










Sentiment Analysis in Education Domain: A Systematic Literature Review

Karen Mite-Baidal¹  , Carlota Delgado-Vera¹ ,
Evelyn Solís-Avilés¹ , Ana Herrera Espinoza¹ ,
Jenny Ortiz-Zambrano² , and Eleanor Varela-Tapia² 

¹ Facultad de Ciencias Agrarias, Universidad Agraria del Ecuador,
Av. 25 de Julio, P.O. Box 09-04-100, Guayaquil, Ecuador
{kmite, cdelgado, esolis, aherrera}@uagraria.edu.ec
² Facultad de Ciencias Matemáticas y Físicas, Universidad de Guayaquil,
Av. Delta, P.O. Box 471, Guayaquil, Ecuador
{jenny.ortizz, eleanor.varelat}@ug.edu.ec

Abstract. E-learning is the delivery of education through digital or electronic methods allowing students to acquire new knowledge and develop new skills. E-learning allows students to expand their knowledge whenever and wherever. Several authors consider sentiment analysis as an alternative to improve the learning process in an e-learning environment since it allows analyzing the opinions of the students in order to better understand their opinion and take more effective, better-targeted actions. In this sense, this work presents a systematic literature review about sentiment analysis in education domain. This review aims to detect the approaches and digital educational resources used in sentiment analysis as well as to identify what are the main benefits of using sentiment analysis on education domain. The results show that Naïve Bayes is the most used technique for sentiment analysis and that forums of MOOCs and social networks are the most used digital education resources to collect data needed to perform the sentiment analysis process. Finally, some of the main benefits of using sentiment analysis in education domain are the improvement of the teaching-learning process and students' performance, as well as the reduction in course abandonment.

Keywords: Sentiment analysis · Opinion mining · Education

1 Introduction

Sentiment analysis, also known as opinion mining [1], is an area of information processing that has been successfully applied in domains such as medicine. For example, there are several works that use opinion mining to analyze the emotional reaction of patients regarding different aspects of diabetes [2] and asthma [3]. The systematic review of the literature presented in this paper focuses on the use of sentiment analysis in the education domain.

Data recovery techniques [4] mainly focus on processing, searching and extracting factual information from digital education resources or learning environments [5], such

as blogs, forums, and social networks. The data have an objective and subjective perspectives. On the one hand, the objective perspective is not influenced by emotions, opinions, or personal feelings, i.e., it is a perspective based in fact, in things quantifiable and measurable. On the other hand, the subjective perspective is one open to greater interpretation based on personal feelings, emotions, aesthetics, etc. Sentiment analysis focuses on analyzing the subjective perspective of data.

The sentiment analysis process is divided into four core phases: data acquisition, data preparation, review analysis, and sentiment classification. There are two main sentiment analysis approaches, namely: (1) machine learning, which is divided into supervised and unsupervised machine learning approaches, and (2) lexicon-based approach, which is divided into two categories dictionary-based and corpus-based approaches [7].

Supervised machine learning uses techniques or algorithms such as Naive Bayes [8], which is the simplest and most used classifier that calculates the posterior probability of a class based on the distribution of words in a document. This algorithm uses the Bayes Theorem to calculate the probability of a word belongs to a particular tag. SVM (Support Vector Machine) classifiers are also used in sentiment analysis. SVM are supervised learning models with associated learning algorithms that analyze data used for classification and regression analysis. On the other hand, neural networks are also used in sentiment analysis [10]. The learning process of neural networks requires a large corpus with positive, negative and neutral opinions collected from data sources such as social networks or forums. Once the training phase is completed, the network will be able to classify a new opinion as positive, negative or neutral. Finally, ME (Maximum Entropy) technique (ME) [11] calculates the probability that a text belongs to a category. To carry out this process, this technique should maximize the entropy in order to avoid introducing a bias in the system. Unlike NB, this method does not assume independence between features or terms.

