ISSN: 2454-9940

INTERNATIONAL JOURNAL OF APPLIED SCIENCE ENGINEERING AND MANAGEMENT

E-Mail :
editor.ijasem@gmail.com
editor@ijasem.org

www.ijasem.org

INTERNATIONAL JOURNAL OF APPLIED SCIENCE ENGINEERING AND MANAGEMENT

ISSN2454-9940

www.ijasem.org

Vol 18, Issue 3, 2024

SENTIMENT ANALSIS OF CUSTOMERS REVIEW USING A HYBRID EVOLUTION ARY SVM-BASED APPROACH IN AN IMBALANCED DATA DISTRIBUTIONS

[1] K VENKATESH, [2] K.SRI DEVI

[1](Assistant Professor), MCA, **DANTULURI NARAYANA RAJU COLLEGE(A) PG COURSES, BHIMAVARAM ANDHRA PRADESH**

[2]MCA, scholar, **DANTULURI NARAYANA RAJU COLLEGE(A) PG COURSES, BHIMAVARAM ANDHRA PRADESH**

ABSTRACT

Online media has an increasing presence on the restaurants' activities through social media websites, coinciding with an increase in customers' reviews of these restaurants. These reviews become the main source of information for both customers and decision-makers in this field. Any customer who is seeking such places will check their reviews first, which usually affect their final choice. In addition, customers' experiences can be enhanced by utilizing other customers' suggestions. Consequently, customers' reviews can influence the success of restaurant business since it is considered the final judgment of the overall quality of any restaurant. Thus, decision-makers need to analyze their customers' underlying sentiments in order to meet their expectations and improve the restaurants' services, in terms of food quality, ambiance, price range, and customer service. The number of reviews available for various products and services has dramatically increased these days and so has the need for automated methods to collect and analyze these reviews. Sentiment Analysis (SA) is a field of machine learning that helps analyze and predict the sentiments underlying these reviews. Usually, SA for customers' reviews face imbalanced datasets challenge, as the majority of these sentiments fall into supporters or resistors of the product or service. This work proposes a hybrid approach by combining the Support

ISSN2454-9940

www.ijasem.org

Vol 18, Issue 3, 2024

Vector Machine (SVM) algorithm with Particle Swarm Optimization (PSO) and different oversampling techniques to handle the imbalanced data problem. SVM is applied as a machine learning classification technique to predict the sentiments of reviews by optimizing the dataset, which contains different reviews of several restaurants in Jordan. Data were collected from Jeeran,a well-known social network for Arabic reviews. A PSO technique is used to optimize the weights of the features, as well as four different oversampling techniques, namely, the Synthetic Minority Oversampling Technique (SMOTE), SVM-SMOTE, Adaptive Synthetic Sampling (ADASYN) and borderline-SMOTE were examined to produce an optimized dataset and solve the imbalanced problem of the dataset. This study shows that the proposed PSO-SVM approach produces the best results compared to different classification techniques in terms of accuracy, F-measure, G-mean and Area Under the Curve (AUC), for different versions of the datasets.

1.INTRODUCTION

The popularity of social media websites has witnessed tremendous growth in the last few years [1]. Social media sites have grown not only in terms of volume but also in their importance to different aspects of life, including business, politics, and education [2]. Nowadays, all businesses are offering their products and services online. These sites allow consumers to share their experiences and recommendations about these businesses' products, places, and services on different platforms such as Trip Advisor, Yelp, Face book and Jeeran [3].

Online reviews represent the electronic version of word of mouth (WOM), which is an important aspect of in traditional marketing. While WOM is restricted to family, friends or close people, online reviews have a worldwide reach [4]. Many websites allow users to rate and review different products and services. These reviews become the main source of information for potential customers who are seeking such products [5]. A survey conducted by Bright Local (2020) found that 79% of customers trust online reviews as much as personal recommendations [6].

These days, whenever a customer wants to buy a new product online, he or she will consider what other people think about it, how they rate it, and their feedback and comments about the product before making

a purchase [7]. According to a Bright Local survey in 2020, 87% of consumers had checked online reviews of local businesses [6]. These reviews may affect the customer's final choice since people trust customers' reviews more than advertisements produced by a company. Furthermore, customers' experiences can be enhanced by utilizing other customers' suggestions [3].

Due to the widespread availability of social websites and applications, the number of reviews available for various products has dramatically increased [8], and so has the need for automated methods to collect and analyze these reviews [9]. These methods are essential to speed up and improve the quality of decision making process [10].

