided a comprehensive study on the fuzzy methods employed in this field. The review portrays different uses of fuzzy logic and summarizes over 120 articles published in the past decade regarding tasks and applications of opinion mining. Moreover, [18] critically evaluates the different modules of a sentiment analysis framework and discusses the limitations of existing methods and systems. Moreover, it proposes potential multidisciplinary applications of sentiment analysis based on data content and suggests future research directions.

In this regard, Support Vector Machines (SVM) have emerged as a prominent technique in the field of sentiment analysis, both independently [19] and in its advanced [20] and combined forms, such as combining with deep learning [21] and Fuzzy Method [22] . Research has shown that SVM outperforms other methods in sentiment mining tasks. The following points can be considered as reasons for the success of SVM in sentiment analysis:

1. High-dimensional input space: Sentiment classifiers dealing with textual features often operate in high-dimensional spaces, with feature dimensions ranging from thousands to even more. SVM's performance is not affected by the number of features, making it potentially capable of handling large document feature spaces.

2. Sparse and irrelevant feature sets: To avoid working with high-dimensional input spaces, it is often assumed that many features are irrelevant and can be reduced to reduce the dimensions of the feature space. Feature selection aims to identify irrelevant features. In sentiment analysis, only a few features are unrelated to the classification task. Most words are relevant for document classification, and this observation can be extended to sentiment analysis. Therefore, feature selection needs to be performed carefully to avoid losing important information.

3. Sparsity of opinion vectors: The feature vector corresponding to each document type has a sparse representation, with only a few non-zero feature values. SVM has shown excellent performance in problems with a large number of concepts and a small number of examples.

These factors contribute to the effectiveness of SVM in sentiment analysis, making it a widely used technique in the field of data mining.

If we aim to provide a concise definition of a Support Vector Machine (SVM), it can be described as a set of points in n-dimensional space that determines the boundary between classes. Data classification and boundary determination are based on these points, and changing one of them can alter the classification output. In a two-dimensional space, support vectors form a line, in a three-dimensional space, they form a plane, and in an n-dimensional space, they form a hyperplane. SVM passes the data through a hyperplane and employs an optimization algorithm to classify them. It starts by forming examples on the class boundary, selecting those samples from the training points that have the shortest distance to the decision boundary as support vectors. The higher the dimensionality of the data, the better the desired outcome. The selection of the boundary between two classes is based on the principle that all samples of the first class are located on one side, and all samples of the second class are on the other side of the boundary. The decision boundary should be in a way that the closest training samples from both classes have the maximum perpendicular distance from each other. This means that in this method, the distance between the closest training samples in two classes is calculated perpendicular to the decision boundaries, and the optimal boundary is determined by analyzing the optimization problem. In this manner, two parallel hyperplanes are created on both sides of the decision boundary, in a way that the boundary hyperplane between the two parallel hyperplanes creates the maximum distance.

Most approaches in semantic analysis revolve around word vector representation. These approaches are utilized in two ways. The first approach involves using deep neural networks to represent words in a way that the vector captures some level of word meaning. These methods are referred to as embedding methods, where words are embedded. The second approach utilizes deep neural networks to detect sentiment polarity in documents. The approach used in [20] consists of four stages: 1) Word-to-vector conversion, 2) Definition of positive and negative polarity, 3) Preparation of training and testing data, and 4) Deep learning process.

The first stage involves converting words into vectors. Various methods, such as word2vec and GloVe, exist for this purpose. These methods map 400,000 words to 200-dimensional vectors. In the second stage, to define polarity, considering the impact of news on stock prices, the average stock price one minute before the news release is represented as $\bar{p}_{b,i}$, and the average stock price one minute after the news release is represented as $\bar{p}_{a,i}$. The polarity of the news is then defined using equation 4 [23].

