Review

Sentiment Analysis of Students' Feedback with NLP and Deep Learning: A Systematic Mapping Study

Zenun Kastrati [1,*], Fisnik Dalipi [1], Ali Shariq Imran [2], Krenare Pireva Nuci [3] and Mudasir Ahmad Wani [2]

[1] Faculty of Technology, Linnaeus University, 351 95 Växjö, Sweden; fisnik.dalipi@lnu.se
[2] Faculty of Information Technology and Electrical Engineering, Norwegian University of Science & Technology (NTNU), 2815 Gjøvik, Norway; ali.imran@ntnu.no (A.S.I.); mudasir.a.wani@ntnu.no (M.A.W.)
[3] Faculty of Computer Science and Engineering, University for Business and Technology, 10000 Prishtine, Kosovo; krenare.pireva@ubt-uni.net
* Correspondence: zenun.kastrati@lnu.se; Tel.: +46-(0)-700-97-87-32

Abstract: In the last decade, sentiment analysis has been widely applied in many domains, including business, social networks and education. Particularly in the education domain, where dealing with and processing students' opinions is a complicated task due to the nature of the language used by students and the large volume of information, the application of sentiment analysis is growing yet remains challenging. Several literature reviews reveal the state of the application of sentiment analysis in this domain from different perspectives and contexts. However, the body of literature is lacking a review that systematically classifies the research and results of the application of natural language processing (NLP), deep learning (DL), and machine learning (ML) solutions for sentiment analysis in the education domain. In this article, we present the results of a systematic mapping study to structure the published information available. We used a stepwise PRISMA framework to guide the search process and searched for studies conducted between 2015 and 2020 in the electronic research databases of the scientific literature. We identified 92 relevant studies out of 612 that were initially found on the sentiment analysis of students' feedback in learning platform environments. The mapping results showed that, despite the identified challenges, the field is rapidly growing, especially regarding the application of DL, which is the most recent trend. We identified various aspects that need to be considered in order to contribute to the maturity of research and development in the field. Among these aspects, we highlighted the need of having structured datasets, standardized solutions and increased focus on emotional expression and detection.

Keywords: sentiment analysis; opinion mining; student feedback; user reviews; teacher assessment; educational platforms; MOOCs; natural language processing; text mining; deep learning; machine learning; polarity assessment; emotion recognition

1. Introduction

The present education system represents a landscape that is continuously enriched by a massive amount of data that is generated daily in various formats and most often hides useful and valuable information. Finding and extracting the hidden "pearls" from the ocean of educational data constitutes one of the great advantages that sentiment analysis and opinion mining techniques can provide. Sentiments and opinions expressed by students are a valuable source of information not only for analyzing students' behavior towards a course, topic, or teachers but also for reforming policies and institutions for their improvement. Although both sentiment analysis and opinion mining seem similar, there is a slight difference between the two: the former refers to finding sentiment words and phrases exhibiting emotions, whereas the latter refers to extracting and analyzing people's opinions for a given entity. For this study, we consider that both techniques are used interchangeably. The sentiment/opinion polarity, which could either be positive, negative, or neutral, represents one's attitude towards a target entity. Emotions, on the other hand,

are one's feelings expressed regarding a given topic. Since the 1960s, several theories about emotion detection and classification have been developed. The study conducted by Plutchik [1] categorizes emotions into eight categories: anger, anticipation, disgust, fear, joy, sadness, surprise, and trust.

Sentiment analysis can be conducted at a word, sentence, or a document level. However, due to the large number of documents, manual handling of sentiments is impractical. Therefore, automatic data processing is needed. Sentiment analysis from the text-based, sentence or document-level corpora is employed using natural language processing (NLP). Most research papers found in the literature published until 2016–2017 employed pure NLP techniques, including lexicon and dictionary-based approaches for sentiment analysis. Few of those papers used conventional machine learning classifiers. Recent years have seen a shift from pure NLP-based approaches to deep learning-based modeling in recognizing and classifying sentiment, and the number of papers published recently on the undertaken topic has increased significantly.

The popularity and importance of students' feedback have also increased recently, especially in the times of the COVID-19 pandemic, when most educational institutions have transcended traditional face-to-face learning to the online mode. Figure 1 shows the country-wise comparison breakdown of interest over the past six years in the use of sentiment analysis for analyzing students' attitudes towards teacher assessment.
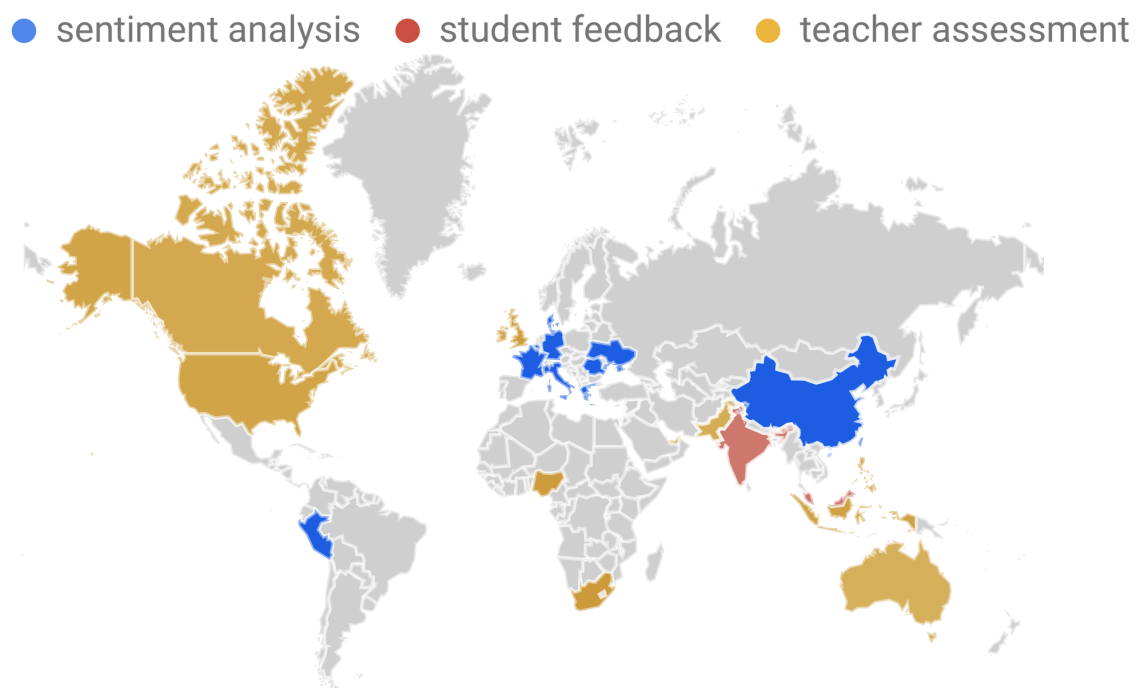

Figure 1. Country-wise comparison breakdown of interest over the past six years towards sentiment analysis, student's feedback, and teacher assessment.

The number of papers published recently indicates a growing interest towards the application of NLP/DL/ML solutions for sentiment analysis in the education domain. However, to the best of our knowledge, in order to establish the state of evidence, the body of literature is lacking a review that systematically classifies and categorizes research and results by showing the frequencies and visual summaries of publications, trends, etc. This gap in the body of literature necessitated a systematic mapping of the use of sentiment analysis to study students' feedback. Thus, this article aims to map how this research field is structured by answering research questions through a step-wise framework to conduct systematic reviews. In particular, we formulated multiple research questions that cover general issues regarding investigated aspects in sentiment analysis, models and

approaches, trends regarding evaluation metrics, bibliographic sources of publications in the field, and the solutions used, among others.

The main contributions of this study are as follows:

- A systematic map of 92 primary studies based on the PRISMA framework;
- An analysis of the investigated educational entities/aspects and bibliographical and research trends in the field;
- A classification of reviewed papers based on approaches, solutions, and data representation techniques with respect to sentiment analysis in the education domain;
- An overview of the challenges, opportunities, and recommendations of the field for future research exploration.

The rest of the paper is organized as follows. Section 2 provides some background information on sentiment analysis and related work, while Section 3 describes the search strategy and methodology adopted in conducting the study. Section 4 presents the systematic mapping study results. Challenges identified from the investigated papers are described in Section 5. Section 6 outlines recommendations and future research directions for the development of effective sentiment analysis systems. Furthermore, in Section 7, we highlight the potential threats to the validity of the results. Lastly, the conclusion is drawn in Section 8.

2. Sentiment Analysis and Related Work

2.1. Overview of Sentiment Analysis

Sentiment analysis is a task that focuses on polarity detection and the recognition of emotion toward an entity, which could be an individual, topic, and/or event. In general, the aim of sentiment analysis is to find users' opinions, identify the sentiments they express, and then classify their polarity into positive, negative, and neutral categories. Sentiment analysis systems use NLP and ML techniques to discover, retrieve, and distill information and opinions from vast amounts of textual information [2].

In general, there are three different levels at which sentiment analysis can be performed: the document level, sentence level, and aspect level. Sentiment analysis at the document level aims to identify the sentiments of users by analyzing the whole document. Sentence-level analysis is more fine-grained as the goal is to identify the polarity of sentences rather than the entire document. Aspect-level sentiment analysis focuses on identifying aspects or attributes expressed in reviews and on classifying the opinions of users towards these aspects.

As can be seen from Figure 2, the general architecture of a generic sentiment analysis system includes three steps [3]. Step 1 represents the input of a corpus of documents into the system in various formats. This is followed by the second step, which is document processing. At this step, the entered documents are converted to text and pre-processed by utilizing different linguistic tools, such as tokenization, stemming, PoS (Part of Speech) tagging, and entity and relation extraction. Here, the system may also use a set of lexicons and linguistic resources. The central component of the system architecture is the document analysis module (step 3) that also makes use of linguistic resources to annotate the pre-processed documents with sentiment annotations. Annotations represent the output of the system—i.e., positive, negative, or neutral—presented using a variety of visualization tools. Depending on the sentiment analysis form, annotations may be attached differently. For document-based sentiment analysis, the annotations may be attached to the entire documents; for sentence-based sentiments, the annotations may be attached to individual sentences; whereas for aspect-based sentiment, they are attached to specific topics or entities.
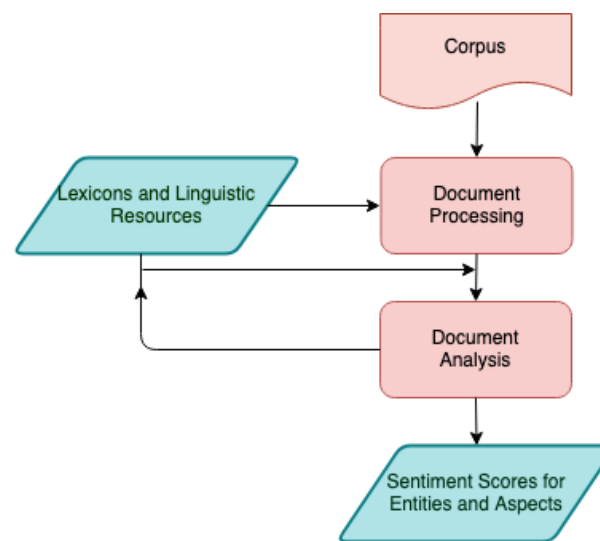
Figure 2. The architecture of a generic sentiment analysis system.

