Pre-print Version. Article published in Mediterranean Conference on Pattern Recognition and Artificial Intelligence
DOI: https://doi.org/10.1007/978-3-031-04112-9_17

An Automated Approach for Analysing Students Feedback Using Sentiment Analysis Techniques

Shpetim Sadriu[1], Krenare Pireva Nuci[1(✉)], Ali Shariq Imran[2], Imran Uddin[3], and Muhammad Sajjad[2,3]

[1] Faculty of Computer Science and Engineering, University for Business and Technology, Pristina, Kosovo
krenare@gmail.com
[2] Faculty of Information Technology and Electrical Engineering, Norwegian University of Science and Technology (NTNU), Gjøvik 2815, Norway
[3] Department of Computer Science, Islamia College University Peshawar, Peshawar 25000, Pakistan

Abstract. Conducting and evaluating continuous student feedback is essential for any quality enhancement cell (QEC) within an education institution. Students' feedback based on their personal opinions can play a vital role in ensuring quality education. However, students' subjective opinions are often ignored due to time constraints or a lack of adequate analysis strategies. Therefore, to automate the quality assurance process, two classification models (i.e., based on Monkey learn API and SentiWord using TextBlob) are proposed to analyze students' feedback data. The results shows that the model employing MonkeyLearn performs nearly 22% points better than the Textblob on the Albanian language dataset obtained from 114 students' responses, achieving 72.12% accuracy.

Keywords: Sentiment analysis · MonkeyLearn · TexBlob · Machine learning · Artificial intelligence · Opinion mining · Emotion classification

1 Introduction

The pandemic of COVID-19 transpose the daily routine and lifestyle of people all over the world. It has also affected the education system and make them shift from in-campus classes into online classes. This switching in the education system induces many unseen consequences including student aptitude, response, learning ability. Therefore, mechanisms should be devised to evaluate the quality of education in online teaching. The traditional approach was to conduct post-class procedures by the quality assurance office. This is usually concretized by collecting feedback from the students about teachers and the courses on prescribed proforma having some predefined questions. This feedback only consists of quantitative data which is not sufficient because qualitative aspects are ignored, such as student's sentiments at the time of feedback. This poses a challenge to combine quantitative data with qualitative data to elucidate the outcome of the current education scenario through more detailed analysis of students' feedback.

The quantitative analysis of student feedback (based on prescribed questions about a course, teacher or assessment) is usually carried out through statistical analysis, whereas the qualitative analysis needs to take into account the students opinion about a particular subject or teacher in their styles, reflecting their state of mind. These opinions can be negative, positive or neutral. To dig out the real essence of their opinion, Sentiment Analysis of opinions is carried out. It is a sub branch of natural language processing (NLP) that aims to extract the sentiments from the text [3,4]. Sentiment Analysis is used to classify the intent of text, whether it is positive, neutral or negative sentiment, and is considered as text classification problem [5,10,21]. Millions of people express their sentiments and opinions in blogs, wikis, social networks and forums [18,20]. These sentiments and opinions are analyzed to monitor public opinion for decision-making [18,21], for example, election results have been predicted in studies by analyzing Twitter posts [16] and social trends related to COVID-19 [8].

In opinion mining, opinions are categorized into positive, negative and neutral while in sentiment analysis different emotions such as love, joy, surprise, anger, sadness, and fear are extracted from the text to evaluate the intended sentiments of the writer. Literature study reveals that there are numerous state-of-the-art machine learning and deep learning based techniques conducting opinion mining and sentiment analysis [3]. The feedback expressed by the students through text-note in the evaluation proforma can be used to evaluate the sentiments of the students. The sentiment analysis plays a vital role to evaluate the overall attitude, feedback and behaviors of the students versus teachers, courses and institutions in the educational system [6]. From the sentiment analysis, it can be easily identified what is the overall status of the students against any course, teacher or institution. Consequently, based on these findings, a different aspect of teaching and learning can be fine-tuned and overall educational policy can be reformed.