(Equation 4)

In the next stage, a deep neural network is employed using RNN (Recurrent Neural Networks) with LSTM (Long Short-Term Memory) cells for news classification. The number of LSTM cells is determined based on the length of the longest news article [24]. The output of this network is then fed into a softmax function, which assigns a value between zero and one to indicate the polarity level [25]. An improved word vector representation is utilized for training the deep network. This enhanced word vector representation is constructed by combining four other word vector representations. The first component is called L2V, which involves creating the corresponding vectors using semantic dictionaries that determine the polarity and positive/negative sentiment for words. Six semantic dictionaries are initially considered, and the polarity of each dictionary is extracted, normalized, and transformed into a vector. By concatenating these vectors together, the final vector representation is constructed. The second component, called POS2VEC, involves converting the syntactic role of words into vectors. For this purpose, the syntactic role of each word is extracted, and a fixed vector is assigned to each role. The corresponding vector is then assigned to the word. The third component represents the distance of each word from both ends of the document. The distance of each word from both ends is measured and converted into a vector. These vectors are then concatenated together to form the position vector, labeled as Wp2V. The fourth component utilizes pre-trained word vector representations, specifically WORD2VEC and GloVe. By combining these two representations, another part of the improved word vector representation is created. Finally, the obtained vector is fed into a deep convolutional neural network model, and the results are obtained.

In [26], several models have been employed for word vector representation and document representation. Subsequently, a bidirectional LSTM network is used to analyze and determine the sentiment polarity of the represented document.

The vector representations used in this model are as follows:

- Transformer-based representation: One of the proposed methods for word vector representation based on the context they appear in is BERT [27]. This model calculates the probability of a word's existence between two neighboring words. In this study, a BERT-based model with 12 encoder layers, 768 hidden layers, and 12 attention heads is utilized.
- Contextualized embedding: Another method for vectorization is ELMO, which considers various aspects of a word, such as the specific context it is used in [28]. This vectorization helps capture different meanings of a word based on the context it appears in.
- GloVe vector: Another word vectorization approach mentioned earlier involves using GloVe vectors.
- Syntactic role vector: Similar to the previous approach, this section first identifies the syntactic role of each word and assigns a fixed-length vector to each of them.
- Semantic vector: In this section, a semantic dictionary consisting of six words is used to transform a word into a semantic vector.
- Character-level embedding: The final part of word vectorization involves character-level representation. A bidirectional LSTM network is utilized to create this vector. Essentially, this approach calculates the probability of a specific character based on the 25 preceding and succeeding characters and represents it as a vector.

Finally, these resulting vectors are fed into a network to measure the document's sentiment polarity. In [29], a semantic dictionary is initially used to label the data, and then support vector machines are employed for classification. TF-IDF is used to determine the importance of words in the text, but it is not used for feature selection. Instead, the K-squared test metric is utilized.

3. The Proposed Method

3-1. Architecture

Due to the mentioned capabilities in the proposed function of reference [16] for multi-kernel fuzzy clustering, this method is the main focus of the current article and is utilized accordingly. The architecture employed in the proposed method is illustrated in Figure 3.

