
MACHINE LEARNING AND DEEP LEARNING FOR SENTIMENT ANALYSIS OVER STUDENTS' REVIEWS: AN OVERVIEW STUDY

A PREPRINT

Ru Yang

Department of Computer Science
Norwegian University of Science & Technology (NTNU)
Norway
ruya@stud.ntnu.no

February 2, 2021

ABSTRACT

Now when the whole world is still under COVID-19 pandemic, many schools have transferred the teaching from physical classroom to online platforms. It is highly important for schools and online learning platforms to investigate the feedback to get valuable insights about online teaching process so that both platforms and teachers are able to learn which aspect they can improve to achieve better teaching performance. But handling reviews expressed by students would be a pretty laborious work if they were handled manually as well as it is unrealistic to handle large-scale feedback from e-learning platform. In order to address this problem, both machine learning algorithms and deep learning models are used in recent research to automatically process students' review getting the opinion, sentiment and attitudes expressed by the students. Such studies may play a crucial role in improving various interactive online learning platforms by incorporating automatic analysis of feedback. Therefore, we conduct an overview study of sentiment analysis in educational field presented in recent research, to help people grasp an overall understanding of the sentiment analysis research. Besides, according to the literature review, we identify three future directions that researchers can focus on in automatically feedback processing: high-level entity extraction, multi-lingual sentiment analysis, and handling of figurative language.

Keywords Sentiment Analysis · Students' feedback · Students' reviews · Natural language processing · Data mining · Deep learning · Machine learning

1 Introduction

As a result of the COVID-19 pandemic, many schools and universities have pivoted from traditional, in-person physical classes towards online courses. Technological development in recent years has led to improvements to the technology behind online courses, and as a result more students in developing countries and remote areas are able to take courses from top universities through their computer or mobile phone. This presents an opportunity for potentially reducing the education inequality across the globe[1]. There are already multiple platforms providing free online courses, such as Massive Open Online Courses (MOOC), Coursera, Khan Academy, Udemy and edX, providing lectures on a vast variety of subjects [2][3].

Many online learners do not primarily rely on online courses to complete their courses, but to enhance the effect of conventional learning techniques, meeting new classmates or to review specific topics [4] As such it is not appropriate to use completion rates as an indicator to evaluate the effectiveness of MOOCs [5] for learners with other needs than completing their course and earning a certificate [6]. The authors in [7] make the point that the effectiveness of MOOCs could be mischaracterized if completion rates are overemphasized. This also includes drop-out rates [8, 9] and other issues relating to the MOOCs [10].

It's therefore very important that learning institutions examine the feedback from students regarding their experiences with online learning platforms, so that both professors and the platforms themselves can learn which aspects need to be altered and improved. For example in [11] the paper indicates that feedback from students helped the platform to implement a co-creation process during the course of the projects life cycle. Additionally, professors can benefit from the feedback from students to help them understand student behavior and refine contents of the courses [12] [13].

Student feedback is usually structured to not only include close-ended questions, but also open questions allowing students to express their thoughts about various aspects of teaching [14]. It's important to examine students' sentiments about specific aspects of this feedback, as seeking out the opinions of others is a very common practice when it comes to decision making [15].

Often the amount of data from these student reviews will be so large that it would be impractical to manually process these reviews. There is also the potential for language barriers to complicate the task, for example with terms and abbreviations used by student-aged people on the web [16]. To analyze this sort of textual feedback accurately would require state-of-the-art technology, such as traditional machine learning or deep neural networks. Additionally, this task comes at a time when the practice of opinion mining is increasingly controversial.