Sentiment analysis has been widely applied in different application domains, especially in business and social networks, for various purposes. Some well-known sentiment analysis business applications include product and services reviews [4], financial markets [5], customer relationship management [6], and marketing strategies and research [5], among others. Regarding social networks applications, the most common application of sentiment analysis is to monitor the reputation of a specific brand on Twitter or Facebook [7] and explore the reaction of people given a crisis; e.g., COVID-19 [8]. Another important application domain is in politics [9], where sentiment analysis can be useful for the election campaigns of candidates running for political positions.

Recently, sentiment analysis and opinion mining has also attracted a great deal of research attention in the education domain [2]. In contrast to the above-mentioned fields of business or social networks, which focus on a single stakeholder, the research on sentiment analysis in the education domain considers multiple stakeholders of education including teachers/instructors, students/learners, decision makers, and institutions. Specifically, sentiment analysis is mainly applied to improve teaching, management, and evaluation by analyzing learners' attitudes and behavior towards courses, platforms, institutions, and teachers.

From the learners' perspective, there are a number of papers [10–12] that have applied sentiment analysis to investigate the correlation of attitude and performance with learners' sentiments as well as the relationship between learners' sentiments and drop-out rates in Massive Open Online Courses (MOOCs). Regarding teachers' perspectives, sentiment analysis has been widely adopted by researchers [13–15] to examine various teacher-associated aspects expressed in students' reviews or comments in discussion forums. These aspects include teaching pedagogy, behavior, knowledge, assessment, and experience, to name a few. Sentiment analysis was also used in a number of studies [16,17] to analyze student's attitudes towards various aspects related to an institution; i.e., tuition fees, financial aid, housing, food, diversity, etc. Regarding courses, aspect-based sentiment analysis systems have been implemented to identify key aspects that play a critical role in determining the effectiveness of a course as discussed in students' reviews and then examine the attitudes and opinions of students towards these aspects. These aspects primarily include course content, course design, the technology used to deliver course content, and assessment, among others.

2.2. Related Work

Referring to past literature, we found that one study [18] on sentiment analysis (SA) in the education domain focused on detecting the approaches and resources used in SA and identifying the main benefits of using SA on education data. Our study is an extended

form of this article; thus a great deal of information is presented from different dimensions including bibliographical sources, research trends and patterns, and the latest tools used to perform SA. Instead of listing the data sources, we present the four categories of education-based data sources that are mostly used for SA. Furthermore, to increase convenience for researchers in this domain, we present groups of studies based on the learning approaches, most frequently used techniques, and most widely used education related lexicons for sentiment analysis.

Another review study [19] provided an overview of sentiment analysis techniques for education. The authors of this study provided a sentiment discovery and analysis (SDA) framework for multimodal fusions. Rather than the text, audio, and visual signals focused in [19], our review article aims to present all aspects related to the sentiment analysis of educational information with a focus on textual information only in a systematic way. Furthermore, we also provide a long list of current approaches employed for sentiment discoveries and the results obtained by them. Similarly, [20] aimed to review the scientific literature of SA on education data and revealed future research prospects in this direction. The authors of [20] focused on the area in more depth, including the design of sentiment analysis systems, the investigation of topics of concern for learners, the evaluation of teachers' teaching performance, etc., from almost 41 relevant research articles. In contrast, to conduct our scientific literature review study, we initially filtered 612 research articles from different journals and conferences. At the final stage of filtering, we finalized and included 92 of the most related and high-quality scientific articles published between 2015 to 2020 in this work. The main aim of this paper is to provide most of the available information regarding the sentiment analysis of educational information in a systematic way in a single place.

Review studies of this kind are greatly helpful for readers in this domain. This review study will assist researchers, academicians, practitioners, and educators who are interested in sentiment analysis with a classification of the approaches to the sentiment analysis of education data, different data sources, experimental results from different studies, etc.

3. Research Design

To conduct this study, we applied systematic mapping as the research methodology for reviewing the literature. Since this method requires an established search protocol and rigorous criteria for the screening and selection of the relevant publications, we utilized the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines, as indicated in [21]. The primary goal of a systematic mapping review (SMR) is to provide an overview of the body of knowledge and the research area and identify the amount of publications and the type of research and results available. Furthermore, an SMR aims to map the frequencies of publications over time to determine trends, forums or venues, and the relevant authors by which the research has been conducted and published. In contrast to the classical systematic literature review (SLR), which focuses on the identification of best practices based on empirical evidence, the focus of an SMR is on establishing the state of evidence. It is also worth mentioning that, from the methodology standpoint, SLR is characterized by narrow and specific research questions, and the studies are evaluated in detail regarding this quality. On the other hand, SMR deals with multiple broader research questions, and studies are not assessed based on details regarding the quality.

To ensure that all relevant studies were located and reviewed, our search strategy involved a stepwise PRISMA approach, consisting of four stages. The overall process of the search strategy is shown in Figure 3. The first stage in the PRISMA entailed the development of a research protocol by determining research questions, defining the search keywords, and identifying the bibliographic databases for performing the search. The second stage involved applying inclusion criteria, which was followed by stage three, in which the exclusion criteria were applied. The last stage was data extraction and analysis.
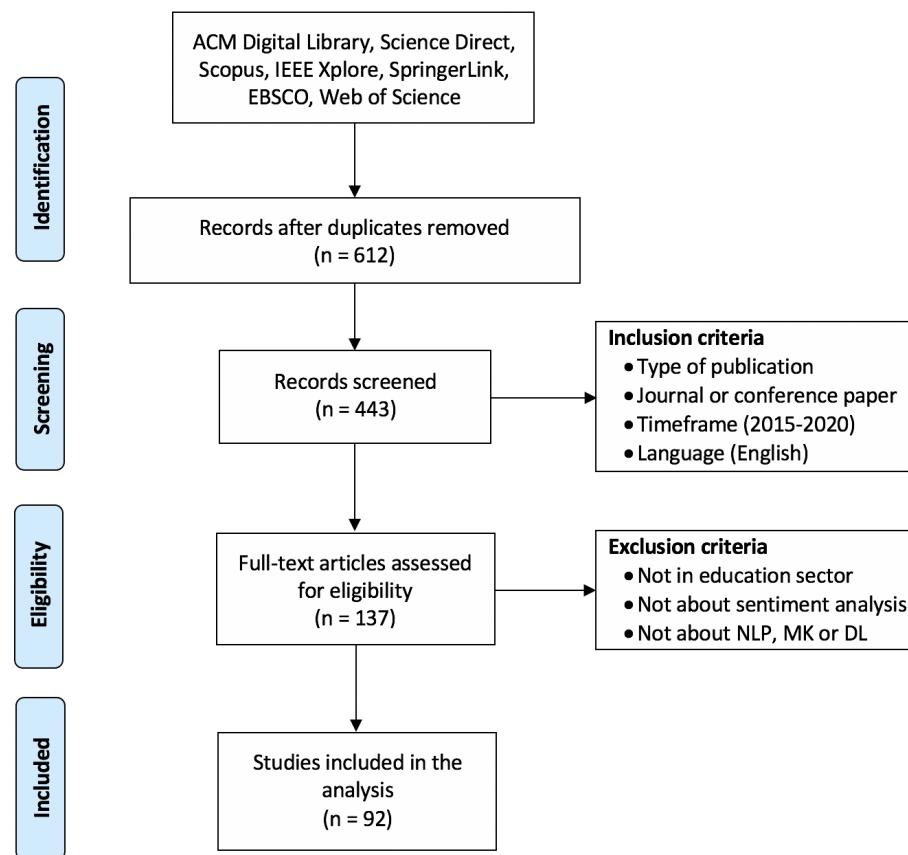
Figure 3. PRISMA search methodology.

The research questions (RQs) devised for this study were as follows:

- RQ1. What are the most investigated aspects in the education domain with respect to sentiment analysis?
- RQ2. Which approaches and models are widely studied for conducting sentiment analysis in the education domain?
- RQ3. What are the most widely used evaluation metrics to assess the performance of sentiment analysis systems?
- RQ4. In which bibliographical sources are these metrics published, and what are the research trends and patterns?
- RQ5. What are the most common sources used to collect students' feedback?
- RQ6. What are the solutions with respect to the packages, tools, frameworks, and libraries utilized for sentiment analysis?
- RQ7. What are the most common data representation techniques used for sentiment analysis?

3.1. Search Strategy

To develop a comprehensive set of search terms, we use the PICO(C) framework. PICO (Population, Intervention, Comparison and Outcomes) aims to help researchers to design a comprehensive set of search keywords for quantitative research in terms of population, intervention, comparison, outcome, and context [22]. As suggested by [23], to avoid missing possible relevant articles, we also added a "context" section to the PICO schema.

First, for all the sections of PICO(C) in Table 1, we identified the adequate keywords, and then we constructed the search string by applying binary operators, as shown in Table 2. To ensure that no possible relevant article would be omitted in the study, we also used the context criterion.

Table 1. PICO(C)-driven keyword framing.

Population	Students
Intervention (Investigation)	Sentiment analysis or opinion mining
Comparison	–
Outcome (What do we measure or evaluate?)	Students' feedback, opinion mining, sentiment analysis, teacher assessment, user feedback, feedback assessment
Context (In what context?)	MOOC, SPOC, distance learning, online learning, digital learning

Table 2. Search string (query).

Context	("MOOC" OR "SPOC" OR "distance learning" OR "online learning" OR "e-learning" OR "digital learning")
	AND
Intervention	("Sentiment analysis" OR "opinion mining")
	AND
Outcome	("Students' feedback" OR "teacher assessment" OR "user feedback" OR "feedback assessment" OR "students' reviews" OR "learners' reviews" OR "learners' feedback")

### 3.1.1. Time Period and Digital Databases

The time period selected for this study was from 2015 to 2020, inclusive. The research was conducted in 2020; therefore, it covered papers published until 30 September 2020.

For our search purposes, we used the following online research databases and engines:

- ACM Digital Library;
- IEEE Xplore;
- ScienceDirect;
- Scopus;
- SpringerLink;
- EBSCO; and
- Web of Science.

3.1.2. Identification of Primary Studies

As of September 2020, the search in Stage 1 yielded 612 papers without duplicates. In Figure 4, we present the total number of selected studies distributed per bibliographic database, identified during the first stage.
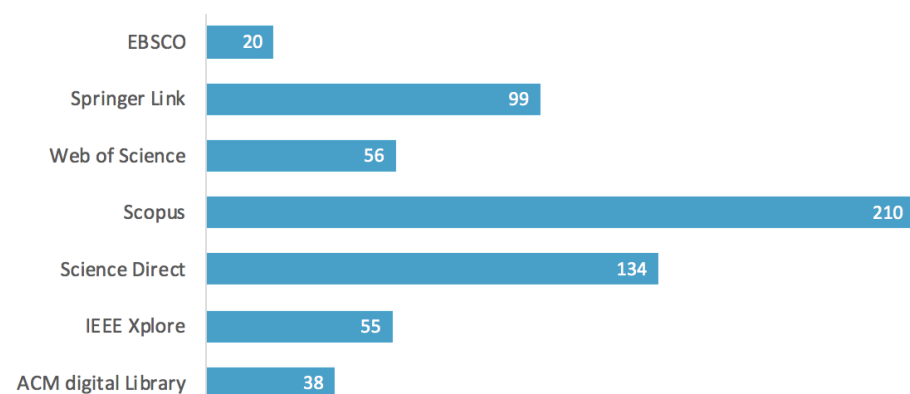
Figure 4. Studies collected from databases during stage 1.