This study aims to facilitate the quality enhancement cell (QEC) for analyzing students qualitative feedback by proposing an automated approach for analyzing the sentiments from the captured data. This study follows the previous study conducted by [2] as part of the project supported by the Ministry of Education, Science, Technology and Innovation for the innovation of quality assurance offices within the Higher Education Institutions amidst COVID-19 pandemic crises. The project aimed to document the transformation of in-campus classes to online classes at Kosovo Universities and identified the technological infrastructure for the online classes during COVID-19 lockdown. The key contribution of this study is the collection of 114 students' feedback in Albanian language and applying two state-of-the-art polarity assessment models.

The rest of this paper is organized as follows: Sect. 2 discusses the related work concerning sentiment analysis in the education domain, while Sect. 3 discusses the methodology approach, including the methods and instruments used in this research paper. In Sect. 4 the results are presented, while Sect. 5 concludes the research study and emphasizes the achieved results.

2 Related Work

Researchers have used the sentiment analysis technique in multiple domains for analysing the users' feedback, particularly, students feedback in the education domain. For instance, [13] presented a comprehensive review of aspect-based sentiment analysis approaches in education domain. The paper focuses on the student's opinions towards teachers, courses and institutions. Moreover, according to a study [1], text classification models based on deep learning techniques gained massive popularity in recent years. Deep neural models can achieve incredible results in sentiment analysis tasks. The authors in [26] emphasized the effectiveness of 3W-CNN for sentiment classification on four benchmark data-sets. The study showed 3W-CNN has achieved higher performance compared to convolution Neural Network(CNN) and Naive Bayes - Support Vector Machine (NB-SVM) with 85.8% accuracy. The accuracy of both CNN and NB-SVM on MR data-set was almost equal, but the convergence of CNN was faster.

Sindhu et al. in [22] applied the Long short-term memory (LSTM) model. In it, they initially performed aspect extraction (six aspects extracted: teaching pedagogy, behaviour, knowledge, assessment, experience and general) attaining accuracy up to 91%, following with the sentiment polarity detection by achieving 93% accuracy. The data-set included the opinions of 40 students. Further, Lee et al. proposed a methodology for identifying keywords by discriminating positive and negative sentences [14]. They classified the word based on weakly supervised learning using CNN. Furthermore, [15,24] proposed models to classify polarity to identify words with high polarity scores using English and Korean language. The models include CNN-Rand, CNN-Static, CNN-Non-Static, CNN-2 channel and CNN-4 channel. Whereas, the work conducted in [15], used deep neural network similar to [7,11]. Moreover, there are certain classification models with pseudo-document generation and self-training modules that use unlabeled data for model refinement. This is a flexible method for handling weak supervision types that can integrate existing deep neural model for text classification. According to Kastrati et al. [11], 80.64% (F1 score) was achieved using weakly supervised framework for aspect-based sentiment analysis whereas for broader course-related aspects the F1 score was 65.64%.

Estradaa et al. in [7] tried to indicate polarity for positive and negative labels (senti-TEXT) as well as polarity with positive, negative learning centered emotions (eduSERE: engaged, excited, bored and frustrated). Moreover, evolutionary algorithm EvoMSA was used to investigate the effectiveness of different architectures (CNN, one LSTM, hybrid between CNN and LSTM and BERT) based on accuracy classification and polarity with learning-centred emotions. The research achieved 95% of accuracy for sentiTEXT and 84% of accuracy for eduSERE. Wang in his work [23] showed that using recurrent Neural network (RNN) the system can achieve significant performance in text categorization. In his study, the author presented disconnected recurrent neural network (DRNN) with positive invariance in combination with RNN. For reaching higher accuracy in text categorization, the author proposed the use of DRNN model, to improve RNN and CNN model.

According to Yang et al. [25], there are two main challenges when classifying sentiment in text:

– Sentiment classification is highly domain-dependent and the efficiency of the trained model is not guaranteed in every domain.
– The quantity of labeled data plays important role in the quality of the classifier, it becomes difficult to evaluate the classifier when there is limited labeled data in a domain.