On the one hand, the development of deep learning technology has made enormous contributions to the Natural Language Processing [17]. For deep learning models, there are several implementations of Neural Networks of Deep learning, such as the Convolutional Neural Networks (CNN), Recursive Neural Networks (RNN), Long Short-Term Memory Networks (LSTM), BERT and so on, which efficiently carry out the task of opinion mining. These deep learning models have been widely employed for opinion mining in various domains, including movies [18], social media platforms [19], e-commerce [20], eLearning [11], tourism [21], to name just a few. However, these sorts of deep learning models will usually need large amount of training data, to achieve competitive classification performance. Additionally, deep learning models usually have to pass through preprocessing steps as the sentences they work with are not structured for direct input of neural models. The preprocessing steps usually includes "tokenizing" and "normalizing". Tokenizing is to split sentences into smaller elements called tokens by using the common delimiters, and then converting the sentence attributes into a set of numeric attributes representing word occurrence information [22]. Normalizing is to replace words that have a similar meaning with a single word – for example could words like "studied", "studies", "studying" be replaced with the word "study". Stemming and lemmatization are two commonly used normalization techniques. On the other hand, there are multiple machine learning algorithms such as Bernoulli Naive Bayes, Support Vector Machine (SVM), LinearSCV, Random Forest, etc. Most of these algorithms can be found in the Python library scikit-learn. Before the data is fed into the neural network or machine learning algorithm, it is necessary to perform the pre-processing step in order to improve the performance of algorithms. Real-world statistics are often strident, incomplete and incompatible, it is important to pre-process and concentrate the data before the dataset can be used for machine learning [23].

Many research studies have been done on automatic processing of student reviews using both conventional machine learning algorithms as well as deep learning models [26-40]. Studies like these could play a vital role in improving various interactive [24] eLearning platforms [25], LMS, and MOOCs [26] by incorporating feedback from various sources into their automatic analysis. It will therefore be useful to perform an overview of recent research that has been done on the topics of different algorithms of machine learning and deep learning.

The rest of this article will be structured as following: Section 2 will present a review of the relevant literature on these topics. Section 3 discusses the challenges inherent to classification and the need to solve them in the future. In section 4 I will present possible avenues for future research. Section 5 will be discussion about the work.

2 Literature Review

In the past ten years, with the rapid development of computation capability of graphics, and more educational resources [27] and frameworks [28], there are more and more sentiment analysis research based on deep learning and traditional learning algorithms.

Sindhu et. [29] built a learning model with two LSTM layers to analyze the sentiment polarization of students' reviews. The two layers works as different classifiers. One layer is used for aspect extraction while another is for classifying the sentiment of the reviews as positive, negative or neutral. The aspects from the output of the first layer would be used as the input of the second layer. The public dataset comprising of restaurant reviews is used to test the model trained by students' reviews from reviews of physical classroom courses of the university. The results of the research indicate that the two-LSTM-layer model used in this paper gets 91% accuracy in extracting aspect and 93% in classifying sentiment of the students' reviews. For the public restaurant reviews, the model achieves 82% accuracy in extracting aspects and 85% accuracy in sentiment classification.

Type	Model
Deep Learning	LSTM_3b
	CNN_5a
	CNN_10a
	CNN_LSTM_7a
	BERT
Machine Learning	Multinomial NB
	KNN
	Decision Tree
	B4MSA
	Bernoulli NB
	SVC
	Linear SVC
	Random Forest
Evolutionary Algorithm	EvoMSA

Table 1: 14 algorithms used in the experiment of [16]

The scientific paper [14] compared the conventional machine learning and deep learning algorithms on sentiment analysis based on 21940 students' reviews from e-learning platform. During their experiments, they used 4 traditional machine learning algorithms: SVM, Naive Bayes, Boosting, Decision Tree, and built one 1D-CNN deep learning model to extract aspect and analyze sentiment. Their research results show that 1D-CNN achieved better performance on sentiment analysis getting 88.2% on F1 scores but traditional machine learning models are better at aspect extraction.

Anna et al. [30] conducted a survey containing several free text questions and got 204 answers from the survey. They classified the 204 students' feedback into 162 positive reviews and 42 negative reviews. In their experiment, the traditional machine learning algorithms Naive Bayes and K -nearest algorithms to predict the students' reviews are positive or negative. Besides, cosine similarity is utilized to measure the similarity and leave-one-out cross in order to validate the result. The authors compared the results from their research with the Recursive Neural Tensor Network (RNTN) [31] method. The paper [30] indicates that RNTN [31] has better performance in Precision but worse Recall and Accuracy.

Katragadda et al. [23] researched sentiment analysis using several supervised machine learning algorithms and one deep learning model in order to classify the feedback as positive, negative or neutral. Their dataset includes thirty thousand feedback containing anonymous personal information, reviews, and students' emotion. Also, the dataset is classified into two categories: linear dataset with same properties, and non-linear dataset. The article shows that machine learning models get better results on linear dataset. More specifically, Naive Bayes model gets 50% accuracy after the precision and recall are calculated inside of it. Besides, SVM algorithms achieves 60.8% accuracy and the deep learning model gets 88.2% accuracy much beyond the conventional machine learning algorithms.

