A Review Paper on the Role of Sentiment Analysis in Quality Education

Pooja¹ · Rajni Bhalla²

Received: 16 April 2022 / Accepted: 4 August 2022
© The Author(s), under exclusive licence to Springer Nature Singapore Pte Ltd 2022

Abstract
Education is a critical indication of progress and a major factor in well-being. The UN’s Sustainable Development Goals establish specific requirements for increasing educational quality and protecting the well-being of children. UN’s agenda for Sustainable Development Goal 4 which aims to “ensure inclusive and equitable quality education and promote lifelong learning opportunities for all” was adopted in India in 2015. Students’ academic success is a vital part of the education system. Predicting student performance has grown more challenging due to the enormous amount of data in educational databases. Low-performing students will experience a variety of difficulties, including delayed graduation and even dropping out. Therefore, educational institutions should closely monitor the academic progress of their students and provide quick assistance to those who have low performance. Using Students’ academic achievement predictions to accomplish that is one method. This method will help educational institutions in identifying and supporting low-performing students at an initial stage. This study presents a systematic review of research on sentiment analysis towards SDG4 quality education through social media platform such as Twitter, Facebook and a review of 21 studies indexed in SCOPUS. Using social media data rather than a conventional survey of the data, evaluation of outspoken opinion and feelings of students towards their institution to obtain Quality Education. In this study, the dataset is taken from kaggle with names as student-performance-data-set which uses two files named as student-math, and student-por which shows the student performance in a Math language course and Portuguese language course, respectively, with 33 attributes and 396 records in each. Of 396 records, 110 records were selected as sample. During the visualization, we analyzed SVM model is stable because even minor data changes have no impact on the hyperplane and it handles the nonlinear data using Kernel techniques.

Keywords Sentiment analysis · Quality education · SDGs · Text mining

Introduction
Sustainable development goals establish priorities and direction for government, corporations and civil society around the world. Education is a right that belongs to everyone. Out of 17 SDGs of the United Nations, the 4th goal is about Quality Education. It aims to ensure inclusive, equitable quality education and promote lifelong learning opportunities for all by 2030. Due to the innovative nature of digital technologies such as Artificial Intelligence and its subsets like Deep Learning and Internet of things (IoT) are rapidly being used by the Government sector and Industry. The knowledge, abilities, attitudes, and values that enable people to contribute to and profit from a more inclusive and sustainable future must be developed through education. Sentiment analysis or opinion mining plays a promising role in Quality Education (SDG4). Most of the information shared by students on social media platforms includes clear sentimental aspects, and it has become an important research field.
Sentiment Analysis and Literature Review

People’s anxiety and resistance to new technology are frequently a result of their development and application. A collective lack of awareness often fuels this hostile attitude. This animosity is frequently caused by a lack of collective awareness/familiarity with the capabilities and the technologies’ potential advantages. AI applications and their related technologies such as Machine learning, deep learning, and natural language processing are enabling the automation and optimization of a wide range of human tasks, many of which would aid in the achievement of the SDGs. Countries improve their educational institutions to grow further in society [1–3]. Education is the key dimension of well-being and a crucial indicator of development. 17 SDGs established by UNs Agenda 2030, in 2015 for SDG 4 which aims to “ensure inclusive and equitable quality education and promote lifelong learning opportunities to all” [4]. According to [5], Artificial intelligence and other emerging technologies are used in almost every aspect of society, particularly in education. Our society is being revolutionized by digital technologies such as artificial intelligence, block chains, cloud computing, big data, and deep learning, among others. In their daily initiatives and endeavors, industry and government administrations are increasingly relying on digital technologies [6–8]. Several studies have emerged in recent years to investigate the problems, benefits, and potential impact of Artificial Intelligence [9–12] and related technologies (e.g. block chain [13, 14]) as drivers for the entire suite of SDGs.

Sentiment analysis is a procedure that automatically analyses natural language utterances, identifies key statements or opinions, and categorizes them based on their emotional attitude. Sentiment techniques can be categorized in Fig. 1.

There are a large number of studies with a great focus [9–14], that give a more general assessment of issues, consequences and advantages of combined applications of AI and digital technology applications to the entire SDG framework. Given the significance of these papers to the scope of this research, I’ve summarized them below:

Over the past few years, numerous studies on understanding the factors influencing a student’s performance have been published. A few of these works incorporate only analyzing the traits which have direct or indirect impacts on students’ result and some also incorporate predicting student’s outcomes based on the considered traits utilizing diverse learning algorithms.