Therefore, the researchers focused on learning high-level features that can generate sentiment classifications in other domains toward global classifiers. The proposed model is based on aggregation between labeled and un-labeled data from multiple domains by learning new feature representations. The experiment used multi-domain sentiment classification by comparing methods within in-domain classifier and multi-domain classifier.

According to [1,7], aspect-based sentiment analysis (ABSA) plays a key role in predicting polarity of text through NLP. To get precise sentiment expression from ABSA, they propose a model named Attention-based Sentiment Reasoner (AS-Reasoner). The study used English and Chinese language datasets for capturing sentiment similarity between two words in a single sentence and computed weights for global attention from a global perspective. Finally, Kastrati et al. in [12] used sentiment rich representation as an impact of deep learning document classification which increased the performance score by five percentage resulting in 78.10% (F1 score).

This research study differs from the previous discussed research by using existing text analysis APIs such as MonkeyLearn and TextBolob in a dataset constructed from data collected using the primary data collection method in Education domain. The students feedback is initially collected in the Albanian Language. The analysis of the captured feedback is performed at the sentence level. Furthermore, the effectiveness of this technique is realized by comparing the accuracy of the classified sentiments using each of the aforementioned APIs.

3 Methodology

For this research paper, a prototype in Python language is developed. The prototype is fed with 624 paragraphs of opinions expressed by 114 student in Albanian language from the Faculty of Computer Science at the University for Business and Technology, Kosovo.

The experiment followed three phases. As shown in Fig. 1 the first phase dealt mainly with the construction of dataset. The dataset includes three sentiment and seven emotion classes. The maximum length of the sentences in the dataset is 108 words, the minimum is 1 word, whereas the average length of the sentences is 28 words.

The students opinion [19] is processed in sentence level, which then is manually labeled into three sentiment classes (see Fig. 1, sentiment column): positive (1), neutral (2), and negative (3). The distribution in both sentiment classes is

Emotions	Sentiment	Data (ALB)	Data (ENG)
4	3	Per shkak të gjendjes së krijuar si pasojë e pandemisë edhe ne sikur shumica e institucioneve në Kosovë kemi qenë të detyruar që të vijojmë ligjëratat online	Due to the situation created as a result of the pandemic, we, like most institutions in Kosovo, have been forced to attend online lectures
5	1	Gjatë ligjëratave nuk kanë munguar as projektet e ndryshme prej të cilëve kemi përfituar shumë dhe këto projekte edhe pse nga distanca mendoj se kanë qenë mjaft të dobishme	During the lectures there were also various projects from which we benefited a lot and these projects, although from a distance I think they were quite useful
2	1	Në përgjithësi mund të them se gjatë kësaj periudhe të vështirë kemi arritur të përfundojmë me sukses procesin mësimor online	In general I can say that during this difficult period we have managed to successfully complete the online learning process

Fig. 1. A snapshot of the labelled dataset depicting emotion and sentiment category along with the actual feedback data.

imbalanced, 75% are with positive polarity, 0.11% neutral, and 13% negative. With respect to emotional labeling, the authors followed the Parrot model [17]. The Parrot model encapsulates the emotional classifications based on primary aspects such as: love (6), joy (5), surprise (4), anger (3), sadness (2), fear (1), and neutral (0) (Fig. 2).

Fig. 2. High-level overview of the proposed methodology.

The second phase follows the creation of the models using two different text analysis approaches i.e., TextBlob API (see Fig. 3) and Monkey Learn API (see Fig. 4) for extracting the sentiment and emotion of the sentences. In both models the Python Flask is used as stack technology for pre-processing dataset and converting data automatically from CSV into JSON format. A Google translation API is also employed to transltate Albanian langauge into English for further processing with MonkeyLearn. An incentive for using two types of APIs is to observe the difference in the classification accuracy using state-of-the-art technology.

To evaluate the proposed model in the third phase, the authors used Precision, Recall, Accuracy and F1 score as evaluation metrics.

TextBlob is an NLP library for Python and it is used for part-of-speech tagging, noun phrase extraction and sentiment analysis. It aims to provide access to a common text processing operation through its interface. In our case, we used the model for performing the sentiment classification of the collected data following the steps depicted in Fig. 3.