3.2. Study Selection/Screening

Screening was stage 2 of the search strategy process and involved the application of inclusion criteria. At this stage, the relevant studies were selected based on the following criteria: (a) the type of publication needed to be a peer-reviewed journal or a conference paper, (b) papers needed to have been published between 2015 and 2020, and (c) papers needed to be in English. Besides, as can be seen in Figure 3, at this stage, we also checked the suitability of papers by examining the keywords, title, and the abstract of each paper. After we applied the mentioned criteria, out of 612 papers, 443 records were accepted as relevant studies for further exploration. Table 3 presents the screened and selected studies distributed according to year and database source.

Table 3. Selected and relevant studies extracted during stage 2.

Year	ACM DL	IEEE Xplore	Science Direct	Scopus	Web Science	SpringerLink	EBSCO	Total
2015	0	3	8	12	5	1	3	**32**
2016	1	7	11	12	11	2	2	**46**
2017	1	9	15	16	9	6	2	**58**
2018	0	10	18	25	10	13	2	**78**
2019	3	9	17	44	6	16	6	**101**
2020	22	10	30	33	9	21	3	**128**
Total	**27**	**48**	**99**	**142**	**50**	**59**	**18**	**443**

The distribution of conference and journal papers reviewed in this study is illustrated in Figure 5. As can be seen from the chart, there has been an increasing trend of research works published in journals in the last two years in contrast to the previous years, where most of the studies were published in conferences.
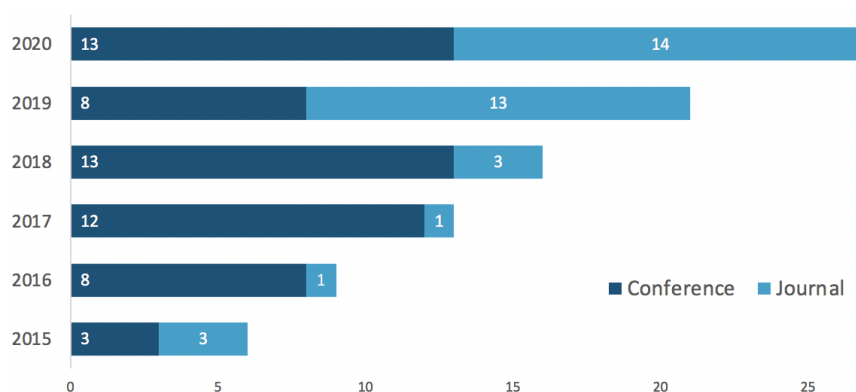

Figure 5. The number of collected conference and journal papers in 2015–2020.

3.3. Eligibility Criteria

In Stage 3, we applied the exclusion criteria in which we eliminated studies that were not (a) within the context of education, (b) about sentiment analysis, and (c) that did not employ the techniques of natural language processing, machine learning, or deep learning. At this stage, all the titles, abstracts, and keywords were also examined once more to determine the relevant records for the next stage. This stage resulted in 137 identified papers, which were divided among the four authors in equal number to proceed to the final stage. The authors agreed to encode the data using three different colors: (i) green—papers that passed the eligibility threshold, (ii) red—papers that did not pass the eligibility threshold, and (iii) yellow—papers that the authors were unsure which category to classify them as (green or red). The authors were located in three different countries, and the whole discussion was organized online. Initially, an online meeting was held to discuss the green and red list of papers, and then the main discussion was focused on papers listed in the

yellow category. For those papers, a thorough discussion among the involved authors took place, and once a consensus was reached, those papers were classified into either the green or red category. In the final stages, a fifth author was invited to increase the level of criticism of the discussion among the authors, to double-check all of the followed stages, and to be able to distinguish the current contribution from the previous ones.

After we applied these criteria, only 92 papers were considered for future investigation in the last stage of analysis.

4. Systematic Mapping Study Results

This section is divided into two parts: the first part presents the findings of the RQs, whereas the second highlights the relevant articles based upon the quality metrics.

4.1. Findings Concerning RQs

For the purposes of the analysis, the 92 papers remaining after the exclusion criteria were reviewed in detail by the five authors; in this section, the results are presented in the context of the research questions listed in Section 3.

RQ1. *What are the most investigated aspects in the education domain with respect to sentiment analysis?*

Students' feedback is an effective tool that provides valuable insights concerning various educational entities including teachers, courses, institutions, etc. and teaching aspects related to these entities. The identification of these aspects as expressed in the textual comments of students is of great importance as it aids decision makers to take the right action to specifically improve them. In this context, we examined and classified the reviewed papers based on the aspects that concerned students and that the authors aimed to investigate. In particular, we found three categories and their related teaching aspects which were objects of investigation in these papers: the first category comprised studies dealing with the comments of students concerning various aspects of the teacher entity, including the teacher's knowledge, pedagogy, behavior, etc; the second category contained papers concerning various aspects of the three different entities, such as courses, teachers, and institutions. Course-related aspects included dimensions such as course content, course structure, assessment, etc., whereas aspects associated to the institution entity were tuition fees, the campus, student life, etc.; the third category included papers dealing with capturing the opinions and attitudes of students toward institution entities. The findings illustrated in Figure 6 show that 81% of reviewed papers focused on extracting opinions, thoughts, and attitudes toward teachers, with 6% corresponding to institutions, whereas 13% presented a more general approach by investigating students' opinions toward teachers, courses, and institutions.

Figure 6. Feedback aspects investigated in the reviewed papers.

RQ2. *Which approaches and models are widely studied for conducting sentiment analysis in the education domain?*

Numerous approaches and models have been employed to conduct sentiment analysis in the education domain, which generally can be categorized into three groups. Table 4 shows the papers grouped based on learning approaches that the authors have applied within their papers. In total, 36 (out of 92) papers used a supervised learning approach, 8 used an unsupervised learning approach, and 20 used a lexicon-based approach.

Table 4. Papers grouped based on the learning approach.

Learning Approach	Papers
Supervised	[14,18,24–50]
Unsupervised	[51–53]
Lexicon-based	[15,54–67]
Supervised and unsupervised	[68–71]
Lexicon-based and supervised	[13,72–86]
Lexicon-based and unsupervised	[12,57,87–89]
Lexicon-based and unsupervised or supervised	[90–92]
N/A	[93–98]

Thus, seven papers used both supervised and unsupervised approaches. Twenty papers used lexicon-based and supervised learning, whereas seven papers used lexicon-based and unsupervised learning.

In total, three (out of 92) articles used all three learning approaches as a hybrid approach, in contrast with five other articles, which did not specify any learning approach.

Table 5 emphasizes that the Naive Bayes (NB) and Support Vector Machines (SVM) algorithms, as part of the supervised learning approach, were used most often in the reviewed studies, followed by Decision Tree (DT), k-Nearest Neighbor (k-NN) and Neural Network (NN) algorithms.

Table 5. Most frequently used algorithms as part of supervised learning.

Supervised Learning Algorithms	Papers
Support Vector Machines (SVM)	[12,18,25,26,28–31,33,35,36,39,42,55,68,71,72,75–78,80–85,90]
Naive Bayes (NB)	[12,25,26,28–30,32–43,55,56,69,71,72,74–80,82,83,85,86,90,91,93]
Decision Trees (DT)	[12,26,29,31,33,36,38,69,75,77,78,84]
k-Nearest Neighbor (k-NN)	[25,29,33,41,70,75,80,82,85,90]
Neural Networks (NN)	[12–14,24,28,33,41,55,73,77,90,95]

Furthermore, the use of a lexicon-based learning approach, also known as rule-based sentiment analysis, was common in a number of studies as shown in Table 4 and very often associated either with supervised or unsupervised learning approaches.

Table 6 lists the most frequently used lexicons elaborated among the reviewed articles, where the Valence Aware Dictionary and Sentiment Reasoner (VADER) and Sentiwordnet were used very often compared to TextBlob, MPQA, Sentistrength, and Semantria.

Table 6. Most frequently used lexicons.

Lexicon-Based	Papers
VADER	[55,60,62,68,99]
Sentiwordnet	[57,78,83,91]
TextBlob	[55,69]
MPQA	[42]
Sentistrength	[94]
Semantria	[61,79]

RQ3. *What are the most widely used evaluation metrics to assess the performance of sentiment analysis systems?*

Information retrieval-based evaluation metrics were widely used to assess the performance of systems developed for sentiment analysis. The metrics include the precision, recall, and F1-score. In addition to this, some studies employed statistical-based metrics to assess the accuracy of systems.

It is very interesting to depict the number of articles that used a specific evaluation metric to assess the performance of systems versus the number of articles that either did not perform any evaluation or decided not to emphasize the used metrics. Figure 7 illustrates the evaluation metrics used and emphasizes the percentage of articles defined for a particular metric.
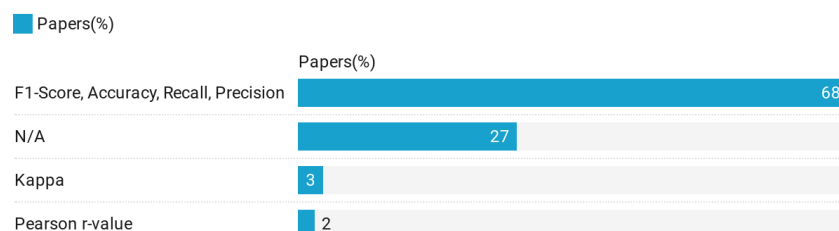

Figure 7. Evaluation metrics applied in the reviewed papers.

As can be seen from Figure 7, 68% of the articles included either only the F1-score or other evaluation metrics including the F1-score, precision, recall, and accuracy. Only 3% of the studies used Kappa, 2% used the Pearson r-value, and the remaining 27% did not specify any evaluation metrics.

RQ4. *In which bibliographical sources are the metrics published and what are the research trends and patterns?*

The publication trend during the review period included in this paper indicated that there was a variation regarding the distribution of publications across years and bibliographic resources. According to our findings, as illustrated in Figure 8, it is obvious that the majority of the papers were published during 2019, where Springer and IEEE were the most represented bibliographical sources. It is also interesting to note that during 2017, there were only three resources in which papers on sentiment analysis were published.

For a better overview, we present the absolute number of publications across years with the publishers' details in Table 7. This will assist readers to swiftly identify the time period and place of publication of the reviewed articles.

Table 7. Number of articles published between 2015 to 2020 by selected publishers.

