Communications on Applied Nonlinear Analysis
ISSN: 1074-133X
Vol 32 No. 7s (2025)

Text-Based Airline Sentiments Analysis Using Deep Learning Ensemble Optimization

Abha Kiran Rajpoot[1], Mayank Kumar Goyal[2], Mandeep Kaur[3], Anubhav Kumar[4], Kimmi Verma[5], Setu Garg[6]

[1,4]School of Computer Science & Engineering, Galgotias University, Greater Noida

[2]Department of Computer Science & Engineering, SSET, Sharda University, Greater Noida

[3]School of Computer Science Engineering and Technology, Bennett University, Greater Noida

[5]Department of Computer Science & Engineering, ABES Engineering College, Ghaziabad

[6]Department of ECE, ITS Engineering College, Greater Noida

Article History:

Received: *24-10-2024*

Revised: *07-11-2024*

Accepted: *15-12-2024*

Abstract:

Text Sentiment Analysis is one of the foundational pillars of sentiment analysis. There hasn't been any research done yet that makes the use of the combined power of modern models and contextual information present in texts, nor has it been studied how contextual information may be controlled in any effective manner. It has been demonstrated that the model proposed in this paper can be used to accurately and efficiently classify sentiments for a given topic. The system is able to capture all of the content contained within the text by combining rule-based methods and deep learning models. In a study undertaken by our group, it has been found that building an embedded representation and attention mechanism is effective in the processing of valence shifting cases. Our proposed mechanism also combines custom-designed rules along with domain-specific sentiment dictionaries. The proposed method has been tested using three datasets and has shown superior performance to other methods, regardless of its higher cost as compared to other methods. PPD (Probability Proportion Difference) is used in this study to pick characteristics that regard each term to be part of a class list. It's a quick and easy way to remove irrelevant words from a feature vector. Categorical Probability Proportion Difference (CPPD) and Probability Proportion Difference (PPD) are used to choose features in a suggested method (CPD). In order to identify features that effectively distinguish between classes and deliver the necessary outcomes, CPPD features selection methods can be used the comparison of the proposed methods' performance with that of the CPD method and the Information Gain (IG) approach shows that all three methods are comparable in their ability to detect sentiment. Two standard datasets (each containing a number of movie reviews and a number of book reviews) have been used as data sources for testing suggested feature selection methods. A study conducted by the authors of this article suggests that the CPPD feature selection method proposed in this article performs better for sentiment classification than other methods. The proposed model is fast and uses the Virgin America Airlines reviews dataset for training and testing. The Monkey learn model achieved 75% precision, 91% recall, and 89% of accuracy, showing the model capabilities in real time Text sentiment analysis.

Keywords: TF-IDF, Textual Classification, Web 2.0, Deep Learning, Machine learning, Sentiment Analysis

1. INTRODUCTION

People are increasingly expressing their thoughts, opinions, experiences, attitudes, feelings, and emotions on the web, thanks to rapid technological advances. Thus, it has become increasingly important to process, organize, and analyze web content in order to find out users' opinions [1]. Natural language processing has incorporated sentiment analysis since 2002 as one of the most intriguing areas. A document's polarity (positive, negative, or neutral) and strength can be determined automatically using sentiment analysis by classifying it from positive to negative to neutral based on its data or meaning [2-4]. To examine sentiment analysis, we have looked at it from two perspectives - by applying machine learning to build classification models based on labeled data; and by analyzing the expression and meaning of words and phrases in a text to calculate overall polarization as both users and companies are interested in understanding the opinions of users, such as reviews for electronic products like laptops, cars, movies, etc [3-5].

We propose a feature selection strategy as a foundation for increasing the performance of machine learning systems [6]. A binary weighting strategy is used to represent the review papers. The reduced and prominent feature set is then subjected to machine learning methods. The process of sentiment analysis is shown in Fig. 1.

