Expert Systems With Applications 209 (2022) 118119

Contents lists available at ScienceDirect

Expert Systems With Applications

journal homepage: www.elsevier.com/locate/eswa

Aspect2Labels: A novelistic decision support system for higher educational institutions by using multi-layer topic modelling approach

Shabir Hussain [a], Muhammad Ayoub [c], Ghulam Jilani [b], Yang Yu [d], Akmal Khan [e], Junaid Abdul Wahid [a], Muhammad Farhan Ali Butt [f], Guangqin Yang [b], Dietmar P.F. Moller [g], Hou Weiyan [a,*]

[a] School of Information Engineering, Zhengzhou University, Zhengzhou, 450001, Henan, China
[b] School of Education, Zhengzhou University, Zhengzhou, 450001, Henan, China
[c] School of Computer Science and Engineering, Central South University, Changsha, 336017, Hunan, China
[d] Distributed Systems Group, University of Duisburg–Essen, Duisburg, 47057, Germany
[e] Department of Data Science, The Islamia University of Bahawalpur, Bahawalpur, 63100, Punjab, Pakistan
[f] Department of Computer Science, NCBA&E, Rahim Yar Khan, 64200, Punjab, Pakistan
[g] Clausthal University of Technology, Clausthal-zellerfeld, 38678, Germany

ARTICLE INFO

Keywords:
Aspect-based sentiment analysis
Machine learning applications in education
Aspect extraction
Topic modelling
Opinion mining
Situational awareness

ABSTRACT

Aspect-based sentiment analysis (ABSA) has gained a rising concentration recently. It aims to provide a set of aspect terms and sentiments from a piece of text. Educational Data Mining (EDM) is now an essential tool for analysing pedagogical data. In academic institutions, student feedback is an influential gauge to measure the quality of the teaching–learning process. It helps higher education institutions to reconsider and improve their policies for student recruitment and retention. This paper proposed a situation awareness multi-layer topic modelling and enhanced hybrid machine learning approach for evaluating students' textual feedback data in academic institutions. The proposed Aspect2Labels (A2L) approach is divided our system into three layers. To preserve semantic information, we extracted general aspects terms in the first layer known as high-level aspects. We pulled low-level aspects terms associated with high-level aspect terms in the second layer and the third layer used for sentiment orientation. We used zero-shot learning, LDA, and different variants of LDA for the aspect extraction process. We performed annotation on unlabelled students' comments using our proposed A2L approach, and we obtained 91.3% accuracy in this process. We developed and tested novel algorithms for aspect terms mapping to label each aspect term to corresponding feedback. Different machine learning algorithms have been used to classify sentiments according to extracted aspects. We have also proposed and used Variable Global Feature Selection Scheme (VGFSS) and Variable Stopwords Filtering (VSF) to improve the performance of classifiers. We have managed to get 97% and 93% accuracy on the test dataset using Support Vector Machine (SVM) and Artificial Neural Networks (ANN), respectively. We highly suggest that our novel approach of aspect-oriented sentiment analysis could provide adequate understanding to analyse students' feedback.

1. Introduction

Natural language processing (NLP) has significantly impacted recent technologies using text mining applications. Nowadays, the adequate analysis of data or information is a challenging task, and it can help to improve analysing an enormous amount of unstructured data into meaningful context. NLP and text mining techniques have been used for knowledge discovery using unstructured textual data. Many studies have been conducted in which text mining applications are influenced effectively and efficiently. In marketing, opinion mining (Kumar, Kar, & Ilavarasan, 2021) has been used to know customer attitudes and possible trends in sales. Businesses squander a considerable amount of money to get consumer sentiments and opinions about their products and services (Kumar, Yadava, & Roy, 2019). In politics, politicians want to know about the voters' views, and voters want to know about the candidate's ability, performance, ideas, and policies (Drus & Khalid,

* Corresponding author.
E-mail addresses: shabir@gs.zzu.edu.cn (S. Hussain), ayoubncbae@gmail.com (M. Ayoub), jilani_101@gs.zzu.edu.cn (G. Jilani), yang.yu@uni-due.de (Y. Yu), akmal.shahbaz@iub.edu.pk (A. Khan), junaid.a.wahid@gs.zzu.edu.cn (J.A. Wahid), farhanfauz007@gmail.com (M.F.A. Butt), guangqinyang@sina.com (G. Yang), dietmar.moeller@tu-clausthal.de (D.P.F. Moller), houwy@zzu.edu.cn, houwy@139.com (H. Weiyan).

https://doi.org/10.1016/j.eswa.2022.118119
Received 21 August 2021; Received in revised form 11 May 2022; Accepted 7 July 2022
Available online 19 July 2022

2019). In social networks, people want to know about people who have common interests (Drus & Khalid, 2019). Sentiment analysis has already impacted different domains. For example, well-known companies are already using this technique to evaluate their customers' feedback (Jurek, Mulvenna, & Bi, 2015; Zhang, Xu, & Wan, 2012). They perform forecasting and monitoring about their products and services using public feedback or sentiments. By using the applications of NLP, Situational Awareness (SA) and Situational Understanding (SU) becomes the most straightforward task to know the actual circumstances in any field of life (da Costa, Meneguette, & Ueyama, 2022; Hussain et al., 2021). The situational awareness approach is a trending domain of research in which the systems take information from the environment, understand it adequately, and predict what may occur in the future (Endsley, 2017).

1.1. Educational data mining

Nowadays, every organization and business uses information systems to store their valuable data in databases. Educational data mining can enrich the student learning experience and outcome. It can impact educational endeavours for student satisfaction at the institution level. Many studies have been conducted in which educators can assess their implemented strategies effectively by evaluating students' feedback (Tondeur, Scherer, Siddiq, & Baran, 2020). Student feedback correlated to teaching–learning activities can help to investigate the strengths, weaknesses, opportunities, and threats in the educational process in all the higher educational institutions (HEIs) programmes and curricula (Boring, 2017; Holmes, Nguyen, Zhang, Mavrikis, & Rienties, 2019). In all higher education academic institutions, usually at the end of each semester, the institution's management conducts an activity called Teacher Performance Evaluation (TPE), and Course Evaluation (CE) (To & Tang, 2019). Students give feedback about the teacher's performance, course outcomes, and university management using online or offline systems in this activity (Constantinou & Wijnen-Meijer, 2022). In the TPE process, students give feedback about a particular instructor regarding his teaching skills, knowledge, delivery of the course, presentation skills, and behaviour (Yu et al., 2016). In CE, students usually give feedback about the course finding, course importance, course objectives, and course benefits in their learning process (Sozer, Zeybekoglu, & Kaya, 2019). Moreover, students also give feedback about the department or institution's policies and overall conduct of the university (Nawaz et al., 2022). This feedback directly or indirectly impacts the instructor's annual appraisal and influences the teacher performance evaluation (Hunter & Springer, 2022). A sample of questions asked from students in TPE and CE Process feedback is shown in Fig. 1.

The course assessment aspires to provide the information with the learning outcomes to the course content development authorities and management. The course organization's board needs to continuously improve the courses and assure university students satisfactory possible conditions for learning and carrying out their studies (Budiharso & Tarman, 2020). On the other hand, in the modern era of the competitive market of academic intuitions, every institute wants to give the best education to their students. Universities and colleges continuously try to enhance the educational environment to compete with other institutes by using situation awareness approaches. Student Recruitment and retention are essential matters for all higher education institutions (Al Shobaki & Abu-Naser, 2017).

1.1.1. The rationale of the study

The current system of evaluating teaching–learning activities outcome by using students' feedback is not adequate and satisfactory (Eng, Ibrahim, & Shamsuddin, 2015). It deals only with closed-ended questions, which consider the answers to students' feedback using the Likert scale. The "comments" as open-ended questions answers are part of the teacher and course evaluation form, which is usually neglected by the management because it is challenging and time-consuming to gain insight into the textual feedback. Knowing the exact Situational Aspect Information (SAI) from the students about educational institutes using students' textual feedback is a very challenging task in situational awareness. Most of the reported research work on the educational domain using student feedback data (Firdaus, Asror, & Herdiani, 2021; Kastrati, Dalipi, Imran, Pireva Nuci, & Wani, 2021), used only sentiments in the term of (positive, negative, neutral) polarities to evaluate the student's feedback. However, students usually give feedback according to some particular aspect or stance. Therefore, it is urgent need to develop a robust and efficient system to explore the situational aspects from students' textual feedback data. Knowing the specific situational aspect behind the unstructured text can help educational institutes explore more specific data insights. Manual annotation of each feedback is impractical (Khan & Ghosh, 2021) because it needs a lot of effort and time. Therefore, an appropriate system is needed to perform annotation automatically. Indexing each review automatically according to its sentiment and providing annotation according to relevant aspects could increase the understanding. Converting textual data into a numeric form for computations creates sparsity (Zhou, Niu, & Yang, 2021) and high dimensional problems in feature vector space that needs to be addressed. Thus, an educational institute needs a robust and adequate system that could efficiently understand information from students' feedback, and the administration could easily trace the problem area where students are not satisfied. The main research questions posed by this study are as follows:

- RQ1: How can we extract the Situational Information Aspects (SIA) and key terms used by students in their textual feedback?
- RQ2: How can we correlate each extracted high-level aspect term with the corresponding low-level aspect term from student's feedback?
- RQ3: How can we efficiently and correctly map and label the extracted aspect terms along with corresponding student feedback?
- RQ4: How can we effectively resolve sparsity and high dimensional feature vector space problems to improve sentiment classification?
- RQ5: How can we predict a student's sentiment and opinion using a textual recommendation dataset, and how HEIs can improve their strategies and decisions?

To answer the research questions mentioned above, we proposed the Aspect2Labels (A2L) machine learning framework, which significantly contributed to the literature.

In this research work, we make five significant contributions, and we proposed and successfully implemented an enhanced hybrid A2L approach to identify and extract Situational aspects using the student's feedback dataset by applying Situational Awareness Theory (SAT).

1. We translated the problem of teaching–learning process evaluation in higher educational institutions to a Situational Awareness (SA) framework using the student feedback dataset.
2. Study shows that machine learning and text mining technologies can effectively be used for educational data analysis to improve the teaching–learning process at all higher educational institutions.
3. We proposed and implemented a multi-layer topic modelling approach to perform annotation on the unlabelled student feedback dataset using the Aspect2Labels (A2L) framework.
4. We also introduced and implemented three novel rule-based algorithms for AE, validation, and mapping.
5. We proposed and implemented two data filtering techniques. We analysed that using the Variable Global Feature Selection Scheme (VGFSS) and Variable Stopwords Filtering (VSF) improved the performance of machine learning classifiers.

Instructor:

1. The Instructor is prepared for each class .
2. The Instructor demonstrates knowledge of the subject.
3. The Instructor has completed the whole course.
4. The Instructor provides additional material apart from the textbook.
5. The Instructor gives citations regarding current situations with reference to Pakistani context.
6. The Instructor communicates the subject matter effectively.
7. The Instructor shows respect towards students and encourages class participation
8. The Instructor maintains an environment that is conducive to learning
9. The Instructor arrives on time
10. The Instructor leaves on time
11. The Instructor is fair in examination
12. The Instructor returns the graded scripts etc. in a reasonable amount of time
13. The Instructor was available during the specified office hours and for after class consultations

Course:

1. The Subject matter presented in the course has increased your knowledge of the subject
2. The syllabus clearly states course objectives requirements, procedures and grading criteria
3. The course integrates theoretical course concepts with realworld applications
4. The assignments and exams covered the materials presented in the course
5. The course material is modern and updated

Fig. 1. Sample of questions asked to students about TPE and CE process.

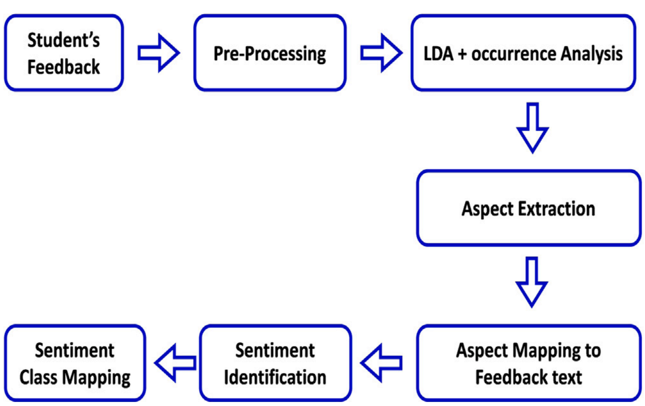


Fig. 2. Workflow of aspects specific sentiment analysis.