Publisher	#Articles Published	Time Period
Elsevier	6	2015–2020
IEEE	41	2015–2020
ACM	6	2016–2020
Springer	17	2015–2020
Wiley	2	2018–2020
Ceur-WS	2	2018–2019
BEIESP, ArXiv	2 (each publisher)	2019
ET and ASR, Erudit, Techscience	1 (each publisher)	2020
Emerald, IAES, JUCS, Res. Trend, T. and Francis	1 (each publisher)	2019
RMI	1	2017
Hindawi, ACL Ant.	1 (each publisher)	2016
Ripublication, TUBITAK	1 (each publisher)	2015

Regarding the applied techniques, there were only two major categories of techniques used to conduct sentiment analysis in the education domain during 2015 and 2017: NLP and ML. The first efforts [12,32] towards applying DL were presented during 2018, as shown in Figure 9. Moreover, an increasing research pattern of DL application appeared in 2019 and 2020—especially during 2020, where an equal distribution of DL versus the other techniques can be observed.
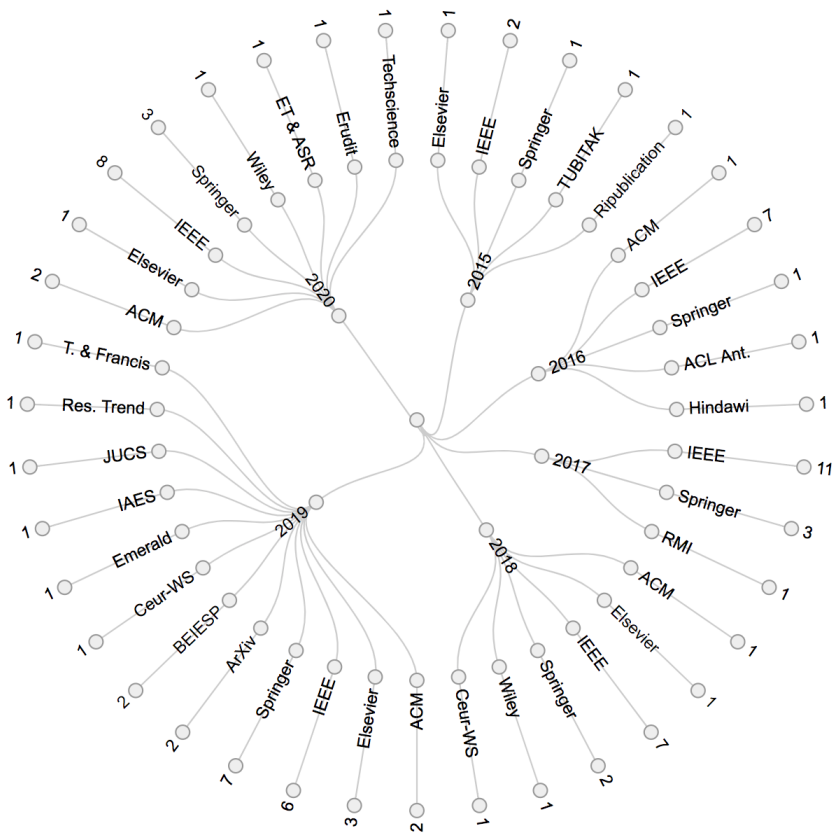
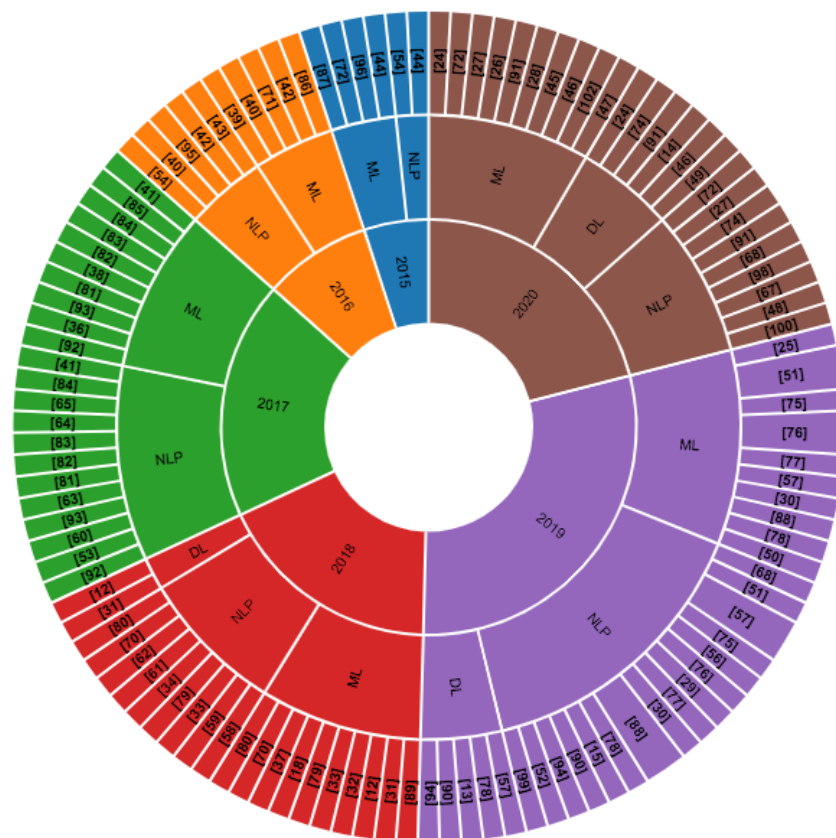**Figure 8.** Distribution of publications across years and bibliographic sources.

Figure 9. Techniques used for sentiment analysis across years.

RQ5. *What are the most common sources used to collect students' feedback?*

Based on the literature review in preparing this study, we came across several data sources, and based on their characteristics, we divided them into the four following categories for the convenience of our readers and the researchers working in this domain. The categories are as follows:

- Social media, blogs and forums: This category of datasets consists of data collected from online social networking and micro-blogging sites, discussion forums etc., such as Facebook and Twitter;
- Survey/questionnaires: This category comprises data that were mostly collected by conducting surveys among students and teachers or by providing questionnaires to collect feedback from the students;
- Education/research platforms: This category contains the data extracted from online platforms providing different courses such as Coursera, edX, and research websites such as ResearchGate, LinkedIn, etc.;
- Mixture of datasets: In this category, we grouped all those studies which used several datasets to conduct their experiments.

As can bee seen in Figure 10, there were only 64 (69.57%) papers that reported the sources from which the data were collected, whereas almost one-third of the papers failed to show any information regarding the sources of datasets. Table 8 shows papers that reported the sources of the datasets used for conducting experiments along with their corresponding categories and description.
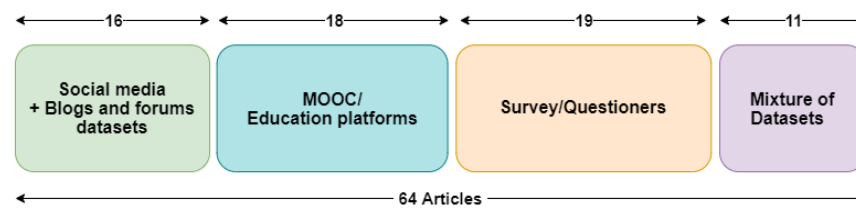

Figure 10. Categories of sources of the datasets.

Table 8. Sources of datasets used across reviewed papers.

Dataset Category	Papers	Description
Social media, blogs, and forums	[12,35,37,38,52,57,59,63,64,68,77,80,81, 87,89,93]	This category of datasets consists of data collected from online social networking and micro-blogging sites, discussion forums etc. such as Facebook and Twitter
Survey/questionnaire	[13,15,32,33,41,51,57,60,62,65,71,77,79, 83,89,94,96,100]	Here, the data were mostly collected by conducting surveys among students and teachers or by providing questioners to collect feedback from the students
Education/research platforms	[14,31,36,40,44–46,48,58,61,70,78,82,84,86,93,95,99,101]	This category contains the data extracted from online platforms providing different courses such as Coursera, edX, and research websites such as ResearchGate, LinkedIn, etc.
Mixture of datasets	[34,42,43,47,49,53,67,68,85,97,98]	In this category, we grouped all those studies which used several datasets to conduct their experiments

RQ6. *What are the solutions with respect to the packages, tools, frameworks and libraries utilized for sentiment analysis?*

Sentiment analysis is still a new field, and therefore there is no single solution/approach that dominates in sentiment analysis systems. In fact, there are dozens of solutions in terms of packages, frameworks, libraries, tools, etc. that are widely used across application

domains in general, and the education domain in particular. Figure 11 shows the findings of articles reviewed in this study with respect to the most commonly used packages, tools, libraries, etc. for the sentiment analysis task.
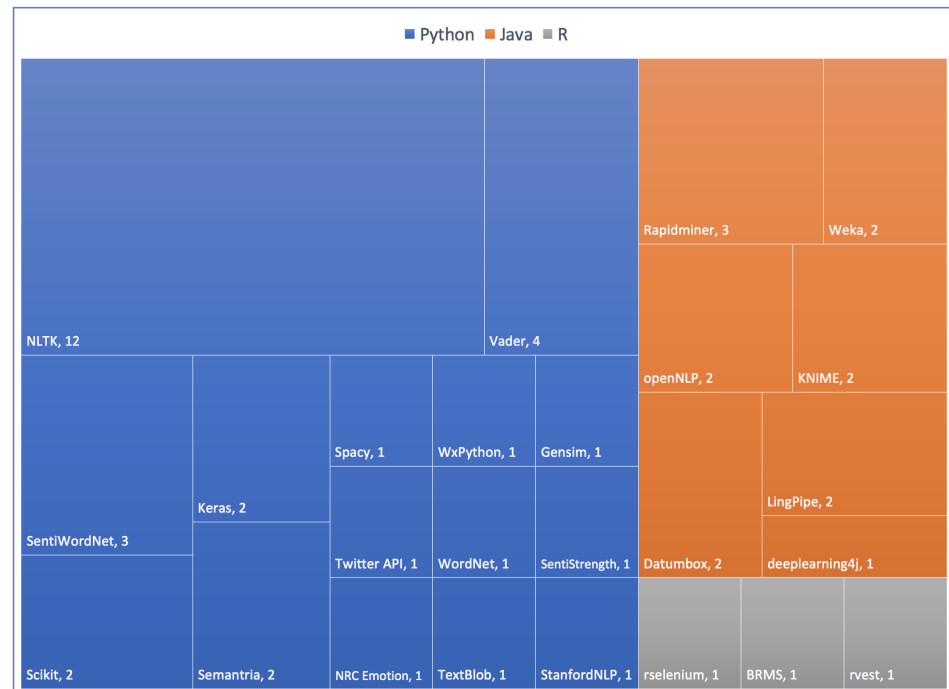

Figure 11. Packages/libraries/tools used to conduct sentiment analysis in the reviewed papers.

As shown in the Treemap illustrated in Figure 11, Python-based NLP and machine learning packages, libraries, and tools (colored in blue) are among the most popular solutions due to the open-source nature of the Python programming language. Specifically, the NLTK (Natural Language Toolkit) package is the dominant solution, and it was used in 12 different articles for pre-processing tasks including tokenizing, part-of-speech, normalization, the cleaning of text, etc.