Sentiment Analysis (SA) can be used to deduct users' feelings about various topics by processing their implicit attitudes and analyzing the underlying sentiments hidden in their comments [11]. Sharif *et al.* [12] defined SA as ``analyzing people's sentiments, opinions, appraisals, attitudes, evaluations and emotions towards such entities as organizations, products, services, individuals, topics, issues, events and their attributes, as presented online via text, video and other means of communication." Sentiment analysis is also referred to as opinion mining based on natural language processing, text analysis, and computational techniques [13]. It can be applied at the document level, sentence level, or aspect level [14]. It aims to classify customers' attitudes towards a product or service as expressed in the comments, reviews, and posts as positive, negative, or neutral comments [15]. Two main SA approaches can be followed, namely, machine learning approaches and the lexicon-based approach [16]. Different machine learning algorithms are used to evaluate results in the sentiment field; the most common ones are Naïve Bayes (NB), SVM, Logistic Regression (LR), Random Forest (RF), and K-Nearest Neighbours (k-NN) [13].

Sentiment analysis is essential for every business as it can be used to improve the decisions of customers, business owners, and service providers [17]. SA is used by business owners to enhance their businesses' image and increase their success [3] since it helps decision-makers improve the quality of their products and services based on their customers' reviews; thus, the business can provide more praiseworthy services. This leads to higher customer satisfaction and more sales and revenues for the business [12]. On the other hand, customers can

utilize these reviews in making thoughtful decisions based on previous customers' experiences [17].

Recently, it has been noticed that almost all restaurant have presence in the online social world. Restaurants are becoming increasingly present on different social websites, and so are customers' reviews of these restaurants [18]. Online restaurant reviews are considered a rich source of information that helps attract new customers. Checking reviews by locals and tourists before visiting restaurants has become a trend [17]. This is supported by 2020 Bright Local survey revealing that 93% of consumers check a restaurant's reviews before visiting it [6]. Consequently, customers' reviews can influence the success of restaurant business [19]. It was found that the more positive comments a restaurant receives the more customers visit its web pages and physical locations, which leads to more popularity and success [4]. In contrast, negative comments lead to the loss of trustworthiness of the restaurant and reduced revenue [17]. According to 2020 Bright Local survey, 94% of consumers are more likely to by from a business if it has received positive reviews, while 92% are less likely to use it if it has been given bad reviews [6]. People tend to post reviews when they either a strong positive or strong negative experience (generally, the number of positive reviews exceeds the number of negative ones) [4].

Customers' reviews and opinions are considered the final judgment of the overall quality of any restaurant. Thus, owners need to analyze their customers' underlying sentiments so they can meet their expectations and offer customized services in terms of food quality, ambiance, price, and customer service [18].

Many studies have followed a Machine Learning (ML) approaches for restaurants sentiment analysis. A study done by Zahoor *et al.* [3] used NB Classifier, logistic regression, SVM, and RF methods to analyze customers' sentiments about restaurants in Karachi. The study annotated 4000 reviews from a well-known Pakistani Face book community called SWOT'S. Random forest gained the highest performance, with an accuracy of 95%. Another study conducted by Sharif *et al.* [19] classified customers reviews for 1000 restaurants (written in Bengali) into positive and negative classes using three machine learning algorithms, namely, Decision Tree (DT), RF, and multinomial NB. The results

showed that the multinomial NB method achieved the best results, with 80.48% accuracy.

Furthermore, sentiment analysis can be used to build a recommender system in different fields including the restaurant industry. Asani *et al.* [11] for example collected people's sentiments from the Trip Advisor website and built a customized restaurant recommender system based on people's opinions and food preferences. The recommender system suggests restaurants according to users' preferences, thus helping them to choose the best option and make an informed decision. Choosing the best restaurant among many unknown options is an important decision, especially for tourists and travelers.

Few studies have followed an evolutionary approach in the restaurant field. Govindarajan [20], for example, applied a hybrid classification method based on restaurant reviews found on Yelp. The study used NB, SVM, and Genetic Algorithm (GA) and then compared their performances with the proposed hybrid model built by coupling all three classification methods. Another study conducted by Somantri *et al.* [21] proposed a hybrid model for restaurant culinary food reviews in Indonesia. The study confirmed the efficiency of PSO, as it used a hybrid model based on Particle Swarm Optimization (PSO) and Information Gain (PSO-IG) with four different classification algorithms, namely SVM, NB, DT, and K-NN. The best results were achieved using the proposed PSO-IG method with the NB classifier. The main limitation of this work was that it ignored the imbalanced
nature of the dataset, as positive reviews were significantly more common than negative reviews. All aforementioned studies were applied to English sentiment; to the best of our knowledge, our work is the first to use such recent evolutionary approach to explore Arabic sentiments.