Generic workflow of proposed of aspects specific sentiment analysis is shown in Fig. 2.

Our proposed approach consists of a three-layered model for situational aspect extraction, aspect mapping, and sentiment analysis. In the first layer of our model, we used LDA and co-occurrence analysis unsupervised models to extract aspects of the student's feedback. We named these extracted terms as high-level aspects. We used an algorithm to automatically annotate each high-level aspect term to its corresponding student feedback. We used a similar method to extract and map the most frequent terms appearing with high-level aspects terms in the second layer. These terms we named low-level aspect, and we mapped them against its relevant high-level aspect term using our proposed low-level aspect extraction and mapping algorithm, and we obtained 91.3% accuracy in this process. We used the third layer to extract polarities and perform sentiment annotation using TextBlob and Syuzhet methods. After the appropriate mapping of aspects and automatic class labelling, we used five machine learning algorithms to perform classification. We used bag-of-words (BOW), TF-IDF, and Word2Vec features to train machine learning models. The study revealed that SVM outperforms other models, achieved 97% accuracy in sentiment classification, and compared our proposed results with other studies. We have proposed and used VGFSS and variable stopwords removal techniques for dataset filtering to improve sentiment classification accuracy. Our results show that our proposed approach could be valuable for higher educational institutions to conduct student feedback analysis with appropriate annotations and summarized results that can be handier for the decision-making process.

This paper is organized as follows. Section 2 provides an overview of related work and technologies investigated as background knowledge. Section 3 defines the methodological framework. Section 4 describes the results and discussion. Finally, Section 5 concludes the paper, and Section 6 outlines future works.

2. Related studies

The role and importance of data and its analysis gained attention globally. Exploring the meaningful hidden pattern from collected data is challenging and laborious. The analyses become complex when collected data is unstructured and heterogeneous. Currently, every well-known organization uses data mining and NLP applications frequently to increase Situation Understanding (SU) to know the actual situational information (Hussain et al., 2021; Tseng, Tran, Ha, Bui,

& Lim, 2021; Wang & Lien, 2019). According to Endsley (2017) Situation Awareness (SA) starts from the perception, which is the first level of proposed mental representation. Perception builds from the environment by fetching data or information from the elements of a particular system by using different attributes (Yu, Qin, Hussain, Hou, & Weis, 2022). Information perceived in the first level of SA is used for comprehension and projection. Various studies used the SA approach to understand and mitigate the risk systems by analysing the situational information (Ardito et al., 2020; Koopmanschap, Hoogendoorn, & Roessingh, 2015; Lagerstrom et al., 2016). This study used SA information in the educational domain to improve the decision-making process for higher educational institutions (HEIs) using machine learning and natural language processing. In this regard, McKenna et al. (2014) explored SA of nursing students while participating in inpatient deterioration simulations. Their experimental results concluded that curriculum planners must consider situational awareness an essential component for undergraduate nursing students. Lee Chang et al. (2017) analysed the SA for simulation and real-time teaching methods in airway management to identify the effective teaching method. Their results showed that the simulation teaching method is better in situational analysis. Using textual data, sentiment analysis regarding situational aspects has excellent worth for various industries and organizations. Sentiment analysis or opinion mining can be performed in two ways, using lexicon-based and machine learning approaches.

Situational aspect-based sentiment analysis has excellent worth in higher educational institutions to evaluate what situational information is floating in students' minds. Nowadays, universities and educational institutes are keen to recruit new students and retain their existing students. Therefore, they are spending hundreds of dollars on this purpose. According to Levitz (2018), a survey has been conducted in the USA. They found that government universities spend $536 an average on a student; in contrast, private universities spend four times more than public universities to retain their existing students and make better educational policies. Different researchers have explored this perspective, proclaimed in the literature, and we reported in this section.

Aspect-specific sentiment analysis gives insight into data. Literature unfolds following many studies about topic modelling discussed in this section. Aspect-based opinion mining is a sub-area of sentiment analysis in which we get insights into data to explore the opinion terms or stance detection.

Teacher and course evaluation survey activity by student feedback has become essential for higher education institutes worldwide. It helps assess faculty performance and the suitability of the course in any academic programme. Aspect-based sentiment analysis is imperative for all higher education institutions' student recruitment and retention issues. The authors (Nikolić, Grljević, & Kovačević, 2020) proposed a system in which they used machine learning, cascade, rule-based and dictionary-based classifiers to identify aspects from student feedback to perform sentiment analysis. By using machine learning and natural language processing approaches, Shaikh and Doudpotta (2019) devised a two-step rule-based system. The first step is to use supervised machine learning to extract the overall topics from the student feedback text, followed by the use of NLP rules to find specific aspects and related opinion words about which the feedback is given, as well as the opinion's orientation, which can be positive, negative, or neutral. Their experiments achieved the recall and precision of 83.89% and 84% on topic identification to classify feedback for teacher and course categories. The Sindhu et al. (2019) used a two-layered LSTM model, which identifies aspects described within the students' feedback and specifies the polarity of those identified aspects. Their proposed model achieved 91% accuracy in the aspect extraction module. In this contrast, the authors (Sangeetha & Prabha, 2020) processed each student's feedback parallel across a multi-head attention layer with Glove and Cove embeddings for the identification of Vietnamese students' emotions from their feedback. Their experimental results revealed that the fusion of multiple layers accompanied with LSTM achieved 94% accuracy in students' emotion identification from their feedback. In another study on higher education, teacher evaluation task, the Kandhro, Wasi, Kumar, Rind, and Ameen (2019) used the LSTM model with word embedding for mapping the words for teacher evaluation from students' feedback. They achieved 92% accuracy in teacher evaluation from sentiment analysis from students' feedback. The lexicon-based approach for sentiment analysis applications uses dictionaries or corpus in which positive and negative values are assigned to positive or negative words. Sentiment scores can be calculated using the aggregated score calculated for each feedback or each opinion word within a sentence (Ray & Chakrabarti, 2017; Taboada, Brooke, Tofiloski, Voll, & Stede, 2011). To assess faculties, Lin et al. (2019) and Mendon, Dutta, Behl, and Lessmann (2021) proposed a knowledge-based system in which they used machine learning and lexicon-based approaches to extract aspects from students' feedback and their sentiments automatically. Their proposed model achieved 84% accuracy in students' feedback sentiment analysis. The Wook et al. (2019) used the lexicon approach with Vader Lexicon as a lexical source to develop a system that could automatically analyse students' feedback to evaluate the performance of teaching faculty. Aung and Myo (2017) proposed a system that automatically analysed students' textual feedback and evaluated the teacher's performance using the lexicon-based approach. They did not consider intensifier and blind negation words in their system. Similarly, Rajput, Haider, and Ghani (2016) introduced a system that automatically analyses text comments and produces quantitative and qualitative metrics that help predict teacher's performance using student feedback.

Machine learning is a sub-area of artificial intelligence in which we get insights and explore hidden patterns from the data (Ayoub et al., 2022). An educational data mining framework (Gutiérrez, Canul-Reich, Zezzatti, Margain, & Ponce, 2018) developed using machine learning 'SocialMining', which purposed to assess teachers' teaching methods and suggest courses for teachers through educator assessment made by students' feedback. Similarly, Asanbe, Osofisan, and William (2016) showed a valuable model for calculating and predicting teachers' performance in educational institutions using data mining techniques. In this study, they used a two-layered approach for classification. First, they used ANN and Decision Tree for classification purposes. The Rudniy and Elliot (2016) investigate the utilization of n-gram examination to analyse teachers' and students' feedback. They evaluated results using uni-gram, bi-gram, digram, trigram, four-gram, and five-gram. Altrabsheh, Cocea, and Fallahkhair (2014) introduced a supervised algorithm to forecast opinion mining from students' comments. The proposed model trained by the n-gram parameter, taken from students' comments, achieved 95% accuracy using the SVM model. The most prominent feature selection from high dimensionality dataset is an excellent approach to reducing the feature space in sentiment classification. This process improves the accuracy and capability of machine learning models. To improve the feature selection process, the Uysal (2016) used a global variable feature selection scheme where the feature selection scheme is modified in order to obtain a more prominent feature set. Their experimental results on the benchmark dataset reported that the classification performance in Micro-F1 and Macro-F1 is improved and achieved high accuracy. Wahid et al. (2021), proposed topic2features (T2F) framework to deal with short and sparse data using the topic distributions of hidden topics gathered from a dataset and converting them into feature vectors to build supervised classifier. Similarly, for the automatic classification of text documents, the Agnihotri, Verma, and Tripathi (2017) used the global variable feature selection system. Their experimental results concluded that variable feature selection vastly enhances the results in a highly unbalanced dataset. In the same contrast in sentiment analysis of short texts, variable stop words filtering is used to deal with the data sparsity problem. Ali, Ehsan-ul Haq, Rauf, Javed, and Hussain (2021) concluded that removing single frequency terms from the lexicon reduces data sparsity significantly while applying mutual information to exclude irrelevant terms improves the classifier's accuracy. The Barrón Estrada, Zatarain Cabada, Oramas Bustillos, and Graff (2020) performed compression among deep learning, machine learning, and evolutionary approach for sentiment classification. They developed two datasets containing students' expressions regarding teachers, exams, homework, and academic projects. Their experimental results revealed that their proposed evolutionary algorithm reported 93% accuracy in sentiment analysis and classification.

Aspect Based Sentiment Analysis technique is also used in various domains to get insight from the data. E-commerce organizations need to analyse feedback and determine the customers' sentiment to give them better products and services. In this contrast, Akhtar, Zubair, Kumar, and Ahmad (2017) and Anoop and Asharaf (2018) performed aspect-specific sentiment analysis of product reviews using LDA. They extracted topics from the LDA data and then mapped them with various aspects of an entity to perform the aspect-specific sentiment analysis on product reviews. Their experiments gave promising results compared to existing methods of sentiment analysis. Similarly, the Dong et al. (2018) and Kumar, Gahalawat, Roy, Dogra, and Kim (2020) proposed unsupervised topic sentiment, a joint probabilistic model, to process and mine the online reviews of products for sentiment classification. Their experimental results achieved significantly better than baseline models such as LDA and JST. García-Pablos, Cuadros, and Rigau (2018) used un-supervised algorithm W2VLDA based on topic modelling combined with some other algorithms to perform aspect classification, aspect-terms separation, and sentiment polarity classification for any given domain and language. They evaluated their model on the multilingual SemEval 2016 task 5 (ABSA) dataset, which produced good accuracy. Similarly, Wahid et al. (2022) used un-supervised deep learning techniques to provide an automated way of labelling the unstructured and unlabelled social media data through LDA and Bert embedding.

Likewise, the current study aims to develop a machine learning and text mining-based educational framework that uses the student's recommendations to improve the strategies and practices within the higher educational institutions domain. Advancement in text mining applications using NLP technologies in HEIs can be a valuable asset to transform students learning experiences to advance the teaching methodologies to improve the quality of services in educational institutions.

3. Proposed framework

The proposed multi-layer Aspect2Labels (A2L) framework for educational data mining for higher educational institutions is explained in this section. To analyse and do experiments on the textual data of students' feedback collected in the teaching–learning process, we build an automated and robust system that evaluates student feedback with proper tagging and gives detailed information from textual data. The A2L framework will provide the summarized annotated results with aspect terms and sentiments. It will present to the management and study board of the institute, and they will further send it to the concerned authorities to analyse and decisions making for the institution. Furthermore, the system receives offline data for the training process to ensure that the system will not delay results. The proposed system depicts the students provided feedback in a meaningful way to understand situational information. The framework of our proposed approach is shown in Fig. 3.

3.1. Dataset

We collected students' feedback datasets from Quality Enhancement Cells (QEC) of the National College of Business Administration and Economics (NCBA&E) Rahim Yar Khan, Punjab, Pakistan. Initially, the dataset consists of 6500 feedback collected from the Quality Enhancement Cell (QEC) department. After filtering and removing the duplicate review, we have 5767 reviews used for experiments. We used these comments as input to our system, and we performed pre-processing (Mehanna & Mahmuddin, 2021) that removes noisy data from the given dataset. The sample of students' feedback is shown in Fig. 4. Pre-processing step involves (tokenization, case conversion, numbers removal, removal of hyphens, removal of punctuation, removal of symbols, removal of stopwords, stemming, and replacing "n't" with "not". All pre-processing steps are briefly explained in this current section.