Huang et al. [12] identified the economic, environmental and social barriers to achieving the SDGs with AI. In e-learning scenarios, the authors highlight the need of implementing personalisation strategies when it comes to customizing or recommending teaching materials. Similarly, they proposed comprehensive conceptual frameworks that depict how data from various sources interact with AI approaches and digital techniques for various SDG perspectives. They conclude that artificial digital technologies have

---

**Fig. 1** Approaches to sentiment analysis
a tremendous deal of promise to help the world’s economic, environmental and social sustainability.

Vinuesa et al. [11] Using a consensus-based method for eliciting opinions, the authors evaluate the emphasis of AI-based services on collectively fulfilling the SDGs. According to the analysis, Artificial Intelligence services can help achieve 79% of SDG targets have been met while hindering 35% of them, 14% leaving a tiny number of “overlapping” ambitions, where depending on the specific uses given, AI could be both an enabler or barrier to the goal(s). Therefore, proper regulation of AI systems could make it more helpful (and reduce the risks associated with it).

For example, social media platforms now give users a profusion of information, the veracity of which is rarely vetted beforehand. As a result, certain social media platforms may experience social and political division, resulting in social and political friction. On the other hand, some artificial intelligence algorithms based on natural language processing (NLP) may be useful in detecting false information on social media [15].

Esparza et al. [16] utilized sentiment analysis to investigate the views of students on teacher performance, and they discovered that the difficulty of assessing teacher evaluation feedback could be solved with the use of social mining.

Altrabsheh et al. [17, 18] SA methods were used to analyze the feedback of students which identify their positive or negative feeling towards the process of teaching. They used sentiment analysis to automatically model student feedback using SVM, NB, and Maximum Entropy. The results revealed that SVM and NB are best for modeling the student’s feedback.

Yadav et al. [19] use multimodal sentiment analysis to extract deeper emotions which sometimes may not be possible from textual analysis. The usage of audio-video equipment aids in the detection of various target users’ moods. Audio and video are gathered simultaneously from audio–video input to infer emotions from facial expressions in multimodal sentiment analysis. Kastrati [20, 21] proposed a model for representing documents with rich semantic content that uses deep learning to automatically categorize financial documents. Also they used different deep learning techniques such as DNN, CNN for managing and classifying educational content for various search and retrieval applications in order to provide a more personalized learning experience. For the purpose of classifying and organizing content in a MOOC scenario, they suggested a video classification system that makes use of multiple transcript feature representations and deep learning. The main aim of their study is to identify the role of higher education in SDGs. They summarized the best practices of QA agencies in promoting SGD's and provided recommendation system internal and external QA system for SDG achievement [22].

Kandhro et al. [23]: Different machine learning methods such as Support Vector machine. Multinomial Naive Bayes, Random Forest and Multilayer Perception classifier were used to propose the SA model for improving teaching quality in HE institutions. The study successfully investigated various SA models to identify the best model for analyzing student’s classroom feedback. They stressed that social media websites such as Twitter and Facebook could be used as a source of information and mining of opinions related to student’s learning activities. El-Sayad et al. [24]: In order to analyze the effects of the COVID-19 epidemic on educational systems, particularly on the psychological health of university students, they used statistics and machine learning techniques. A variety of data were gathered using an online questionnaire, including demographic data, digital tools, sleeping patterns, social contact, academic performance, psychological condition, and a scale for anxiety and sadness.

An automated evaluation approach for assessing student performance has been suggested, along with an analysis of student achievement. To accurately estimate student performance, the author employed the Tree algorithm here. Naive bayes is used frequently to find parameter that affects the performance of student. But Naive Bayes faced issue of zero probability [25]. Author proposed new algorithm RB-Bayes to solve this issue.

A Brief Interpretation of Sentiment Analysis in SDG4 (Quality Education)