This work uses the Jeeran website to collect people's comments about different restaurants in Jordan. The Jeeran website is a social platform on which customers can post their reviews about more than 300,000 different places, including shopping centers, cafes, restaurants, or doctors' . People use Jeeran to find the best places and services in their cities and avoid bad experiences. Customers post thousands of comments on different social platforms every day, mainly in the Arabic language. This study is conducted on Arabic sentiment since it is the

fifth-most widely spoken language in the world and the first language of more than 422 million people [22]. Moreover, about 185 million web users are Arabic speakers [23]. The Arabic language is a more challenging language to study than to English for many reasons. Firstly, Arabic has a dialectal variety; people often post their comments in dialectical Arabic rather than Modern Standard Arabic, thus requiring more complex prepossessing [24]. Another reason is the morphology of the Arabic language, meaning the same word may have a different meaning, even if it has the same root. Also suffixes, affixes, and prefixes added to the same word may carry essential information [23]. Moreover, the richness of synonyms in the Arabic language plays a key role in its complexity. Furthermore, the same word may have different meanings according to the context, and a word can _t into more than one lexical category [22].

In addition, the problem of imbalanced datasets is very common since the majority of the sentiments of customers fall into either the supporter or resistor category. Parameter tuning of the oversampling technique is another challenge. Moreover, a large number of features is generated by tokenizing the sentiments of the customers. Thus, leading methodologies can be applied to feature analysis to achieve the best outcome from the classification process. Optimization techniques are well-suited for feature analysis and parameter tuning, making the classification process more reliable. Another challenge relates to the choice of tokenization method used to form the dataset for the classification process, and it can be recognized through experimental practice. This research proposes an evolutionary approach to analyzing people's sentiments regarding restaurants' reviews in the Arabic language. Furthermore, this work followed an evolutionary hybrid approach by combining the PSO evolutionary algorithm with different oversampling techniques and the SVM algorithm to automatically detect the sentiment in the customers' comments. Four different oversampling techniques are applied to handle the problem of imbalance in the dataset. Additionally, the applied evolutionary algorithm helps reduce the effort and time needed to tune the parameters and optimize the classification by finding the best feature weights and best *k* value for the oversampling technique, thereby resulting in better performance measures.

This work collects the reviews for almost 3000 restaurants from the Jeeran website. After the data preparation process, four different versions of the dataset are presented using different tokenization methods. The initial individual of the study is created of random weights and a random k value for the oversampling parameter. The weighted oversampled data is then classified using the SVM classification technique and the results are evaluated using G-mean. A Particle Swarm Optimizer (PSO) evolutionary algorithm is then used to optimize the values of the individual and achieve a better G-mean. Finally, the proposed approach is compared with different standard and powerful classification models, including SVM, XGBoost, DT, RF, NB, k-NN, and LR based on Accuracy, $F1P$, $F1N$, G-mean, and AUC evaluation measures.

The main contributions of the study can be summarized as follows V

_ Collecting the dataset from the Jeeran website with approximately 3000 restaurant reviews. The dataset is cleaned, labelled, formatted, and stemmed.

_ Oversampling the dataset using four different oversampling techniques to solve the imbalanced problem.

_ Applying the PSO optimization technique to find the best weights for the dataset features and the best *k* value for each oversampling technique, then applying the SVM classification technique to the oversampled and weighted dataset to find the sentiments of the restaurant reviews.

The remainder of the paper is divided as follows: Section II presents a review of the literature on restaurants' sentiment analysis. Section III introduces the backgrounds of different methods and concepts that have been used. The proposed approach is described in Section IV. Section V discusses the experiments and provides the results achieved by the proposed approach and other models. Finally, the conclusion and future directions are offered in Section VI.

2. EXISTING SYSTEM

As explained by Tubishat *et al.* [26], SA, which is also referred to as opinion mining, is a text-classification field in which people's opinions, evaluations, attitudes, moods, and emotions regarding a service or product are analyzed to detect orientations. It is conducted computationally by using natural language processing, linguistics, or text analysis to detect the feelings expressed within informal text posted online. Recently,

due to the popularity of social networks and online review websites, people tend to check a restaurant's reviews before visiting it. As a result, customer's impressions have become a vital factor influencing the success of restaurants; the interest among decision-makers toward customers' experience about services provided has also increased as stated by Sharif *et al.* [19]. SA has been applied to online reviews about restaurants in the literature. For instance, Gan *et al.* [5] studied the attributes representing consumers' reviews of restaurants. This study found that the attributes derived from previous studies such as food, service, ambiance, and price were not enough to affect restaurants' ratings and that context should be added as a significant attribute. Meanwhile, Aye and Aung [1] proposed a Myanmar language resource for lexicon-based SA as a solution to language-specific problems since most studies have considered the English language for SA. Restaurant review data were used, but informal expressions were not addressed.