3.2. Pre-processing

In the pre-processing process, we cleaned up our data to perform computation. We used the "quanteda" package of R. We did all quantitative analysis on textual data using this package for pre-processing. This process is also called data pipelining.

Word Tokenization: In word tokenisation (Prasad, 2021), we split our feedback into separate words or chunks. We used a word tokeniser to break the sentence into words or tokens for this process. Example: He has great communication skills in his explanations. "He", "has", "great", "communication", "skills", "in", "his", "explanations", ".".

Lowercase Conversion: We performed case conversion after tokenisation of text. This step increased the value of manipulating the textual data.

Removal of Numbers: The next step of data pre-processing is removing all the digits from the text, such as "12", "555".

Removal of Punctuation's: We removed all the punctuation from the text in this process, such as commas, exclamations, full stop, and question marks.

Remove Symbols: In this process, we removed all the symbols from text such as "hashtags", "@", "&", "$", "/" etc.

Remove Hyphens: This process involves removing hyphens in text and turning them into space. For example, "long-term "into long term and "pre-processing" into "pre-processing".

Identifying Stop Words: Stopwords are words in the text that are generally considered worthless. In the English language, many words appear very frequently like "is", "and", "the", and "a". We need to exclude stopwords before doing any processing on text data. It improves the feature extraction process. We used the stopwords function of R to remove all stopwords from our dataset. Example: He is a good teacher. In this phrase, "is" and "a" are useless stopwords in text mining.

Stemming: The stemming process involves normalizing the words into their base or root forms. For example, ran, run, runs, and running, all these words originated with a single root word, "run" occasionally, it yields the root word that may not have any meaning. e.g., stem operation bundles "response" and "respond" into a joint "respond", and this word has no meaning in the English language.

Lemmatization: Lemmatisation is moderately similar to stemming. The purpose of lemmatisation is to produce a meaningful root word. Both techniques are usually included in morphological normalization.

3.3. Aspect extraction and mapping

The aspect modelling module comprises three layers shown in Fig. 5. The first layer is used for high-level aspects extraction. We used Latent Dirichlet Allocation (Nanda, Douglas, Waller, Merzdorf, & Goldwasser, 2021) (LDA) with its different variants and co-occurrence analysis (Chen, Li, & Zhou, 2021) methods to extract topics and related terms. Our rule-based algorithm verifies all aspects of topics and terms to ensure the topic suggested by LDA and co-occurrence analysis. The second layer of our proposed aspect modelling approach is used for low-level aspect extraction, and terms are confirmed with a low-level aspect extraction algorithm. At last, we used the third layer for sentiment extraction and labelling of each comment, and we mapped and assigned labels according to given polarities.

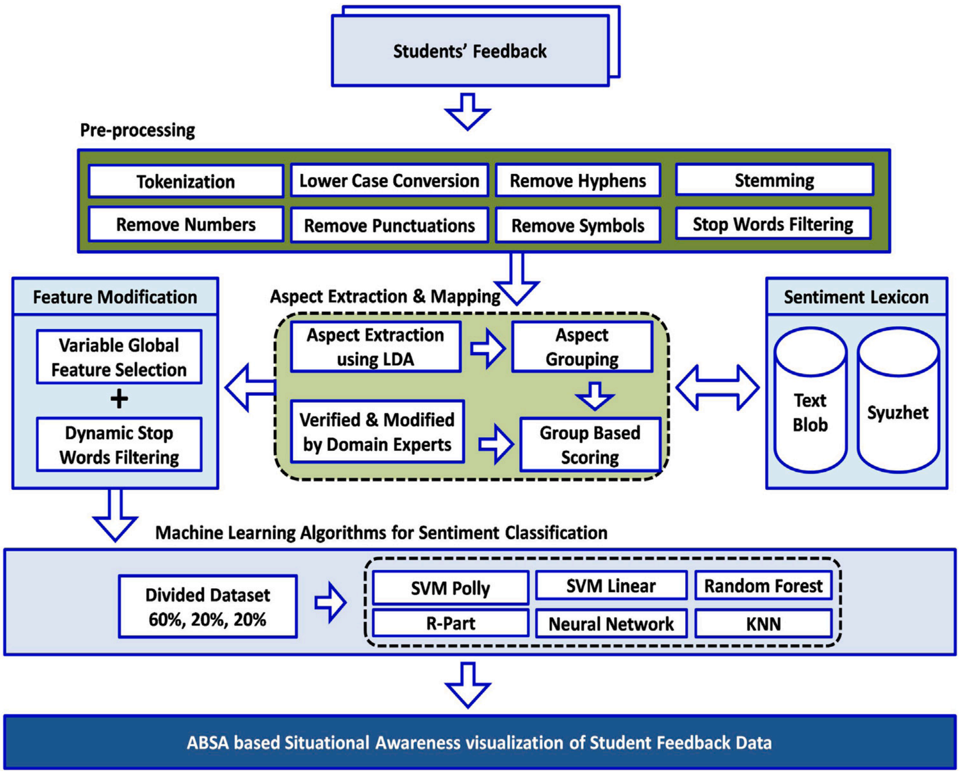


Fig. 3. An enhanced hybrid machine learning framework for aspect specific sentiment analysis.

	feedback
1	Best Regards to him and the department of computer scien...
2	I learned all the stuff and the explanation provided was good.
3	Lecture content is useful for the most part
4	Very poorly designed course material. Everything has been ...
5	The instructor is way too dull.?
6	I've tried hard to learn these lectures at least thrice, but I lo...
7	There's a lack of energy and no connection with the audie...
8	The final quiz was ridiculous
9	Confusion with the deadline for assignments.
10	Instructor examples are practical in all contexts.
11	Assumptions about the student are not explicit before star...
12	The quality of the presentation is not very good
13	Understandable lectures with challenging assignments
14	The course assignments are a bit time consuming without ...
15	Good material but sometimes useful part are after homew...
16	Professor and the team behind the course did a great job
17	The teacher is very good, and he explain very well and dee...
18	A great lecturer, thank you!
19	Great assignments, it is more than just theoretical knowledge
20	Still, there is a scope of improvement.

Fig. 4. Sample of students feedback dataset before pre-processing.

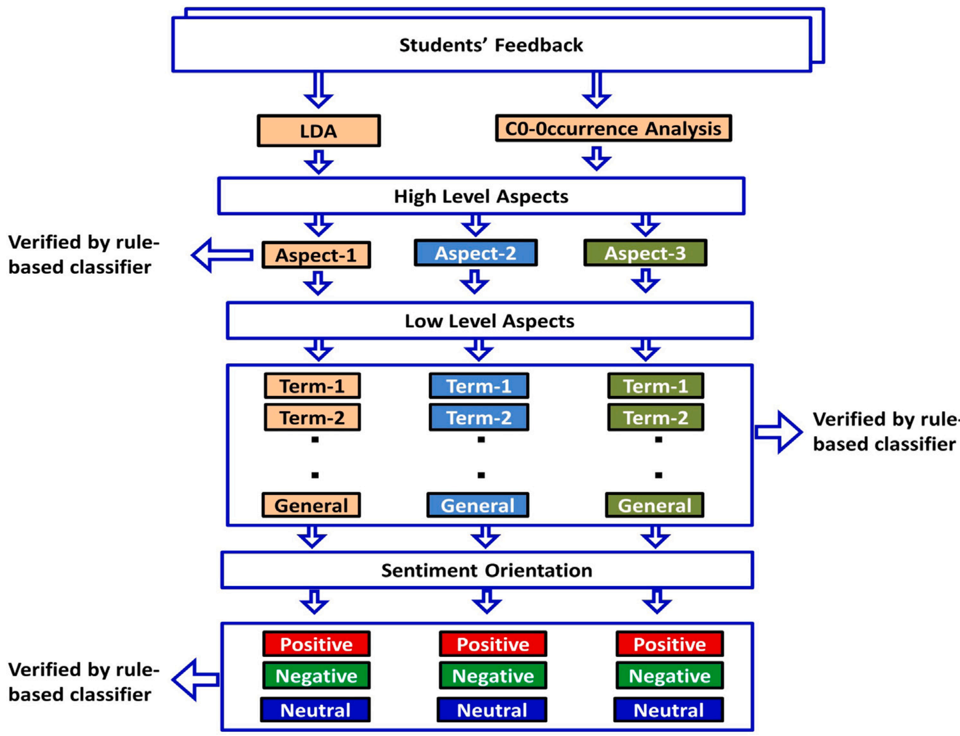


Fig. 5. Proposed multi-layer approach for aspect extraction and annotation.

3.3.1. Co-occurrence analysis

Co-occurrence analysis is the interconnection of paired terms present in a specific document; this co-occurrence network is based on the defined criteria of co-occurrence (Harshini & Gobi, 2020). For instance, terms A and B might be said to co-occur if both exist in a specific document. Similarly, another document may contain terms B and C. Connecting A to B and B to C makes a co-occur organization of these three terms. Co-occurrence analysis (Buzydlowski, 2015) is simply the counting of paired terms within a document. Co-occurrence networks (Özgür, Cetin, & Bingol, 2008) are created for a given list of terms concerning the collection of texts. Co-occurring pairs of terms can be called 'neighbors', and these often group into 'neighborhoods' based on their interconnections. Co-occurrence analysis depends on the variant of expected mutual information (Larkey, Ballesteros, & Connell, 2002) which is used to measure the proportion of word co-occurrences. For instance, two terms a and b, then variant is defined as shown in Eq. (1).

$$em(a,b) = max(\frac{n_{ab} - En(a,b)}{n_a + n_b}, 0) \tag{1}$$

As shown in Eq. (1), n_{ab} represent the number of times in which a and b co-occur in a text. While n_a and n_b are the numbers of occurrences of a and b in the corpus, and $En(a,b)$. The expected number of co-occurrences of a and b, is kn_an_b which shown in Eq. (2).

$$k = \sum n_{ab} / \sum n_a n_b \tag{2}$$

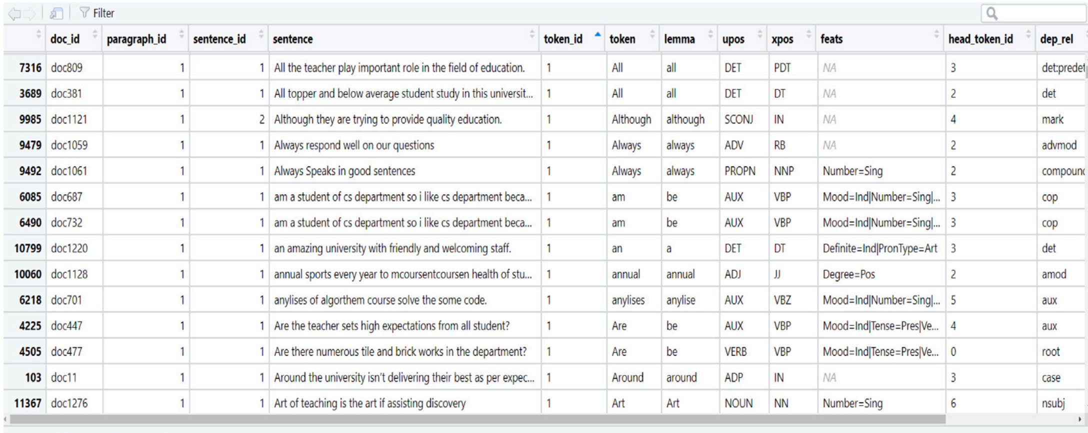
As shown in Eq. (2), k is the constant value based on the corpus size. We used part of speech tagging (Buoy, Taing, & Kor, 2021) to perform co-occurrence analysis. We extracted part of the speech against each comment. Sample of extracted parts of speech for each comment are shown in Fig. 6.