Java-based NLP and machine learning packages, frameworks, libraries, and tools constitute the second group of solutions used for sentiment analysis. These solutions are colored in orange in Figure 11. Rapidminer is the most common Java-based framework and was used in three articles.

The third group is composed of NLP and machine learning solutions based on the R programming language. Only three studies used solutions in this group to conduct the sentiment analysis task.

RQ7. *What are the most common data representation techniques used for sentiment analysis?*

To provide our readers with more information on sentiment discoveries and analysis, we briefly present the commonly used word embedding techniques for the sentiment analysis task.

From the related reviewed articles, we observed that very few studies employed word embedding techniques to represent textual data collected from different sources. Only one article [48] employed the Word2Vec embedding model to learn the numeric representation and supply it as an input to the long short-term memory (LSTM) network. In addition to Word2Vec, GloVe and FastText models were used in two articles [14,45] to generate the embeddings for an input layer of CNN and compare the performance of the proposed aspect-based opinion mining system.

As presented above, word embedding techniques were seen in very few papers (3) out of all the references (92), particularly regarding sentiment analysis in the education domain for students' feedback. Therefore, more focus is needed to bridge this gap by incorporating

and testing different embedding techniques while analyzing the sentiment, emotion, or aspect of a student-related text.

4.2. Most Relevant Articles

To present the readers with a selection of the good-quality articles presented in this survey paper, we further narrowed down and short-listed 19 journal and conference articles. In particular, only articles published from 2018 to 2020 in Q1/Q2 level (https://www.scimagojr.com/journalrank.php) journals and A/B ranked (http://www.conferenceranks.com) conferences were identified as relevant, and these are summarized in Table 9.

Table 9. A summary of relevant articles.

Ref.	Year	Type	Techn.	Appr.	Models/Algorithms	Evaluation Metrics	Dataset	Rank
[73]	2020	J	NLP, DL	LB, Sup	Glove, LSTM	F1 = 83%, R = 78%, P = 90%, Acc = 86%	16,175 sentences	Q1
[24]	2020	J	ML, DL	Sup	NB, SVC, LSCV, RF, LSTM, CNN, CNN_LSTM, BERT, EvoMSA	Acc = 93%	24,552 opinions, 9712 opinions	Q1
[90]	2020	J	NLP, ML, DL	LB, UnS	w2v, tf*idf, GloVe, fastText, LDA2Vec, NB, SVM, LR, K-NN, RF, AdaBoost, Bagging, CNN, RNN, GRU, LSTM	F1 = 96%, Acc = 98.29%	154,000 reviews	Q1
[14]	2020	J	DL	Sup	LSTM, CNN	F1 = 86.13%	Coursera (104 K reviews)	Q1
[25]	2020	J	ML	Sup	NB, SVM, k-NN, GBT	F1 = 88%	Class central	Q1
[68]	2020	J	NLP	UnS	E-LDA, SVM, kMeans, tf*idf	F1 = 89%	Questionnaire (10 students)	Q1
[51]	2019	J	NLP, ML	UnS	LDA	N/A	Survey	Q1
[56]	2019	J	NLP, ML, DL	LB	SPPM + ID3, NB, SCM, BFTree, LR, BayeNEt, Stacking, AdaBoost	F1 = 93%, Acc = 88%, P = 92%, R = 97.5%	30,500 sentences	Q1
[87]	2019	J	NLP	LB, UnS	VADER, Topic Modeling, Ensemble LDA	F1 = 79.54%, P = 79.69%, R = 79.84%	Niche.com (100 K)	Q1
[13]	2019	J	DL	LB, Sup	Glove, LSTM	F1 = 86%, P = 88%, R = 85%, Acc = 93%	Questionnaire (5015)	Q1
[89]	2019	J	NLP	LB, UnS	Sentiment topic models-LDA	Acc = 86.5%	Feedback form (4895)	Q2
[51]	2019	J	NLP	UnS	LDA	N/A	Survey (2254)	Q1
[61]	2019	J	NLP	LB	Semantria	N/A	Survey	Q2
[12]	2018	C	ML, DL	LB, UnS	BiNB, BiSVM, LSTM, DT-LSTM, L-SVM, D-SVM, LD-SVM	F1 = 89.77%, Acc = 90.12%, Pearson = 0.095	RSelenium and rvest (36,646)	B
[32]	2018	C	ML	Sup	NB, ME	F1 = 87.94%	Survey (16,000)	B
[58]	2018	J	DL, ML	Sup	CNN, SVM	Acc = 76%, Kappa = 85%	Feedback form (73 reviews)	Q1
[60]	2018	C	NLP	LB	VADER	N/A	Survey (16,000)	B
[69]	2018	C	NLP, ML	UnS	DT, NB, GLM, CT, LDA	F1 = 79.3%, P = 67.5%, R = 96.2%	Questionnaire	B
[79]	2018	C	NLP, ML	LB, Sup	NB, ME	F1=87%	SFMS (5341)	B

Label: Techn: technique, Appr: approach, J: journal, C: conference, LB: lexicon based, Sup: supervised, UnS: unsupervised

Table 9 depicts pivotal aspects that were examined in the reviewed articles, including publication year and type, techniques, approaches, models/algorithms, evaluation metrics, and the sources and size of the datasets used to conduct the experiments. It can be seen that it is almost impossible to directly compare the articles in terms of performance due to the variety of algorithms/models and datasets applied to conduct the sentiment analysis task. However, it is interesting to note that the performance of sentiment analysis systems has generally improved over the years, achieving an accuracy of up to 98.29% thanks to the recent advancements of deep learning models and NLP representation techniques.

5. Identified Challenges and Gaps

Based on the systematic mapping study, we found that there is still a wide gap in some areas concerning the sentiment analysis of students' feedback that need further research and development. The following list shows some of the prominent issues, as presented in Table 10.

Table 10. Challenges linked to research questions.

Research Question	Identified Challenges
RQ1	Fine-grained sentiment analysis
RQ1	Figurative language
RQ2	Generalization
RQ2	Complex language constructs
RQ2	Representation techniques
RQ5	Scarcity of datasets
RQ5	Limited resources
RQ5	Unstructured format
RQ6	Unstandardized solutions/approaches

- Fine-grained sentiment analysis: Most studies have focused their attention on a complete review to determine a sentiment rather than going deeper into identifying fine-grained teaching/learning-related aspects and sentiments associated with them;
- Figurative language: Identifying figurative speech, such as sarcasm and irony, from student feedback text in particular is lacking and needs further exploration;
- Generalization: Most of the techniques are domain-specific and thus do not perform well in different domains;
- Complex language constructs: There is an incapability to handle complex language involving constructs such as double negatives, unknown proper names, abbreviations, and words with dual and multiple meanings;
- Representation techniques: There is a lack of research effort on the use of general-purpose word embedding as well as contextualized embedding approaches;
- Scarcity of publicly available benchmark datasets; there is a lack of benchmark datasets and an insufficient dataset size. Although there are a few open datasets available, there is no benchmark dataset that is useful for testing deep learning models due to the small number of samples those datasets provide;
- Limited resources: There is a lack of resources such as lexica, corpora, and dictionaries for low-resource languages (most of the studies were conducted in the English or Chinese language);
- Unstructured format: most of the datasets found in the studies discussed in this survey paper were unstructured. Identifying the key entities to which the opinions were directed is not feasible until an entity extraction model is applied, which makes the existing datasets' applicability very limited;
- Unstandardized solutions/approaches: We observed in this review study that a vast variety of packages, tools, frameworks, and libraries are applied for sentiment analysis.

6. Recommendations and Future Research Directions

This section provides various recommendations and proposals for suitable and effective systems that may assist in developing generalizable solutions for sentiment analysis in the education domain. We consider that the recommendations appropriately address the challenges identified in Section 5. An illustration of the proposed recommendations is given in Figure 12.
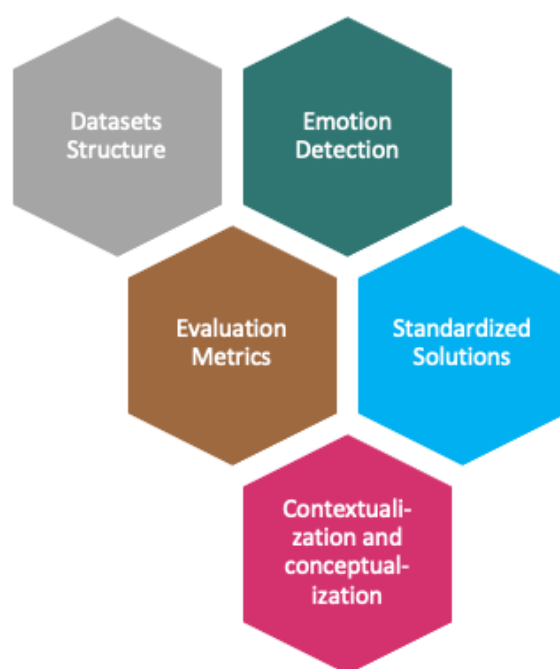
Figure 12. Recommendations for developing effective sentiment analysis systems.

6.1. Datasets Structure and Size

There is a need for a structured format to represent feedback datasets, whether they are captured at the sentence level or document level via a survey or a questionnaire form. A structured format in either an XML or a JSON file would be highly useful to standardize dataset generation for sentiment analysis in this domain. Furthermore, there is a need to associate the meta-data acquired at the time of the feedback responses. The meta-data would help to provide a descriptive analysis of the opinions expressed by a group of people for a given subject (aspect). Moreover, more than half (56.7%) of the datasets used in the reviewed papers were of a small-size, with merely 5000 samples or less, which affects the reliability and relevance of the results [102]. Additionally, most of these datasets are not publicly available, meaning that the results are not reproducible. Therefore, we recommend the collection of large-scale labeled datasets [14] to develop generalized deep learning models that could be utilized for various sentiment analysis tasks and for big data analysis in the education domain.

6.2. Emotion Detection

We found only a small number of articles focused on emotion detection. We feel that there is a greater need to take into consideration the emotions expressed in opinions to better identify and address the issues related towards the target subject, as has been investigated in many other text-based emotion detection works [103]. Furthermore, there are standard publicly available datasets such as ISEAR (https://www.kaggle.com/shrivastava/isears-dataset), and SemEval-2019 [104] that can be used to train deep learning models for text-based emotion detection tasks utilizing the Plutchik model [1] coupled with emoticons [8]. People often use emoticons to address emotions; thus, one aspect that researchers could explore is to make use of emoticons to identify the emotions expressed in an opinion.

6.3. Evaluation Metrics

Our study showed that researchers have used various evaluation metrics to measure the performance of sentiment analysis systems and models. Additionally, a considerable number of papers (27%) failed to report the information regarding the metrics used to assess the accuracy of the their systems. Therefore, we consider that a special focus and emphasis should be placed on including the utilized metrics in order to enhance the transparency of

the research results. Information retrieval evaluation metrics such as the precision, recall, and F1-score would be a good practice for the performance evaluation of sentiment analysis systems relying on imbalanced datasets. Accuracy would be another metric that could be used to evaluate the performance of systems trained on balanced datasets. Statistic metrics such as the Kappa statistic and Pearson correlation are other metrics that can be used to measure the correlation between the output of sentiment analysis systems and data labeled as ground truth. Moreover, this could help and benefit other researchers when conducting comprehensive and comparative performance analyses between different sentiment analysis systems.

