Journal of **N**onlinear **A**nalysis and **O**ptimization
Vol. 15, Issue. 2 : 2024
ISSN :**1906-9685**

CROSS PLATFORM REPUTATION GENERATION SYSTEM BASED ON ASPECT-BASED SENTIMENT ANALYSIS

[1]T. RAVI KIRAN KUMAR, [2]CH.ABHINAYA ,[3]C. AKHILA,[4]V. MITHIN RAJ,[5]K. RAJU

[1]ASSISTANT PROFESSOR, [2,3,4&5] UG STUDENTS

DEPARTMENT OF CSE, MNR COLLEGE OF ENGG. & TECHNOLOGY, MNR NAGAR, FASALWADIGUDA, SANGA REDDY-502294

ABSTRACT

The active growth of Internet-based applications such as social networks and e-commerce websites leads people to generate a tremendous amount of opinions and reviews about products and services. Thus, it becomes very crucial to automatically process them. Over the last ten years, many systems have been proposed to generate and visualize reputation by mining textual and numerical reviews. However, they have neglected the fact that online reviews could be posted by malicious users that intend to affect the reputation of the target product. Besides, these systems provide an overall reputation value toward the entity and disregard generating reputation scores toward each aspect of the product. Therefore, we developed a system that incorporates spam filtering, review popularity, review posting time, and aspect-based sentiment analysis to generate accurate and reliable reputation values. The proposed model computes numerical reputation values for an entity and its aspects based on opinions collected from various platforms. Our proposed system also offers an advanced visualization tool that displays detailed information about its output. Experiment results conducted on multiple datasets collected from various platforms (Twitter, Facebook, Amazon) show the efficacy of the proposed system compared with state-of-the-art reputation generation systems. Machine learning is an important component of the growing field of data science. Through the use of statistical methods, different type of algorithms is trained to make classifications or predictions, and to uncover key insights in this project. These insights subsequently drive decision making within applications and businesses, ideally impacting key growth metrics. Machine learning algorithms build a model based on this project data, known as training data, in order to make predictions or decisions without being explicitly programmed to do so. Machine learning algorithms are used in a wide variety of datasets, where it is difficult or unfeasible to develop conventional algorithms to perform the needed tasks.

1.INTRODUCTION Having easy access to the web has radically changed the way people interact with brands and products. From physical products to online services, people tend to instantly share their opinions and reviews on various platforms on the Internet. A recent research experiment1 shows that consumers are more willing to share a review when the experience they have had evokes emotions, whether positive or negative. This large volume of consumers' reviews holds insightful information about the quality of the product/service, therefore analyzing them will help consumers make a better judgment toward the targeted item. In the past few years, a new sub- field of natural language processing (NLP) called reputation generation has been well-established as an area of interest. The main focus of reputation generation systems is to produce a numerical value in which an entity is held based on mining customer reviews and their numerical ratings. Over the last decade, many reputation generation systems have been proposed [1]_[8] to generate and visualize reputation of online products and services based on fusing and mining textual and numerical reviews. However, these systems have not taken into consideration (1) extracting and processing reviews from various platforms, (2) filtering reviews written by potential spammers, (3) generating a numerical reputation value toward each aspect of the target product, and, (4) providing an advanced reputation visualization tool for a better decision-making process. Thereby, we designed and built an upgraded reputation generation model that overcomes the shortcomings

https://doi.org/10.36893/JNAO.2024.V15I2.117

of the previous systems in order to compute and visualize the reputation of an entity (product, movie, hotel, restaurant, service) with consistent reliability. The proposed system collects and processes data from both e-commerce and social media platforms. Then, a spam filtering system is applied to eliminate spam reviews and prepare the cleaned output for aspect-based sentiment analysis (ABSA), where aspects of the target entity are extracted from the reviews with their sentiment polarities. Later, the time and popularity features of the reviews are exploited along with the ASBA results to finally generate a reputation value of each aspect of the target entity as well as the overall reputation value using mathematical formulas. The system also proposes an analytical dashboard that displays in-depth information about the reputation of the target entity.In this manner, this study addresses the following research question: with the consideration of review popularity, review time, spam filtering, and ABSA, can the proposed reputation model offer better results in terms of generating and visualizing reputation than state-of-the-art (SOTA) systems?