Since online booking websites gained substantial interest recently, and since people now check hundreds of reviews before making any booking decisions, Agüero-Torales *et al.* [27] proposed a cloud-based software tool to analyze data from the TripAdvisor website by conducting SA on them in the province of Granada. The SA task was accomplished by using various datasets, such as the Yelp dataset, to examine the approaches proposed by other researchers as explained by Hegde *et al.* [28] and made public for research and academic studies. The Zomato Restaurant Dataset is derived from the online multinational restaurant aggregator in which reviews are posted alongside information, menus, and delivery options. Also, Taneja *et al.* [29] discussed that Zomato is a very rich database that includes information on more than 20,000 restaurants. Zomato API enables users to access the most up-to-date content and generate information about nearby restaurants. Furthermore, SemEval Datasets are high-quality annotated datasets generated through a series of international workshops; different versions of these datasets (e.g., SemEval-2015, SemEval-2016) have been used in literature as stated by Khan *et al.* [30].

Various ML algorithms have been used to conduct SA in the restaurants domain. NB, LR, and DT. ML algorithms were applied by Hassan *et al.* [32] to conduct SA on three different datasets,

ISSN2454-9940

www.ijasem.org

Vol 18, Issue 3, 2024

namely, the Yelp dataset, IMDB dataset, and Arabic qaym.com restaurant reviews dataset. Performance was measured in terms of accuracy and recall. NB and LR recorded the best results. Similarly, NB, SVM, multilayer perceptron, DT, k-NN, and fuzzy logic were applied by Kumar and Jaiswal [33] on data extracted from Twitter and Tumblr, which are widely used micro-blogging social networks. A comparative analysis of performance is presented in terms of precision, recall, and accuracy. Besides, a deep learning model called DOC-ABSADeepL was proposed by Zuheros *et al.* [34] and applied on the Trip Advisor dataset for restaurants to categorize the aspects included in an expert review while also extracting opinions and criteria. The tripR-2020 dataset was built, manually annotated, and released before being used in the same study.

DISADVANTAGES

The system is not implemented Pseudo-Code of Particle Swarm Optimizer to optimize datasets.

In The support vector machine (SVM) algorithm is a supervised classifier that is not applied widely to solve classification and regression problems.

3. PROPOSED SYSTEM

This research proposes an evolutionary approach to analyzing people's sentiments regarding restaurants' reviews in the Arabic language. Furthermore, this work followed an evolutionary hybrid approach by combining the PSO evolutionary algorithm with different oversampling techniques and the SVM algorithm to automatically detect the sentiment in the customers' comments. Four different oversampling techniques are applied to handle the problem of imbalance in the dataset. Additionally, the applied evolutionary algorithm helps reduce the effort and time needed to tune the parameters and optimize the classification by finding the best feature weights and best *k* value for the oversampling technique, thereby resulting in better performance measures.

This work collects the reviews for almost 3000 restaurants from the Jeeran website. After the data preparation process, four different versions of the dataset are presented using different tokenization methods. The initial individual of the study is created of random weights and a random k value for the oversampling parameter. The weighted oversampled data is then classified

using the SVM classification technique and the results are evaluated using G-mean. A Particle Swarm Optimizer (PSO) evolutionary algorithm is then used to optimize the values of the individual and achieve a better G-mean. Finally, the proposed approach is compared with different standard and powerful classification models, including SVM, XGBoost, DT, RF, NB, k-NN, and LR based on Accuracy, *F*1*P*, *F*1*N* , G-mean, and AUC evaluation measures.

ADVANTAGES

Collecting the dataset from the Jeeran website with approximately 3000 restaurant reviews. The dataset is cleaned, labeled, formatted, and stemmed.

Oversampling the dataset using four different oversampling techniques to solve the imbalanced problem.

Applying the PSO optimization technique to find the best weights for the dataset features and the best *k* value for each oversampling technique, then applying the SVM classification technique to the oversampled and weighted dataset to find the sentiments of the restaurant reviews.