In the research job [32] the big data mining framework were built to achieve the instant monitoring of the students' satisfaction on online learning platforms. The framework contains both Data Management and Data Analytic techniques aiming to fill the gap between limited focus of big data on educational fields and rapid development of big data techniques. The framework is made to be able to connect to various kinds of data sources easily. The discussion of the classes in forum and students' reviews are used as 2 main data sources used in this research. The significant part of the their framework is called Analysis Engine which is used to analyze sentiment, cluster, and classify the reviews. During the pre-processing steps, the text features are collected using the below TF-IDF [33] equation:

$$TF * \log\left(\frac{N}{DF}\right) \quad (1)$$

where TF is the number of keywords occurrences in the current processing file, N is the counting of keywords occurrences in all files, and DF is the counting of total files in the experiment. The dataset contains 15000 balanced textual reviews. Then the balanced dataset is fed into two machine learning models Linear SVM in order to train the machine learning models, and Cross Validation is used to validate the results. The statistical supervised learning algorithm (SVM) that "one-against-one" strategy has been used with a "max wins" selection strategy. When clustering the dataset, TF-IDF algorithm is implemented by the authors to transform textual reviews to number firstly. Then they the K-means models to extract clusters for survey form and feedback collected from the forum. Finally, the controlled experiment involving few e-learning students and a small test data is conducted showing the functionalities of the framework and its potential value. Besides, the authors point out the future direction of the research. On the one hand, the connection between sentiment of the lessons and students' final mark could be built. On the other hand, other information, e.g. login information, admitted lessons, as well as contents posted on social media could be incorporated.

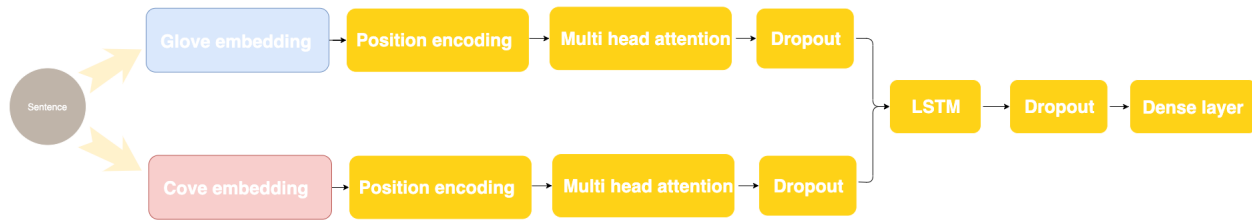


Figure 1: Multi-head attention model architecture

The research work in [6] investigated the factors influencing students' MOOC satisfaction and give us an extended general understanding of those factors. In their research, students' satisfaction is regarded as an crucial metric defining the success of MOOC. They classify the independent variables into learner-level and course-level variables and utilize those two aspects variables to predict students satisfaction. How those independent variables of student and course level affect the dependent variable, – MOOC students satisfaction , is evaluated by the authors. The dataset is downloaded from a public course website *Class Central* where people are able to download class metadata and feedback regarding the respective course. They collected feedback from 6391 students from this website. During the experiment, several traditional machine learning models, e.g k-nearest neighbors regression [34], gradient boosting trees [35], support vector machines [36], logistic regression [37], and naive Bayesian [38], are used to classify the emotion polarization. The algorithm achieving the best performance among all models was then chosen to predict aspect labels for the left part of unlabeled reviews. Their research results show that the gradient boosting tree got best results among all traditional machine learning models. As for the computation of sentiment polarity scores of the input text, the TextBlob3 is utilized that is a public free text processing software. The output scores for each review from TextBlob3 ranges from -1.0 to 1.0. The authors identified three crucial factors having statistically strong associations with learner satisfaction regarding learner sentiment in the conclusion part, which are content, assessment, as well as instructor. But there are no directive connections between course structure, video, and interactions and MOOC students' satisfaction [6]. There are two disadvantages of this sentiment analysis research. On the one hand, eighty percent of the feedback from the *Class Central* is written by those students who finished the whole courses. On the other hand, because of the intrinsic difference of the data the real randomness between the dependent variable –MOOC satisfaction, and those independent variables was not addressed.