2.LITERATURE REVIEW

1. **Cross-Platform Reputation Generation System Based on Aspect-Based Sentiment Analysis**

   Achraf Boumhidi, Abdessamad Benlahbib, E. Nfaoui

   Published in IEEE Access 2022

   The active growth of Internet-based applications such as social networks and e-commerce websites leads people to generate a tremendous amount of opinions and reviews about products and services. Thus, it becomes very crucial to automatically process them. Over the last ten years, many systems have been proposed to generate and visualize reputation by mining textual and numerical reviews. However, they have neglected the fact that online reviews could be posted by malicious users that intend to affect the reputation of the target product. Besides, these systems provide an overall reputation value toward the entity and disregard generating reputation scores toward each aspect of the product. Therefore, we developed a system that incorporates spam filtering, review popularity, review posting time, and aspect-based sentiment analysis to generate accurate and reliable reputation values. The proposed model computes numerical reputation values for an entity and its aspects based on opinions collected from various platforms. Our proposed system also offers an advanced visualization tool that displays detailed information about its output. Experiment results conducted on multiple datasets collected from various platforms (Twitter, Facebook, Amazon $\dots$) show the efficacy of the proposed system compared with state-of-the-art reputation generation systems.

2. **LCF: A Local Context Focus Mechanism for Aspect-Based Sentiment Classification**

   Biqing Zeng, Heng Yang, +2 authors Xuli Han

   Published in Applied Sciences 17 August 2019

   Aspect-based sentiment classification (ABSC) aims to predict sentiment polarities of different aspects within sentences or documents. Many previous studies have been conducted to solve this problem, but previous works fail to notice the correlation between the aspect's sentiment polarity and the local context. In this paper, a Local Context Focus (LCF) mechanism is proposed for aspect-based sentiment classification based on Multi-head Self-Attention (MHSA). This mechanism is called LCF design, and utilizes the Context features Dynamic Mask (CDM) and Context Features Dynamic Weighted (CDW) layers to pay more attention to the local context words. Moreover, a BERT-shared layer is adopted to LCF design to capture internal long-term dependencies of local context and global context. Experiments are conducted on three common ABSC datasets: the laptop and restaurant datasets of SemEval-2014 and the ACL twitter dataset. Experimental results demonstrate that the LCF baseline model achieves considerable performance. In addition, we conduct ablation experiments to prove the significance and effectiveness of LCF design. Especially, by incorporating with BERT-shared layer, the LCF-BERT model refreshes state-of-the-art performance on all three benchmark datasets.