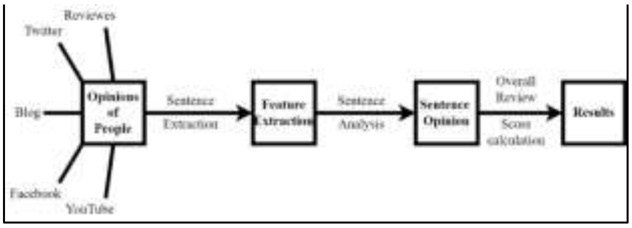


Figure 1: Feature extraction of text sentiment analysis process

This paper offers the following main contributions in the sentimental analysis include:

An ensemble model for sentiment classification that incorporates various sub-models of sentiment acquired from a range of sub-datasets has been suggested. In the field of linguistics, contextual factors are fully exploited. Deep learning methods enhance the adaptability of feature extraction tasks using rule-based methods that capture contextual valence shifting. As well as embedding representation and attention mechanisms, we use inductive sentiment dictionaries tailored to each domain. When evaluated on several datasets, our suggested model outperforms standard state-of-the-art approaches. PPD and CPPD, two new features selection methods, are presented to classify sentiment.

2. LITERATURE REVIEW

Pang et al. 2002 employed the Unigram and Bigram elements of movie reviews to analyse sentiment analysis of movie reviews. Then they applied SVM, NB, and a technique known as maximum entropy (ME). There was a consensus that SVMs outperformed all other classifiers amongst the authors. Further, the authors concluded that binary weighting schemes offer more satisfactory results than doing term frequencies (TFs) as a way to describe sentiment in the text [8]. A study conducted by SVM and IG researchers showed that the SVM algorithm provides the best results when comparing machine learning algorithms, so that is worth your attention [9].

Many researchers have investigated feature selection in the context of improving machine-learning methods to classify emotions through reducing the dimensions of the feature vectors [1-10]. Entropy weighted genetic algorithms (EWGA) have been developed to improve sentiment classification accuracy by combining IG and genetic algorithms. Their aggregated classifier was constructed based on multiple classifiers constructed for different feature vectors using different weightings on specific sentiment features [11]. Taking into account different scoring methods, the authors examined three methods to select features that could indicate sentiment. One of them used Categorical Proportional Differences (CPD) while another study used Sentiment Orientation (SO) values. In 2009, it was proposed the concept of Fisher's discriminant ratios for the classification of text review sentiment against other studies [12-15].

2.1 Contextual Valance Shifter

The subjectivity of texts is usually found within the descriptions of events as well as the objective descriptions; this is called the attitude of the writer or participant. The words are chosen and the manner in which they are arranged indicates an emotional attitude [13].

Defective verbs become semantically meaningful when used in combination with other verbs and modifiers that have semantic associations [16]. The sentimental value of the sentence can be negated in cases where the defective word in the sentence is accompanied by a sentimental phrase or word.

The following statement would apply if the person who commented did not know anything about the car or its characteristics [11]. A subject may say, for example, "I see it differently, the car isn't bad at all." Inconsistency (incompatibility): Sentence and paragraph scores alone cannot accurately describe the emotional nature of a text when the context is present in the text, and emotional weight is largely influenced by the context. For example, [12] Neither he nor his math prodigy was able to solve that problem. His work is horrible. Studying kept him up until 2 am every day. [13]

2.2 Attention Mechanism

There has been considerable interest in deep learning methods, such as Word2Vec [17], Glove [18], which use automatic feature representation learning methods to learn representations of features. Attention enables machine translation to overcome this problem by scanning the source sentence and creating an appropriate word based on the most recent work shown within the context as well as the most recent work on which the machine translation is operating [20-23].

One of the main research topics in machine learning in recent years, deep learning, has resulted in scientific advances that have revolutionized the usage of artificial intelligence in people's daily lives and totally reshaped technology [25]. By employing deep learning techniques in fields such as speech recognition and medical imaging, Artificial Intelligence systems have been able to overcome many limitations of the past [26]. With the help of LSTM models as well as featured representations and word embedding’s, a hierarchical attention mechanism was used for analyzing the feature set. In comparison to other models that can achieve a higher score, the approach proposed here can achieve a higher score [27].

2.3 Hybrid Approach

Despite recent studies showing that semantic analysis and machine learning approaches are capable of working, both have advantages and limitations in different fields. Based on the use of a mutual information dictionary, an algorithm was developed that detects sentiment polarization based on the mutual information. Researchers studied documents that were not using SentiWordNet [28], accuracyon average than SentiWordNet.