Lwin et al. [22] conducted the research on not only open text reviews but also on rating scores. The dataset is gotten from the online survey form from the students of the university, which all the questions are just rating value questions except the last question that is free text based and used to collect feedback on classes and teachers. In textual comments analysis, the textual feedback were classified into two categories: negative, and positive, while rating scores were classified into five types: Worse, Bad, Neutral, Good and Excellent. The labeling work of the dataset utilizes the K-means clustering algorithms to pre-label the huge amount of the feedback data. Then the labeled dataset is fed into multiple conventional machine learning models. The six algorithms, – Logistic Regression, Multiplayer Perceptron, Simple Logistic Regression, Support Vector Machine, LMT and Ransom Forest, are selected to make comparisons in terms of performances. The situation is different for the sentiment analysis of textual comments. Firstly, people label each sentence as positive or negative manually, and conduct pre-processing steps on those reviews using the open-source library NLTK in order to analyze textual comments. The authors conclude that SVM gets the best results on rating score classification and Naive Bayers algorithms yields best performance for textual comment analysis.

The research work [16] conducted the experiments using 8 conventional machine learning models, 5 deep learning models and one evolutionary model. These fourteen algorithms they have used in their research have been presented in the Table 1. Two different kinds of dataset that is crawled from the HTML code of YouTube and other online learning platforms, is used: eduSERE and SentiTEXT. eduSERE is able to represent learning-center sentiments like engaged, excited, disappointed, and bored while the other is only with two polarities: *positive* and *negative*. The authors then build a sentimental dictionary connecting between words in text and emotion of the reviews. An algorithm based on the word count to classify the sentiment and the learning-centered emotion is proposed in order to pre-label the feedback they have collected. There are some reviews which were hard to classify and so were removed by the authors when checking the pre-label results. Finally, they choose the accuracy as the metrics to evaluate the performance of the models and algorithms. Their research presents that BERT and EvoMSA get better performance with 93% accuracy on SentiTEXT classification and descent accuracy of 84% and 83% on EduSERE classification. In the last, the integration between the models and an intelligent learning environment is performed by them. The intelligent learning environment is developed using Java. On the last part of the article, the authors concluded that genetic EvoMSA model have the best performance after adopting more knowledge and being optimized by macro-F1 aiming to solve the unbalanced dataset problem.

The researchers came up with a fusion deep learning model in [39] to analyze learners' reviews. The data set is the Vietnamese Student Feedback Corpus (UIT-VSFC) [40], which contains 16,000 reviews from Vietnamese students, which are then machine-translated into English for later use. The fusion model contains of a multi-head layer and an LSTM layer, and its structure is shown in the Fig.1. At the beginning, the feedback is fed into two different training embedding: Glove embedding and Cove embedding. Through embedding, the word information in the sentence would be changed to position information. Then use multiple attention blocks to calculate the weighted sum of multiple attentions rather than only concerning a single attention. The attention mechanism is used to assign weights to context words to find the words that determine the sentiment of the input sentence. Then the researchers utilize different dropout rates in order to avoid over-fitting problem, as well as combine the outputs of two different embedding models (1024 features) as the input of LSTM. Finally, through dropout and intensive layer, one of three emotions is classified: positive, negative, and neutral. The results show that the performance of the proposed multi-head attention fusion model is better than the single LSTM model, LSTM model + attention model and multi-head attention model.

3 Research Challenges

In the paper [22], the authors point out that many students would not to give directive negative reviews to the lessons or teacher even though there are difficulties with studying process. The learners are used to give many good feedback firstly and point out the shortcomings of the aspects at the end of the review. So many feedback may contain two aspects in only one review. e.g., considering such a review: *She is extremely good at the techniques of computer science but becomes impatient easily*. The above feedback evaluates both the teacher's knowledge and behavior. The authors in [29] indicate one of the reasons regarding misclassifications could be the occurrence of more than one aspect within one feedback because of the lack of the connectors e.g. "but", "and". So often this kind of review would be classified with only one aspect. This is a major setback when calculating the accuracy of the system. There are some public and free library such as OpenNLP that is able to break a review sentence on connectors. But when there were no any connectives such as "AND" & "BUT", those tools would not work well anymore. There is a research study [41] trying to use BERT to solve this multiple aspects problem. The model is called target-dependent Bert(TD-BERT) [41] using the the positioned output at the target terms instead of only the first tag since one review may have multiple aspects with their own background. After that, there is a max-pooling layer before the output is used in the next fully-connected layer. The D-BERT model is able to extract several aspects from one review at the same time and then analyze the sentiment of the feedback by combining predicted aspects. Their research study shows there is not too much value if just combining TD-BERT with complicated neural network, sometimes even leading to worse results than vanilla BERT-FC (fully connected). Also, the accuracy is improved when the aspects information is adopted. The study of [17, 42] indicates that it is not common for models paying much attention to semantics. We can see that BERT and other deep learning algorithms are not all good at natural language processing tasks, not much when it comes to natural language understanding¹. Such issues can be addressed employing ontologies, better vector space representation models [43, 44], and objective and semantic metrics [45, 46].