| Study       | Objectives                                                                 | Platform            | Types of sentiment analysis tools | Outcomes                                                                 |
|-------------|---------------------------------------------------------------------------|---------------------|----------------------------------|--------------------------------------------------------------------------|
| Esparza et al. [16] | To examine students’ reviews of teacher performance using SA | Polytechnic University (1040 comments of students) | Support Vector Machine          | The difficulty of analyzing teacher evaluation comments can be solved with the use of social mining |
| Study                        | Objectives                                                                 | Platform                  | Types of sentiment analysis tools       | Outcomes                                                                                                                                 |
|------------------------------|-----------------------------------------------------------------------------|---------------------------|-----------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------|
| Altrabsheh et al. [17]       | To use SA to automatically student's feedback                                | Dataset (1036 comments)  | NB, SVM, Maximum Entropy and            | Both NB and SVM can be utilised for feedback analysis. Whenever insufficient data in the neutral class, CNB can be useful. It can be the right solution for irregular training sessions. |
|                             | To use SA to automatically student's feedback                                |                           |                                         |                                                                                                                                            |
|                             |                             | Twitter, Facebook           | NB and SVM                            |                                                                                                                                            |
| Altrabsheh et al. [18]       | To use SA to analyze the feedback of students' feeling towards the teaching process |                           |                                         |                                                                                                                                            |
| Yadav et al. [19]            | To take audio-visual input and extract emotions from audio and video in parallel | Twitter and Facebook      | NB, SVM and, Maximum Entropy Classification (MEC) | Using multimodal sentiment analysis to extract deeper emotions which sometimes may not be possible from textual analysis. The usage of audio-video equipment aids in the detection of various target users' moods. |
| Kandhro et al. [23]          | For increasing the Quality of HEI Instruction                                | Dataset (950 posts)       | SVM, SGD, MNB, RF and Multi-layer Perceptron (MLP) machine | Analyzing teacher evaluation comments can be solved with the use of social mining.                                                                                                                                 |
|                             | Develop SA model                                                           |                           |                                         |                                                                                                                                            |
|                             | Analysis to examine students' reviews of teacher performance                | University (1040 comments of students) |                                         |                                                                                                                                            |
| Altrabsheh et al. [17]       | To use SA to analyses students' feedback automatically                      | Dataset (1036 comments)  | NB, SVM and Maximum Entropy             | Both NB and SVM can be utilised for feedback analysis. When there is insufficient data in the neutral class, CNB can be useful. It can be the right solution for irregular training classes. |
| Altrabsheh et al. [18]       | To use SA to analyse feedback of students to identify their positive and negative feeling towards the teaching process | Twitter, Facebook        | NB and SVM                            | The use of SA in educational systems has a lot of promise both SVM and NB techniques were better for institutional data.                                                                       |
| Study | Objectives | Platform | Types of sentiment analysis tools | Outcomes |
|-------|------------|----------|-----------------------------------|----------|
| Yadav et al. [19] | To take audio-visual input and extract emotions from audio and video in parallel | Twitter and facebook | NB, SVM and, maximum entropy classification (MEC) | Using multimodal sentiment analysis to extract deeper emotions which sometimes may not be possible from textual anaysis. The usage of audio-video equipment aids in the detection of various target users' moods |
| Kandhro et al. [23] | To develop a sentiment analysis model for increasing the quality of HEI instruction | Dataset (950 posts) | SVM, stochastic gradient descent, MNB, random forest and multilayer perceptron (MLP) | In the context of educational study, MLP and MNB were more effective in terms of text classification |
| [26] | For predicting the performance of Moroccan students through a recommendation system using AI techniques during COVID-19 | They used AI and ML for predicting Student performance during COVID-19 | Random Forest, Decision Tree, Logistic Regression used | Random Forest gives highest accuracy in best prediction as compared to Decision tree, Logistic regression |
| [27] | In order to predicting the students’ performance they discussed BP, SVR, LSTM algorithms. Gradient Boosting classifier is implemented in classification phase | They used 2 datasets and 18 experiments done for predicting the performance | Support Vector Regression with 87.88% accuracy |
| [28] | There are several aspects on which students’ performance depends such as personal, academic etc | Python 3 tool is used for comparison between ML algorithms | SVM, Regression Linear and Regression Linear Algorithm provided superior prediction |

After a thorough review of the literature, many researchers used supervised machine learning techniques such as classification and regression algorithms for the prediction of students’ performance. Classification and Regression algorithms such as Support Vector Machine (SVM), Naïve Bayes (NB), Decision Tree (DT), K- Nearest Neighbors(KNN), Random Forest(RF), Artificial Neural Network (ANN), and Linear Regressions (LR) are the most of the researchers choice. Some researchers also used unsupervised machine learning techniques such as K-means, Fuzzy C-Means, Hierarchical, and Apriori algorithms.
**Research Gap Identification**

In student academic performance prediction, many researchers have developed models for the same using machine learning. They focused on academic, marital, psychological and demographic factors. But there are still some important and unknown factors that may influence the performance of students.

These factors are:

- Eating Habits, Sleeping Hours
- Mobile addiction
- Availing any scholarship (Single Girl Child, EWS, SC, ST, BC etc.)
- Place they live (Urban/Rural)
- Transportation Facility
- Language Barrier
- Mode of Admission (Regular/ Private/Distance)
- Gap year/ Pass out school background
- Hostler or Day scholar
- Effect of friend circle
- Parents intention and cooperation
- High expectation of parents
- Career Awareness
- Self employment

**Purpose of Proposed Model**

- To study and analyze existing research pertaining to the performance of students for identifying the parameters affecting the performance of the students.
- To propose an ensemble model for predicting the performance of the students.
- To test and validate the proposed model on various performance metrics.