A bonding model that integrates models of varying capabilities can be stronger than one built using only one model. Similarly, the idea of an assemblage of people also follows Condorcet's theory of voting in which the voting probability for each individual is $p > 1/2$ (i.e., every individual vote according to their preferences), so more voters increase majority accuracy [33]. When there are more voters, the probability of securing a majority rises. Combining prediction models can be achieved in three ways like bagging, boosting and stacking.

Fig. 2 demonstrates how we construct our classification model by employing a stacking Strategy. Ensembles made up of more than one learner can be more effective than individual learners. For example, bioinformatics is one sector where assembly learning can be seen in Table 3. For many classification issues, ensemble learning has been shown to work well.

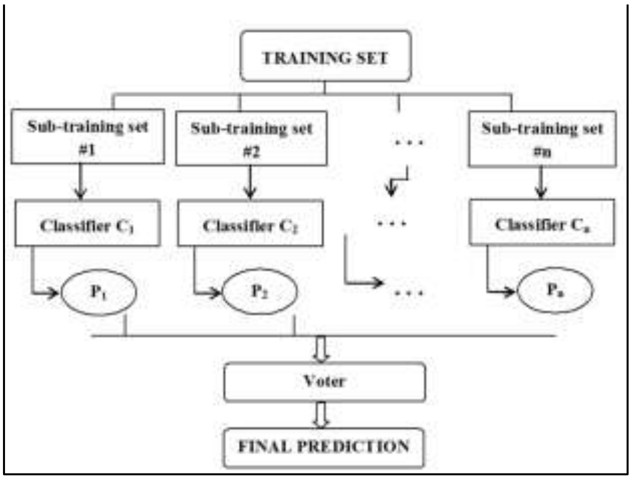


Figure 2: Stacking process used in Ensemble Learning

3. PROPOSED METHODOLOGY

3.1 Dataset

The proposed feature selection approaches are evaluated using a publicly available dataset of movie review https://catalog.data.gov/dataset/nist-campus-photovoltaic-pv-arrays-and-weather-station-data-sets. Each of the 2000 reviews has been tagged with one positive and one negative. There were 1000 positive reviews and 1000 negative reviews among the 1000 books that were reviewed. To prepare documents for pre-sale, it is necessary to follow the steps outlined below:

(i) A negation word's inverted meaning is caused by the addition of "NOT_" to words following it (no, not, isn't, can't).

(ii) The list of features is trimmed when there are few documents containing the term in question. The feature vector that is generated following preprocessing is then used as the basis for classifying the data. All experiments using the Weka tool (WEKA) are configured by default to use Linear SVM and Naive Bayes as the machine learning algorithms. For many classification issues, ensemble learning has been shown to work well. Also the three datasets with positive and negative reviews are shown in Table 1.

Table 1. The three datasets are described in detail

Corpus	Training data		Testing data	
	Positive reviews	**Negative reviews**	**Positive reviews**	**Negative reviews**
HOTEL- reviews	928	928	928	928
UIT-VSFC	2632	2632	2632	2632
FOODY- reviews	11,000	11,000	11,000	11,000

The models are trained by using a variety of base-learners using a method known as ensemble learning. To classify each piece of text, we extract individual features, such as negative and positive contrasts, inconsistencies, and shifts, by extracting and synthesizing attributes [26]. There are a number of ways to represent text in the TF-IDF, and subsequently, Meta-learners use data gathered from the base-learners to make predictions about whether a given outcome would be favorable or negative. Ensemble learning takes place over several stages, as illustrated in Figure 3.

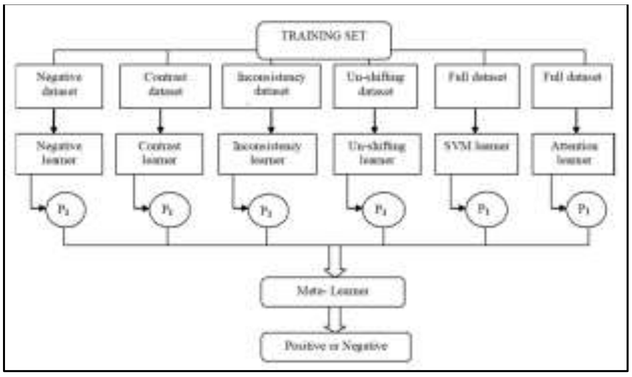


Figure 3: The architecture of the Proposed Methodology

3.2 Probability Proportion Difference (PPD)

According to the Probability Proportion Difference (PPD), terms belong to a particular category within a given number of degrees of belonging.