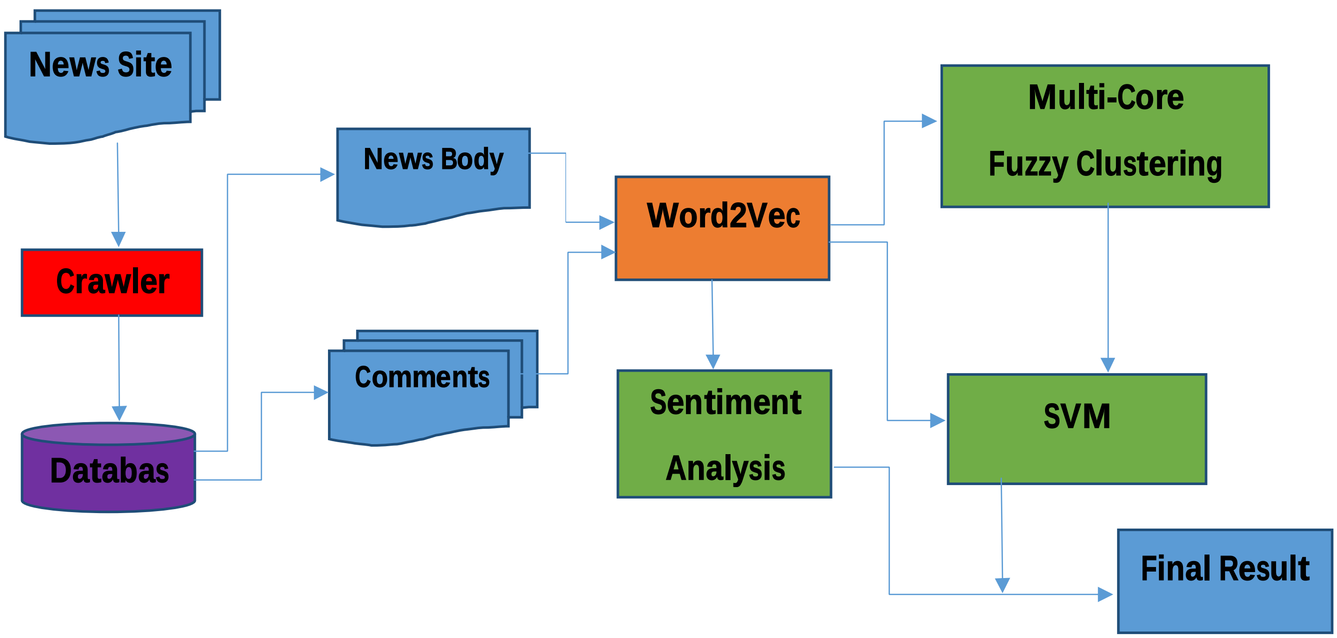

Figure 3: The architecture utilized in the proposed method.

In the proposed model, news articles and relevant comments are first fetched from a website using a web crawler and stored in a local database. Then, in the word-to-vector conversion step, the text of the news articles is preprocessed by using the "hazm" library, which includes tasks such as removing stop words, stemming, and tokenization. The comments, on the other hand, are vectorized using ParsBERT. The news sentences are then sent to the fuzzy multi-core clustering component for aspect extraction. Various aspects of the news are extracted in this step. Subsequently, an SVM model is trained for pairwise classification of different aspects using the extracted aspects. Each cluster's members are considered as one aspect, and the members of a cluster are used as a class for training the SVM model. The vectorized aspects related to the comments are used as input to this model, and the corresponding class, which represents the aspect related to the comment, is determined using this model. Additionally, the comments are semantically analyzed by converting them to vectors and applying weighted scoring using a semantic dictionary. The output of each comment, along with the related aspect, is then reported as the final output.

3-2. Data Extraction from the Target Website

The initial step involves extracting the text of user comments and news articles from the target website. To accomplish this, we have developed a Python-based web crawler. The crawler is specifically designed to retrieve a set of news articles from the Tabnak news website (tabnak.ir) along with the associated published comments. Since the objective of this article is to extract various aspects and analyze them, we only consider news articles that have a minimum of 10 published comments. News articles with fewer comments are excluded from the analysis. The pseudocode for the data extraction crawler from the target website is illustrated in Figure 4.

```
For i in range(news_code):
    If count(comment) ≥ 10:
        Extract title and abstract and body of news
            Foreach comments as comment:
                Extract comment_text
```

Figure 4: Pseudocode for data extraction crawler from the target website.

3-3. Map-Redeuce Framework

The proposed method consists of three stages. In the first stage, the number of clusters and cluster representatives are randomly determined, and the data is divided into subgroups. In the second stage, the mapping operation is performed. The news sentences are separated and each one is transformed into a vector. The same process is done separately for the comments as well. In the key-value pair, the intermediate key represents the news or comment code, and the value represents the vector representation of the news or comment, which is emitted. Then, using the fuzzy multi-cluster clustering method, different aspects of the news are extracted. Next, the intermediate key is set as the news or comment code, and the value is set as their vector representation. In the third stage, the reduction operation is performed. In this stage, using the fuzzy multi-cluster clustering, the aspects related to the news are extracted, and using the Support Vector Machine, each comment is associated with the relevant aspect. In the reduction stage, the matrix V, with each column representing a cluster representative, and the matrix U, which is the fuzzy membership matrix, are updated. The termination condition is based on minimizing the objective function and ensuring that the change in its numerical value is below a predefined threshold. The general logic of the proposed method can be summarized as follows:

- Input: Text of relevant news and comments, number of clusters, weight vector of kernels W, cluster representative matrix V, fuzzy degree controller m, feasibility membership matrix T, weight factor for feasibility membership degree, and μ_c as a scaling parameter.
- Receive inputs.
- Randomly create the cluster representative matrix V.
- Randomly divide the input data into subgroups within the framework of mapping-reduction: a) In the mapping stage, separate the news sentences and transform each one into a vector. Similarly, perform this process separately for the comments. Set the intermediate key as the news or comment code, and the value as the vector representation of the news or comment, and emit it. Then, extract different aspects of the news using the fuzzy multi-cluster clustering method. Next, set the intermediate key for the news and comment as the news or comment code, and set the value as their vector representation. b) In the reduction stage, using the fuzzy multi-cluster clustering, extract the aspects related to the news, and associate each comment with the relevant aspect using the Support Vector Machine.
- Check the termination condition.
- End.