6.4. Standardized Solutions

We have shown that the current landscape of sentiment analysis is characterized by a wide range of solutions that are yet to mature as the field is obviously novel and rapidly growing. These solutions were generally (programming) language-dependent and have been used to accomplish specific tasks—i.e., tokenizing, part-of-speech, etc.—in different scenarios. Thus, standardization will play an important role as a means for assuring the quality, safety, and reliability of the solutions and systems developed for sentiment analysis.

6.5. Contextualization and Conceptualization of Sentiment

Machine learning/deep learning approaches and techniques developed for sentiment analysis should pay more attention to embedding the semantic context using lexical resources such as Wordnet, SentiWordNet, and SenticNet, or semantic representation using ontologies [105] to capture users' opinions, thoughts, and attitudes from a text more effectively. In addition, state-of-the-art static and contextualized word embedding approaches such as fastText, GloVe, BERT, and ELMo should be further considered for exploration by researchers in this field as they have proven to perform well in other NLP-related tasks [106,107].

7. Potential Threats to Validity

There are several aspects that need to be taken into account when assessing this systematic mapping study as they can potentially limit the validity of the findings. These aspects include the following:

- The study includes papers collected from a set of digital databases, and thus we might have missed some relevant papers due to them not being properly indexed in those databases or having been indexed in other digital libraries;
- The search strategy was designed to search for papers using terms appearing in keywords, titles, and abstracts, and due to this, we may have failed to locate some relevant articles;
- Only papers that were written in English were selected in this study, and therefore some relevant papers that are written in other languages might have been excluded;
- The study relies on peer-reviewed journals and conferences and excludes scientific studies that are not peer-reviewed—i.e., book chapters and books. Furthermore, a few studies that conducted a systematic literature review were excluded as they would not provide reliable information for our research study;
- Screening based on the title, abstract, and keyword of papers was conducted at stage 2 to include the relevant studies. There are a few cases in which the relevance of an article cannot be judged by screening these three dimensions (title, abstract, keyword) and instead a full paper screening is needed; thus, it is possible that we might have excluded some papers with valid content due to this issue.

8. Conclusions

In the last decade, sentiment analysis enabled by NLP, machine learning, and deep learning techniques has also been attracting the attention of researchers in the educational domain in order to examine students' attitudes, opinions, and behavior towards numerous teaching aspects. In this context, we provided an analysis of the related literature by applying a systematic mapping study method. Specifically, in this mapping study, we selected 92 relevant papers and analyzed them with respect to different dimensions such as the investigated entities/aspects on the education domain, the most frequently used bibliographical sources, the research trends and patterns, what tools were utilized, and the most common data representation techniques used for sentiment analysis.

We have shown an overall increasing trend of publications investigating this topic throughout the studied years. In particular, there was a significant growth of articles published during the year 2020, where the DL techniques were mostly represented.

The mapping of the included articles showed that there is a diversity of interest from researchers on issues such as the approaches/techniques and solutions applied to develop sentiment analysis systems, evaluation metrics to assess the performance of the systems, and the variety of datasets with respect to their size and format.

In light of the findings highlighted by the body of knowledge, we have identified a variety of challenges regarding the application of sentiment analysis to examine students' feedback. Consequently, recommendations and future directions to address these challenges have been provided. We believe that this study's results will inspire future research and development in sentiment analysis applications to further understand students' feedback in an educational setting.

In future work, our plan is to further deepen the analysis that we performed in this mapping study by conducting systematic literature reviews (SLRs), as also suggested by [108].

References

1. Plutchik, R. The Nature of Emotions. *Am. Sci.* **2001**, *89*, 344–350. [CrossRef]
2. Cambria, E.; Schuller, B.; Xia, Y.; Havasi, C. New Avenues in Opinion Mining and Sentiment Analysis. *IEEE Intell. Syst.* **2013**, *28*, 15–21. [CrossRef]
3. Feldman, R. Techniques and Applications for Sentiment Analysis. *Commun. ACM* **2013**, *56*, 82–89. [CrossRef]
4. Yang, L.; Li, Y.; Wang, J.; Sherratt, R.S. Sentiment analysis for E-commerce product reviews in Chinese based on sentiment lexicon and deep learning. *IEEE Access* **2020**, *8*, 23522–23530. [CrossRef]
5. Carosia, A.; Coelho, G.P.; Silva, A. Analyzing the Brazilian financial market through Portuguese sentiment analysis in social media. *Appl. Artif. Intell.* **2020**, *34*, 1–19. [CrossRef]
6. Capuano, N.; Greco, L.; Ritrovato, P.; Vento, M. Sentiment analysis for customer relationship management: An incremental learning approach. *Appl. Intell.* **2020**, *50*, 1–14. [CrossRef]
7. Sharma, S.K.; Daga, M.; Gemini, B. Twitter Sentiment Analysis for Brand Reputation of Smart Phone Companies in India. In *Proceedings of ICETIT 2019*; Springer: Berlin/Heidelberg, Germany, 2020; pp. 841–852.
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* **2020**, *8*, 181074–181090. [CrossRef]
9. Chauhan, P.; Sharma, N.; Sikka, G. The emergence of social media data and sentiment analysis in election prediction. *J. Ambient. Intell. Humaniz. Comput.* **2020**, *11*, 1–27. [CrossRef]

10.   Wen, M.; Yang, D.; Rosé, C.P. Sentiment Analysis in MOOC Discussion Forums: What does it tell us? In Proceedings of the 7th International Conference on Educational Data Mining, EDM 2014, London, UK, 4–7 July 2014; pp. 130–137.

11.   Chaplot, D.S.; Rhim, E.; Kim, J. Predicting Student Attrition in MOOCs using Sentiment Analysis and Neural Networks. In Proceedings of the AIED Workshops, Madrid, Spain, 22–26 June 2015; Volume 53, pp. 54–57.

12.   Nguyen, V.D.; Van Nguyen, K.; Nguyen, N.L.T. Variants of Long Short-Term Memory for Sentiment Analysis on Vietnamese Students' Feedback Corpus. In Proceedings of the 2018 10th International Conference on Knowledge and Systems Engineering (KSE), Ho Chi Minh City, Vietnam, 1–3 November 2018; pp. 306–311.

13.   Sindhu, I.; Daudpota, S.M.; Badar, K.; Bakhtyar, M.; Baber, J.; Nurunnabi, M. Aspect-based opinion mining on student's feedback for faculty teaching performance evaluation. *IEEE Access* **2019**, *7*, 108729–108741. [CrossRef]

14.   Kastrati, Z.; Imran, A.S.; Kurti, A. Weakly supervised framework for aspect-based sentiment analysis on students' reviews of moocs. *IEEE Access* **2020**, *8*, 106799–106810. [CrossRef]

15.   Chauhan, G.S.; Agrawal, P.; Meena, Y.K. Aspect-based sentiment analysis of students' feedback to improve teaching–learning process. In *Information and Communication Technology for Intelligent Systems*; Springer: Berlin/Heidelberg, Germany, 2019; pp. 259–266.

16.   Moreno-Marcos, P.M.; Alario-Hoyos, C.; Muñoz-Merino, P.J.; Estévez-Ayres, I.; Kloos, C.D. Sentiment analysis in MOOCs: A case study. In Proceedings of the 2018 IEEE Global Engineering Education Conference (EDUCON), Santa Cruz de Tenerife, Spain, 17–20 April 2018; pp. 1489–1496.

17.   Bogdan, R.; Pop, N.; Holotescu, C. Using web 2.0 technologies for teaching technical courses. In *AIP Conference Proceedings*; AIP Publishing LLC: Melville, NY, USA, 2019; Volume 2071, p. 050003.

18.   Mite-Baidal, K.; Delgado-Vera, C.; Solís-Avilés, E.; Espinoza, A.H.; Ortiz-Zambrano, J.; Varela-Tapia, E. Sentiment Analysis in Education Domain: A Systematic Literature Review. In *International Conference on Technologies and Innovation*; Valencia-García, R., Alcaraz-Mármol, G., Del Cioppo-Morstadt, J., Vera-Lucio, N., Bucaram-Leverone, M., Eds.; Springer International Publishing: Berlin/Heidelberg, Germany, 2018; pp. 285–297.

19.   Han, Z.; Wu, J.; Huang, C.; Huang, Q.; Zhao, M. A review on sentiment discovery and analysis of educational big-data. *Wiley Interdiscip. Rev. Data Min. Knowl. Discov.* **2020**, *10*, e1328. [CrossRef]

20.   Zhou, J.; min Ye, J. Sentiment analysis in education research: A review of journal publications. *Interact. Learn. Environ.* **2020**, 1–13. [CrossRef]

21.   Moher, D.; Liberati, A.; Tetzlaff, J.; Altman, D.G. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *PLoS Med.* **2009**, *6*, e1000097. [CrossRef] [PubMed]

22.   Schardt, C.; Adams, M.B.; Owens, T.; Keitz, S.; Fontelo, P. Utilization of the PICO framework to improve searching PubMed for clinical questions. *BMC Med. Inform. Decis. Mak.* **2007**, *7*, 1–16. [CrossRef] [PubMed]

23.   Gianni, F.V.; Divitini, M. Technology-enhanced smart city learning: A systematic mapping of the literature. *Interact. Des. Archit. J.* **2016**, *27*, 28–43.

24.   Estrada, M.L.B.; Cabada, R.Z.; Bustillos, R.O.; Graff, M. Opinion mining and emotion recognition applied to learning environments. *Expert Syst. Appl.* **2020**, *150*, 113265. [CrossRef]

25.   Hew, K.F.; Hu, X.; Qiao, C.; Tang, Y. What predicts student satisfaction with MOOCs: A gradient boosting trees supervised machine learning and sentiment analysis approach. *Comput. Educ.* **2020**, *145*, 103724. [CrossRef]

26.   Giang, N.T.P.; Dien, T.T.; Khoa, T.T.M. Sentiment Analysis for University Students' Feedback. In *Future of Information and Communication Conference*; Springer: Berlin/Heidelberg, Germany, 2020; pp. 55–66.

27.   Nikolić, N.; Grljević, O.; Kovačević, A. Aspect-based sentiment analysis of reviews in the domain of higher education. *Electron. Libr.* **2020**, *38*, 44–64. [CrossRef]

28.   Katragadda, S.; Ravi, V.; Kumar, P.; Lakshmi, G.J. Performance Analysis on Student Feedback using Machine Learning Algorithms. In Proceedings of the 2020 6th International Conference on Advanced Computing and Communication Systems (ICACCS), Coimbatore, India, 6–7 March 2020; pp. 1161–1163.