Negative or positive, this signifies the importance of a term (positive or negative) when it comes to identifying new categories. As long as a term has a high degree of overlap with both categories, it is useless for classifying people. [27]. According to statistical analysis, a term will be more likely to belong to a group depending on how many documents appear in which it appears and how many unique terms appear in that group [28].

It was assessed the degree to which categorical proportional differences (CPDs) enabled a term to discriminate between classes. Values from CPD have been used for feature selection. A CPD value can be calculated for a feature using equation 1 [28].

$cpd = |posD - negD|/(posD + negD)$ $cpd = |posD - negD|/(posD + negD)$ ------------ (1)

When using the posD and negD numbers, you can see how many positive or negative papers each term appears in. A CPD value is a number between 0 and 1. Positive or negative terms are important for sentiment classification if they appear predominantly in positive or negative class, and if they occur equally in both positive and negative class, they are not useful for sentiment classification [6].

3.3 Categorical Probability Proportion Difference (CPPD)

Methods for blending CPD and PPD feature selection, the Category-based Proportional Difference (CPPD), combine both methods' advantages and minimize their disadvantages. CPD is an extremely beneficial method of measuring a prominent feature since it provides the ability to measure the degree of class differentiation, which is crucial to the measurement of that feature. CPD research method, however, is favorable because it can identify uncommon yet unimportant terms which are not selected in PPD analysis, allowing the researcher to include these terms that are rarely found in documents. According to CPD's methods, terms that are very frequently occurring in documents but do not hold great significance can be excluded. These terms might be displayed on the PPD based on the method used by PPD [31].

4 Results

This section is divided into three parts 1. Pre-processing, 2. Analyzing the system (Feature extraction) 3. Performance measure (Feature Classification)

4.1 Pre-Processing

As a dataset of this size, UIT-VSFC has lots of positive comments that bear the test of time. Deep learning models outperform classic machine learning approaches such as LSTM, SVM, and BiLSTM, as well as attention mechanisms. Ems. (WLLR-6C-ATT-6C-ATT-VNSD) and (6C-ATT-6C-ATT-WLLR) achieved excellent results when training on sufficiently large data. It should be noted that, when compared with the best model in the literature so far, the proposed model is 1.65% better than CEM(6C-LSTM-WLLR).

4.2 Analyzing the System (Feature Extraction)

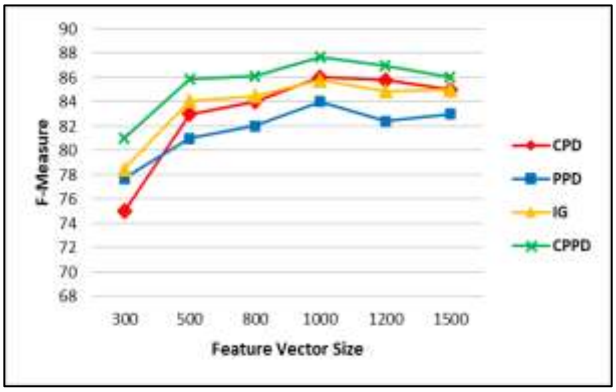
Ensemble learning systems that use CEM(6C-ATT-WLLR) rather than CEM(6C-LSTM-WLLR) frequently get superior outcomes to those that use standard techniques. This indicates that consideration of attention (attention-based) does not provide superior sentiment classification to linear SVM.

4.3 Performance Measures (Feature Classification)

This measure, which is frequently used, combines both precision and recall, both of which reveal a lot about a thing. Precision is defined in class C as the difference between the total number of correctly classified documents and the total number of classified documents (sum of True Positives (TP) and False Positives (FP)).

In CPD feature selection methods, features with very low document frequencies, which do not play a significant role in classification, can result in extremely high CPD values. Due to their low PPD values, these terms were excluded using the PPD Features Selection Method. The terms "poor" and "despicable" are also highly disproportionally represented in the data set that contains movie reviews. Compared to the CPD method, the PPD method includes these, while the CPD method eliminates them. Some terms with high PPD value, nevertheless, are not significant, even though they have strong DF value.