The lexicon-based approach classifies a text according to the positive, negative and neutral words contained in it. This approach does not require a training phase. As was mentioned earlier, the lexicon-based approach can be divided into two categories: dictionary-based and corpus-based approaches. On the one hand, the corpus-based approach tries to find co-occurring word patterns to determine the polarity of a text. On the other hand, the dictionary-based approach uses synonyms, antonyms, and hierarchies that are found within the lexical database. The lexicon-based approach uses techniques such as specialized vocabularies [12] and dictionary construction techniques. For instance, in [13], the authors propose an emotional dictionary for sentiment analysis applied to online news. Another example of dictionary construction is presented in [14], where authors propose a dictionary for sentiment analysis based on common-sense knowledge.

E-learning is the delivery of education through digital or electronic methods allowing students to acquire new knowledge and develop new skills. E-learning allows students to expand their knowledge whenever and wherever. Kechaou [15] considers the sentiment analysis as an alternative to improve the learning process in an e-learning environment since it allows analyzing the opinions of the students in order to better understand their opinion and take more effective, better-targeted actions. Hence, it is important to analyze the use of sentiment analysis in the education domain. Despite there are currently several works that present literature reviews of sentiment analysis,

there is still no proposal that presents a systematic literature review of sentiment analysis in education domain.

The remainder of this work is structured as follows: Sect. 2 presents the research methodology followed in this literature review. Section 3 describes the systematic review execution, while, Sect. 4 presents our results. Finally, our conclusions are presented in Sect. 5.

2 Systematic Review Planning

The literature review presented in this work has three main objectives: (1) to identify the techniques and classification algorithms used by sentiment analysis in education domain; (2) to identify digital educational resources or learning environments that serve as data sources for the sentiment analysis; and (3) to identify the most used techniques and data sources by the sentiment analysis in education domain.

2.1 Research Questions

For the purposes of this literature review, three research questions were defined to guide us throughout the research and help us to meet the established objectives. The research questions are listed below:

- RQ1. What is the sentiment analysis process?
- RQ2. What approaches and digital educational resources are used in sentiment analysis?
- RQ3. What are the main benefits of using sentiment analysis on education domain?

2.2 Digital Libraries

Table 1 shows the digital libraries that were used to perform the systematic literature review. Also, this table presents the type of bibliographic source, language, the period of publication, and search strategy used in this work. As can be observed, a keyword-based search strategy was used to search for research works focused on sentiment analysis in education domain. This strategy is described in detail in the next section.

2.3 Search Strategy

To answer the research questions, we use a keyword-based search strategy. For this purpose, we identified a set of keywords related to sentiment analysis in education domain as well as synonyms for the set of keywords identified. Once these terms were defined, we combined these terms with the connectors “AND” and “OR”, resulting in the following search chain:

```
(sentiment analysis) AND (sentiment classification OR sentiment
analysis techniques OR opinion mining OR education domain)
AND/OR (digital educational resource) AND/OR (students)
AND/OR (university)
```

Table 1. Digital libraries.

| Digital library | Type | Approach | Language | Period |
|--|---|----------|----------|---------------|
| IEEE Xplore | e-books, scientific journals, conferences, scientific articles | Keywords | English | 2013– 2018 |
| Science Direct | | | | |
| Springer Link Wiley | | | | |
| Google Scholar | | | | |
| ERIC Institute of Education of Science | | | | |
| Elsevier | | | | |
| ACM Digital Library | | | | |

Finally, it should be mentioned that only the works published in the 2013–2018 period were considered in this work, such as was specified in Table 1.

2.4 Exclusion Criteria

We discarded those papers that were not directly related with sentiment analysis and education domain. Also, we use next exclusion criteria:

- Research works not written in English.
- Master and doctoral dissertations.
- Duplicated research works obtained from Google Scholar and Web of Science.

3 Systematic Review Execution

This section presents the systematic review execution which consisted in searching for research works relates to sentiment analysis and education domain in the digital libraries selected and evaluating the obtained studies considering the inclusion and exclusion criteria. Also, this review allowed responding to the research questions presented in Sect. 2.1. These responses are discussed in next sections.