29.   Kavitha, R. Sentiment Research on Student Feedback to Improve Experiences in Blended Learning Environments. *Int. J. Innov. Technol. Explor. Eng. (IJITEE)* **2019**, *8*. [CrossRef]

30.   Lalata, J.A.P.; Gerardo, B.; Medina, R. A Sentiment Analysis Model for Faculty Comment Evaluation Using Ensemble Machine Learning Algorithms. In Proceedings of the 2019 International Conference on Big Data Engineering, Hong Kong, China, 11–13 June 2019; pp. 68–73.

31.   Sultana, J.; Sultana, N.; Yadav, K.; AlFayez, F. Prediction of sentiment analysis on educational data based on deep learning approach. In Proceedings of the 2018 21st Saudi Computer Society National Computer Conference (NCC), Riyadh, Saudi Arabia, 25–26 April 2018; pp. 1–5.

32.   Van Nguyen, K.; Nguyen, V.D.; Nguyen, P.X.; Truong, T.T.; Nguyen, N.L.T. Uit-vsfc: Vietnamese students' feedback corpus for sentiment analysis. In Proceedings of the 2018 10th International Conference on Knowledge and Systems Engineering (KSE), Ho Chi Minh City, Vietnam, 1–3 November 2018; pp. 19–24.

33.   Spatiotis, N.; Perikos, I.; Mporas, I.; Paraskevas, M. Evaluation of an educational training platform using text mining. In Proceedings of the 10th Hellenic Conference on Artificial Intelligence, Patras, Greece, 9–12 July 2018; pp. 1–5.

34.   Aung, K.Z.; Myo, N.N. Lexicon Based Sentiment Analysis of Open-Ended Students' Feedback. *Int. J. Eng. Adv. Technol. (IJEAT)* **2018**, *8*, 1–6.

35. Esparza, G.G.; de Luna, A.; Zezzatti, A.O.; Hernandez, A.; Ponce, J.; Álvarez, M.; Cossio, E.; de Jesus Nava, J. A sentiment analysis model to analyze students reviews of teacher performance using support vector machines. In *International Symposium on Distributed Computing and Artificial Intelligence*; Springer: Berlin/Heidelberg, Germany, 2017; pp. 157–164.

36. Ibrahim, Z.M.; Bader-El-Den, M.; Cocea, M. A data mining framework for analyzing students' feedback of assessment. In Proceedings of the 13th European Conference on Technology Enhanced Learning Doctoral Consortium, Leeds, UK, 3 September 2018; p. 13.

37. Barrón-Estrada, M.L.; Zatarain-Cabada, R.; Oramas-Bustillos, R.; González-Hernández, F. Sentiment analysis in an affective intelligent tutoring system. In Proceedings of the 2017 IEEE 17th international conference on advanced learning technologies (ICALT), Timisoara, Romania, 3–7 July 2017; pp. 394–397.

38. Pong-Inwong, C.; Kaewmak, K. Improved sentiment analysis for teaching evaluation using feature selection and voting ensemble learning integration. In Proceedings of the 2016 2nd IEEE international conference on computer and communications (ICCC), Chengdu, China, 14–17 October 2016; pp. 1222–1225.

39. Ullah, M.A. Sentiment analysis of students feedback: A study towards optimal tools. In Proceedings of the 2016 International Workshop on Computational Intelligence (IWCI), Dhaka, Bangladesh, 12–13 December 2016; pp. 175–180.

40. Krishnaveni, K.; Pai, R.R.; Iyer, V. Faculty rating system based on student feedbacks using sentimental analysis. In Proceedings of the 2017 International Conference on Advances in Computing, Communications and Informatics (ICACCI), Udupi, India, 13–16 September 2017; pp. 1648–1653.

41. Koufakou, A.; Gosselin, J.; Guo, D. Using data mining to extract knowledge from student evaluation comments in undergraduate courses. In Proceedings of the 2016 International Joint Conference on Neural Networks (IJCNN), Vancouver, BC, Canada, 24–29 July 2016; pp. 3138–3142.

42. Terkik, A.; Prud'hommeaux, E.; Alm, C.O.; Homan, C.; Franklin, S. Analyzing gender bias in student evaluations. In Proceedings of the COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers, Osaka, Japan, 11–16 December 2016; pp. 868–876.

43. Tewari, A.S.; Saroj, A.; Barman, A.G. E-learning recommender system for teachers using opinion mining. In *Information Science and Applications*; Springer: Berlin/Heidelberg, Germany, 2015; pp. 1021–1029.

44. Ortega, M.P.; Mendoza, L.B.; Hormaza, J.M.; Soto, S.V. Accuracy'Measures of Sentiment Analysis Algorithms for Spanish Corpus generated in Peer Assessment. In Proceedings of the 6th International Conference on Engineering & MIS 2020, Larnaka, Cyprus, 9–11 June 2020; pp. 1–7.

45. Kastrati, Z.; Arifaj, B.; Lubishtani, A.; Gashi, F.; Nishliu, E. Aspect-Based Opinion Mining of Students' Reviews on Online Courses. In Proceedings of the 2020 6th International Conference on Computing and Artificial Intelligence, Tianjin, China, 23–26 April 2020; pp. 510–514.

46. Lwin, H.H.; Oo, S.; Ye, K.Z.; Lin, K.K.; Aung, W.P.; Ko, P.P. Feedback Analysis in Outcome Base Education Using Machine Learning. In Proceedings of the 2020 17th International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology (ECTI-CON), Phuket, Thailand, 24–27 June 2020; pp. 767–770.

47. Karunya, K.; Aarthy, S.; Karthika, R.; Deborah, L.J. Analysis of Student Feedback and Recommendation to Tutors. In Proceedings of the 2020 International Conference on Communication and Signal Processing (ICCSP), Chennai, India, 28–30 July 2020; pp. 1579–1583.

48. Kandhro, I.A.; Jumani, S.Z.; Ali, F.; Shaikh, Z.U.; Arain, M.A.; Shaikh, A.A. Performance Analysis of Hyperparameters on a Sentiment Analysis Model. *Eng. Technol. Appl. Sci. Res.* **2020**, *10*, 6016–6020. [CrossRef]

49. Asghar, M.Z.; Ullah, I.; Shamshirband, S.; Kundi, F.M.; Habib, A. Fuzzy-based sentiment analysis system for analyzing student feedback and satisfaction. *Preprints* **2019**. [CrossRef]

50. Mostafa, L. Student sentiment analysis using gamification for education context. In *International Conference on Advanced Intelligent Systems and Informatics*; Springer: Berlin/Heidelberg, Germany, 2019; pp. 329–339.

51. Cunningham-Nelson, S.; Baktashmotlagh, M.; Boles, W. Visualizing student opinion through text analysis. *IEEE Trans. Educ.* **2019**, *62*, 305–311. [CrossRef]

52. Sivakumar, M.; Reddy, U.S. Aspect based sentiment analysis of students opinion using machine learning techniques. In Proceedings of the 2017 International Conference on Inventive Computing and Informatics (ICICI), Coimbatore, India, 23–24 November 2017; pp. 726–731.

53. Nitin, G.I.; Swapna, G.; Shankararaman, V. Analyzing educational comments for topics and sentiments: A text analytics approach. In Proceedings of the 2015 IEEE Frontiers in Education Conference (FIE), El Paso, TX, USA, 21–24 October 2015; pp. 1–9.

54. Rajput, Q.; Haider, S.; Ghani, S. Lexicon-based sentiment analysis of teachers' evaluation. *Appl. Comput. Intell. Soft Comput.* **2016**, *2016*, 1–12. [CrossRef]

55. Cobos, R.; Jurado, F.; Blázquez-Herranz, A. A Content Analysis System that supports Sentiment Analysis for Subjectivity and Polarity detection in Online Courses. *IEEE Rev. Iberoam. Technol. Aprendiz.* **2019**, *14*, 177–187. [CrossRef]

56. Pong-Inwong, C.; Songpan, W. Sentiment analysis in teaching evaluations using sentiment phrase pattern matching (SPPM) based on association mining. *Int. J. Mach. Learn. Cybern.* **2019**, *10*, 2177–2186. [CrossRef]

57. Iram, A. Sentiment Analysis of Student's Facebook Posts. In *International Conference on Intelligent Technologies and Applications*; Springer: Berlin/Heidelberg, Germany, 2018; pp. 86–97.

58. Yu, L.C.; Lee, C.W.; Pan, H.; Chou, C.Y.; Chao, P.Y.; Chen, Z.; Tseng, S.; Chan, C.; Lai, K.R. Improving early prediction of academic failure using sentiment analysis on self-evaluated comments. *J. Comput. Assist. Learn.* **2018**, *34*, 358–365. [CrossRef]

59. Liu, Z.; Yang, C.; Peng, X.; Sun, J.; Liu, S. Joint exploration of negative academic emotion and topics in student-generated online course comments. In Proceedings of the 2017 International Conference of Educational Innovation through Technology (EITT), Osaka, Japan, 7–9 December 2017; pp. 89–93.

60. Newman, H.; Joyner, D. Sentiment analysis of student evaluations of teaching. In *International Conference on Artificial Intelligence in Education*; Springer: Berlin/Heidelberg, Germany, 2018; pp. 246–250.

61. Santos, C.L.; Rita, P.; Guerreiro, J. Improving international attractiveness of higher education institutions based on text mining and sentiment analysis. *Int. J. Educ. Manag.* **2018**, *32*, 431–447. [CrossRef]

62. Raj, A.G.S.; Ketsuriyonk, K.; Patel, J.M.; Halverson, R. What Do Students Feel about Learning Programming Using Both English and Their Native Language? In Proceedings of the 2017 International Conference on Learning and Teaching in Computing and Engineering (LaTICE), Hong Kong, China, 20–23 April 2017; pp. 1–8.

63. Aung, K.Z.; Myo, N.N. Sentiment analysis of students' comment using lexicon based approach. In Proceedings of the 2017 IEEE/ACIS 16th international conference on computer and information science (ICIS), Wuhan, China, 24–26 May 2017; pp. 149–154.

64. Rani, S.; Kumar, P. A sentiment analysis system to improve teaching and learning. *Computer* **2017**, *50*, 36–43. [CrossRef]

65. Bano, M.; Zowghi, D.; Kearney, M. Feature based sentiment analysis for evaluating the mobile pedagogical affordances of apps. In *IFIP World Conference on Computers in Education*; Springer: Berlin/Heidelberg, Germany, 2017; pp. 281–291.

66. Martínez-López, J.I.; Gonzalez, C.M.A. Evaluation of Decision-making Quality using Multi-attribute Decision Matrix and Design Thinking. In Proceedings of the 2020 IEEE Global Engineering Education Conference (EDUCON), Porto, Portugal, 27–30 April 2020; pp. 622–629.

67. Watkins, J.; Fabielli, M.; Mahmud, M. Sense: A student performance quantifier using sentiment analysis. In Proceedings of the 2020 International Joint Conference on Neural Networks (IJCNN), Glasgow, UK, 19–24 July 2020; pp. 1–6.