This paper presents a new feature selection technique based on a combination of PPD and CPD that takes into account when selecting features the class distinguishability of terms as well as their relevance based on probability and the importance of documents. In Table 5 we have shown values of F measures, Accuracy and Precision for movie and book reviews across different feature sizes, using various Feature Selection (FS) methods as a training set for SVM classifiers.



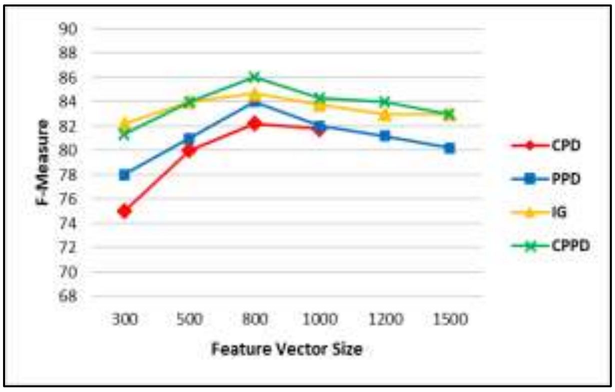
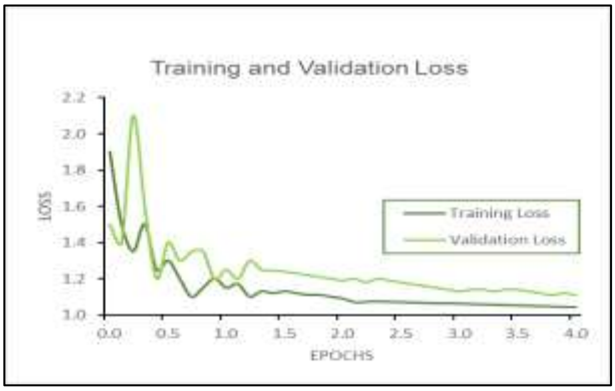


Figure 4 (a): The F-measure for Hotel Reviews. **Figure 4(b):** The F-measure for movie

The first graph 5(a) shows the Training and Validation Loss with regards to epochs. The second graph 5(b) shows the Actual Validation and Training Accuracy with five different thresholds.



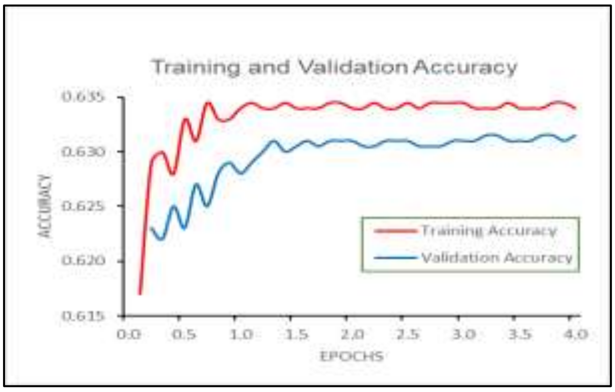


Figure 5(a): Training and validation loss **Figure 5(b):** Actual validation and training accuracy

5. Conclusion

Comparing the PPD methodology with different feature selection techniques, CPPD methodology produces superior results. Additionally, book review and movie review datasets revealed the most meaningful F-measures when 1000 and 800 features are measured, i.e. approximately 10-15% of the total unigram features. This proposed method utilizes the concept of CPPD feature selection in order to eliminate irrelevant features in addition to being highly computationally efficient. A categorization system is created by selecting features that can be used in the system. Evaluation of the proposed schemes is done using two datasets that have been standardized. The proposed strategy clearly increased the experiment's categorization accuracy.

REFERENCES

[1] Abid, F., Li, C., &Alam, M. (2020). Multi-source social media data sentiment analysis using bidirectional recurrent convolutional neural networks. Computer Communications, 157, 102-115.https://doi.org/10.1016/j.comcom.2020.04.002

[2] Ahmad, M., Aftab, S., Ali, I., & Hameed, N. J. I. J. M. S. E. (2017). Hybrid tools and techniques for sentiment analysis: a review. Int. J. Multidiscip. Sci. Eng, 8(3), 29-33.