3.1 RQ1. What is the Sentiment Analysis Process?"

Sentiments

Sentiments are attitudes, thoughts or judgments triggered by sensations or mental processes. Sentiments are defined according to the experiences of each person and are generated in the subconscious. Also, sentiments are durable and recurrent since they remain in the emotional memory [16]. Sentiment analysis aims to assign a sentiment polarity to a text, in this case to texts generated by students. Sentiment polarity indicates whether the message has a positive, negative or neutral sentiment [17]. Sentiment analysis can be performed at three levels: document, sentence, and entity level.

Sentiment Analysis Process

Figure 1 shows the sentiment analysis process which is divided into four main phases: data acquisition, data preparation, review analysis, and sentiment classification. These phases are described below:

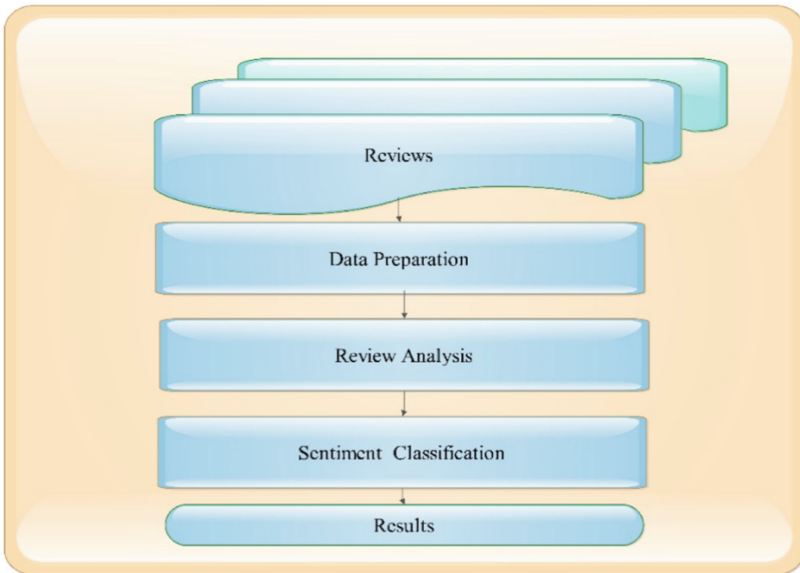


Fig. 1. Sentiment analysis process.

- Data acquisition can implement data mining techniques used in education domain [18] since data can be extracted from digital educational resources such as forums of MOOCs.
- Data preparation phase, also known as data preprocessing [19], is a necessary step for sentiment classification [20]. This phase consists of cleaning and preparing the text for classification. For instance, online texts contain usually lots of noise and uninformative parts such as HTML tags, scripts, and advertisements. In addition, on words level, many words in the text do not have an impact on the general orientation of it.
- Review analysis phase analyzes the linguistic features of reviews so that interesting information can be identified. This phase aims also to select the words that will be used in the last phase of sentiment analysis process.
- Sentiment classification phase classifies a new opinion as positive, negative or neutral based by implementing the machine learning, lexicon-based or hybrid approaches.

3.2 RQ2. What Approaches and Digital Educational Resources are Used in Sentiment Analysis?

Figure 2 shows the sentiment analysis approaches used in education domain according to the literature review performed. There are two main sentiment analysis approaches used in this domain: the machine learning and lexicon-based approaches. On the one hand, machine learning approach can be divided into supervised and unsupervised machine learning approaches. Regarding supervised machine learning approach, there are several classifiers used in education domain such as decision tree, linear, rule-based, and probabilistic classifiers. On the other hand, the lexicon-based approach uses techniques such as dictionary-based and corpus-based approaches.