3. **Enhancing Aspect-Based Sentiment Analysis With Capsule Network**
   Jindian Su, Shanshan Yu, Da Luo
   Published in IEEE Access 2020
   Existing feature-based neural approaches for aspect-based sentiment analysis (ABSA) try to improve their performance with pre-trained word embeddings and by modeling the relations between the text sequence and the aspect (or category), thus heavily depending on the quality of word embeddings and task-specific architectures. Although the recently pre-trained language models, i.e., BERT and XLNet, have achieved state-of-the-art performance in a variety of natural language processing (NLP) tasks, they still subject to the aspect-specific, local feature-aware and task-agnostic challenges. To address these challenges, this paper proposes a XLNet and capsule network based model XLNetCN for ABSA. XLNetCN firstly constructs auxiliary sentence to model the sequence-aspect relation and generate global aspect-specific representations, which enables to enhance aspect-awareness and ensure the full pre-training of XLNet for improving task-agnostic capability. After that, XLNetCN also employs a capsule network with the dynamic routing algorithm to extract the local and spatial hierarchical relations of the text sequence, and yield its local feature representations, which are then merged with the global aspect-related representations for downstream classification via a softmax classifier. Experimental results show that XLNetCN outperforms significantly than the classical BERT, XLNet and traditional feature-based approaches on the two benchmark datasets of SemEval 2014, Laptop and Restaurant, and achieves new state-of-the-art results.
4. **Out of Context: A New Clue for Context Modeling of Aspect-based Sentiment Analysis**
   Bowen Xing, I. Tsang
   Published in Journal of Artificial… 21 June 2021
   Aspect-based sentiment analysis (ABSA) aims to predict the sentiment expressed in a review with respect to a given aspect. The core of ABSA is to model the interaction between the context and given aspect to extract aspect-related information. In prior work, attention mechanisms and dependency graph networks are commonly adopted to capture the relations between the context and given aspect. And the weighted sum of context hidden states is used as the final representation fed to the classifier. However, the information related to the given aspect may be already discarded and adverse information may be retained in the context modeling processes of existing models. Such a problem cannot be solved by subsequent modules due to two reasons. First, their operations are conducted on the encoder-generated context hidden states, whose value cannot be changed after the encoder. Second, existing encoders only consider the context while not the given aspect. To address this problem, we argue the given aspect should be considered as a new clue out of context in the context modeling process. As for solutions, we design three streams of aspect-aware context encoders: an aspect-aware LSTM, an aspect-aware GCN, and three aspect-aware BERTs. They are dedicated to generating aspect-aware hidden states which are tailored for the ABSA task. In these aspect-aware context encoders, the semantics of the given aspect is used to regulate the information flow. Consequently, the aspect-related information can be retained and aspect-irrelevant information can be excluded in the generated hidden states. We conduct extensive experiments on several benchmark datasets with empirical analysis, demonstrating the efficacies and advantages of our proposed aspect-aware context encoders..
5. **Quantum-Inspired Complex-Valued Language Models for Aspect-Based Sentiment Classification**
   Qin Zhao, Chenguang Hou, Ruifeng Xu
   Published in Entropy 29 April 2022
   Aiming at classifying the polarities over aspects, aspect-based sentiment analysis (ABSA) is a fine-grained task of sentiment analysis. The vector representations of current models are generally constrained to real values. Based on mathematical formulations of quantum theory, quantum

language models have drawn increasing attention. Words in such models can be projected as physical particles in quantum systems, and naturally represented by representation-rich complex-valued vectors in a Hilbert Space, rather than real-valued ones. In this paper, the Hilbert Space representation for ABSA models is investigated and the complexification of three strong real-valued baselines are constructed. Experimental results demonstrate the effectiveness of complexification and the outperformance of our complex-valued models, illustrating that the complex-valued embedding can carry additional information beyond the real embedding. Especially, a complex-valued RoBERTa model outperforms or approaches the previous state-of-the-art on three standard benchmarking datasets.