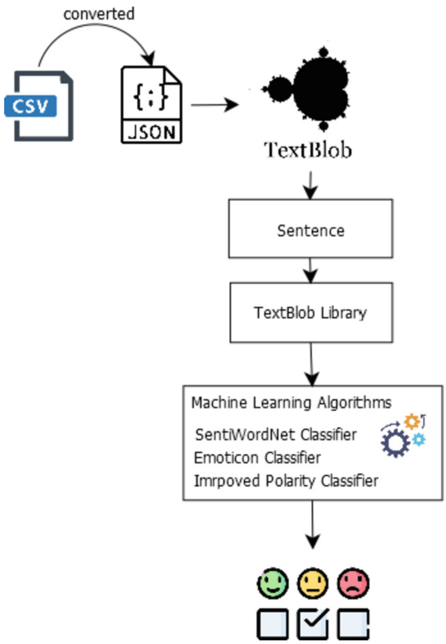

Fig. 3. Sentiment analysis using TextBlog

As shown in Fig. 3, the dataset is initially converted from CSV format into JSON. This is followed by SentiWord Classifier (SWNC), which implements NaiveBayesAnalyzer. NaiveBayesAnalyzer uses Natural Language ToolKit (NLTK) with Emoticon Classifier (EEC) and Improved Polarity Classifier (IPC).

In the second model using MonkeyLearn API (see Fig. 4) on the other hand, the authors integrated Google translate API for language translation, as the MonekyLearn API can not process Albanian language. Additionally, combining MonkeyLearn API and Google-Translate API [9] for Albanian to English translation in real-time shows higher precision (see Table 4). MonkeyLearn API is text-analysis API that allows developers to process textual data for classification. It implements machine learning models for sentiment analysis, keyword extraction and topic detection.

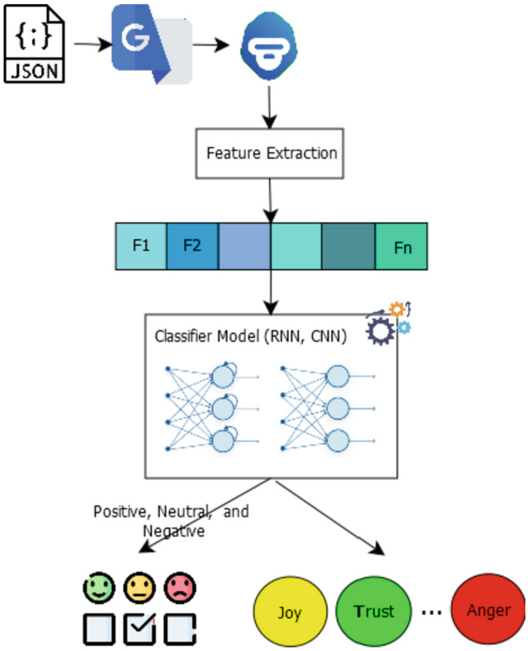

Fig. 4. Sentiment analysis using MonkeyLearn

4 Results

This section presents the results obtained with TextBlob and MonkeyLearn API for sentiment polarity assessment. In the following subsection, we also present the metrics used to evaluate the models.

4.1 Evaluation Metrics

To evaluate the models' performance for predicting sentiment polarity results, we employed precision, accuracy, recall, and F1 score. These are derived from the confusion matrix which is composed of true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN), as described below:

Accuracy - It is the percentage of correctly predicted instances from total instances.

$$Accuracy = \frac{(\mathrm{TP} + \mathrm{TN})}{\mathrm{TP} + \mathrm{FP} + \mathrm{TN} + \mathrm{FN}} \tag{1}$$

Precision - Precision is the percentage of correctly classified samples for the particular class out of all predicted labels for that class.

$$Precision = \frac{\mathrm{TP}}{(\mathrm{TP} + \mathrm{FP})} \tag{2}$$

Recall - The recall is the percentage of all predicted samples for the particular class relation with actual labels for that class.