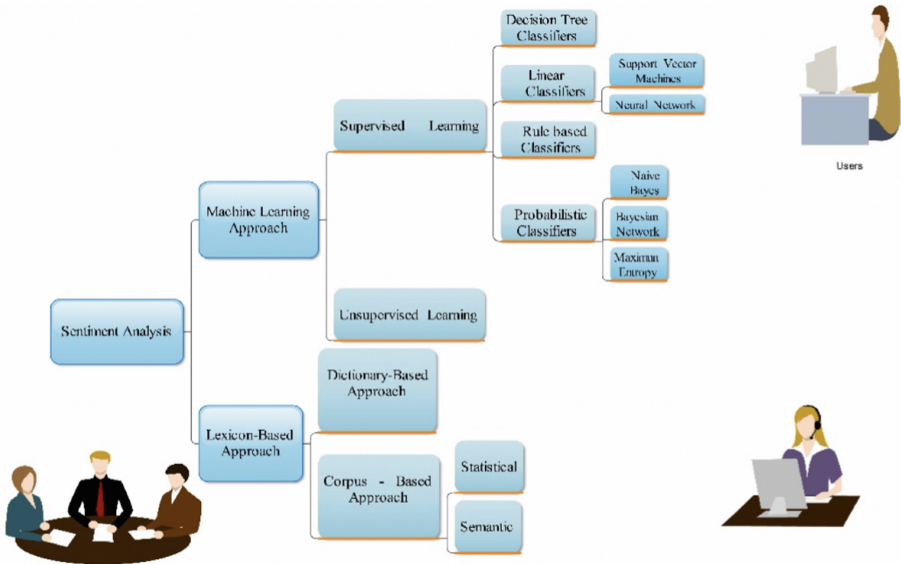


Fig. 2. Sentiment analysis approaches.

Table 2 shows the works analyzed in this literature review. This table presents the year of publication, sentiment analysis approach, classifier, and techniques used by the authors, the sentiment analysis level (document, sentence, and entity) adopted, and the precision achieved by the sentiment analysis process proposed by authors.

Japtap [22], who employed different techniques for sentiment analysis at the sentence level, concluded that it is not reliable to determine the sentiment of a user based on a brilliant or boring sentence. In this sense, the author analyzed the sentiment analysis techniques and established that each technique has a percentage of accuracy when the sentiment of a person is determined.

Table 3 present a set of the works analyzed in this literature review. This table aims to identify what are the digital educational resources most used for sentiment analysis

Table 2. Approaches, techniques and levels of sentiment analysis.

| Work | Year | Approach | Classifier | Technique | Level | Precision |
|------|------|-----------------------------|--------------------------------------|---|------------------------|------------------------|
| [22] | 2013 | Supervised machine learning | Probabilistic and linear classifiers | Naïve Bayes, maximum entropy, SVM | Sentence | (82.9% SVM) |
| [23] | 2014 | Supervised machine learning | Probabilistic and linear classifiers | Naïve Bayes, complement Naïve Bayes (CNB), maximum entropy, SVM | Document | (94% SVM) (84% CNB) |
| [24] | 2016 | Supervised machine learning | Linear classifiers | Linear discriminant analysis (LDA) | Entity | 86% |
| [4] | 2013 | Lexicon-based | Dictionary-based approach | Dictionary-based approach | Entity | N/P |
| [25] | 2014 | Lexicon-based | Dictionary-based approach | Dictionary-based approach | Entity | 90% |
| [26] | 2013 | Supervised machine learning | Probabilistic classifiers | Naïve Bayes, SVM | Document | 75–86%, 77–57% |
| [27] | 2013 | Supervised machine learning | Rule-based classifiers | Sentiment fuzzy classification | Document | N/P |
| [28] | 2013 | Supervised machine learning | Linear classifiers | Hybrid approach (BLSTM neural networks and SVM) | Audio, video, document | 73% |

in education domain. As can be seen, the most used resources are the forums of MOOCs followed by social networks such as Facebook and Twitter.