[3] Ahmad, M., Aftab, S., Bashir, M. S., & Hameed, N. (2018). Sentiment analysis using SVM: a systematic literature review. International Journal of Advanced Computer Science and Applications, 9(2).

[4] Ibrahim, M. A., & Salim, N. (2013). Opinion analysis for twitter and arabic tweets: A systematic literature review. Journal of Theoretical & Applied Information Technology, 56(3).

[5] Alam, M., Abid, F., Guangpei, C., &Yunrong, L. V. (2020). Social media sentiment analysis through parallel dilated convolutional neural network for smart city applications. Computer Communications, 154, 129-137.https://doi.org/10.1016/j.comcom.2020.02.044

[6] Alarifi, A., Tolba, A., Al-Makhadmeh, Z., & Said, W. (2020). A big data approach to sentiment analysis using greedy feature selection with cat swarm optimization-based long short-term memory neural networks. The Journal of Supercomputing, 76, 4414-4429.https://doi.org/10.1007/s11227-018-2398-2

[7] Alexandridis, G., Michalakis, K., Aliprantis, J., Polydoras, P., Tsantilas, P., &Caridakis, G. (2020, May). A deep learning approach to aspect-based sentiment prediction. In IFIP International Conference on Artificial Intelligence Applications and Innovations (pp. 397-408). Cham: Springer International Publishing.https://doi.org/10.1007/978-3-030-49161-1_33

[8] Al-Moslmi, T., Omar, N., Abdullah, S., &Albared, M. (2017). Approaches to cross-domain sentiment analysis: A systematic literature review. Ieee access, 5, 16173-16192.https://doi.org/10.1109/ACCESS.2017.2690342

[9] Almotairi, M. (2009, July). A framework for successful CRM implementation. In European and Mediterranean conference on information systems (pp. 1-12).

[10] Bondielli, A., &Marcelloni, F. (2019). A survey on fake news and rumour detection techniques. Information Sciences, 497, 38-55.https://doi.org/10.1016/j.ins.2019.05.035

[11] Budgen, D., Brereton, P., Drummond, S., & Williams, N. (2018). Reporting systematic reviews: Some lessons from a tertiary study. Information and Software Technology, 95, 62-74.https://doi.org/10.1016/j.infsof.2017.10.017

[12] Cai, Y., Huang, Q., Lin, Z., Xu, J., Chen, Z., & Li, Q. (2020). Recurrent neural network with pooling operation and attention mechanism for sentiment analysis: A multi-task learning approach. Knowledge-Based Systems, 203, 105856.https://doi.org/10.1016/j.knosys.2020.105856

[13] Cambria, E., Das, D., Bandyopadhyay, S., &Feraco, A. (2017). Affective computing and sentiment analysis. A practical guide to sentiment analysis, 1-10.https://doi.org/10.1007/978-3-319-55394-8_1

[14] Cambria, E., Hussain, A., Havasi, C., &Eckl, C. (2009). Common sense computing: From the society of mind to digital intuition and beyond. In Biometric ID Management and Multimodal Communication: Joint COST 2101 and 2102 International Conference, BioID_MultiComm 2009, Madrid, Spain, September 16-18, 2009. Proceedings 2 (pp. 252-259). Springer Berlin Heidelberg.https://doi.org/10.1007/978-3-642-04391-8_33

[15] Cambria, E., Poria, S., Gelbukh, A., &Thelwall, M. (2017). Sentiment analysis is a big suitcase. IEEE Intelligent Systems, 32(6), 74-80.https://doi.org/10.1109/MIS.2017.4531228

[16] Cambria, E., Li, Y., Xing, F. Z., Poria, S., & Kwok, K. (2020, October). SenticNet 6: Ensemble application of symbolic and subsymbolic AI for sentiment analysis. In Proceedings of the 29th ACM international conference on information & knowledge management(pp.105114).https://doi.org/10.1145/3340531.3412003

[17] Can, E. F., Ezen-Can, A., & Can, F. (2018). Multilingual sentiment analysis: An RNN-based framework for limited data. arXiv preprint arXiv:1806.04511.https://doi.org/10.48550/arXiv.1806.04511