$$Recall = \frac{\mathrm{TP}}{(\mathrm{TP} + \mathrm{FN})} \tag{3}$$

F1-score - F1 score is a combination of both precision and recall, it can be interpreted as the harmonic mean of precision and recall.

$$F1 - score = \frac{2 \times (\mathrm{Precision} \times \mathrm{Recall})}{(\mathrm{Precision} + \mathrm{Recall})} \tag{4}$$

4.2 Polarity Prediction Using Textblob and MonkeyLearn

The TextBlob API uses a range from −1 to 1 for sentiment classification. The negative responses are assigned −1, 0 for neutral, and +1 for positive sentiment. Table 1 shows that the TextBlob model predicted 165 sentences as negative out of 624 in total (i.e. 26.44%), 147 responses with neutral polarity (23.56%), and the rest of 312 sentences as positive (50%). Whereas, the MonkeyLearn API predicted 98 negative sentences (15.71%), 85 responses with neutral polarity (13.62%), and the rest of 441 sentences with positive polarity (70.67%).

Table 1. MonkeyLearn versus TextBlob API prediction

Sentiment classification	MonkeyLearn (%)	TextBlob (%)
Positive	70.67	50
Neutral	13.62	23.56
Negative	15.71	26.44

The validation of the results for the model using MonkeyLearn for predicting the sentiment polarity (positive, neutral and negative) is performed using the confusion matrix, and Table 2 shows the classification report. It can be seen that the model based on MonkeyLearn was 72% accurate.

The same approach for validating the results for the model using TextBlob for predicting the three classes of sentiment polarity is performed, and Table 3 shows the classification report. TextBlob achieved an accuracy of 50%.

Table 2. Classification report for MonkeyLearn model

	Precision	Recall	F1-Score	Support
−1	0.44	0.51	0.47	84
0	0.28	0.33	0.31	72
1	0.87	0.82	0.84	468
Accuracy			0.72	624
Macro avg	0.53	0.55	0.54	624
Weighted avg	0.74	0.72	0.73	624

Table 3. Classification report for TextBlob model

	Precision	Recall	F1-Score	Support
−1	0.19	0.38	0.26	84
0	0.11	0.22	0.15	72
1	0.86	0.57	0.68	468
Accuracy			0.50	624
Macro avg	0.39	0.39	0.36	624
Weighted avg	0.68	0.50	0.56	624

Table 4, presents the comparison results of the prediction for both models with respect to accuracy, precision (avg.), recall (avg.) and F1-Score (avg.). The model using MonkeyLearn resulted with higher accuracy (72%) compared to TextBlob (50%).

Table 4. Accuracy (%), Precision (Avg%), Recall (Avg%) and F1-Score results for MonkeyLearn and TexBlob API

Model	Accuracy	Precision	Recall	F1-Score
MonkeyLearn API	0.72	55	53	54
TextBlob API	0.50	39	39	36

4.3 Emotion Identification Using MonkeyLearn

In addition to sentiment classification, we further analyszed the emotions expressed in the feedback by students employing MonkeyLearn emotion classifier API. We classified the emotions into distinct categories defined by the Parrot model i.e. love, joy, surprise, anger, sadness, and fear [17]. We further had few instances in our dataset that didn't depict any significant emotions, which we classified as neutral.