Also, the model that are able to handle multi-lingual reviews is absent and most current sentiment analysis model can only process reviews written in English. For example, the paper [22] points out that Myanmar language also appears in the students' feedback. Machine-translation translating all reviews to English is used in their paper but the sentiment is also lost a lot during translation process. The authors in [39] also just translated all the Vietnamese reviews into English by using machine translation. The above examples demonstrates there is a lack of effective model to handle multi language feedback which is still a research challenge that needs to be solved.

4 Future Directions

- Better automatic high-level entity extraction models [47]. For example, teacher, lesson, content and fine curriculum including lesson structure and teacher experience. As mentioned above, a comment may contain more than is already defined in the model. The authors of [47] point out that one of the main challenges of natural texts is to extract entities mentioned in texts, which can be named entities that refer to individuals or abstract concepts. In most study of sentiment analysis, the research conduct survey in advance to get all possible aspects and select the most common aspects occurring in the reviews. How to automatically extract important aspects is the focus of researchers in the future.
- Extend the most advanced model to include multilingual sentiment analysis. Most student reviews are not in English and may involve multiple languages. For convenience, many research just use machine translation

¹<https://medium.com/ontologik/time-to-put-an-end-to-bertology-or-ml-dl-is-not-even-relevant-to-nlu-e5ba6fc53403>

to translate all comments into English, and emotions may be lost in the translation process. Therefore, it is necessary to implement a multi-language processing model.

- Ability to handle figurative language such as sarcasm, irony. According to [48], irony and sarcasm are sophisticated forms of speech in which authors use the language signifying the opposite meaning. Even for humans, the recognizing irony can be quite difficult and complex, which leads to much misunderstanding in our daily life. It is a challenging work to detect sarcasm and irony for natural language processing, especially for sentiment analysis. Data mining activities and result can be misled by sarcastic sentences to wrong classification [49]. The authors in [50] built the model using millions of emoji occurrences to learn any-domain representations for detecting sentiment, emotion and sarcasm. In the future, the slang dictionary and the emoji can be integrated to the model to better detect sarcasm or irony to get more accurate sentiment analysis results. Able to deal with figurative language e.g. satire and satire. In the paper [48], irony and satire are a complex form of speech, and the language used by the author expresses the opposite meaning. Even for humans, understanding irony is quite difficult and complicated, which can lead to many misunderstandings in our life, work and study. In natural language processing, especially in sentiment analysis, how to find irony and irony is a challenging task. Sentiment analysis and data mining results can be misled into incorrect classifications by satirical statements. In the book [50], millions of emojis are utilized to build the model aiming to learn any domain notation to detect emotions, sentiments and sarcasm. In the future, slang dictionaries and emoticons can be adopted into the model to get better results in detecting irony.
- Addressing data imbalance problem by adopting GAN. Using GAN network to generate students' reviews is also a future direction to improve accuracy of sentiment analysis. Deep learning method usually requires huge amount of datasets while the datasets acquired from real world is generally unbalanced. For example, there may be more reviews about teacher or subject which may influence the accuracy of training result, or there are often more positive reviews while no enough negative and neutral reviews.

5 Conclusion

In this paper, we conducted a literature review study of sentiment analysis research using machine learning methods and deep learning methods. The research scope we investigated is restricted on educational field and based on students' reviews containing feedback both from physical classroom and online learning platform, such as MOOC.