[18] Catal C, Mishra D (2013) Test case prioritization: a systematic mapping study. SoftwQual J 21(3):445–478 Chandra Y, Jana A (2020) Sentiment analysis using machine learning and deep learning. In: 2020 7th Inter- national conference

on computing for sustainable global development (INDIACom). IEEE, pp. 1–4 Chapelle O, Zien A (2005) Semi-supervised classification by low density separation. In: AISTATS vol 2005, pp 57–64https://doi.org/10.1007/s11219-012-9181-z

[19] Che, S., Li, X. RETRACTED ARTICLE: HCI with DEEP learning for sentiment analysis of corporate social responsibility report. CurrPsychol 41, 3298 (2022). https://doi.org/10.1007/s12144-020-00789-y

[20] Chen, I. J. (2003). i Popovich, K.,(2003). Understanding Customer Relationship Management (CRM): People, Process and Technology. Business Process Management Journal, 9(5), 672-688.

[21] ChenL, C., & Wang, F. (2015). Recommender systems based on user reviews the state of the art. User Modeling and User-Adapted Interaction, 25, 99-154.

[22] de Oliveira Lima, T., Colaco Junior, M., &Nunes, M. A. S. (2018). Mining on line general opinions about sustainability of hotels: a systematic literature mapping. In Computational Science and Its Applications–ICCSA 2018: 18th International Conference, Melbourne, VIC, Australia, July 2-5, 2018, Proceedings, Part I 18 (pp. 558-574). Springer International Publishing.https://doi.org/10.1007/978-3-319-95162-1_38

[23] Dessí, D., Dragoni, M., Fenu, G., Marras, M., &ReforgiatoRecupero, D. (2020). Deep learning adaptation with word embeddings for sentiment analysis on online course reviews. deep learning-based approaches for sentiment analysis, 57-83.https://doi.org/10.1007/978-981-15-1216-2_3

[24] Devlin, J., Chang, M. W., Lee, K., &Toutanova, K. (2018). Bert: Pre-training of deep bidirectional transformers for language understanding. arXiv preprint arXiv:1810.04805.https://doi.org/10.48550/arXiv.1810.04805

[25] Dietterich, T. G. (2002). Machine learning for sequential data: A review. In Structural, Syntactic, and Statistical Pattern Recognition: Joint IAPR International Workshops SSPR 2002 and SPR 2002 Windsor, Ontario, Canada, August 6–9, 2002 Proceedings (pp. 15-30). Springer Berlin Heidelberg.https://doi.org/10.1007/3-540-70659-3_2

[26] Ji, C., & Wu, H. (2020). Cascade architecture with rhetoric long short-term memory for complex sentence sentiment analysis. Neurocomputing,405,161172.https://doi.org/10.1016/j.neucom.2020.04.055

[27] Jia, Z., Bai, X., & Pang, S. (2020). Hierarchical gated deep memory network with position-aware for aspect-based sentiment analysis. IEEE Access, 8, 136340-136347.https://doi.org/10.1109/ACCESS.2020.3011318

[28] Jiang, T., Wang, J., Liu, Z., & Ling, Y. (2020). Fusion-extraction network for multimodal sentiment analysis. In Advances in Knowledge Discovery and Data Mining: 24th Pacific-Asia Conference, PAKDD 2020, Singapore, May 11–14, 2020, Proceedings, Part II 24 (pp. 785-797). Springer International Publishing.https://doi.org/10.1007/978-3-030-47436-2_59

[29] A Ligthart, C Catal, B Tekinerdogan Systematic reviews in Sentiment Analysis: A Tertiary Study Pacific-Asia conference on knowledge discovery and data mining. Springer, Cham, pp 785–797https://doi.org/10.1007/978-3-339-95162-3_39

[30] Jin, N., Wu, J., Ma, X., Yan, K., & Mo, Y. (2020). Multi-task learning model based on multi-scale CNN and LSTM for sentimentclassification.IEEEAccess,8,7706077072.https://doi.org/10.1109/ACCESS.2020.2989428

[31] Josiassen, A., Assaf, A. G., &Cvelbar, L. K. (2014). CRM and the bottom line: Do all CRM dimensions affect firm performance? International Journal of Hospitality Management, 36, 130-136.https://doi.org/10.1016/j.ijhm.2013.08.005