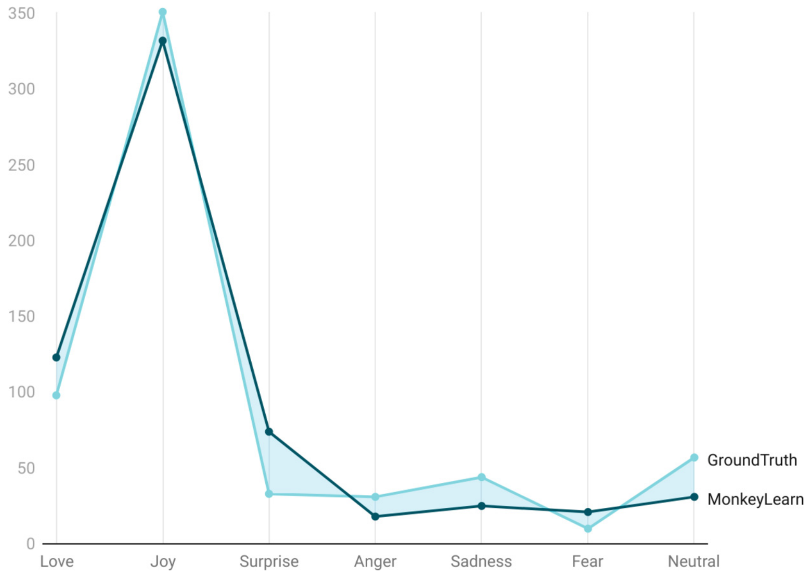

Fig. 5. Emotion classification using MonkeyLearn model

The results in Fig. 5 shows that Joy emotion dominated with 351 sentences (56%), followed by Love emotion with 98 responses (16%). The model predicted 32 (5%) responses as Surprise and Anger, While Sadness was recorded in 44 responses (7%) and the last emotion class Fear, appeared in 10 response sentences (2%). Finally, we have included Neutral emotion category for the dataset instances that are not satisfied with Parrot emotions, containing 57 sentences (9%).

5 Conclusion

This research paper studied two automated models for sentiment analysis of student's feedback to predict the sentiment polarity and identify emotions. To predict the sentiment polarity, the first model uses TextBlob API, whereas the second model uses MonkeyLearn API. The latter model is further utilized to identify seven emotional categories from the students' feedback, following the Parrot Model classification. Results from both TextBlob and MonkeyLearn models are validated using a ground truth dataset, which is constructed from 114 students' feedback (624 sentences with average length of 28 words) as part of the quality enhancement cell procedure for the Faculty of Computer Science and Engineering at The University for Business and Technology in Kosovo. The results shows that the model using MonkeyLearn achieved a sentiment polarity classification accuracy of 72.12% compared to the model using Texblob, with 50.48%.

Acknowledgment. The authors would like to thank the Ministry of Education, Science, Technology and Innovation in Kosovo for supporting this research, as part of the project: The impact of COVID-19 pandemic in teaching and learning in Higher Education Institutions in Kosovo.

References

1. Apuk, V., Nuci, K.P.: Classification of pedagogical content using conventional machine and deep learning model. WiPiEC J. Works Progr. Embed. Comput. **7**(1) (2021)
2. Baftiu, E., Nuci, K.P.: The transformation process from in-campus classes into online classes due to the Covid-19 situation-the case of higher education institutions in Kosovo. arXiv preprint arXiv:2104.03896 (2021)
3. Birjali, M., Kasri, M., Beni-Hssane, A.: A comprehensive survey on sentiment analysis: approaches, challenges and trends. Knowl. Based Syst. 107134 (2021)
4. Chaturvedi, I., Cambria, E., Welsch, R.E., Herrera, F.: Distinguishing between facts and opinions for sentiment analysis: survey and challenges. Inf. Fusion **44**, 65–77 (2018). https://doi.org/10.1016/j.inffus.2017.12.006, https://www.sciencedirect.com/science/article/pii/S1566253517303901
5. Choi, Y., Lee, H.: Data properties and the performance of sentiment classification for electronic commerce applications. Inf. Syst. Front. **19**(5), 993–1012 (2017). https://doi.org/10.1007/s10796-017-9741-7
6. Edalati, M., Imran, A.S., Kastrati, Z., Daudpota, S.M.: The potential of machine learning algorithms for sentiment classification of students' feedback on MOOC. In: Arai, K. (ed.) IntelliSys 2021. LNNS, vol. 296, pp. 11–22. Springer, Cham (2022). https://doi.org/10.1007/978-3-030-82199-9_2
7. Estradaa, M.L.B., Cabada, R.Z., Bustillos, R.O., Graff, M.: Opinion mining and emotion recognition applied to learning, ELSEVIER. Expert Syst. Appl. **150**, 12 (2020)
8. Imran, A.S., Daudpota, S.M., Kastrati, Z., Batra, R.: Cross-cultural polarity and emotion detection using sentiment analysis and deep learning on Covid-19 related tweets. IEEE Access **8**, 181074–181090 (2020)
9. Google Cloud Translate, INC.: Translation. https://cloud.google.com/translate. Accessed 28 2006
10. Ju, S., Li, S.: Active learning on sentiment classification by selecting both words and documents. In: Ji, D., Xiao, G. (eds.) CLSW 2012. LNCS (LNAI), vol. 7717, pp. 49–57. Springer, Heidelberg (2013). https://doi.org/10.1007/978-3-642-36337-5_6
11. Kastrati, Z., Imran, A.S., Kurti, A.: Weakly supervised framework for aspect-based sentiment analysis on students reviews of MOOCs. IEEE Access **4**, 2016(4(10.)) (2017)
12. Kastrati, Z., Imran, A.S., Yayilgan, S.Y.: The impact of deep learning on document classification using semantically rich representations, research gate, no. 10, 25 (2019)
13. Kastrati, Z., Dalipi, F., Imran, A.S., Pireva Nuci, K., Wani, M.A.: Sentiment analysis of students' feedback with NLP and deep learning: a systematic mapping study. Appl. Sci. **11**(9), 3986 (2021)
14. Lee, G., Jeong, J., Seo, S., Kim, C., Kang, P.: Sentiment classification with word attention based on, knowledge-based systems. Knowl. Based Syst. **152**(10) (2018)