3-4. Selection of Base Kernels

In various experiments, different types of kernel functions with different numbers can be used as base kernels. In real-world problems, there is no definitive guide for selecting an appropriate set of kernels, and different sets of kernels yield different clustering performances. In our proposed method, we utilize three categories of kernel functions: spectral-based kernel functions, Gaussian kernel functions, and polynomial kernel functions to construct the set of kernels $\{K\}_{k=1}^{M}$.

Let's assume $X=[x_1,x_2,\ldots,x_N]_{l\times N}$ is an $l\times N$ matrix where each column represents a data vector from the R^l space, and $V=[v_1,v_2,\ldots,v_N]_{N\times N}$ are the eigenvectors of the linear kernel matrix X^TX. In the first category of kernels, the M_v kernel matrix $\{k_1^v, k_2^v, \ldots, k_{M_v}^v\}$ is constructed according to Equation 5:

$$k_k^V = V_k^T V_k, \quad k = 1,2,\ldots,M_v \qquad \text{(Equation 5)}$$

In the second category, the M_g kernel matrix $\{k_1^v, k_2^v, \ldots, k_{M_v}^v\}$ is constructed using Gaussian mapping according to Equation 6:

$$k_k^V(x_i, x_j) = \exp\left(\frac{(x_i - x_j)^T (x_i - x_j)}{2^{(M_g - k)}\sigma_X}\right), k = 1,2,\ldots,M_g \qquad \text{(Equation 6)}$$

Where σ_X is the standard deviation $\binom{N}{2}$ of pairwise distances between points in the data set. The multiplier $2^{(M_g-k)}$ in the denominator of the fraction leads to the generation of kernel matrices with different scales. In the last category, polynomial kernel functions are used as shown in Equation 7:

$$k_k^P(x_i, x_j) = \left(x_i^T x_j - C\right)^T, k = 1,2, \dots, M_P \qquad \text{(Equation 7)}$$

Therefore, the matrix $M=M_v+M_g+M_p$ represents the kernel matrix, as described in Equation 8. These kernels will be used as the base kernels in our experiments.

$$\left\{k_1^v, k_2^v, \dots, k_{M_v}^v, k_1^g, k_2^g, \dots, k_{M_g}^g, k_1^P, k_2^P, \dots, k_{M_P}^P\right\} \qquad \text{(Equation 8)}$$

4. Implementation and Evaluation

Due to the unavailability of Persian datasets on the specific topic under investigation, efforts were made to collect a dataset from the Tabaanak news website during the period of April 2022. This dataset was used for labeling the submitted comments. It consisted of 4,610 news articles and 74,336 comments. Among these articles, 1,593 were without comments, and the remaining 3,017 articles contained user comments. The collected dataset for training the reaction prediction model included news articles from September 2020 to September 2021, totaling 75,069 articles and 7,422,204 comments. The dataset was obtained using a Python-based web crawler that retrieved all published and unpublished comments from the target website's database. It's important to note that the collected dataset is specific to the Tabaanak news website and may not represent a diverse range of topics or sources. Therefore, caution should be exercised when generalizing the findings or applying the trained models to other contexts. Additionally, the dataset collection process and the performance of the web crawler should be validated and verified to ensure the reliability and accuracy of the collected data. In the next step, various aspects of each news article were extracted, and each comment was associated with one or more relevant aspects. To ensure the accuracy of this process, two experts identified the aspects of each news article. In case of disagreement, a third expert determined the final aspects based on the extracted aspects provided by the previous two experts. Similarly, for comments, three experts assigned them to the corresponding aspects, and in case of disagreement, the majority opinion was considered. To evaluate the proposed method, a confusion matrix was utilized. The ambiguity matrix is one of the most important evaluation metrics, and its components can be seen in Table 1.

Table 1: Confusion Matrix

		Actual class	
		Positive	**Negative**
Predicted class	**Positive**	TP(True Positive)	FP(False Positive)
	Negative	FN(False Negative)	TN(True Negative)

The confusion matrix is a crucial evaluation measure used to assess the performance of the proposed method in aspect extraction and sentiment analysis. It consists of a 2x2 matrix that captures the following components:

1. True Positive (TP): This component represents the instances where the proposed method correctly identifies and predicts positive aspects or sentiments.
2. False Negative (FN): This component refers to the cases where the proposed method fails to recognize positive aspects or sentiments, despite their presence in the actual data.
3. False Positive (FP): This component indicates the situations where the proposed method incorrectly identifies positive aspects or sentiments that do not exist in the actual data.
4. True Negative (TN): This component represents the instances where the proposed method correctly recognizes the absence of positive aspects or sentiments.