3.3 RQ3. What are the Main Benefits of Using Sentiment Analysis on Education Domain?

Sentiment analysis in education domain [45] goes beyond just knowing what the students' sentiments are. Table 4 presents the benefits that can be provided by sentiment analysis to education domain. Some of these benefits are learning process improvement, performance improvement, reduction in course abandonment, teaching process improvement, and satisfaction with a course. Furthermore, Analytical learning refers to the collection and analysis of students' information and their context aiming to

Table 3. Education resources used for sentiment analysis.

| Work | Year | Approach | Technique | Digital educational resource | Precision |
|------|------|--|--|---|----------------------|
| [29] | 2014 | Hybrid (machine learning and lexicon-based approach) | SVM and dictionary-based approach | Facebook | 83.27% |
| [30] | 2014 | Lexicon-based approach | Brown clustering | Forums | N/P |
| [31] | 2013 | Probabilistic classifiers | Naïve Bayes, SVM. | Twitter | 84.63% |
| [32] | 2013 | Subjective classification | N/A | Learning journals | N/P |
| [23] | 2014 | Supervised machine learning | Naïve Bayes, complement Naive Bayes (CNB), maximum entropy classification, SVM | Computer texts | (94% SVM), (84% CNB) |
| [33] | 2015 | Unsupervised machine learning | Clustering | Forum of a MOOC (massive open online courses) | 75.1% |
| [34] | 2016 | Supervised machine learning | Linear regression | Forum of a MOOC | 56.1% |
| [35] | 2014 | Lexicon-based approach | Dictionary-based approach | Forum of a MOOC | N/P |
| [36] | 2015 | Supervised machine learning | Naïve Bayes | Forum of a MOOC | 83.65% |
| [37] | 2015 | Supervised machine learning | Neural networks | Computer texts | 84.8% |
| [38] | 2016 | Supervised machine learning | Decision tree classifiers | Computer program | 99.7% |
| [39] | 2013 | Unsupervised machine learning | Clustering | N/P | 80% |
| [40] | 2013 | Supervised machine learning | Naïve Bayes | Computer program, Facebook | 87% |

(continued)

Table 3. (continued)

| Work | Year | Approach | Technique | Digital educational resource | Precision |
|------|------|-----------------------------|-------------------------|------------------------------|-----------|
| [41] | 2015 | Lexicon-based approach | Statistical analysis | Forum of a MOOC | 67.8% |
| [42] | 2013 | Lexicon-based approach | Semantic-based approach | Lexical database | 75.8% |
| [43] | 2014 | Lexicon-based approach | Semantic-based approach | Lexical database | 84% |
| [44] | 2014 | Supervised machine learning | Naïve Bayes | Twitter | 75% |
| [45] | 2017 | Supervised machine learning | SVM | Documents | 72.79% |

understand and optimize the learning process and the environment in which it occurs. This information is especially important for e-learning systems, which guide students through the learning process according to their particular needs and preferences. Hence, this information is also important for teachers since it allows them to know the emotional state of their students.

Table 4. Benefits that can be provided by sentiment analysis to education domain.

| Sentiment analysis advantage | Works |
|---------------------------------|------------------------------|
| Learning process improvement | [37, 38, 39, 40, 42, 43, 44] |
| Performance improvement | [35, 36] |
| Reduction in course abandonment | [41] |
| Teaching process improvement | [23] |
| Satisfaction with a course | [23] |

4 Results

Table 5 shows that forums of MOOCs [46] are the most used resources for sentiment analysis in education domain. In other words, in the education domain, datasets and lexicons are mainly built from forums of MOOCs. These data are provided as input to the sentiment analysis system to allow it to classify a new opinion.

Table 6 shows the most used techniques for sentiment analysis in education domain. These techniques are grouped according to the sentiment analysis approach to which they belong. According to the literature review presented in this work, the most used technique under the supervised machine learning approach is Naive Bayes, which commonly provides higher precision than other techniques used under this approach.

Table 5. Digital educational resources used in sentiment analysis on education context.

| Digital educational resource | Work |
|------------------------------|----------------------|
| Forum of a MOOC | [35, 34, 33, 36, 41] |
| Computer texts | [23, 37] |
| Learning journals | [32] |
| Twitter | [31, 44] |
| Forums | [30] |
| Facebook | [29, 40] |
| Computer programs | [38, 39] |
| Lexical databases | [42, 43] |

Regarding lexicon-based approach, dictionary-based techniques are the most used for sentiment analysis. Finally, Table 6 also reflects that machine learning and lexicon-based approaches can be used in conjunction to perform the sentiment analysis process.