In the beginning of the paper, we introduced the reasons why it is necessary to analyze the feedback of students for both conventional and online learning settings. It is an effective way to get effective suggestions for both teachers and platforms so that both can know where they need to improve to give better teaching experience, especially when the whole world is still under pandemic and many schools have transferred the physical teaching to online teaching. Then the common techniques of automatically processing reviews are also introduced including traditional machine learning and deep learning. For the machine learning methods, Bernoulli Naive Bayes, Support Vector Machine(SVM), Linear SCV, Random Forest, Decision Tree, K-means clustering are the common used algorithms to classify the sentiment of students' reviews. Most of these algorithms can be implemented easily using Python library scikit-learn. For the deep learning models, RNN, LSTM, CNN, BERT, Word Embeddings, TF-IDF, BERT are the common and popular techniques or models.

Finally, further research directions arising from this article concern the better models for automatic high-level entity extraction because one review often contains more than one entity. Besides, the ability to handle figurative language such as sarcasm, irony can significantly improve the sentiment analysis accuracy since figurative language signifies the opposite meaning and human sometimes even cannot recognize the irony correctly in sentences. Finally, in terms of multiple languages challenges, further analysis could be devoted to implement model that can handle reviews in multiple language since sentiment may be lost during translation by machine.

References

- [1] Jong-Wha Lee. Education for technology readiness: Prospects for developing countries. *Journal of human development*, 2(1):115–151, 2001.
- [2] Ali Shariq Imran and Faouzi Alaya Cheikh. Multimedia learning objects framework for e-learning. In *2012 International Conference on E-Learning and E-Technologies in Education (ICEEE)*, pages 105–109. IEEE, 2012.
- [3] Joi L Moore, Camille Dickson-Deane, and Krista Galyen. e-learning, online learning, and distance learning environments: Are they the same? *The Internet and Higher Education*, 14(2):129–135, 2011.

- [4] Dan Davis, Ioana Jivet, René F Kizilcec, Guanliang Chen, Claudia Hauff, and Geert-Jan Houben. Follow the successful crowd: raising mooc completion rates through social comparison at scale. In *Proceedings of the seventh international learning analytics & knowledge conference*, pages 454–463, 2017.
- [5] Fisnik Dalipi, Ali Shariq Imran, Florim Idrizi, and Hesat Aliu. An analysis of learner experience with moocs in mobile and desktop learning environment. In *Advances in human factors, business management, training and education*, pages 393–402. Springer, 2017.
- [6] Khe Foon Hew, Xiang Hu, Chen Qiao, and Ying Tang. What predicts student satisfaction with moocs: A gradient boosting trees supervised machine learning and sentiment analysis approach. *Computers & Education*, 145:103724, 2020.
- [7] Isaac Chuang and Andrew Ho. Harvardx and mitx: Four years of open online courses–fall 2012–summer 2016. Available at SSRN 2889436, 2016.
- [8] Fisnik Dalipi, Ali Shariq Imran, and Zenun Kastrati. MOOC dropout prediction using machine learning techniques: Review and research challenges. In *2018 IEEE Global Engineering Education Conference (EDUCON)*, pages 1007–1014. IEEE, 2018.
- [9] Ali Shariq Imran, Fisnik Dalipi, and Zenun Kastrati. Predicting student dropout in a MOOC: An evaluation of a deep neural network model. In *Proceedings of the 2019 5th International Conference on Computing and Artificial Intelligence*, pages 190–195, 2019.
- [10] Fisnik Dalipi, Sule Yildirim Yayilgan, Ali Shariq Imran, and Zenun Kastrati. Towards understanding the MOOC trend: pedagogical challenges and business opportunities. In *International Conference on Learning and Collaboration Technologies*, pages 281–291. Springer, 2016.
- [11] Zenun Kastrati, Ali Shariq Imran, and Arianit Kurti. Weakly supervised framework for aspect-based sentiment analysis on students’ reviews of MOOCs. *IEEE Access*, 8:106799–106810, 2020.
- [12] Anna D Rowe. Feelings about feedback: the role of emotions in assessment for learning. In *Scaling up assessment for learning in higher education*, pages 159–172. Springer, 2017.
- [13] Anna D Rowe and Julie Fitness. Understanding the role of negative emotions in adult learning and achievement: A social functional perspective. *Behavioral sciences*, 8(2):27, 2018.
- [14] Zenun Kastrati, Blend Arifaj, Arianit Lubishtani, Fitim Gashi, and Engjëll Nishliu. Aspect-based opinion mining of students’ reviews on online courses. *Proceedings of the 2020 6th International Conference on Computing and Artificial Intelligence*, 2020.
- [15] Erik Cambria. Affective computing and sentiment analysis. *IEEE intelligent systems*, 31(2):102–107, 2016.
- [16] María Lucía Barrón Estrada, Ramón Zatarain Cabada, Raúl Oramas Bustillos, and Mario Graff. Opinion mining and emotion recognition applied to learning environments. *Expert Systems with Applications*, 150:113265, 2020.
- [17] Zenun Kastrati, Ali Shariq Imran, and Sule Yildirim Yayilgan. The impact of deep learning on document classification using semantically rich representations. *Information Processing & Management*, 56(5):1618–1632, 2019.
- [18] Cicero Dos Santos and Maira Gatti. Deep convolutional neural networks for sentiment analysis of short texts. In *Proceedings of COLING 2014, the 25th International Conference on Computational Linguistics: Technical Papers*, pages 69–78, 2014.
- [19] Ali Shariq Imran, Sher Muhammad Daudpota, Zenun Kastrati, and Rakhi Batra. Cross-cultural polarity and emotion detection using sentiment analysis and deep learning on covid-19 related tweets. *IEEE Access*, 8:181074–181090, 2020.
- [20] Satuluri Vanaja and Meena Belwal. Aspect-level sentiment analysis on e-commerce data. In *2018 International Conference on Inventive Research in Computing Applications (ICIRCA)*, pages 1275–1279. IEEE, 2018.
- [21] Muhammad Afzaal, Muhammad Usman, and Alvis Fong. Tourism mobile app with aspect-based sentiment classification framework for tourist reviews. *IEEE Transactions on Consumer Electronics*, 65(2):233–242, 2019.
- [22] Htar Htar Lwin, Sabai Oo, Kyaw Zaw Ye, Kyaw Kyaw Lin, Wai Phyo Aung, and Phyo Paing Ko. Feedback analysis in outcome base education using machine learning. In *2020 17th International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology (ECTI-CON)*, pages 767–770. IEEE, 2020.
- [23] Sharnitha Katragadda, Varshitha Ravi, Prasanna Kumar, and G Jaya Lakshmi. Performance analysis on student feedback using machine learning algorithms. In *2020 6th International Conference on Advanced Computing and Communication Systems (ICACCS)*, pages 1161–1163. IEEE, 2020.