These components provide valuable insights into the accuracy, precision, recall, specificity, and a combination of these measures called the F-score. Table 2 provides the definitions of these metrics:

Table 2: Evaluation Metrics for Machine Learning Models

Metric	Formula
Accuracy	
Precision	
Recall	
F-score	
Specificity	

- Accuracy: It measures the overall correctness of the aspect extraction and sentiment analysis results by considering the ratio of correctly predicted instances (TP and TN) to the total number of instances.
- Precision: It quantifies the proportion of correctly predicted positive aspects or sentiments (TP) to the total number of instances predicted as positive (TP and FP). It reflects the model's ability to correctly identify positive aspects or sentiments without misclassifying negatives.
- Recall (Sensitivity): Also known as True Negative Rate (TNR), it calculates the ratio of correctly predicted positive aspects or sentiments (TP) to the total number of instances that are actually positive (TP and FN). It represents the model's ability to capture all positive aspects or sentiments without missing any.
- Specificity: It measures the ability of the model to correctly predict negative aspects or sentiments by considering the ratio of correctly predicted negatives (TN) to the total number of instances that are actually negative (TN and FP).
- F-score: It combines precision and recall into a single measure that balances both metrics. The F-score is calculated using the harmonic mean of precision and recall, providing a comprehensive evaluation of the model's performance.

These metrics offer valuable insights into different aspects of the aspect extraction and sentiment analysis tasks and help assess the effectiveness of the proposed method.

After conducting preliminary analysis on user feedback and conducting a survey through the website's editorial team, designated labels were considered for the comments, including emotional labels. Based on the model presented in [30], six basic emotions were defined: happiness, sadness, anger, disgust, fear, and surprise. Due to the difficulty in accurately distinguishing between the emotions of anger and disgust, and the emotions of fear and surprise, they were grouped together.

A mobile application was designed and developed for comment labeling. In this application, each user can view a list of news articles and by clicking on each article, they can see the content of the article, including the title, summary, and full text, along with the comments on that article. In this application, the following emojis were used to represent emotions: for happiness, 😔 for sadness, 😟 for anger and disgust, 😨 for fear and surprise, and 😶 for comments without any specific emotion. Users were requested to select the relevant emotion for each comment based on the provided descriptions. The instructions provided to users for labeling comments are as follows:

- Each comment should be categorized into one of the following emotion categories based on the commenter's or writer's emotion:
    - for happiness, satisfaction, and agreement

- 😔 for sadness, grief, and sorrow
- 😧 for anger, frustration, and resentment
- 😶 for comments without any specific emotion

A total of 170,823 labels were assigned by 16 users, corresponding to 67,887 unique comments. Based on the distribution of labels among users, some of them assigned one label, some assigned two labels, and some assigned three or more labels to the comments. The statistical chart illustrating this distribution is shown in Figure 5.

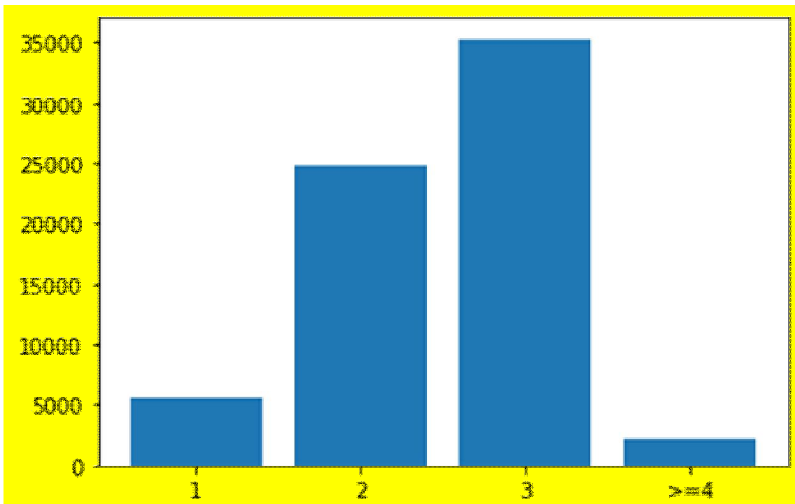

Figure 5: Statistics of the number of comments and the number of labels assigned to them.