Table 6. Sentiment analysis approaches.

| Approach | Work |
|--------------------------------------|--------------------------|
| <i>Supervised machine learning</i> | |
| Naïve Bayes | [44, 40, 36, 31, 26, 22] |
| SVM | [23, 31, 26, 22] |
| Maximum entropy | [22] |
| Decision tree classifiers | [38] |
| Neural networks | [37] |
| Linear regression | [34] |
| Sentiment fuzzy classification | [27] |
| <i>Unsupervised machine learning</i> | |
| Clustering | [39, 33] |
| Brown clustering | [30] |
| <i>Lexicon-based approach</i> | |
| Dictionary-based approach | [35, 25, 4] |
| Semantic-based approach | [43, 42] |
| Statistical analysis | [41] |
| <i>Hybrid approaches</i> | |
| SVM and dictionary-based approach | [29] |
| BLSTM neural networks and SVM | [28] |

5 Conclusions and Future Work

The systematic literature review presented in this work revealed that there are several works that use sentiment analysis to improve different aspects of education domain such as learning process, students' performance, reduction in course abandonment,

teaching process, and satisfaction with a course. This review also revealed that forums of MOOCs and social networks such as Facebook and Twitter are the most used digital education resources to collect data needed to perform the sentiment analysis process. Other educational resources used in sentiment analysis are learning journals, computer texts, software programs, lexical databases, and other electronic documents. Regarding sentiment analysis techniques, Support Vector Machine (SVM) and Naive Bayes are the most used techniques. Finally, we note that there is a trend to combine both machine learning approach and lexicon-based approach to perform the sentiment analysis process.

As future work, we plan to extend this literature review by including a wider set of digital libraries such as the Wiley Online Library. Furthermore, we plan to establish more research questions that help domain experts to obtain a better perspective on the use of sentiment analysis in education domain. This information could help experts to propose solutions that address challenges and limitations in education domain.

References

1. Vinodhini, G., Chandrasekaran, R.: Sentiment analysis and opinion mining: a survey. *Int. J. Adv. Res. Comput. Sci. Softw. Eng.* **2**, 282–292 (2012)
2. Salas-zárate, M.P., Medina-moreira, J., Lagos-ortiz, K., Luna-aveiga, H., Rodríguez-garcía, M.Á., Valencia-garcía, R.: Sentiment analysis on tweets about diabetes: an aspect-level approach. *Hindawi Comput. Math. Methods Med.* **2017**, 9 (2017)
3. Luna-Aveiga, H., et al.: Sentiment polarity detection in social networks: an approach for asthma disease management. In: Le, N.-T., Van Do, T., Nguyen, N.T., Thi, H.A.L. (eds.) *ICCSAMA 2017. AISC*, vol. 629, pp. 141–152. Springer, Cham (2018). https://doi.org/10.1007/978-3-319-61911-8_13
4. Feldman, R.: Techniques and applications for sentiment analysis. *Commun. ACM* **56**, 82 (2013)
5. Mandinach, E.B., Cline, H.F.: *Classroom Dynamics: Implementing a Technology-Based Learning Environment*. Taylor & Francis, New York (2013)
6. Mayer, J.D., Salovey, P., Caruso, D.R.: Emotional intelligence: new ability or eclectic traits? *Am. Psychol.* **63**, 503–517 (2008)
7. Anitha, N., Anitha, B.: Sentiment classification approaches – a review. *Int. J. Innov. Eng. Technol.* **3**, 22–31 (2013)
8. Zhang, H.: The optimality of Naive Bayes. *Am. Assoc. Artif. Intell.* **19** (2004)
9. Varghese, R., Science, C.: Aspect based sentiment analysis using support vector machine classifier. In: *International Conference on Advances in Computing, Communications and Informatics (ICACCI)*, pp. 1581–1586. IEEE (2013)
10. Tang, D., Qin, B., Liu, T.: Document modeling with gated recurrent neural network for sentiment classification. In: *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*, pp. 1422–1432 (2015)
11. Batista, F., Ribeiro, R.: Sentiment analysis and topic classification based on binary maximum entropy classifiers. *Proces. del lenguaje Nat.* **50**, 77–84 (2013)
12. Rice, D.R.: Corpus-based dictionaries for sentiment analysis of specialized vocabularies. In: *Annual Meeting of Midwest Political Science Association* (2015)
13. Rao, Y., Lei, J., Wenyan, L., Li, Q., Chen, M.: Building emotional dictionary for sentiment analysis of online news. *World Wide Web* **17**, 723–742 (2014)