4.OUTPUTSCREENS

LOGIN PAGE :

SERVICE PROVIDER LOGIN PAGE

SENTIMENT TYPE PREDICTION :

VIEW SENTIMENT PREDICTION

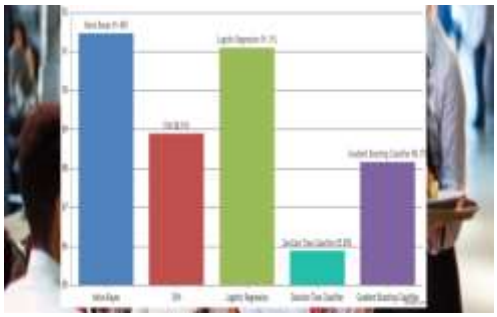
BAR CHART

INTERNATIONAL JOURNAL OF APPLIED SCIENCE ENGINEERING AND MANAGEMENT

ISSN2454-9940

www.ijasem.org

Vol 18, Issue 3, 2024



5.CONCLUSION

Sentiment analysis has witnessed increased interest in the academic field in the last few years. Many people post reviews of different services and products. The analysis of customers' attitudes and feedback is essential for all businesses, including restaurants. Thus, this research proposed a new hybrid evolutionary technique that aims to analyze people's sentiment towards various restaurants across Jordan. The data were collected from a popular social network, namely Jeeran. The proposed approach consisted of collecting more than 3000 restaurant reviews and labeling them using the crowd sourcing technique. Oversampling techniques were then applied to solve the problem of imbalanced data in the dataset. We produced four versions of the collected dataset using different tokenization methods, including 1-Gram, 2-Gram, 3-Gram, and bag-of-words. Further, we implemented a hybrid optimization technique comprising PSO and SVM to find the best weights while also finding the k values of four different oversampling techniques to predict the sentiments of reviews. The study demonstrates that the proposed PSO-SVM approach is effective and outperforms the other approaches in all investigated measures (accuracy, F-measure, g-means, and AUC). In more detail, the PSO-SVM provided better results than the standard SVM, LR, RF, DT, k-NN, and XGBoost in all versions of the datasets .We plan to employ different algorithms on this data in future work. Moreover, other applications can be applied to predict the sentiments of reviews for other products, such as medical and engineering products.

6. REFERENCES

[1] Y. M. Aye and S. S. Aung, ``Senti-lexicon and analysis for restaurant reviews of Myanmar text," *Int. J. Adv. Eng., Manage. Sci.*, vol. 4, no. 5, Jan. 2018, Art. NO 240004.

[2] P. P. Rokade and A. K. D, ``Business intelligence analytics using sentiment Analysis A survey," *Int. J. Electr. Comput. Eng.*, vol. 9, no. 1, p. 613, Feb. 2019.

[3] K. Zahoor, N. Z. Bawany, and S. Hamid, ``Sentiment analysis and classi-

_cation of restaurant reviews using machine learning," in *Proc. 21st Int.*

Arab Conf. Inf. Technol. (ACIT), Nov. 2020, pp. 1_6.

[4] M. Nakayama and Y. Wan, ``The cultural impact on social commerce:

A sentiment analysis on yelp ethnic restaurant reviews," *Inf. Manage.*,

vol. 56, no. 2, pp. 271_279, Mar. 2019.

[5] Q. Gan, B. H. Ferns, Y. Yu, and L. Jin, ``A text mining and multidimensional

analysis of online restaurant reviews," *J. Quality*

Assurance Hospitality Tourism, vol. 18, no. 4, pp. 465_492, Oct. 2017.

[6] R. Murphy. (Dec. 9 2020). *Local Consumer Review Survey*

2020. BrightLocal. Accessed: Nov. 5, 2021. [Online]. Available:

https://www.brightlocal.com/research/local-consumer-review-survey/

[7] R. Feldman, ``Techniques and applications for sentiment analysis," *Com-. ACM*, vol. 56, no. 4, pp. 82_89, 2013.

[8] H. Kang, S. J. Yoo, and D. Han, ``Senti-lexicon and improved Naïve

Bayes algorithms for sentiment analysis of restaurant reviews," *Expert*

Syst. Appl., vol. 39, no. 5, pp. 6000_6010, 2012.

[9] L. Li, L. Yang, and Y. Zeng, ``Improving sentiment classification of restaurant

reviews with attention-based bi-GRU neural network," *Symmetry*,

vol. 13, no. 8, p. 1517, Aug. 2021.

[10] O. Oueslati, A. I. S. Khalil, and H. Ounelli, ``Sentiment analysis for helpful

reviews prediction," *Int. J. Adv. Trends Comput. Sci. Eng.*, vol. 7, no. 3,

pp. 34_40, Jun. 2018.

[11] E. Asani, H. Vahdat-Nejad, and J. Sadri, ``Restaurant recommender system