- [24] Ali Shariq Imran and Stewart James Kowalski. HIP—a technology-rich and interactive multimedia pedagogical platform. In *International Conference on Learning and Collaboration Technologies*, pages 151–160. Springer, 2014.
- [25] Ali Shariq Imran, Krenare Pireva, Fisnik Dalipi, and Zenun Kastrati. An analysis of social collaboration and networking tools in elearning. In *International Conference on Learning and Collaboration Technologies*, pages 332–343. Springer, 2016.
- [26] Krenare Pireva, Ali Shariq Imran, and Fisnik Dalipi. User behaviour analysis on LMS and MOOC. In *2015 IEEE Conference on e-Learning, e-Management and e-Services (IC3e)*, pages 21–26. IEEE, 2015.
- [27] Zenun Kastrati, Arianit Kurti, and Ali Shariq Imran. WET: Word embedding-topic distribution vectors for MOOC video lectures dataset. *Data in Brief*, 28:105090, 2020.
- [28] Zenun Kastrati, Ali Shariq Imran, and Arianit Kurti. Integrating word embeddings and document topics with deep learning in a video classification framework. *Pattern Recognition Letters*, 128:85–92, 2019.
- [29] Irum Sindhu, Sher Muhammad Daudpota, Kamal Badar, Maheen Bakhtyar, Junaid Baber, and Mohammad Nurunnabi. Aspect-based opinion mining on student’s feedback for faculty teaching performance evaluation. *IEEE Access*, 7:108729–108741, 2019.
- [30] Anna Koufakou, Justin Gosselin, and Dahai Guo. Using data mining to extract knowledge from student evaluation comments in undergraduate courses. In *2016 International Joint Conference on Neural Networks (IJCNN)*, pages 3138–3142. IEEE, 2016.
- [31] Richard Socher, Alex Perelygin, Jean Wu, Jason Chuang, Christopher D Manning, Andrew Y Ng, and Christopher Potts. Recursive deep models for semantic compositionality over a sentiment treebank. In *Proceedings of the 2013 conference on empirical methods in natural language processing*, pages 1631–1642, 2013.
- [32] Gianluca Elia, Gianluca Solazzo, Gianluca Lorenzo, and Giuseppina Passiante. Assessing learners’ satisfaction in collaborative online courses through a big data approach. *Computers in Human Behavior*, 92:589–599, 2019.
- [33] Jin-Cheon Na, Haiyang Sui, Christopher SG Khoo, Syin Chan, and Yunyun Zhou. Effectiveness of simple linguistic processing in automatic sentiment classification of product reviews. 2004.
- [34] Naomi S Altman. An introduction to kernel and nearest-neighbor nonparametric regression. *The American Statistician*, 46(3):175–185, 1992.
- [35] Jerome H Friedman. Greedy function approximation: a gradient boosting machine. *Annals of statistics*, pages 1189–1232, 2001.
- [36] Corinna Cortes and Vladimir Vapnik. Support vector machine. *Machine learning*, 20(3):273–297, 1995.
- [37] David R Cox. The regression analysis of binary sequences. *Journal of the Royal Statistical Society: Series B (Methodological)*, 20(2):215–232, 1958.
- [38] Kevin P Murphy. *Machine learning: a probabilistic perspective*. MIT press, 2012.
- [39] K Sangeetha and D Prabha. Sentiment analysis of student feedback using multi-head attention fusion model of word and context embedding for lstm. *Journal of Ambient Intelligence and Humanized Computing*, pages 1–10, 2020.
- [40] Kiet Van Nguyen, Vu Duc Nguyen, Phu XV Nguyen, Tham TH Truong, and Ngan Luu-Thuy Nguyen. Uit-vsfc: Vietnamese students’ feedback corpus for sentiment analysis. In *2018 10th International Conference on Knowledge and Systems Engineering (KSE)*, pages 19–24. IEEE, 2018.
- [41] Zhengjie Gao, Ao Feng, Xinyu Song, and Xi Wu. Target-dependent sentiment classification with BERT. *IEEE Access*, 7:154290–154299, 2019.
- [42] Ali Shariq Imran, Laksmi Rahadiani, Faouzi Alaya Cheikh, and Sule Yildirim Yayilgan. Semantic tags for lecture videos. In *2012 IEEE Sixth International Conference on Semantic Computing*, pages 117–120. IEEE, 2012.
- [43] Zenun Kastrati and Ali Shariq Imran. Performance analysis of machine learning classifiers on improved concept vector space models. *Future Generation Computer Systems*, 96:552–562, 2019.
- [44] Zenun Kastrati and Ali Shariq Imran. Adaptive concept vector space representation using markov chain model. In *International Conference on Knowledge Engineering and Knowledge Management*, pages 203–208. Springer, 2014.
- [45] Zenun Kastrati, Ali Shariq Imran, and Sule Yildirim-Yayilgan. SEMCON: a semantic and contextual objective metric for enriching domain ontology concepts. *International Journal on Semantic Web and Information Systems (IJSWIS)*, 12(2):1–24, 2016.