14. Tsai, A.C.: Building a Concept-Level Sentiment on Commonsense Knowledge, pp. 22–30. IEEE Computer Society, Washington, D.C. (2013)
15. Kechaou, Z., Alimi, A.M.: Improving e-learning with sentiment analysis of users' opinions. In: IEEE Global Engineering Education Conference – Learning Environment Ecosystem for Engineering Education, pp. 1032–1038 (2011)
16. Munezero, M., Montero, C.S., Sutinen, E., Pajunen, J.: Are they different? Affect, feeling, emotion, sentiment, and opinion detection in text. *IEEE Trans. Affect. Comput.* **5**, 101–111 (2014)
17. Hoffmann, P., Wilson, T., Wiebe, J.: Recognizing contextual polarity: an exploration of features for phrase-level sentiment analysis. *Comput. Linguist.* **35**, 399–433 (2009)
18. Romero, C., Ventura, S.: Data mining in education. *Wiley Interdiscip. Rev. Data Min. Knowl. Discov.* **3**, 12–27 (2013)
19. Haddi, E., Liu, X., Shi, Y.: The role of text pre-processing in sentiment analysis. *Procedia Comput. Sci.* **17**, 26–32 (2013)
20. Ravi, K., Ravi, V.: *A Survey on Opinion Mining and Sentiment Analysis: Tasks, Approaches and Applications*. Elsevier B.V., New York City (2015)
21. Medhat, W., Hassan, A., Korashy, H.: Sentiment analysis algorithms and applications: a survey. *Ain Shams Eng. J.* **5**, 1093–1113 (2014)
22. Jagtap, V.S., Pawar, K.: Analysis of different approaches to sentence-level sentiment classification. *Int. J. Sci. Eng. Technol.* **2**, 164–170 (2013)
23. Altrabsheh, N., Cocea, M., Fallahkhair, S.: Learning sentiment from students' feedback for real-time interventions in classrooms. In: Bouchachia, A. (ed.) *ICAIS 2014. LNCS (LNAI)*, vol. 8779, pp. 40–49. Springer, Cham (2014). https://doi.org/10.1007/978-3-319-11298-5_5
24. Rana, T.A., Cheah, Y., Letchmunan, S.: Topic modeling in sentiment analysis: a systematic review. *J. ICT Res. Appl.* **10**, 76–93 (2016)
25. Gonçalves, P., Araújo, M., Benevenuto, F., Cha, M.: Comparing and combining sentiment analysis methods. *Comput. Appl. Soc. Behav. Sci. ACM.* 27–37 (2014)
26. Aliandu, P.: Sentiment analysis on Indonesian tweet. In: *Proceedings of International Conferences of Information, Communication, Technology, and Systems*, pp. 203–208 (2013)
27. Mouthami, K., Devi, K.N., Bhaskaran, V.M.: Sentiment analysis and classification based on textual reviews. In: *2013 International Conference on Information Communication and Embedded Systems*, pp. 271–276 (2013)
28. Wöllmer, M., Weninger, F., Knaup, T., Schuller, B.: YouTube movie reviews: sentiment analysis in an audio-visual context. *IEEE Intell. Syst.* 46–53 (2013)
29. Ortigosa, A., Martín, J.M., Carro, R.M.: Sentiment analysis in Facebook and its application to e-learning. *Comput. Human Behav.* **31**, 527–541 (2014)
30. Wen, M., Yang, D., Rosé, C.: Sentiment analysis in MOOC discussion forums: what does it tell us? In: *Proceedings of the Educational Data Mining*, pp. 1–8 (2014)
31. Neves-Silva, R., Watada, J., Phillips-Wren, G.E.: Intelligent decision technologies. In: *Proceedings of the 5th KES International Conference on Intelligent Decision Technologies (KES-IDT 2013)*. IOS Press (2013)
32. Munezero, M., Mozgovoy, M.: Exploiting sentiment analysis to track emotions in students' learning diaries. *Nat. Lang. Process. ACM.* 145–152 (2013)
33. Wang, X., Yang, D., Wen, M., Koedinger, K., Rosé, C.P.: Investigating how student' s cognitive behavior in MOOC discussion forums affect learning gains. In: *Proceedings of the 8th International Conference on Educational Data Mining*, pp. 226–233 (2015)
34. Robinson, C., Yeomans, M., Reich, J., Gehlbach, H.: Forecasting student achievement in MOOCs with natural language processing. In: *LAK 2016*, pp. 383–387. ACM (2016)