3.SYSTEM ANALYSIS

3.1. EXISTING SYSTEM

Poria et al. presented the first deep learning approach for the AE task in opinion mining. The authors employed a 7-layer deep convolutional neural network to tag each word in the textual opinions as either aspect or non-aspect word. The authors also proposed a set of heuristic linguistic patterns and integrated them with the deep learning classifier which significantly improves the accuracy compared with the previous SOTA methods. In [19], the authors proposed an attention-based long short-term memory (LSTM) [20] for aspect-level sentiment classification. The idea is to learn aspect embeddings and let aspects participate in computing attention weights. The proposed model can focus on different parts of a sentence when different aspects are given so that they are more competitive for aspect-level classification. The proposed model achieved better results compared with the standard LSTM on the SemEval 2014 Task 4 dataset [21]. In [22], Wei and Toi improved the deficiencies of the previous LSTM approaches by proposing convolutional neural networks [23] and gating mechanisms (GCAE) based model, which has been proved to be more accurate and efficient. The novel Gated Tanh-ReLU Units can selectively output the sentiment features according to the provided aspect or entity. The architecture of the proposed model is much simpler than the attention layer used in the previously existing models. The experiments on SemEval datasets show a performance improvement compared with the LSTM based models. The authors in [24] proposed an interactive multi-task learning network (IMN) capable of jointly learning multiple related tasks simultaneously at both the token-level as well as the document-level. The IMN introduces a message passing mechanism that allows informative interactions between tasks, enabling the correlation to be better exploited. Experiments on three benchmark datasets, taken from SemEval2014 and SemEval 2015 [25] show that IMN outperforms other baselines by large margins. Since most existing methods ignore the position information of the aspect when encoding the sentence, authors in [26] proposed a hierarchical attention-based position-aware network (HAPN), which includes position embeddings to learn the position-aware representations of sentences to generate the target-specific representations of contextual words. HAPN achieved the SOTA performance on SemEval 2014 dataset compared with the previous methods. Xu et al. [27] presented a review reading comprehension (RRC) task where they adopted BERT [28] as a base model, and proposed a joint post-training and fine-tuning approach for ATE, APC. Experimental results show that the proposed post-training approach is very effective. Later in [29], the authors proposed a novel architecture called BERT Adversarial Training (BAT) to employ adversarial training for AE and APC by generating artificial data which is carried out in the embedding space. The proposed model outperforms the standard BERT as well as the in-domain post-trained BERT in both AE and APC tasks. In [30], the authors exploit domain-specific BERT language model fine tuning in addition to supervised task-specific fine tuning to produce a new SOTA performance on the SemEval 2014 Task 4 restaurants dataset. The authors also showed that cross-domain adapted BERT model performs better than strong baseline models such as XLNet-base [31] and vanilla BERT-base. In [32], the authors compared the induced trees from pre-trained models and the dependency parsing trees on various popular models for the ABSA task. They found that the induced tree from fine tuned RoBERTa [33] (FT-RoBERTa) outperforms the parser-provided tree. The experiments show that the RoBERTa-based model can outperform or approximate the previous SOTA performances on six datasets across four languages including SemEval 2014 task 4. Recently, authors in [34] proposed a multi-task learning model named LCF-ATEPC for ABSA based on the multi-head self-attention and the local context focus (LCF) [35] mechanisms. The proposed model is multilingual and applicable to the classic English review SA task, such as the SemEval-2014 task4. The proposed model can automatically extract aspects and determine their sentiment polarities. Since the LCF-ATEPC model currently achieves SOTA performance on AE and APC tasks,2 it was selected to be employed in this paper.

Disadvantages:

- An existing system not implemented Aspect term extractor which performs the basic token-level classification for each token, which means that each token will be given a label, and a classification is performed to predict the aspects in the sentence.
- An existing system is not implemented Local Context Focus in which Local context is a new technique that can be adapted to most ne-grained NLP tasks.

3.2. PROPOSED SYSTEM

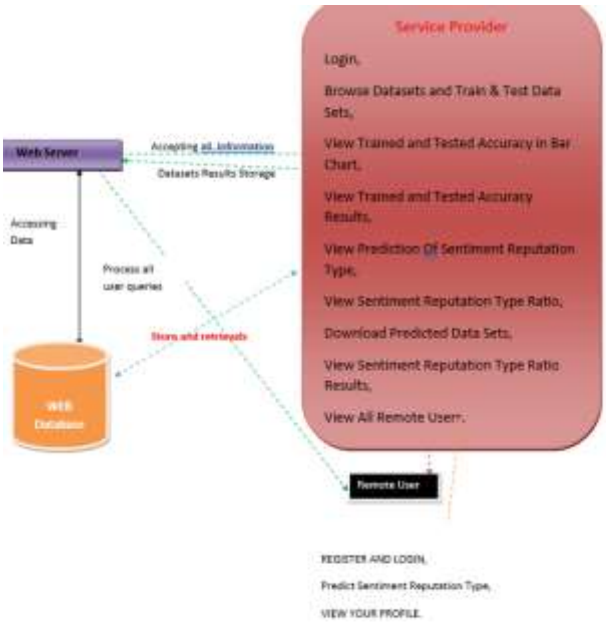
This system aims at generating a reputation value toward online entities (movies, hotels, restaurants, services, etc.) and computing a satisfaction score toward each aspect of the target entity by processing textual and numerical data collected from multiple platforms. Proposed system presents its architecture. First, we start by collecting users' reviews from different platforms such as Twitter, Amazon, YouTube, etc. Next, an automatic spammers filtering system is employed to detect and eliminate unwanted spam reviews. Then, we apply a SOTA ABSA model to users' textual reviews in order to compute a score based on the sentiment orientation of the extracted aspects from those reviews. Further, we calculate a popularity score and a time score based on statistical features extracted with the textual reviews. Finally, we compute a reputation value based on the previously calculated scores, and we propose a new user-friendly visualization interface that displays in-depth details about the reputation of the target entity. One of the important features of the proposed system is the ability to collect and process data from various platforms. Previous reputation generation systems gather necessary data from either e-commerce websites such as Amazon, TripAdvisor, or social media platforms such as Twitter and Facebook. In this work, we decided to normalize the features of all platforms in order to create a single merged dataset by classifying the platforms on the Internet into two types: the first type provides the accessibility of extracting the textual review with the number of likes received for that review such as Amazon, YouTube, etc. The second type provides the accessibility of extracting the textual review with the number of likes received for that review along with the number of times the review was shared among the network such as Twitter, Facebook, etc.