From the selected comments, those that have been labeled by three or more users were chosen. Among these comments, the ones that had a majority agreement among users regarding the corresponding emotion were selected. The distribution of comments in emotion types is shown in Figure 6.

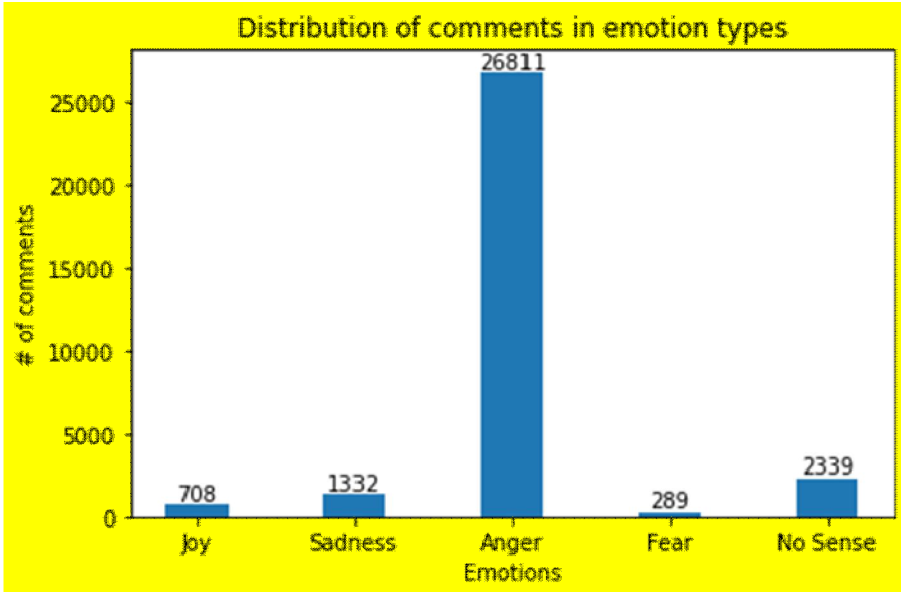

Figure 6: Distribution of comments in emotion types.

In the following, the obtained results from the proposed model were compared with labeled data. In one stage, the accuracy of aspect detection was measured, and in the subsequent stage, the correctness of identifying the aspect for each opinion was determined using accuracy, precision, recall, F-score, and specificity metrics, as described in Table 2. If an aspect was not extracted from a news article, all associated opinions were considered incorrect model responses. Based on these explanations, the results presented in Table 3 were obtained.

Table 3: Evaluation Results in Percentage

Description	Accuracy	Precision	Recall	F-score	Specificity
Aspect Extraction	87.09	87.25	82.12	86.12	84.37
Determining Aspects Related to Each Opinion	75.65	74.93	79.60	75.14	76.08

Indeed, it is worth mentioning that the subject of this article has a strong dependency on the language under study. Due to the innovative nature of the proposed approach, no research has been conducted on the specific topic in relation to the Persian language on websites. Therefore, it is not possible to compare it with similar works. However, similar works have been carried out on English websites, as mentioned in the research background. As previously mentioned, this method heavily relies on the semantic dictionary of each language for labeling the data, making it impossible to use a dataset from one language (e.g., Persian) for another language (e.g., English). The creation of a Persian dataset is another innovation of this article. The comparison of the obtained results for the evaluation metrics indicates the satisfactory performance of the proposed method in this article.

5. Conclusion and Future Work

Establishing a connection between news articles and audience reactions regarding different aspects of the news provides valuable insights for analysts and top-level managers. In this article, considering the time and financial costs associated with obtaining labeled data, we utilized a fuzzy clustering method, which is an unsupervised approach, to extract various aspects of a news article and associate the opinions with the extracted aspects. To implement this approach, given the absence of a reference dataset in the Persian language, we took the initiative to create a gold-standard dataset. This dataset is one of the key achievements of this article. For this purpose, we employed the assistance of three experts to extract aspects, label them, and establish correspondence with the opinions. To evaluate the performance, accuracy, precision, recall, specificity, the confusion matrix, and the composite metric F-score were utilized. A confusion matrix that summarizes the model's predictions and actual labels, showing the counts of true positives, true negatives, false positives, and false negatives. It provides insights into the model's performance for each class. The obtained values for these metrics indicate the favorable performance of the proposed method. The reason behind this can be attributed to the utilization of the fuzzy clustering model and its ability to model nonlinear relationships in the data, the use of dimensionality reduction techniques to enhance scalability, the unsupervised nature of the approach due to the unavailability of labeled data, and the focus of this article on considering aspects in news articles. It is recommended that in future work, in addition to considering the internal structure of the news article for aspect extraction, the submitted opinions for that article and other related news articles should also be taken into account to better identify the main and sub aspects of the news. This is the intended direction for future work as proposed by the authors of this article. Lastly, it is necessary to emphasize that, as observed, the topic of this article has a strong dependence on the lexical semantics of each language. Since no similar studies have been conducted on Persian-language websites, this article was unable to compare its results with other studies. Therefore, it is hoped that this article can serve as a basis for comparing and evaluating the studies of other researchers in order to improve the results in the future.