35. Tucker, C.S.: Mining student-generated textual data in MOOCS and quantifying their effects on student performance and learning outcomes. In: Proceedings of the 121st ASEE Annual Conference and Exposition, vol. 5 (2014)
36. Merceron, A.: Educational data mining/learning analytics: methods, tasks and current trends. In: Proceedings of the 13th e-Learning Conference of the German Computer Society (DeLFI 2015), pp. 101–109 (2015)
37. Bowman, S.R., Potts, C., Manning, C.D.: Learning distributed word representations for natural logic reasoning. In: Proceedings Knowledge Representation, Reasoning, Integration Symbolic Neural Approaches Paper from 2015 of the Association for the Advancement of Artificial Intelligence Spring Symposium (AAAI) Spring Symposium—Lea, pp. 10–13 (2015)
38. Darcy, A., Louie, A., Weiss, L.: Machine learning and the profession of medicine. *Am. Med. Assoc. Innov. Heal. CARE Deliv.* **5719**, 2–3 (2016)
39. Blikstein, P.: Multimodal learning analytics. In: LAK 2013, pp. 102–106. ACM (2013)
40. Troussas, C., Virvou, M., Espinosa, K.J., Llaguno, K., Caro, J.: Sentiment analysis of Facebook statuses using Naive Bayes classifier for language learning. *IEEE* (2013)
41. Crossley, S., Danielle, S., Baker, R., Wang, Y., Barnes, T.: Language to completion: success in an educational data mining massive open online class. In: Proceedings of 8th International Conference on Educational Data Mining Society, ERIC, pp. 8–11 (2015)
42. Chen, D., Socher, R., Manning, C.D., Ng, A.Y.: Neural tensor networks and semantic word vectors. *Comput. Sci. Comput. Lang. Cornell Univ. Libr.* 1–4 (2013)
43. Bowman, S.R.: Can recursive neural tensor networks learn logical reasoning? *Comput. Sci. Comput. Lang. Cornell Univ. Libr.* 1–10 (2014). arXiv: 1312.6192v4 [cs. CL]. Accessed 15 Feb 2014
44. Chen, X., Member, S., Vorvoreanu, M., Madhavan, K.: Mining social media data for understanding students' learning experiences. *IEEE Trans. Learn. Technol.* **7**, 246–259 (2014)
45. Peña-ayala, A.: Expert systems with applications educational data mining: a survey and a data mining-based analysis of recent works. *Expert Syst. Appl.* **5G**, 31 (2013)
46. Clow, D., Hall, W., Keynes, M.: MOOCs and the funnel of participation. In: Proceedings of the 3rd International Conference on Learning Analytics and Knowledge, pp. 185–189. ACM (2013)