3.3.2. Latent Dirichlet Allocation (LDA)

Using the LDA and Co-occurrence analysis techniques is to perform the aspect-based sentiment analysis to explore students' aspect or stance terms in their textual feedback. We performed annotation on the unlabelled data using these extracted terms and each feedback sentence by exploring aspect terms. We have used different topic modelling approaches for the aspect extraction from the student textual feedback. Topic modelling is a type of statistical modelling for discovering the abstract "topics" that occur in a collection of documents. There are many approaches for obtaining topics from textual data, such as Term Frequency and Inverse Document Frequency (TF-IDF) and Probabilistic Latent Semantic Analysis (pLSA) techniques. The Latent Dirichlet Allocation (LDA) is a popular topic modelling technique to extract topics from a given textual data. The term latent means something that exists but has not yet been discovered. In contrast, student textual feedback's high- and low-level aspect terms are also "hidden topics". It is yet to be discovered. Hence, we used unsupervised LDA and co-occurrence analysis models to extract hidden aspects of the student's feedback. Lisena, Harrando, Kandakji, and Troncy (2020). The basic idea is that Vidaurre, Kawanabe, von Bünau, Blankertz, and Müller (2010) the documents are represented as random mixtures over latent topics, where a distribution over words characterizes a topic. The words with the highest probabilities in each topic usually give a good idea of the topic of word probabilities from LDA. As shown in Fig. 7, the generative workflow of LDA in which D represents the corpus having multiple documents and z denotes topics, and w denotes words in a document. In Fig. 7 α and β represents Dirichlet distribution and θ and ϕ represents multinomial distributions.

$$P(W,Z,\theta,\varphi,\alpha,\beta) = \prod_{M}^{j=1} P(\theta_j;\alpha) \prod_{K}^{i=1} P(\varphi_i;\beta) \prod_{N}^{t=1} P(Z_j,t|\theta_j) P(W_j,t|\varphi z_j,t) \tag{3}$$

In the left side of Eq. (3) we have a probability of the document that appears for the mixture over latent topics. On the right-hand side of the equations, there are four factors, in which the first two factors work as the setting of LDA and the last two factors work as gears of the LDA. Each factor calculates the probabilities and multiply with each factor's probability, and gets the article's final probability as output. The first factor on the right-hand side of the equation represents the topics, the



Filter

	doc_id	paragraph_id	sentence_id	sentence	token_id	token	lemma	upos	xpos	feats	head_token_id	dep_rel
7316	doc809	1	1	All the teacher play important role in the field of education.	1	All	all	DET	PDT	NA	3	det:predet
3689	doc381	1	1	All topper and below average student study in this universit...	1	All	all	DET	DT	NA	2	det
9985	doc1121	1	2	Although they are trying to provide quality education.	1	Although	although	SCONJ	IN	NA	4	mark
9479	doc1059	1	1	Always respond well on our questions	1	Always	always	ADV	RB	NA	2	advmod
9492	doc1061	1	1	Always Speaks in good sentences	1	Always	always	PROPN	NNP	Number=Sing	2	compound
6085	doc687	1	1	am a student of cs department so i like cs department beca...	1	am	be	AUX	VBP	Mood=Ind|Number=Sing|...	3	cop
6490	doc732	1	1	am a student of cs department so i like cs department beca...	1	am	be	AUX	VBP	Mood=Ind|Number=Sing|...	3	cop
10799	doc1220	1	1	an amazing university with friendly and welcoming staff.	1	an	a	DET	DT	Definite=Ind|PronType=Art	3	det
10060	doc1128	1	1	annual sports every year to mcoursentcoursen health of stu...	1	annual	annual	ADJ	JJ	Degree=Pos	2	amod
6218	doc701	1	1	anylises of algorthem course solve the some code.	1	anylises	anylise	AUX	VBZ	Mood=Ind|Number=Sing|...	5	aux
4225	doc447	1	1	Are the teacher sets high expectations from all student?	1	Are	be	AUX	VBP	Mood=Ind|Tense=Pres|Ve...	4	aux
4505	doc477	1	1	Are there numerous tile and brick works in the department?	1	Are	be	VERB	VBP	Mood=Ind|Tense=Pres|Ve...	0	root
103	doc11	1	1	Around the university isn't delivering their best as per expec...	1	Around	around	ADP	IN	NA	3	case
11367	doc1276	1	1	Art of teaching is the art if assisting discovery	1	Art	Art	NOUN	NN	Number=Sing	6	nsubj

Fig. 6. Part of speech tagging on students feedback.

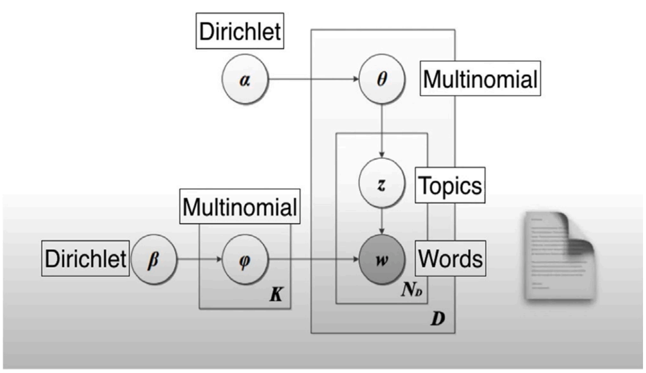


Fig. 7. Graphical representation of LDA.

second represents the words, and both showed Dirichlet distributions. Similarly, the third factor in the equation represents the topics; the fourth represents the words, representing the Multinomial distribution.

We have used the Gensim implementation of LDA and optimized it according to our dataset. Hyperparameters α and β are tuned, and we also optimized the number of topics T corresponding to our student feedback. The details of hyper-parameters of the LDA model according to our dataset are shown in Table 1.

Table 1
Summary of hyper-parameters for each topic modelling models.

Model	No. of components	No. of iteration	Random state	Batch size	No. of topics
LDA Genism	10	50	None	128	5
LDA2VEC	10	50	None	128	5
SS-LDA	10	50	None	128	5
Zero Shot	50	20	–	64	5

Table 2
Extracted high and low level aspect terms using LDA and co-occurrence analysis.

Sr. No	High level aspect category/class	Low level aspects categories/classes
1	Teacher	Subject matter knowledge Experience Behaviour General
2	Course	Objective & Goals Course content Assessment General
3	University	Environment Policy General

3.3.3. LDA2Vec

Furthermore, we used the LDA2Vec word embedding variant, which is obtained by modifying the skip-gram word2Vec variant. We used fifty iterations and ten number of components to extract the topics from the student feedback.

3.3.4. SS-LDA

Similarly, we also used a semi-supervised LDA (SS-LDA) model that supports multiple-topic labels per document (Hughes et al., 2018). This improvement enables the alignment of the resulting model with human expectations for topic modelling and extraction. We also fine-tuned and adjusted hyper-parameters of SS-LDA variant models. Like the LDA2VEC, we used fifty iterations and ten components to extract the robust topics from the student feedback. The complete details of fine-tuned hyper-parameters of the LDA2VEC model are reported in Table 1.

3.3.5. Zero shot learning

To validate and compare the performance of LDA models, we used the zero-shot learning technique for topic extraction (Bianchi, Terragni, Hovy, Nozza, & Fersini, 2020). This technique recently comes from transfer learning, and its testing and training samples are independent of each other. We also adjusted hyper-parameters of zero-shot learning according to our dataset. We used fifty iterations and fifty number of components for the topic extraction. Furthermore, the complete hyper-parameters values are shown in Table 1. Furthermore, the experimental results in terms of extracted topics, their relevant terms and coherence score by using LDA, SS-LDA, LDA2VEC, and zero-shot learning are shown in Tables 7, 5, 4, and 6 respectively.

The extracted high and low-level aspects by using LDA and co-occurrence analysis are shown in Table 2. Two domain experts analyse all these aspects. One expert has expertise in English linguistics, and the other has expertise in text mining.

3.3.6. High-level aspects mapping

We have designed and implemented an algorithm to verify the high-level aspects extracted by LDA with co-occurrence analysis using Algorithm 1. In Algorithm 1 we used aspect terms extracted by LDA and its different variants with co-occurrence analysis and scanned each student's feedback using a loop and matched the terms and their synonyms terms. Synonyms terms are collected for each term manually and stored all relative synonyms terms in separate files. We also used Part of Speech (POS) tagging to map each noun and adjective to assign specific aspect terms to the relevant comments, which are shown in

	Feedback	POS.Tagging
1	The classrooms of my university are small and not well venti...	DT | NNS | IN | PRP | NN | VBP | JJ | CC | RB | RB | VBN
2	This course give us limited knowledge	DT | NN | VB | PRP | JJ | NN
3	University arranging very less Events for Students	NNP | VBG | RB | JJR | NNS | IN | NNS
4	Some course becomes difficult to understand because of te...	DT | NN | VBZ | JJ | TO | VB | IN | IN | NNS
5	There is need an improvements in courses	EX | VBZ | MD | DT | NNS | IN | NNS
6	Sometimes teachers discourage students from asking the ...	RB | NNS | JJ | NNS | IN | VBG | DT | NN

Fig. 8. Part of speech tagging for each comment.

Fig. 8. Each effective term is mapped to each corresponding feedback and stored in the high-level aspects column. Results of Algorithm 1 shown in results and discussion sections.

3.3.7. *Sentiment extraction and mapping*

We developed and tested a rule-based sentiment classification algorithm, as shown in Algorithm 3, and it was used to assign the sentiment labels according to sentiment orientation.

Algorithm 1: High Level Aspect Extraction & Mapping

h **Data:** Student Feedback Corpus
Result: Sentences Co-relation Aspects
initialization;
$Pre - processed\,StudentFeedbackCorpus \leftarrow$ LowerCaseConversion&RemovePunctuation&RemoveSymbols;
$Topics \leftarrow$ LDA + co − occurrenceanalysis(Pre − processedStudentFeedbackCorpus);

$ListofAspect \leftarrow$ StudentFeedbackAspects;
for *each "Feedback"* **do**
  Topic Aspect Mapping;
end
for *each "Aspect[i]"* **do**
  Scan Feedback text lines;
  if *line contains "Aspect Words"* **then**
    $HighLevelAspect[i] \leftarrow$ write(line);
  else
    skip the lines;
  end
end

We used sentiment dictionary (Hu & Liu, 2004) consisting of 2006 positive terms and 4783 negative terms to calculate each feedback's sentiment scores. Using the above sentiment lexicon, we calculated the sentiment score, and according to the sentiment score, we assigned positive, negative, and neutral class labels against each feedback. The sentiment orientation according to high and low level aspects is shown in Figs. 18 and 19 respectively. We also calculated the sentiment score of each feedback using the "afinn", and "syuzhet (Jockers, 2017)" lexicon. Still, we analysed that the "bing" lexicon gave more accurate results than others. After the tagging with each lexicon approach, our class labels were verified and modified by two experts' recommendations. One has a background in English linguistics, and the other has an experience with NLP. So, at last, we computed the sentiment score by subtracting the score of negative words from the score of positive words from each feedback using Eq. (4) and implemented this equation in R and with the negation handling process. Results of Algorithm 3 are shown in the results and discussion section.

3.3.8. *Low-level aspects mapping*

The second layer of our aspect-specific sentiment analysis approach is to extract low-level aspects extraction. For this module, we used LDA with its different variants and co-occurrence analysis with seed words and synonyms provided by the two annotators with a background in the education sector. We used the second Algorithm 2 with minor changes to map the low-level aspects correlated with high-level aspects and stored them in the third column of the data frame. Working of low-level aspects mapping is shown in Algorithm 2 and results are shown in the results and discussion section.

Algorithm 2: Low Level Aspect Extraction & Mapping

h **Data:** Student Feedback Corpus
Result: Sentences Co-relation Aspects
initialization;
$Pre - processed\,StudentFeedbackCorpus \leftarrow$ LowerCaseConvertion&RemovePunctuation&RemoveSymbols;
$Topics \leftarrow$ LDA + co − occurrenceanalysis(Pre − processedStudentFeedbackCorpus);

$ListofAspect \leftarrow$ StudentFeedbackAspects;
for *each "Feedback"* **do**
  Topic Aspect Mapping;
end
for *each "Aspect[i]"* **do**
  Scan Feedback text lines;
  if *line contains "Aspect Words"* **then**
    $LowLevelAspect[i] \leftarrow$ write(line);
  else
    skip the lines;
  end
end

$$Senti\_Score = \sum_{i=1}^{n} Ovrallattitude(i) \tag{4}$$

where i represents the specific word in a sentence and n represents the number of positive and negative words. The overall attitude is the product of word attitude and the word frequency in particular feedback.

3.4. *Features extraction and selection for sentiment classification using ML*

After dividing our dataset into a training set, validation set, and test set in the ratio of 60%, 20%, and 20%, respectively. We encoded textual data into a numeric format to do processing on text. The numeric vectors represent the feature representation of words or sentences. For feature extraction we used various word embedding techniques, includes Bag-of-Words (BOW) (Levy & Goldberg, 2014), TF-IDF, Word2Vec (Goldberg & Levy, 2014) to train machine learning models used for sentiment classification.