Declarations

Ethical Approval

All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards. This article does not contain any studies with animals performed by any of the authors.

Competing interests

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Authors' contributions

Seyed Omid Azarkasb, Seyed Hossein Khasteh and Mostafa Amiri. These authors contributed equally to this work.

Authors and Affiliations

K.N. Toosi University of Technology, Tehran, Iran.

Seyed Omid Azarkasb

K.N. Toosi University of Technology, Tehran, Iran.

Seyed Hossein Khasteh

K.N. Toosi University of Technology, Tehran, Iran.

Mostafa Amiri

Corresponding author

Correspondence to Seyed Omid Azarkasb

Funding

This research was not funded.

Availability of data and materials

A significant amount of data is addressed in this article. The remaining data that support the findings of this study are available on request from the corresponding author. The data are not publicly available due to privacy or ethical restrictions. The authors declare that all the experimental data in this paper are true and valid. Moreover, The authors declare that all experimental data are obtained from detailed experiments.

References

[1] E. Georgiadou, S. Angelopoulos, H. Drake, **"Big Data Analytics and International Negotiations: Sentiment Analysis of Brexit Negotiating Outcomes"**, International Journal of Information Management, Vol. 51, pp. 1–9, 2020.

[2] E.M. Clark, T. James, C.A. Jones, A. Alapati, P. Ukandu, C.M. Danforth, P.S. Dodds, **"A Sentiment Analysis of Breast Cancer Treatment Experiences and Healthcare Perceptions Across Twitter",** http://arxiv.org/abs/1805.09959, 2018.

[3] R-P Shen, H-R Zhang, H. Yu, F. Min, **"Sentiment Based Matrix Factorization with Reliability for Recommendation"**, Expert Systems with Applications, Vol. 135, pp. 249-258, 2019.

[4] S. Biswas, G. poornalatha, "**Opinion Mining using Multi-Dimensional Analysis**", IEEE Access, Vol. 11, pp. 25906-25916, 2023.

[5] M. Birjali, M. Kasri, A. Beni-Hssane, **"A Comprehensive Survey on Sentiment Analysis: Approaches, Challenges and Trends",** Knowledge-Based Systems, Vol. 226, pp. 107-134, 2021.

[6] B. Lio, **"Sentiment Analysis and Opinion Mining",** Morgan & Claypool Publishers, Vol. 5, pp. 1-167, 2012.

[7] A.G. Lopes, L. Fonseca, **"Tools and Techniques for Opinion Mining"**, 9th International Symposium on End-User Development, Cagliari, Italy, 2023.

[8] S.P. Algur, J.G. Biradar, **"Opinion Mining and Review Spam Detection: Issues and Challenges",** International Journal of Advanced Research in Computer Science and Software Engineering, Vol. 7, pp. 243-250, 2019.

[9] X. Liu, M. Li, L. Wang, Y. Dou, J. Yin, E. zhu, **"Multiple Kernel K-Means with Incomplete Kernels"**, Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence, pp.2259-2265, 2017.

[10] R. K. Sanodiya, S. Saha, J. Mathew, **"A Kernel Semi-Supervised Distance Metric Learning with Relative Distance: Integration with a MOO Approach"**, Expert Systems with Applications, Elsevier, Vol. 125, pp. 233-248, 2019.

[11] M. Soleymani Baghshah, S. Bagheri Shouraki, **"Efficient Kernel Learning from Constraints and Unlabeled Data"**, 20th International Conference on Pattern Recognition, Istanbul, Turkey, pp. 3364-3367, 2010.