Advantages:

- Multi-Head Self-Attention (MHSA): The multi-head attention mechanism helps the model to learn the words' relevant information in different presentation subspaces. MHSA is based on multiple scale-dot attention that can be used to extract deep semantic features in the context.
- Aspect Polarity Classifier: To perform the sentiment polarity classification, the LCF-ATEPC model combines the local context features and the global context features. Then, the aspect polarity classifier performs a head-pooling on the learned concatenated context features from the feature interactive learning layer.

4.IMPLEMENTATION

4.1.SYSTEMARCHITECTURE



4.2. MODULES

Service Provider

- In this module, the Service Provider has to login by using valid user name and password. After login successful he can do some operations such as

- Login, Browse Datasets and Train & Test Data Sets, View Trained and Tested Accuracy in Bar Chart, View Trained and Tested Accuracy Results,
- View Prediction Of Sentiment Reputation Type, View Sentiment Reputation Type Ratio, Down load Predicted Data Sets, View Sentiment Reputation Type Ratio Results, View All Remote Users.

View and Authorize Users

In this module, the admin can view the list of users who all registered. In this, the admin can view the user's details such as, user name, email, address and admin authorizes the users.

Remote User

In this module, there are n numbers of users are present. User should register before doing any operations. Once user registers, their details will be stored to the database. After registration successful, he has to login by using authorized user name and password. Once Login is successful user will do some operations like REGISTER AND LOGIN, Predict Sentiment Reputation Type, VIEW YOUR PROFILE.

5.ALGORITHAMS

5.1.K-NEAREST NEIGHBORS (KNN)

- Simple, but a very powerful classification algorithm
- Classifies based on a similarity measure
- Non-parametric
- Lazy learning
- Does not “learn” until the test example is given
- Whenever we have a new data to classify, we find its K-nearest neighbors from the training data

Example

- Training dataset consists of k-closest examples in feature space
- Feature space means, space with categorization variables (non-metric variables)
- Learning based on instances, and thus also works lazily because instance close to the input vector for test or prediction may take time to occur in the training dataset