Algorithm 3: Sentiment Extraction & Mapping

```
Data: Aspect Oriented Feedback's
Result: Sentiment Orientation
SentimentExtractor() ← File;
initialization;
total_positive, total_negative, total_neutral = 0 ;
for each "feedback[i]" do
    Calculate_PositiveScoer(feedback[i]);
    Calculate_NegativeScore(feedback[i]);
    if feedbakc[i] == NegationWord_list then
        interchange Positive and Negative Score;
    end
    Sentiment_Score ← Calculate_Overall_SentiScore();
    if Sentiment_Score[feedback[i]] > 0 then
        feedback[i] ← Positive;
        total_positive ++;
    else
        if Sentiment_Score[feedback[i]] < 0 then
            feedback[i] ← Negative;
            total_negative ++;
        else
            feedback[i] ← Neutral;
            total_neutral ++;
        end
    end
end
```

Algorithm 4: Features modifications for sentiment classification

```
Data: Data_train + Data_test
Result: Modified reduced feature space
DataPre − processing() ← Data_train;
vocabulary_of_terms() ← TF − IDF_vectorizer(Data_train);
a = FSS_Score(t_j) ← DFS(t_j, C_j);
if a == Maximum_threshold then
    C_j ← t_j;
    N ← Total_No_of_features(C_j);
    if N == Total_No_of_features then
        C_j ← t_j;
        N ← Total_No_of_features(C_j);
    end
end
```

3.4.1. DTF/DFM

We created a bag of words (BOW) using the function DFM. Document Feature/Frequency Matrix (DFM) (Choi, Kwon, & Kim, 2004) function took our pre-processed tokens and created a matrix. Document frequency matrices use the bag of words model, took our free-form textual data, and ran it through a pre-processing data pipeline consisting of stop words removal, tokenization, and stemming. However, there is some issue in the DTF model (Wilbur & Kim, 2009) in which longer documents have a higher terms count, and those terms frequently appearing across the corpus are not as significant. So, we improved our model accuracy by document normalization based on their length and penalizing the terms that often occur across the corpus. So, we used the TF-IDF feature to get more accurate results for our model's more precise and accurate training

3.4.2. Term Frequency (TF)

In-text mining Vector Space Model (Jing, Ng, & Huang, 2010) (VSM) is a well-known representation technique for textual data. VSM is an algebraic model in which textual data can be converted into a vector of numbers instead of string or text using the Term Frequency (TF) and Inverse Document Frequency (Robertson, 2004) (IDF). Using this method, we evaluated the importance of the word in a particular document. We calculated TF-IDF using Eq. (5).

$$tf(t,d) = \frac{freq(t,d)}{\sum_i^n freq(ti,d)} \tag{5}$$

In the Eq., (5) $freq(t,d)$ is the function that counting the instances of the term t in document d, and $tf(t,d)$ is the function that calculates the proportion of the count of the term t in document d.

3.4.3. Inverse term frequency (IDF)

In Eq. (6) N represents the count of different documents in the corpus, and $count(t)$ is the function that counts the documents in the corpus in which the term t is present.

$$idf(t) = log(\frac{N}{count(t)}) \tag{6}$$

3.4.4. TF-IDF

In Eq. (7), we combined the TF and IDF and enhanced the document term frequency matrix, which is further used as a feature vector during training and testing of the sentiment analysis model.

$$tf - idf(t,d) = tf(t,d) * idf(t) \tag{7}$$

3.4.5. Word2Vec

Word2Vector word embedding technique (Ge & Moh, 2017) converts textual data as a multidimensional array. We used CBoW and Skip-gram unsupervised models to generate word embedding. We used Gensim library (Vorontsov, Frei, Apishev, Romov, & Dudarenko, 2015) from python, which has many pre-trained models for training models and extracting word vectors.

3.5. Features modifications for sentiment classification

In this research work, we have used variable global feature selection and variable stop words filtering for prominent features modification for better sentiment classification. We extracted lexicon-based features (Zou, Gui, Zhang, & Huang, 2018) from our text using (Hu & Liu, 2004) lexicon. We also modified this lexicon according to the student and teacher domain's context and added some words that were not in the original lexicon. We also did modifications in polarities of different words.

3.5.1. Negation words

In the proposed approach, whenever a negation word is encountered in feedback, the opinion score is reversed by a certain amount as good (+1) to not good (−1). We reversed that particular sentence's polarity by identifying (no, not, neither, nothing, never, none). This method preserved sentence consistency in a better way.

3.5.2. Blind negation words

In the proposed system, we also tried to handle some blind negation words (e.g., Need, needed, require, required). These words in the sentence have a significant impact, for example, "He needs to prepare his lecture", so we converted its label into negative and assigned a (−1) score to this sentence.

3.5.3. Resolving sparsity issue

Characterizing textual data using BoW produced sparsity, which leads the model to produce low accuracy, especially when dealing with unbalanced data. Hence, we used the variable stop words filtering technique to resolve sparsity issues created by using BoW. Variable stop word filtering relates to the fact that the stop words list is collected from the data and is not pre-determined. According to our data's vocabulary analysis, 83% of the terms appeared to have less than five times in the corpus. A closer examination of these terms revealed that almost all terms with a frequency of fewer than five are unimportant; hence these terms can be ostracized from the feature space.

3.5.4. Resolving high dimensionality issue

To select the most prominent features to reduce the feature space, we used the variable global feature selection technique as described in Algorithm 4.

In this algorithm pre-processing is applied to the training and testing dataset. Vocabulary V of terms is generated by applying TF-IDF vectorizer to the training dataset. The distinguishing features are selected from each term V to compute the final feature score. Based on a maximum number of local feature scores of each feature t_j, it is assigned to class C_j. After that, the total number of the feature of class C_j is calculated and assigned to N. Final Feature Set (FFS) comprising of N features is shaped by eliminating the features other than the variable split of the class Cj from the vocabulary of terms V as shown in Algorithm 4. This method selects the most relevant features from each dataset class, which minimizes the feature space and improves the machine learning classifiers' performance.

3.6. Machine learning models used for sentiment classification

3.6.1. SVM

Support vector machine classifier is a supervised learning model that attempts to find the optimal hyperplane separating two different classes of data. It is a very efficient algorithm for solving linear and nonlinear problems and is particularly applicable to large and high-dimensional classification problems (Goudjil, Koudil, Bedda, & Ghoggali, 2018). The main theme of SVM is to create a line known as a hyperplane to separate the data into different classes. It first uses the kernel method for unstructured data, transforming data into 2D or 3D to perfect the hyperplane. As described in Table 3, we used polynomial and linear kernels and seven iterations for sentiment analysis. The complete hyper-parameters details of the SVM are shown in Table 3

3.6.2. KNN

We also used a non-parametric supervised machine learning algorithm for the context of text classification (Tan, 2005). As described in Table 3, we used Brute force and KD Tree as a parameter and three numbers of nearest neighbours according to our classification task for three topics.

3.6.3. Random forest

We used Random Forest (RF) classifiers as it is suitable for dealing with the high dimensional noisy data in text classification (Salles, Gonçalves, Rodrigues, & Rocha, 2018). We used fifty trees and controlled the randomness of the bootstrapping of the samples used when building trees. Table 3, reported the parameters values for the RF according to our dataset.

3.6.4. RPart

We have implemented Classification and Regression Tree (CART) algorithm for sentiment classification. It is binary tree algorithm and the R implementation of this algorithm is known as Rpart. The cart method of Denison, Mallick, and Smith (1998) and Liu, Wang, and Zhang (2012) solves the regression and classification problems. It create multiple decision trees that follow the splitting rule based on the forecaster variable.

3.6.5. ANN

We used ANN (Prasanna & Rao, 2018) by having a layer of input neurons where we feed in our feature vectors, and the values are then fed forward to a hidden layer. At each connection, we feed the value forward, which also contains multiple biases and weights. We adjust the hyper-parameters of ANN according to our classification task for three topics, which are described in Table 3.

Table 3
Summary of hyper-parameters of machine learning algorithms for text classification.

Algorithm	Parameters	Values
SVM	Kernel	Polynomial
	No. of iteration	7
	Degree	3
	C	3
	Gamma	1
	Random state	None
SVM	Kernel	Linear
	No. of iteration	7
	Degree	3
	C	3
	Gamma	1
	Random state	None
KNN	Algorithm	Brute force, KD Tree
	Leaf size	10
	No. of K neighbours	3
	Weights	Uniform
	Metric	Correlation
RF	No. of tree	50
	Random state	0
	Maximum depth	8
	N estimators	10
ANN	Sequential3 dense layers	17942826 Trainable Parameters
	Batch size	128
	Epochs	7
	Verbose	0
	Validation split	0.2
CART	Criterion	MSE
	Maximum depth	8
	Random state	None
	Min. samples split	3
	Splitters	Best

4. Results and discussion

In our proposed framework of Aspect2Labels (A2L) for Higher Educational institutions (HEIs), we performed Educational Data mining (EDM). To transform the teaching–learning process into a text mining domain, we used students' textual recommendations to perform aspect-based sentiment analysis to uncover the emotions and sentiments of students towards their teachers or intuitions. We used different state-of-the-art topic modelling techniques to explore students' aspects terms that they have expressed in the student feedback survey activity.

We used five machine learning models; Artificial Neural Network (Arras, Montavon, Müller, & Samek, 2017), Random Forest (Al Amrani, Lazaar, & El Kadiri, 2018), Recursive Partitioning (Van, Thai, & Nghiem, 2017) and Regression Trees (rpart), Support Vector Machine (Zainuddin & Selamat, 2014) (SVM) by linear kernel, and Support Vector Machine (SVM) by a polynomial kernel for aspect-based sentiment classification. As we mentioned before, our dataset was collected by us from the Quality Enhancement Cell (QEC) of NCBA&E University Rahim Yar Khan campus. From this feedback data, we only select 6500 students reviews from the computer science department of NCBA&E university. For our experiments, we used students' feedback from fall semester 2016 to fall-spring semester 2020. The first part of the result section provided Aspect-Based Sentiment Analysis (ABSA) results. The second part explained our sentiment classification results using machine learning algorithms.

We have divided our A2L process into a three-layers. Using the first layers of A2L, we extracted high-level aspect terms using LDA with different variants and co-occurrence analysis techniques. Our proposed rule-based algorithm performed mapping or annotation of high-level aspects terms as labels. We specified the number of topics 3 with ten passes and id2word vocabulary in the corpus as parameters of LDA. In the second attempt, we used Part of Speech (POS-tag) as "NN" to extract nouns only from text and created vocabulary according to

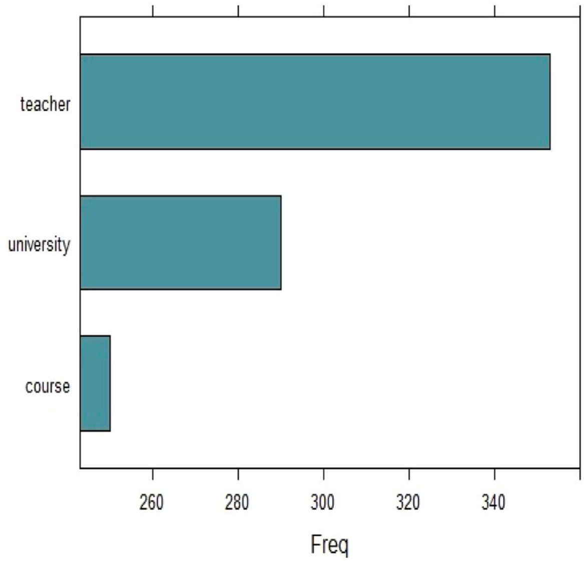


Fig. 9. Most occurring nouns using co-occurrence analysis.

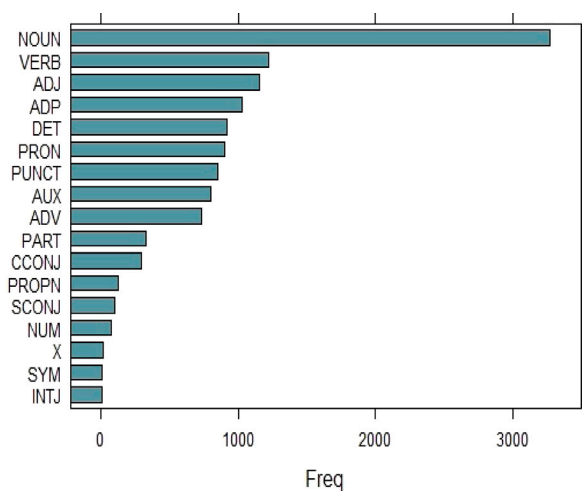


Fig. 10. Universal parts of speech frequency extracted using co-occurrence analysis.

extracted nouns. Most occurring nouns from students' feedback were extracted using co-occurrence analysis shown in Fig. 9. Fig. 10 shows the universal part of speech extracted from students' feedback using the co-occurrence analysis process. Co-occurrence analysis results in the form knowledge graph using Pointwise Mutual Information (PMI), which shows the relationship between nouns and adjectives can be seen in Fig. 11.