**Proposed Framework**

**Source of Dataset**

In this paper, we took the dataset named as student-performance-dataset from kaggle. It uses 2 files named student-math and student-por which show the student performance in the Math language course and Portuguese language course, respectively. Sizes of these files are 56,993 bytes and 93,220 bytes. In this dataset, there are 33 attributes (like student's school, age, sex, address, family size, mother_education(medu), father_education,(fedu),mother’s job(mjob), Father’s job(fjob), parent’s cohabitation status grade1(G1),grade2(G2), final_grade(G3) etc. ) and 396 records in each file.

We took sample of 110 records to visualize the relationship between age and grade. [http://www.kaggle.com/datasets/larsen0966/student-performance-data-set](http://www.kaggle.com/datasets/larsen0966/student-performance-data-set)

The architecture of the proposed model is shown in Fig. 2.

(1) **Data Collection: Data** will be collected from online and offline mode through Google form and questionnaire.

(2) **Data Preprocessing: Data** will be collected from online/offline sources and did not have a proper structure, the data needed to be converted into suitable data for injecting into the proposed model using appropriate preprocessing methods.

(3) **Split the dataset: After that we will split the dataset into 2 parts.**

   (a) Training the dataset
   (b) Testing the dataset

(4) **Proposed Method (SVM/Naive Bayes):** Our proposed model is based on a hybrid approach of sentiment analysis (i.e. SVM and Naive Bayes)

(5) **Performance of the model:** We will compare the performance of the proposed hybrid model with an existing model which is based on either SVM or Naive Bayes.
Visualization of the Dataset

Figure 3a and b show the graphical representation of the relationship between age and grade attributes. After the visualization, we analyzed SVM model is stable because even minor data changes have no impact on the hyperplane and it handles the nonlinear data using Kernel techniques and works quickly with a clear margin of separation, but SVM takes a long training time on a large dataset. On the other hand, Naive Bayes approach works as all the features are independent, but if any one categorical value is missing in the dataset, then it assigns zero probability and would not be able to make any prediction.

In future, we will apply ensemble techniques and will perform comparative analysis to find the best model for prediction.
Future Plan

To make learning and teaching about sustainable development a reality, it is critical to coordinate the activities of stakeholders in university education. To make learning and teaching about sustainable development a reality at all levels to begin and maintain a systematic discussion on innovation in SDG and enhancement in learning and teaching, as well as to encourage/support student/teacher feedback issues related to learning outcomes, assessment, and quality assurance reviews, while taking into account the skills needed to add value [29].

- In future, we will ensemble such a hybrid model based on SVM and NB to overcome the problem of the existing model.
- Improved accuracy of classification techniques.

To capture a user’s opinion or thoughts from a text more effectively, ML/DL sentiment analysis approaches and techniques should place a great emphasis on embedding the semantic context using lexical sources such as SentiWordNet, WordNet as well as semantic representation using ontologies [19].

Conclusion

In this study, we have read and reviewed 21 studies indexed in Scopus based on Sentiment Analysis towards Quality Education and how it can be used for various purposes. In this paper, I could find methods to implement algorithms in the predictive analysis of students’ performance. Machine learning algorithms such as Support Vector classifier (SVC) and Naïve Bayes (NB) were utilized to predict the students’ performance.

We visualized the students’ performance according to existing methods. We will obtain the result of different dimensions separately, and compare them with the proposed model. Then, after this step, to show the superiority of the proposed hybrid model over the traditional approaches of sentiment analysis to achieve quality education which is the fourth goal of SDGs and its main aim towards individual/society is to ensure that all boys and girls complete free, equitable and primary and secondary education leading to relevant and effective learning outcomes. It also empowers people everywhere to live more healthy and sustainable lives.

Declarations

Conflict of Interest Both authors declare that they have no conflict of interest.

Ethical Approval This article does not contain any studies with human participants or animals performed by any of the authors.