15. Meng, Y., Shen, J., Zhang, C., Han, J.: Weakly-supervised neural text classification, department of computer science, university of at Illinois Urbana-champaign, IL, USA, no. 10, pp. 983–992 (2018)
16. O'Connor, B., Balasubramanyan, R., Routledge, B.R., Smith, N.A.: From tweets to polls: linking text sentiment to public opinion time series. In: Fourth International AAAI Conference on Weblogs and Social Media (2010)
17. Parrott, W.: Emotions in Social Psychology. Psychology Press, Philadelphia (2001). 978-0863776830
18. Ramírez-Tinoco, F.J., Alor-Hernández, G., Sánchez-Cervantes, J.L., Olivares-Zepahua, B.A., Rodríguez-Mazahua, L.: A brief review on the use of sentiment analysis approaches in social networks. In: CIMPS 2017. AISC, vol. 688, pp. 263–273. Springer, Cham (2018). https://doi.org/10.1007/978-3-319-69341-5_24
19. Rrustemi, V., Nuci, K.P.: Aspect based sentiment analysis using deep learning techniques. Thesis, Work in Progress (2021)
20. Saleh, M.R., Martín-Valdivia, M.T., Montejo-Ráez, A., Ureña-López, L.: Experiments with SVM to classify opinions in different domains. Expert Syst. Appl. **38**(12), 14799–14804 (2011)
21. Schuller, B., Mousa, A.E.D., Vryniotis, V.: Sentiment analysis and opinion mining: on optimal parameters and performances. Wiley Interdisc. Rev. Data Min. Knowl. Discov. **5**(5), 255–263 (2015)
22. Sindhu, I., Daudpota, S.M., Badar, K., Bakhtyar, M., Baber, J., Nurunnab, D.M.: Aspect based opinion mining on student s feedback for faculty teaching performance evaluation. IEEE Access **7**(10), 108729–108741 (2019)
23. Wang, B.: Disconnected recurrent neural networks for text categorization. Joint Laboratory of HIT and iFLYTEK, iFLYTEK Research, Beijing, China (20118)
24. Wojcik, K., Tuchowski, J.: Feature based sentiment analysis. In: 3rd International Scientific Conference on Contemporary Issues in Economics, Business and Management EBM 2014, no. 269093520 (2014)
25. Yang, F., Mukherjee, A., Zhang, Y.: Leveraging multiple domains for sentiment classification. In: Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers, pp. 2978–2988 (2016)
26. Zhang, Y., Zhang, Z., Miao, D., Wang, J.: Three-way enhanced convolutional neural networks for sentence-level sentiment classification. ELSEVIER (10), 55–64 (2018)