Results of the first layer are shown in Table 8 in which the first column of the table consists of student feedback and the second column shows the high-level aspects terms according to three topics which are Teacher, Course, and University suggested by LDA and co-occurrence analysis using id2word vocabulary from our corpus. Extracted topics, keywords, and topic contribution percentages along representative text are shown in Fig. 12. We also demonstrated topic modelling techniques results in the form of word cloud shown in Fig. 14, and extracted topics along correlated terms results shown in Fig. 13. Furthermore, we used a coherence score for accessing the quality of learned topics through different models. We used a coherence score as it is essential to find the optimal number of topics because these topics will ultimately convert into labels for annotating the dataset.

Tables 7, 4, 5, and 6 reported the extracted topics with relevant terms and coherence scores using LDA Gensim, LDA2Vec, SS-LDA, and

Table 4
Topics with their relevant terms and coherence score using LDA2Vec.

Topics	Relevant terms	Coherence value
Topic 0	Behaviour	0.056
	Method	0.029
	Satisfaction	0.023
	Capability	0.017
Topic 1	Curriculum	0.038
	Lab work	0.032
	Learning material	0.026
	Assignment	0.019
Topic 2	Library	0.061
	Workshop	0.022
	Scholarship	0.020
	Cafeteria	0.018

Table 5
Topics with their relevant terms and coherence score using SS-LDA.

Topics	Relevant terms	Coherence value
Topic 0	Capability	1.726
	Method	0.188
	Behaviour	0.116
	Skill	0.074
Topic 1	Course content	1.617
	Lab work	0.260
	Learning material	0.061
	Objective	0.033
Topic 2	Cafeteria	1.588
	Library	0.252
	Scholarship	0.063
	Concert	0.055

zero-shot learning respectively. To handle the instability of the LDA model, we have used other different topic modelling approaches like SS-LDA, LDA2Vec and zero-shot learning for the topic extraction from student feedback data. As shown in Table 1, the details of hyper-parameters of all these models are reported, and similarly, we can see the results of these models in terms of extracted topics in Tables 7, 4, 5, 6 by using LDA Genism, LDA2Vec, SS-LDA, and zero-shot learning respectively. These results concluded that models like zero-shot learning and different variants of LDA almost extracted the same topics which LDA Gensim extracts in its first run. The LDA is one of the most robust topic modelling approaches, and as we can see in Table 7, it extracted more relevant and robust topics according to our student textual feedback data. Furthermore, to handle the instability of the LDA model, we used the stability method proposed by Mantyla, Claes, and Farooq (2018), in which we select k number of topics, replicate the runs of LDA n times, and hence LDA generate topics as shown in Eq. (8).

$$Topics = n * k \tag{8}$$

Further, we make a cluster of $n * k$ topics into k number of clusters by using K-medoids and Rank-Biased Overlap as stability metrics. Finally, each k cluster represents extracted LDA topics with stability. This approach makes LDA stability transparent and is complementary rather than an alternative to many prior works focusing on LDA parameter tuning. Hence, we resolved the instability issue of the LDA model and used LDA Genism as our topic modelling approach.

We also visualized the result of high-level aspects using pyLDAvis visualization. PyLDAvis (Sievert & Shirley, 2014) is a web-based interactive python library that envisioned topics estimated by using Latent Dirichlet Allocation. The package extracts information from a fitted LDA topic model to inform an interactive web-based visualization. As shown in Fig. 15, we visualized the pyLDAvis visualization for high-level aspect teachers. Our visualization in Fig. 15, has two sections. The left side of our visualization shows a high-level topic view of the topic model. In this view, we plotted the topics as circles in the two-dimensional, and then we used multidimensional scaling to project the

Fig. 11. Co-occurrence within a sentence as knowledge graph using nouns and adjectives.

	Topic_Num	Topic_Perc_Contrib	Keywords	Representative Text
0	0.0	0.9251	course, learn, study, teach, program, interesting, achieve, skill, understand, help	[learn, research, base, course, dip, machine, learning, vision]
1	1.0	0.9390	teacher, empower, overlook, often, bad, due, love, give, job, time	[teacher, bad, habit, tell, teacher, thing, tell, parent, exact, opposite]
2	2.0	0.9209	university, provide, experience, big, opportunity, develop, game, offer, staff, library	[educational, institution, play, key, role, city, region, operate]

Fig. 12. Extracted topics and their relevant terms from LDA.

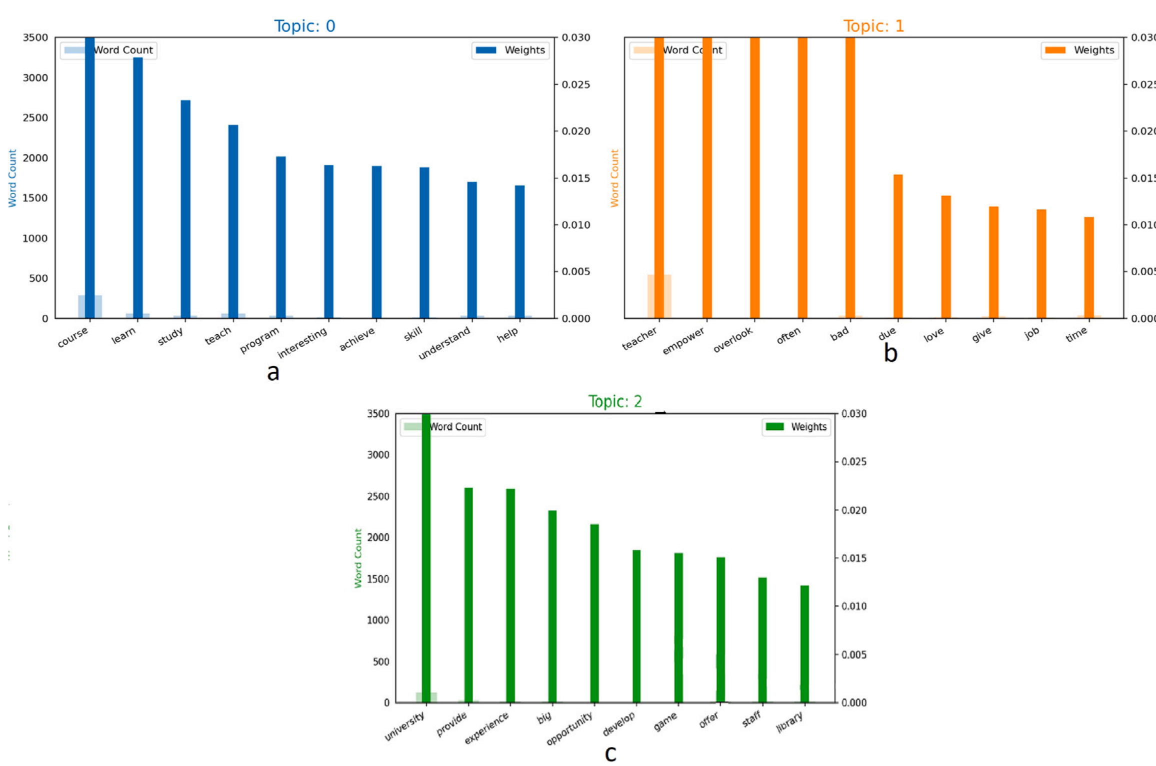


Fig. 13. Word count and aspects terms with correlated terms.

Fig. 14. Word cloud depiction of high-level aspects (topic) terms along low-level aspect terms.

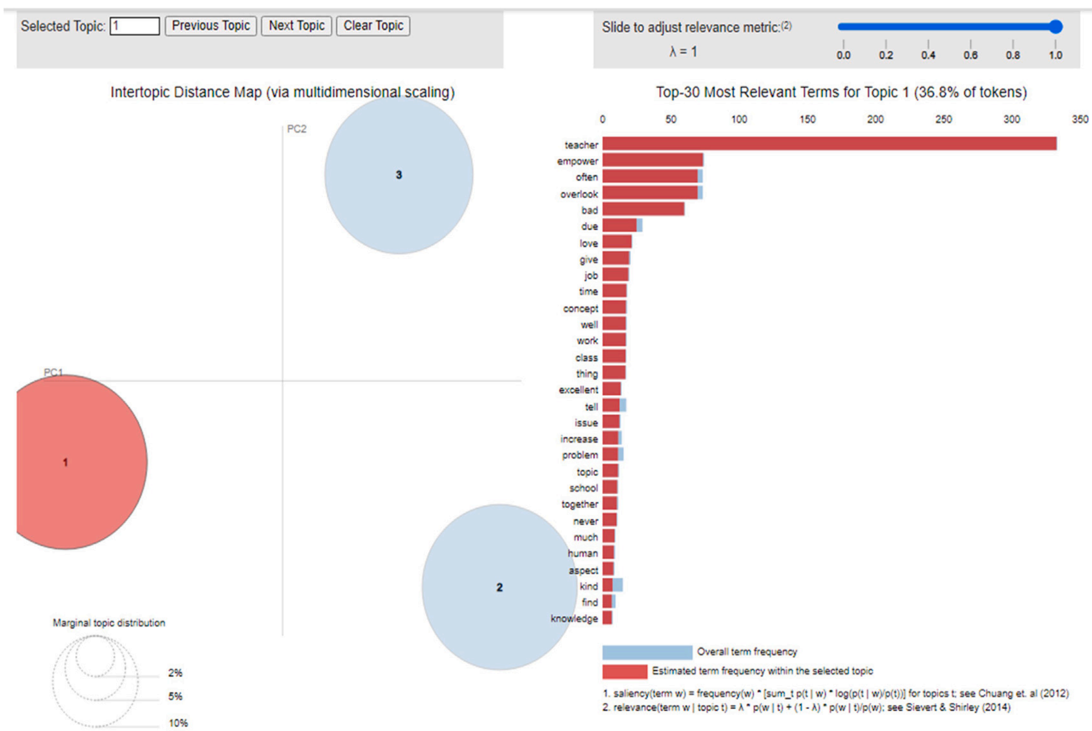


Fig. 15. The visualization of high-level aspect term: 'teacher' as topic view using pyLDAvis on the left, and the associated terms and topic on the right.

Table 6
Topics with their relevant terms and coherence score using Zero Shot learning.

Topics	Relevant terms	Coherence value
Topic 0	Behaviour	0.056
	Skill	0.029
	Satisfaction	0.188
	Capability	0.116
Topic 1	Achievement	1.617
	Helpful	0.032
	Learning outcome	0.026
	Goals	0.019
Topic 2	Fee	0.061
	Library	0.022
	Cafeteria	0.020
	Auditorium	0.015

Table 7
Topics with their relevant terms and coherence score using LDA Genism.

Topics	Relevant terms	Coherence value
Topic 0	Capability	1.726
	Satisfaction	0.188
	Behaviour	0.056
	Skill	0.017
Topic 1	Curriculum	0.038
	Lab work	0.032
	Way of Teaching	0.026
	Knowledgeable	0.019
Topic 2	Library	0.061
	Workshop	0.022
	Cafeteria	0.020
	Auditorium	0.015

inter-topic distances onto two dimensions. While on the right side of the visualization, a horizontal bar chart represents the individual terms that are the most useful for interpreting the selected topic on the left side. The layout of LDAvis, shows the high-level topic view on the left side, and the most relevant terms are shown as a bar chart on the right side of the figure by selecting thirty terms. It is constructive for any organization to visualize what Situation information the students give.

Similarly, we plotted the pyLDAvis visualization of the high-level aspect course. Fig. 16, shows a high-level topic course extracted from student's feedback, and relevant terms are also visualized in this Figure.