[12] S. Zhu, D. Wang, T. Li, **"Data Clustering with Size Constraints"**, Knowledge-Based Systems, Elsevier, Vol. 23, pp. 883-889, 2010.

[13] O. Ozdemir, A. Kaya, **"Comparison of FCM, PCM, FPCM and PFCM Algorithms in Clustering Methods"**, Afyon Kocatepe University Journal of Science and Engineering, pp. 92-102, 2019.

[14] F. Hai Jun, W. Xiao Hong, M. Han Ping, W. Bin, **"Fuzzy Entropy Clustering Using Possibilistic Approach",** Advanced in Control Engineering and Information Science, Elsevier, Procedia Engineering Vol. 15, pp.1993-1997, 2011.

[15] M. Bouzbida, L. Hassine, A. Chaari, **"Robust Kernel Clustering Algorithm for Nonlinear System Identification"** Hindawi, Mathematical Problems in Engineering, pp. 1-11, 2017.

[16] A.A. Abin, H. Beigy, **"Active Constrained Fuzzy Clustering: A Multiple Kernels Learning Approach"**, Pattern Recognition, Vol. 48, Issue. 3, pp. 935-967, 2015.

[17] J. Serrano-Guerrero, F. P. Romero, J. A. Olivas, "**Fuzzy Logic Applied to Opinion Mining: A Review**", ELSEVIER, Knowledge-Based Systems, Vol. 222, 2021.

[18] M. Bordoloi, S.K. Biswas, "**Sentiment Analysis: A survey on Design Framework, Applications and Future Scopes**", Artificial Intelligence Review, 2023.

[19] Q. Ye, Z. Zhang, R. Law, **"Sentiment Classification of Online Reviews to Travel Destinations by Supervised Machine Learning Approaches"**, Expert Systems with Applications, Vol. 36, Issue 3, Part 2, pp.6527-6535, 2009.

[20] K. Xu, SS. Liao, J. Li, Y. Song, **"Mining Comparative Opinions from Customer Reviews for Competitive Intelligence"**, Decision support systems, Vol. 50, No. 4, pp.743-754, 2011.

[21] P. Hajek, A. Barushka, M. Munk, "**Opinion Mining of Consumer Reviews using Deep Neural Networks with Word-Sentiment Associations"**, Artificial Intelligence Applications and Innovations, Vol. 583, pp 419-429, 2020.

[22] F. Ali, K.S. Kwak, Y-G. Kim, **"Opinion Mining Based on Fuzzy Domain Ontology and Support Vector Machine: A Proposal to Automate Online Review Classification"**, Applied Soft Computing, Vol. 47, pp. 235-250, 2016.

[23] W. Souma, I. Vodenska, H. Aoyama, **"Enhanced News Sentiment Analysis using Deep Learning Methods"**, Journal of Computational Social Science, Vol. 2, pp. 33-46, 2019.

[24] J. Jiang, , K. Zhou, C. Xong, G. Du, C. Ou, J. Zhang, **"KSCB: A Novel Unsupervised Method for Text Sentiment Analysis"**, Applied Intelligence, Vol. 53, No. 1, PP. 301-311, 2023.

[25] S.M. Rezaeinia, R. Rahmani, A. Ghodsi, H. Veisi, **"Sentiment Analysis Based on Improved Pre-Trained Word Embeddings"**, Expert Systems with Applications, Vol. 117, pp. 139-147, 2019.

[26] U. Naseem, I. Razzak, K. Musial, M. Imran, **"Transformer based Deep Intelligent Contextual Embedding for Twitter sentiment analysis"**, Future Generation Computer Systems, Vol. 113, pp. 58-69, 2020.

[27] J. Devlin, M-W Chang, K. Lee, K. Toutanova, **"BERT: Pre-Training of Deep Bidirectional Transformers for Language Understanding"**, Conference of the North American Chapter of the Association for Computational Linguistics, Vol. 1, pp. 4171-4186, 2019.

[28] M. Ulcar, M. Robnik-Sikonja, **"Cross-Lingual Alignments of ELMo Contextual Embeddings"**, Neural Computing and Applications, Vol. 34, pp.13043–13061, 2022.

[29] A. Borg, M. Boldt, **"Using VADER sentiment and SVM for Predicting Customer Response Sentiment"**, Expert Systems with Applications, Vol. 162, 2020.

[30] H. Zhuang, F.Guo, C. Zhang, L. Liu, **"Joint Aspect-Sentiment Analysis with Minimal User Guidance"**, 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval, 2020.