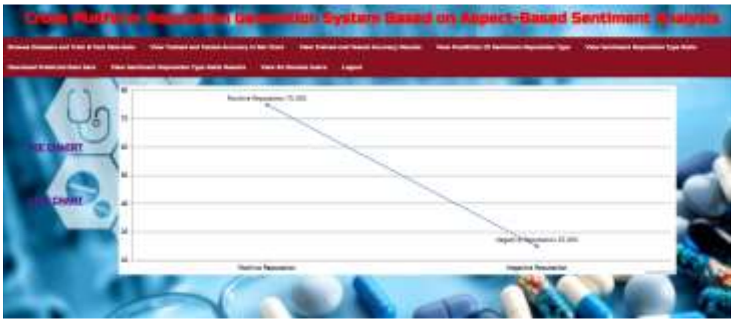
5.2. NAÏVE BAYES

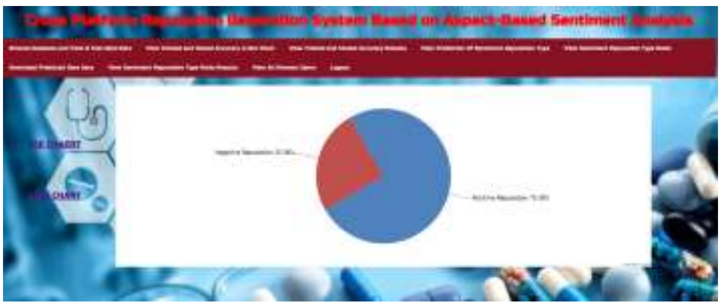
The naive bayes approach is a supervised learning method which is based on a simplistic hypothesis: it assumes that the presence (or absence) of a particular feature of a class is unrelated to the presence (or absence) of any other feature .Yet, despite this, it appears robust and efficient. Its performance is comparable to other supervised learning techniques. Various reasons have been advanced in the literature. In this tutorial, we highlight an explanation based on the representation bias. The naive bayes classifier is a linear classifier, as well as linear discriminant analysis, logistic regression or linear SVM (support vector machine). The difference lies on the method of estimating the parameters of the classifier (the learning bias). While the Naive Bayes classifier is widely used in the research world, it is not widespread among practitioners which want to obtain usable results. On the one hand, the researchers found especially it is very easy to program and implement it, its parameters are easy to estimate, learning is very fast even on very large databases, its accuracy is reasonably good in comparison to the other approaches. On the other hand, the final users do not obtain a model easy to interpret and deploy, they does not understand the interest of such a technique. Thus, we introduce in a new presentation of the results of the learning process. The classifier is easier to understand, and its deployment is also made easier. In the first part of this tutorial, we present some theoretical aspects of the naive bayes classifier. Then, we implement the approach on a dataset with Tanagra. We compare the obtained results (the parameters of the model) to those obtained with other linear approaches such as the logistic regression, the linear discriminant analysis and the linear SVM. We note that the results are highly consistent. This largely explains the good performance of the method in comparison to others. In the second part, we use various tools on the same dataset (Weka 3.6.0, R 2.9.2, Knime 2.1.1, Orange 2.0b and RapidMiner 4.6.0). We try above all to understand the obtained results.

5.3. RANDOM FOREST

Random forests or random decision forests are an ensemble learning method for classification, regression and other tasks that operates by constructing a multitude of decision trees at training time. For classification tasks, the output of the random forest is the class selected by most trees. For regression tasks, the mean or average prediction of the individual trees is returned. Random decision forests correct for decision trees' habit of overfitting to their training set. Random forests generally outperform decision trees, but their accuracy is lower than gradient boosted trees. However, data characteristics can affect their performance. The first algorithm for random decision forests was created in 1995 by Tin Kam Ho[1] using the random subspace method, which, in Ho's formulation, is a way to implement the "stochastic discrimination" approach to classification proposed by Eugene Kleinberg. An extension of the algorithm was developed by Leo Breiman and Adele Cutler, who registered "Random Forests" as a trademark in 2006 (as of 2019, owned by Minitab, Inc.).The extension combines Breiman's "bagging" idea and random selection of features, introduced first by Ho[1] and later independently by Amit and Geman[13] in order to construct a collection of decision trees with controlled variance. Random forests are frequently used as "blackbox" models in businesses, as they generate reasonable predictions across a wide range of data while requiring little configuration.