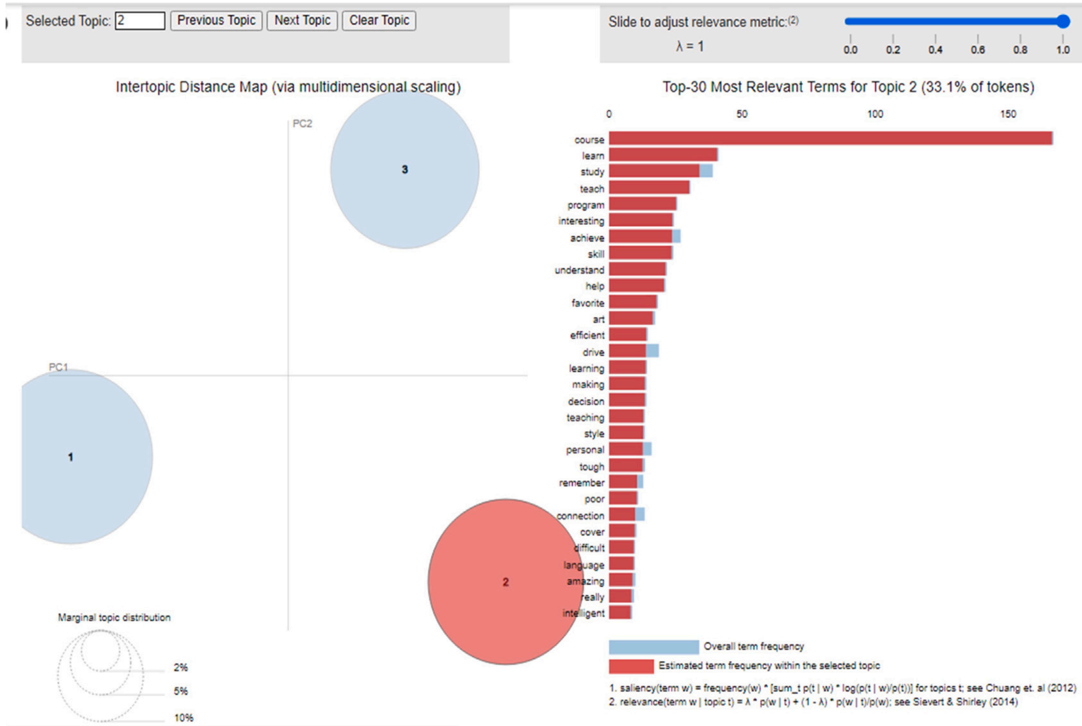


Fig. 16. The visualization of high-level aspect term: 'course as a topic view using pyLDAvis on the left, and the associated terms along topic on the right.

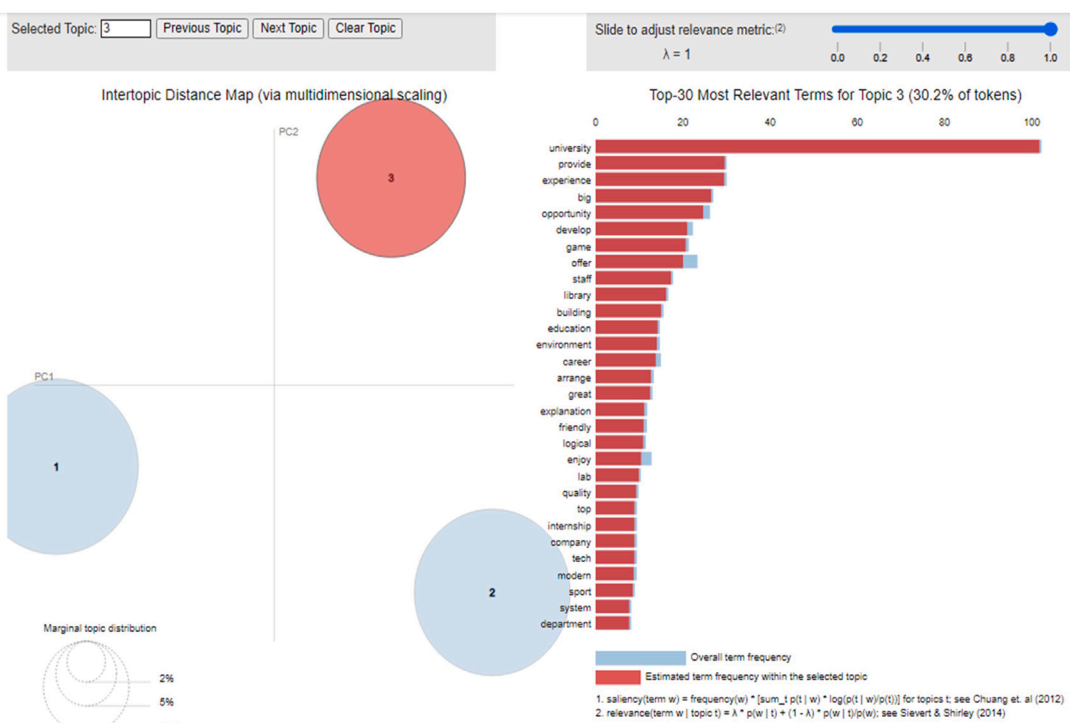


Fig. 17. The visualization of high level aspect term 'university' as topic view using pyLDAvis on the left, and the associated terms along topic on the right.

A similar visualization approach is used by the authors of Sievert and Shirley (2014) to provide a global view of the topics while at the same time allowing for a deep inspection of the terms most highly associated with each topic. In addition, they used LDAvis visualization to fully understand the fitted LDA model, allowing users to explore topic-term relationships flexibly. Finally, just like Fig. 16, we visualized the pyLDAvis visualization of high-level aspect university, which is shown in Fig. 17, the high-level topic university and relevant terms are extracted from student's feedback and also visualized in this pyLDAvis visualization. Further, we also visualized the sentiments of student's comments for each high-level aspect and low-level aspect, which are shown in Figs. 18 and 19 respectively.

To extract low-level aspects from students' feedback, we altered the vocabulary of our corpus. We used the nltk library in python

Table 8
High-level aspect extraction and mapping with relevant feedback.

	Feedback	High level aspect
1	Teacher having accurate and up-to-date knowledge in his lectures.	Teacher
2	Teacher provide effective examples with real-time scenarios to understand his course.	Teacher, course
3	Teacher behaviour atrocious for all students.	Teacher
4	This university is not best for students.	University
5	University should make policy to conduct extra activities rather than studies.	University
6	University environment is not good for students.	University
7	I am unable to understand the objective and goals of this course.	Course
8	I am unsatisfied with the content of this course.	Course
9	I like the course programming	Course
10	Teacher did not do much explanation of the topics.	Teacher
11	Teacher's method is very bad for practical explanation.	Teacher
12	The teacher experience for teaching is not good.	Teacher
13	Sometimes the teacher could not arrive in class on time.	Teacher
14	Need so much improvement in cafeteria environment of university.	University
15	The management of university is so good they have managed all the things in efficient manners.	University
16	University should focus on making more labs and improve the current labs.	University
17	There is Just a name of internet servers but have no access, it should be improved.	University
18	Policy for the course selection is bad.	Course
19	Every course gives us a skill to achieve new goals.	Course
20	Some course becomes difficult to understand because of less knowledge in the domain of teachers	Teacher, course
21	The course content not up to the mark.	Course

Table 9
Low level aspect extraction and mapping according to high-level aspects.

	Feedback	High level aspect	Low level aspect
1	Teacher having accurate and up-to-date knowledge in his lectures.	Teacher	Subject matter knowledge
2	Some course becomes difficult to understand because of less knowledge in the domain of teachers	Teacher, course	Subject matter knowledge
3	Teacher behaviour is atrocious for all students.	Teacher	Behaviour
4	University is not best for students.	University	General
5	University should make policy to conduct extra activities rather than studies.	University	General, policy
6	University environment is not good for students.	University	Environment
7	I am unable to understand the objective and goals of this course.	Course	Objective and goals
8	I am unsatisfied with the content of this course.	Course	Course content
9	I like the course programming	Course	General
10	There was no such material and explanation given by teacher.	Teacher	Methods
11	Teacher's method is very bad for practical explanation.	Teacher	Methods
12	The teacher experience for teaching is not good.	Teacher	Experience
13	Sometimes the teacher could not arrive in class on time.	Teacher	General
14	Need so much improvement in cafeteria environment of university.	University	Environment
15	Policy for the course selection is bad.	Course	Policy
16	Every course gives us a skill to achieve new goals.	Course	Objectives & goals, content
17	Some course becomes difficult to understand because of less knowledge in the domain of teachers.	Teacher, course	Subject matter knowledge
18	The course content not up to the mark.	Course	Course content

and Wordnet corpus to match synonyms, hypernyms, holonyms, and meronyms. In Wordnet, all these terms are defined, and we created a corpus vocabulary according to part of speech that we needed to extract low-level topics or aspects. We used expert opinion to determine the seed word for low-level aspects extraction. We successfully extracted and mapped low-level aspects using the algorithm mentioned above, and the results are shown in Table 9.

We used our proposed rule-based classifier to calculate the sentiment score from each comment, and we assigned positive, negative, and neutral labels against each comment according to sentiment polarity. The results of sentiment labelling using the Algorithm 3 are shown in Table 10. Two domain experts analysed all labels. One expert has expertise in English linguistics, and the other has expertise in text mining. After refining our data, we used VGFS for dimension reduction to reduce the feature space and avoid the sparsity problem, and we used variable stop words filtering. We build our classifiers using five different classification algorithms. Moreover, our multi-core parallel processing technique increased efficiency and decreased the overall processing time.

4.1. Improvement of performance by resolving the sparsity and high dimensionality problems

We used Variable Stop Words Filtering (VSF) to reduce the sparsity problem in our dataset, as described in the above section. After training the SVM and ANN, we got improved accuracy over the baseline model, as shown in Table 21. This improvement eliminates low-frequency words from the feature space, which helps the models train more effectively. Similarly, after resolving the problem of high sparsity, we used the VGFS technique to select the more distinguishing features to shrink the feature space. This process minimizes the feature space and improves the machine learning classifiers. After applying the global variable feature selection technique, we again train our SVM and ANN algorithms and perform testing, and hence, we got more accurate results as shown in Table 21. Moreover, we plotted the confusion matrix to depict ANN and SVM model's performances on test data as shown in Figs. 20 and 21 respectively.

Table 10
Results of proposed Algorithm 3 for sentiment orientation and mapping.

	Feedback	High level aspect	Low level aspect	Sentiment orientation
1	Teacher behaviour atrocious for all students.	Teacher	Behaviour	Negative
2	University should make policy to conduct extra activities rather than studies.	University	General, policy	Neutral
3	University environment is not good for students.	University	Environment	Negative
4	I am unable to understand the objective and goals of this course.	Course	Objective and goals	Negative
5	I am unsatisfied with the content of this course.	Course	Course content	Negative
6	I like the course programming	Course	General	Positive
7	There was no such material and explanation given by teacher.	Teacher	Methods	Neutral
8	Teacher's method is very bad for practical explanation.	Teacher	Methods	Negative
9	The teacher experience for teaching is not good.	Teacher	experience	Negative
10	Sometimes the teacher could not arrive in class on time.	Teacher	General	Negative
11	Need so much improvement in cafeteria environment of university.	University	Environment	Negative
12	University should policy on making more labs and improve the current labs.	University	Policy	General
13	Policy for the course selection is bad.	Course	Policy	Negative

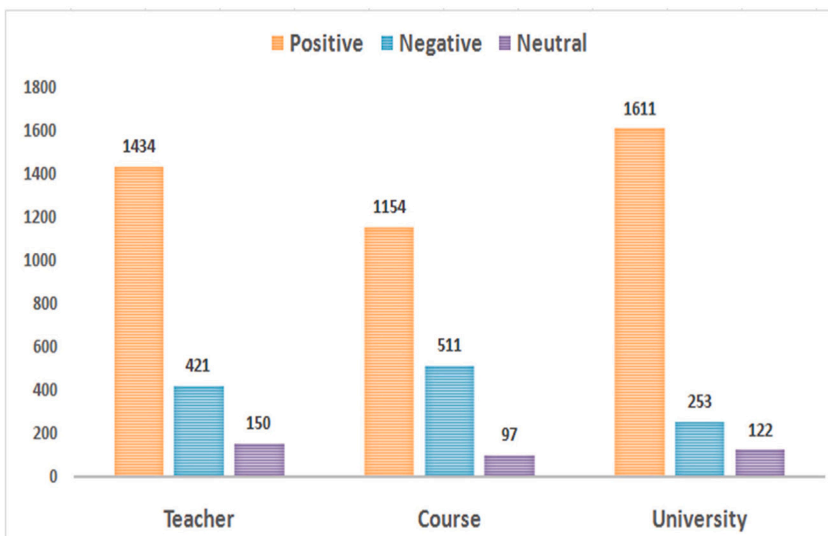


Fig. 18. Sentiment orientation according high-level aspects.