- [46] Zenun Kastrati, Ali Shariq Imran, and Sule Yildirim Yayilgan. SEMCON: semantic and contextual objective metric. In *Proceedings of the 2015 IEEE 9th International Conference on Semantic Computing (IEEE ICSC 2015)*, pages 65–68. IEEE, 2015.
- [47] Tareq Al-Moslmi, Marc Gallofré Ocaña, Andreas L Opdahl, and Csaba Veres. Named entity extraction for knowledge graphs: A literature overview. *IEEE Access*, 8:32862–32881, 2020.
- [48] DI Hernández Farias and Paolo Rosso. Irony, sarcasm, and sentiment analysis. In *Sentiment Analysis in Social Networks*, pages 113–128. Elsevier, 2017.
- [49] Anukarsh G Prasad, S Sanjana, Skanda M Bhat, and BS Harish. Sentiment analysis for sarcasm detection on streaming short text data. In *2017 2nd International Conference on Knowledge Engineering and Applications (ICKEA)*, pages 1–5. IEEE, 2017.
- [50] Bjarke Felbo, Alan Mislove, Anders Søgaard, Iyad Rahwan, and Sune Lehmann. Using millions of emoji occurrences to learn any-domain representations for detecting sentiment, emotion and sarcasm. *arXiv preprint arXiv:1708.00524*, 2017.