6.RESULT





CONCLUSION

In this paper, we proposed a reputation system capable of generating numerical reputation values for a specific item (product, movie, service, hotel, etc.) and its aspects based on opinions and reviews expressed online. The contribution of this work revolves around four components that were not exploited in previous systems. The first one is cross-platform compatibility, where the proposed system can collect and process opinions from different platforms (Face book, Amazon, Twitter, Trip Advisor, etc.) as well as managing and standardizing those platforms' features. The second one is opinion spam filtering, where the spam opinions are detected and eliminated based on spammers' behavior features, keeping only authentic opinions. The third one is employing a SOTA aspect-based sentiment-analysis model named LCF-ATEPC in order to extract and analyze the aspects within the textual opinions. Finally, we incorporated the previous results with a calculated review time score and review popularity score using mathematical formulas to obtain a reputation value for the targeted entity as well as the reputation values of the entities' aspects. In addition a holistic reputation visualization is provided within the system that displays the detailed output results of the reputation generation process. To assess the effectiveness of our reputation system, we invited 32 participants and 3 experts to choose the best performing system out of four SOTA reputation systems by giving numerical satisfaction scores to each system. Our reputation system achieved the highest average satisfaction scores from both users and experts. In the future, we propose to investigate the effectiveness of our proposed system by attempting to generate more than the numerical reputation values, such as extending the system to automatically generate a textual summary of the benefits and drawbacks of the targeted entity. Also, we intend to extend this system to be used in multilingual content.

REFERENCES

[1]Abdel-Hafez, Y. Xu, and D. Tjondronegoro, ``Product reputation model: An opinion mining based approach,'' in *Proc. 1st Int. Workshop Sentiment Discovery Affect. Data*, vol. 917, London, U.K., Jun. 2013, pp. 16_27. [Online]. Available: https://eprints.qut.edu.au/58118/

[2] U. Farooq, A. Nongaillard, Y. Ouzrout, and M. A. Qadir, ``A featurebased reputation model for product evaluation,'' *Int. J. Inf. Tech- nol. Decis. Making*, vol. 15, no. 6, pp. 1521_1553, Nov. 2016, doi: 10.1142/S0219622016500358.

[3] Z. Yan, X. Jing, and W. Pedrycz, ``Fusing and mining opinions for reputation generation,'' *Inf. Fusion*, vol. 36, pp. 172_184, Jul. 2017, doi: 10.1016/j.inffus.2016.11.011.

[4] A. Benlahbib and E. H. Nfaoui, ``A hybrid approach for generating reputation based on opinions fusion and sentiment analysis,'' *J. Organizational Comput. Electron. Commerce*, vol. 30, no. 1, pp. 9_27, 2020, doi: 10.1080/10919392.2019.1654350.

[5] E. I. Elmurngi and A. Gherbi, ``Building sentiment analysis model and compute reputation scores in E-commerce environment using machine learning techniques,'' *Int. J. Organizational Collective Intell.*, vol. 10, no. 1, pp. 32_62, Jan. 2020.

[6] A. Benlahbib and E. H. Nfaoui, ``Aggregating customer review attributes for online reputation generation,'' *IEEE Access*, vol. 8, pp. 96550_96564, 2020, doi: 10.1109/ACCESS.2020.2996805.

[7] A. Gupta, S. Priyani, and R. Balakrishnan, ``Customized reputation generation of entities using sentiment analysis,'' *World J. Eng.*, vol. 18, no. 4, pp. 596_605, Jul. 2021, doi: 10.1108/WJE-09-2020-0470.

[8] A. Boumhidi and E. Nfaoui, ``Leveraging lexicon-based and sentiment analysis techniques for online reputation generation,'' *Int. J. Intell. Eng. Syst.*, vol. 14, no. 6, pp. 274_289, 2021, doi: 10.22266/ijies2021.1231.25.

[9] V. M. Pradhan, J. Vala, and P. Balani, ``A survey on sentiment analysis algorithms for opinion mining,'' *Int. J. Comput. Appl.*, vol. 133, no. 9, pp. 7_11, Jan. 2016.

[10] A. Tripathy, A. Anand, and S. K. Rath, ``Document-level sentiment classification using hybrid machine learning approach," *Knowl. Inf. Syst.*, vol. 53, no. 3, pp. 805_831, 2017.