Informed Consent Not applicable

References

1. Rivzi P, Lingard B. Globalising education policy. Routledge; 2009.
2. Costa EB, Fonseca B, Santana MA, de Araujo FF, Rego J. Evaluating the effectiveness of educational data mining techniques for early prediction of students’ academic failure in introductory programming courses. Comput Hum Behav. 2017;73:247–56.
3. Liao SN, Zingaro D, Thai K, Alvarado C, Griswold WG, Porter L. A robust machine learning technique to predict low- performing students. ACM Trans Comput Educ (TOCE). 2019;19:1–19.
4. Kaurav RP (2020) Theoretical extension of the new education policy 2020 using twitter mining.
5. Palomares I. Reciprocal recommender system: analysis of state-of-art literature, challenges and opportunities on social recommendation. Inf Fusion Press. 2021. https://doi.org/10.1016/j.inffus.2020.12.001.
6. Dunis C. Artificial intelligence in financial markets. Berlin: Springer; 2019.
7. Lytras MD. The recent development of artificial intelligence for smart and sustainable energy systems and applications. Energies. 2019;12(16):3108.
8. Mao C. Real time carbon emissions monitoring tool for prefabricated construction: an IoT based system framework. In: ICCREM 2018: Sustainable construction and prefabrication. American Society of Civil Engineers Reston, VA; 2018; pp. 121–7.
9. Goralski MA. Artificial intelligence and sustainable development. Int J Manag Educ. 2020;18(1):10030.
10. Truby J. Governing artificial intelligence to benefit the UN sustainable development goals. Sustain Dev. 2020;28(4):946–59.
11. Vinuesa R. The role of artificial intelligence in achieving the sustainable development goals. Nat Commun. 2020;11(1):1–10.
12. Huang. Information and communication technologies for sustainable development goals: state-of-the-art, needs and perspectives. IEEE Commun Surv Tut. 2018;20(3):2389–406.
13. Nguyen QK (2016) Blockchain—a financial technology for future sustainable development. In: 2016 3rd International Conference on green technology and sustainable development (GTSD); pp 51–4
14. Zwitter A and Herman J (2018) Blockchain for sustainable development goals. University of Groningen, Report 2018 7–2018 Ed. 2018
15. Zhang X, Ghorbani AA. An overview of online fake news: characterization, detection and discussion. Inf Process Manag. 2020;57(2): 102025.
16. Esparza GG. A sentiment analysis model to analyze students reviews of teacher performance using support vector machines. In: International Symposium on Distributed Computing and Artificial Intelligence. 2017. Springer.
17. Altrabsheh N. Sentiment analysis towards a tool for analysis real time students feedback. In: IEEE 26th International Conference on tools with artificial intelligence IEEE, 2014.
18. Altrabsheh N. SA-E: Sentiment Analysis for Education International conference on Intelligent Decision technologies, 2013.
19. Yadav SK. Multimodal sentiment analysis: sentiment analysis using audiovisual format. In: 2nd International Conference on Computing for Sustainable Global Development. (INDIACom) 2015.
20. Kastrati Z. The impact of deep learning on document classification using semantically rich representations. Inf Process Manag. 2019;2019(56):1618–32.
21. Kastrati Z. Integrating word embedding and document topics with deep learning in a video classification framework. Pattern Recognit Lett. 2019;128:85–92.
22. Rome Communique. 2020. https://erasmusplus.org.ua/novyny/3131-bologna-conference-in-rome-19-nov-2020.html
23. Kandhro IA. Student feedback sentiment analysis model using various Machine Learning schemes: a review. Indian J Sci Technol. 2019; 12(14).
24. El-Sayad A, Ewis A, Abdel Rauof MM, Ghoneim O. A new approach in identifying the psychological impact of COVID-19 on university Students’ academic performance. Alexandria Eng J. 2022. https://doi.org/10.1016/j.aej.2021.10.046.
25. Bhalla R (2019) A comparative analysis of application of proposed and the existing methodologies on a mobile phone survey. In: International conference on futuristic trends in networks and computing technologies. Springer, Singapore
26. Tarik A, Aissa H, Yousef F. Artificial Intelligence and Machine Learning to predict Student performance during COVID-19. In: The 3rd International workshop on Big Data and Business Intelligence(BDBI 2021) March 23–26, 2021; Warsaw, Poland. https://doi.org/10.1016/j.procs.2021.03.104
27. Sekeroglu B, Dimililer K, Tuncal K. Student performance prediction and classification using machine learning algorithms. ICEIT 2019, March 24, Cambridge. 2019. https://doi.org/10.1145/3318396.3318419.
28. Dabhade P, Agarwal R, Alameen KP, Fatima AT, Sridharan R, Gopukumar G. Educational Data Mining for predicting student’s academic performance using Machine learning algorithms. Mater Today. 2021. https://doi.org/10.1016/j.matpr.2021.05.646.
29. Shuang K. Convolution deconvolution word embedding: an end-to-end multi-prototype fusion embedding method for natural language processing. Inf Fusion. 2020;2020(53):112–22.