68. Elia, G.; Solazzo, G.; Lorenzo, G.; Passiante, G. Assessing learners' satisfaction in collaborative online courses through a big data approach. *Comput. Hum. Behav.* **2019**, *92*, 589–599. [CrossRef]

69. Pyasi, S.; Gottipati, S.; Shankararaman, V. SUFAT—An Analytics Tool for Gaining Insights from Student Feedback Comments. In Proceedings of the 2018 IEEE Frontiers in Education Conference (FIE), San Jose, CA, USA, 3–6 October 2018; pp. 1–9.

70. Lubis, F.F.; Rosmansyah, Y.; Supangkat, S.H. Experience in learners review to determine attribute relation for course completion. In Proceedings of the 2016 International Conference on ICT For Smart Society (ICISS), Surabaya, Indonesia, 20–21 July 2016; pp. 32–36.

71. Colace, F.; Casaburi, L.; De Santo, M.; Greco, L. Sentiment detection in social networks and in collaborative learning environments. *Comput. Hum. Behav.* **2015**, *51*, 1061–1067. [CrossRef]

72. Spatiotis, N.; Perikos, I.; Mporas, I.; Paraskevas, M. Sentiment Analysis of Teachers Using Social Information in Educational Platform Environments. *Int. J. Artif. Intell. Tools* **2020**, *29*, 2040004. [CrossRef]

73. Sangeetha, K.; Prabha, D. Sentiment analysis of student feedback using multi-head attention fusion model of word and context embedding for LSTM. *J. Ambient. Intell. Humaniz. Comput.* **2020**, *12*, 4117–4126. [CrossRef]

74. Soe, N.; Soe, P.T. Domain Oriented Aspect Detection for Student Feedback System. In Proceedings of the 2019 International Conference on Advanced Information Technologies (ICAIT), Yangon, Myanmar, 6–7 November 2019; pp. 90–95.

75. Hariyani, C.A.; Hidayanto, A.N.; Fitriah, N.; Abidin, Z.; Wati, T. Mining Student Feedback to Improve the Quality of Higher Education through Multi Label Classification, Sentiment Analysis, and Trend Topic. In Proceedings of the 2019 4th International Conference on Information Technology, Information Systems and Electrical Engineering (ICITISEE), Yogyakarta, Indonesia, 20–21 November 2019; pp. 359–364.

76. Lalata, J.A.P.; Gerardo, B.; Medina, R. A correlation analysis of the sentiment analysis scores and numerical ratings of the students in the faculty evaluation. In Proceedings of the 2nd International Conference on Artificial Intelligence and Pattern Recognition, Beijing, China, 16–18August 2019; pp. 140–144.

77. Sahu, G.T.; Dhyani, S. Dynamic Feature based Computational model of Sentiment Analysis to Improve Teaching Learning System. *Int. J. Emerg. Technol.* **2019**, *10*, 17–23.

78. Korkmaz, M. Sentiment analysis on university satisfaction in social media. In Proceedings of the 2018 Electric Electronics, Computer Science, Biomedical Engineerings' Meeting (EBBT), Istanbul, Turkey, 18–19 April 2018; pp. 1–4.

79. Muollo, K.; Basavaraj, P.; Garibay, I. Understanding Students' Online Reviews to Improve College Experience and Graduation Rates of STEM Programs at the Largest Post-Secondary Institution: A Learner-Centered Study. In Proceedings of the 2018 IEEE Frontiers in Education Conference (FIE), San Jose, CA, USA, 3–6 October 2018; pp. 1–7.

80. Abed, A.A.; El-Halees, A.M. Detecting subjectivity in staff perfomance appraisals by using text mining: Teachers appraisals of palestinian government case study. In Proceedings of the 2017 Palestinian International Conference on Information and Communication Technology (PICICT), Gaza, Palestine, 8–9 May 2017; pp. 120–125.

81. Nasim, Z.; Rajput, Q.; Haider, S. Sentiment analysis of student feedback using machine learning and lexicon based approaches. In Proceedings of the 2017 international conference on research and innovation in information systems (ICRIIS), Langkawi, Malaysia, 16–17 July 2017; pp. 1–6.

82. Kaur, W.; Balakrishnan, V.; Singh, B. Social Media Sentiment Analysis of Thermal Engineering Students for Continuous Quality Improvement in Engineering Education. *J. Mech. Eng.* **2017**, *4*, 263–272.

83. Mandal, L.; Das, R.; Bhattacharya, S.; Basu, P.N. Intellimote: A hybrid classifier for classifying learners' emotion in a distributed e-learning environment. *Turk. J. Electr. Eng. Comput. Sci.* **2017**, *25*, 2084–2095. [CrossRef]

84. Spatiotis, N.; Paraskevas, M.; Perikos, I.; Mporas, I. Examining the impact of feature selection on sentiment analysis for the Greek language. In *International Conference on Speech and Computer*; Springer: Berlin/Heidelberg, Germany, 2017; pp. 353–361.

85. Dhanalakshmi, V.; Bino, D.; Saravanan, A.M. Opinion mining from student feedback data using supervised learning algorithms. In Proceedings of the 2016 3rd MEC international conference on big data and smart city (ICBDSC), Muscat, Oman, 15–16 March 2016; pp. 1–5.

86. Kumar, A.; Jain, R. Sentiment analysis and feedback evaluation. In Proceedings of the 2015 IEEE 3rd International Conference on MOOCs, Innovation and Technology in Education (MITE), Amritsar, India, 1–2 October 2015; pp. 433–436.

87. Srinivas, S.; Rajendran, S. Topic-based knowledge mining of online student reviews for strategic planning in universities. *Comput. Ind. Eng.* **2019**, *128*, 974–984. [CrossRef]

88. Nguyen, P.X.; Hong, T.T.; Van Nguyen, K.; Nguyen, N.L.T. Deep learning versus traditional classifiers on Vietnamese students' feedback corpus. In Proceedings of the 2018 5th NAFOSTED Conference on Information and Computer Science (NICS), Ho Chi Minh City, Vietnam, 23–24 November 2018; pp. 75–80.

89. Nimala, K.; Jebakumar, R. Sentiment topic emotion model on students feedback for educational benefits and practices. *Behav. Inf. Technol.* **2019**, *40*, 311–319. [CrossRef]

90. Onan, A. Mining opinions from instructor evaluation reviews: A deep learning approach. *Comput. Appl. Eng. Educ.* **2020**, *28*, 117–138. [CrossRef]

91. Balachandran, L.; Kirupananda, A. Online reviews evaluation system for higher education institution: An aspect based sentiment analysis tool. In Proceedings of the 2017 11th International Conference on Software, Knowledge, Information Management and Applications (SKIMA), Malabe, Sri Lanka, 6–8 December 2017; pp. 1–7.

92. Gottipati, S.; Shankararaman, V.; Gan, S. A conceptual framework for analyzing students' feedback. In Proceedings of the 2017 IEEE Frontiers in Education Conference (FIE), Indianapolis, IN, USA, 18–21 October 2017; pp. 1–8.

93. Lubis, F.F.; Rosmansyah, Y.; Supangkat, S.H. Topic discovery of online course reviews using LDA with leveraging reviews helpfulness. *Int. J. Electr. Comput. Eng.* **2019**, *9*, 426. [CrossRef]

94. Jiranantanagorn, P.; Shen, H. Sentiment analysis and visualisation in a backchannel system. In Proceedings of the 28th Australian Conference on Computer-Human Interaction, Launceston, Australia, 29 November–2 December 2016; pp. 353–357.

95. Kaewyong, P.; Sukprasert, A.; Salim, N.; Phang, F.A. A correlation analysis between sentimental comment and numerical response in students' feedback. *ARPN J. Eng. Appl. Sci.* **2015**, *10*, 18054–18060.

96. Banan, T.; Sekar, S.; Mohan, J.N.; Shanthakumar, P.; Kandasamy, S. Analysis of student feedback by ranking the polarities. In Proceedings of the Second International Conference on Computer and Communication Technologies, Hyderabad, India, 24–26 July 2015; Springer: Berlin/Heidelberg, Germany, 2016; pp. 203–214.

97. Marcu, D.; Danubaanu, M. Sentiment Analysis from Students' Feedback: A Romanian High School Case Study. In Proceedings of the 2020 International Conference on Development and Application Systems (DAS), Suceava, Romania, 21–23 May 2020; pp. 204–209.

98. Mukhtar, N.; Khan, M.A. Effective lexicon-based approach for Urdu sentiment analysis. *Artif. Intell. Rev.* **2019**, *53*, 1–28. [CrossRef]

99. Lundqvist, K.; Liyanagunawardena, T.; Starkey, L. Evaluation of Student Feedback Within a MOOC Using Sentiment Analysis and Target Groups. *Int. Rev. Res. Open Distrib. Learn.* **2020**, *21*, 140–156. [CrossRef]

100. Balahadia, F.F.; Fernando, M.C.G.; Juanatas, I.C. Teacher's performance evaluation tool using opinion mining with sentiment analysis. In Proceedings of the 2016 IEEE Region 10 Symposium (TENSYMP), Bali, Indonesia, 9–11 May 2016; pp. 95–98.

101. Moharil, A.; Singh, S.; Dravid, Y.; Dharap, H.; Bhanuse, V. Integrated Feedback Analysis And Moderation Platform Using Natural Language Processing. In Proceedings of the 2020 Fourth International Conference on Inventive Systems and Control (ICISC), Coimbatore, India, 8–10 January 2020; pp. 872–877.

102. Kagklis, V.; Karatrantou, A.; Tantoula, M.; Panagiotakopoulos, C.T.; Verykios, V.S. A learning analytics methodology for detecting sentiment in student fora: A case study in Distance Education. *Eur. J. Open Distance E-Learn.* **2015**, *18*, 74–94. [CrossRef]

103. Acheampong, F.A.; Wenyu, C.; Nunoo-Mensah, H. Text-based emotion detection: Advances, challenges, and opportunities. *Eng. Rep.* **2020**, *2*, e12189. [CrossRef]

104. Chatterjee, A.; Narahari, K.N.; Joshi, M.; Agrawal, P. SemEval-2019 Task 3: EmoContext Contextual Emotion Detection in Text. In Proceedings of the 13th International Workshop on Semantic Evaluation, Minneapolis, MN, USA, 6–7 June 2019; Association for Computational Linguistics: Minneapolis, MN, USA, 2019; pp. 39–48.

105. Kastrati, Z.; Imran, A.S.; Yayilgan, S.Y. The impact of deep learning on document classification using semantically rich representations. *Inf. Process. Manag.* **2019**, *56*, 1618–1632. [CrossRef]

106. Kastrati, Z.; Imran, A.S.; Kurti, A. Integrating word embeddings and document topics with deep learning in a video classification framework. *Pattern Recognit. Lett.* **2019**, *128*, 85–92. [CrossRef]

107. Shuang, K.; Zhang, Z.; Loo, J.; Su, S. Convolution–deconvolution word embedding: An end-to-end multi-prototype fusion embedding method for natural language processing. *Inf. Fusion* **2020**, *53*, 112–122. [CrossRef]

108. Petersen, K.; Vakkalanka, S.; Kuzniarz, L. Guidelines for conducting systematic mapping studies in software engineering: An update. *Inf. Softw. Technol.* **2015**, *64*, 1–18. [CrossRef]