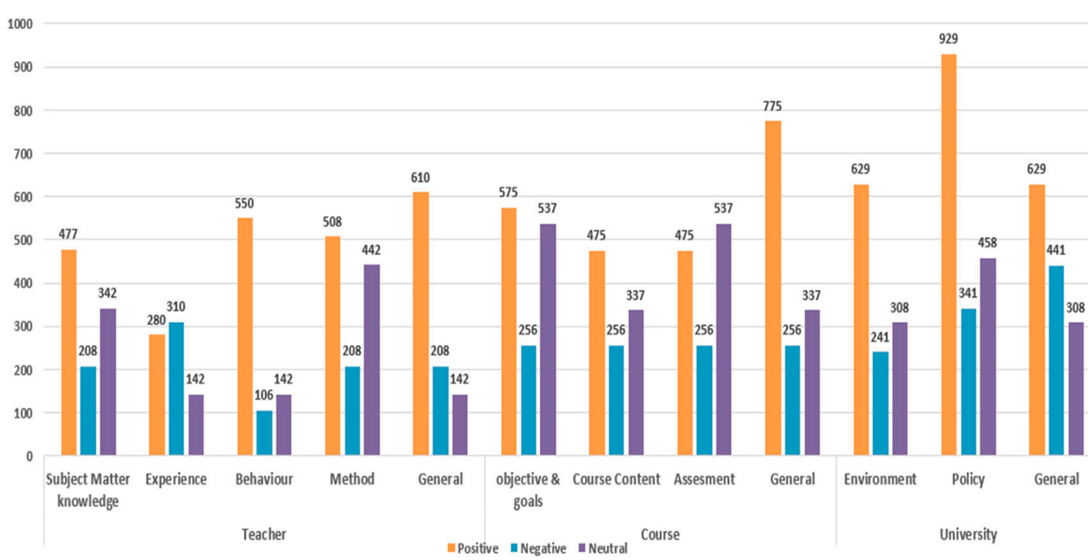


Fig. 19. Sentiment orientation according low-level aspects w.r.t high-level aspects.

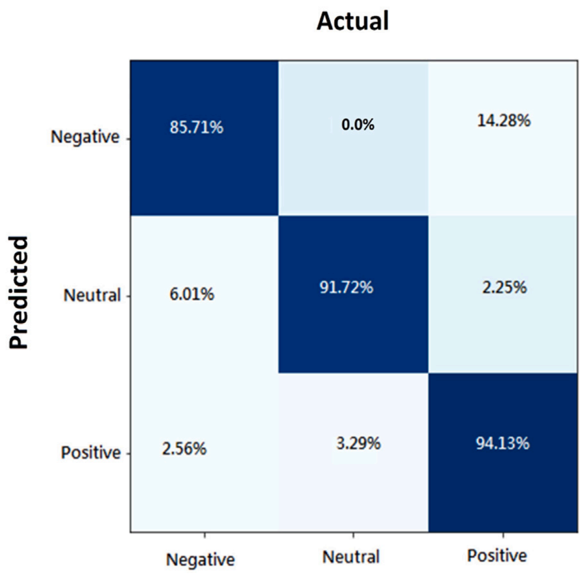


Fig. 20. Confusion matrix of ANN.

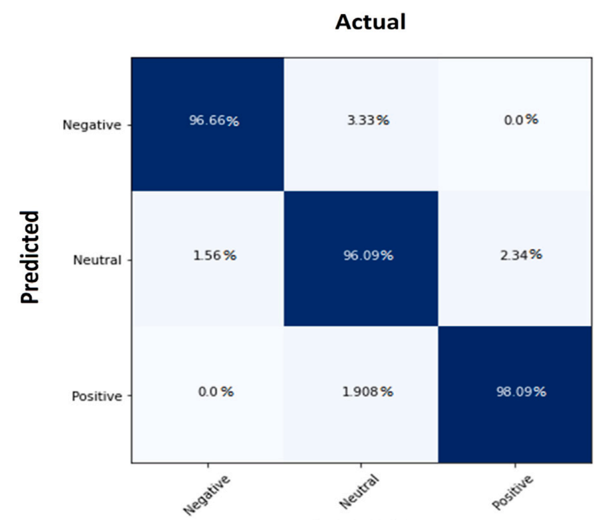


Fig. 21. Confusion matrix of SVM using linear kernel.

Accuracy: Accuracy is a model evaluation metric that indicates the correct prediction in percentage. We can calculate it from the confusion matrix using the following formula.

$$accuracy = \frac{Tp + Tn}{Tp + Tn + Fp + Fn} \tag{9}$$

Precision: It is also called positive predictive value. It is the fraction of related instances between the recovered instances. Precision is the ratio of system-generated results that correctly predicted positive observations "True Positive" to the total predicted positive observations.

$$precision = \frac{Tp}{Tp + Fp} \tag{10}$$

Recall (Sensitivity) : The recall is the evaluation metric in which the system's recall is the ratio of system-generated correctly predicted positive observations to all observations in the actual class.

$$Recall = \frac{Tp}{Tp + Fn} \tag{11}$$

F1 Measure : The F1 score is the weighted average or harmonic mean of precision and recall in the classification evaluation process. Moreover, it is used to gain a balance between recall and precision.

$$F1Measure = \frac{2 * (Recall * Precision)}{(Recall + Precision)} \tag{12}$$

However, Artificial Neural Network (ANN) outperformed after SVM linear and reported 93% accuracy, 93% Confidence Interval, and Cohen's kappa statistic 85%. The confusion matrix of ANN is shown in Fig. 21, the classification results calculated using the test dataset. Moreover, we calculated the comparative accuracy of five models on cross-validation data shown in Fig. 22, results showed that the SVM model performed better than other models. To check the class imbalanced data problem, we performed stratified ten-fold cross-validation using the createMultiFolds method from the caret package, repeated three times. We created 30 random samples, which gave us more robust estimates.

On the other hand, the comparative accuracy of models is also computed on the test dataset as shown in Fig. 23. As we discussed earlier, SVM performed very well with its linear kernel. We achieved 97% accuracy, 95% recall, and a precision 96.94% which is considered the best of all other previously reported approaches in the same field of sentiment analysis. Detailed information of SVM Linear classification results shown in Table 11.

Table 11
Statistics by classes using SVM via linear.

Performance metrics	Negative	Neutral	Positive
Sensitivity	0.93548	0.9531	0.9885
Specificity	0.99742	0.9828	0.9686
Pos Pred value	0.96667	0.9606	0.9809
Neg Pred value	0.99486	0.9795	0.9809
Prevalence	0.07399	0.3055	0.6205
Detection rate	0.06921	0.2912	0.6134
Detection prevalence	0.07160	0.3031	0.6253
Balanced accuracy	0.96645	0.9680	0.9785

Table 12
Statistics by classes using ANN.

Performance metrics	Negative	Neutral	Positive
Sensitivity	0.42308	0.9313	0.9809
Specificity	0.99491	0.9618	0.8981
Pos Pred value	0.84615	0.9173	0.9414
Neg Pred value	0.96305	0.9685	0.9658
Prevalence	0.06205	0.3126	0.6253
Detection rate	0.02625	0.2912	0.6134
Detection prevalence	0.03103	0.3174	0.6516
Balanced accuracy	0.70899	0.9466	0.9395

As we can see, SVM performs better than the other algorithms on our dataset. Liu, Lv, Liu, and Shi (2010) SVM can efficiently perform a non-linear classification of the kernel trick, implicitly mapping their inputs into high-dimensional feature spaces in student textual feedback classification, which has high dimensional input space. Joachims (1998) SVM use over-fitting protection, which does not necessarily depend on the number of features that have the potential to handle these large feature spaces. Furthermore, SVM does not require parameter tuning since it can automatically find good parameter settings from the training dataset. Hence SVM accuracy is better than other algorithms, and we also generalize this claim to other datasets as SVM is one of the most robust and accurate algorithms among the other classification algorithms. Moreover, our framework can work correctly for all textual datasets after applying the proposed feature modification techniques.

The second highest overall accuracy, 93.079%, is obtained using an artificial neural network, the Kappa statistic is 0.859, precision is 90.161%, and 77.84% recall. The detailed results of the Artificial Neural Network (ANN) are shown in Table 12.

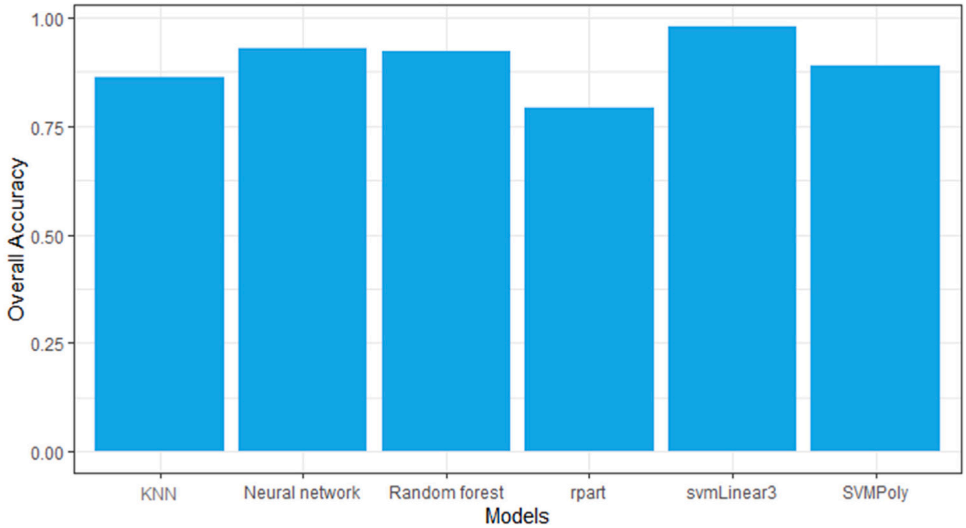


Fig. 22. Classification models performance on cross validation data.

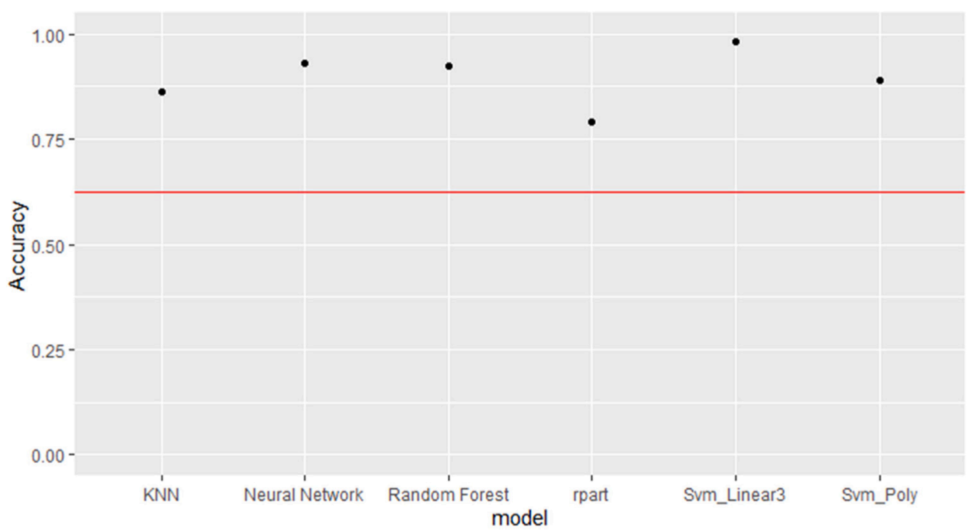


Fig. 23. Result of sentiment classification using different models on test dataset.

Table 13
Statistics by classes using random forest algorithm.

Performance metrics	Negative	Neutral	Positive
Sensitivity	0.76923	0.8168	0.9924
Specificity	1.00000	0.9896	0.8153
Pos Pred value	1.00000	0.9727	0.8997
Neg Pred value	0.98496	0.9223	0.9846
Prevalence	0.06205	0.3126	0.6253
Detection rate	0.04773	0.2554	0.6205
Detection prevalence	0.04773	0.2625	0.6897
Balanced accuracy	0.88462	0.9032	0.9038

Table 14
Statistics by classes using recursive partitioning and regression tree algorithm.

Performance metrics	Negative	Neutral	Positive
Sensitivity	0.87097	0.8047	0.9115
Specificity	0.98711	0.9278	0.8365
Pos Pred value	0.84375	0.8306	0.9011
Neg Pred value	0.98966	0.9153	0.8526
Prevalence	0.07399	0.3055	0.6205
Detection rate	0.06444	0.2458	0.5656
Detection prevalence	0.07637	0.2959	0.6277
Balanced accuracy	0.92904	0.8663	0.8740

Similarly, the Random forest algorithm gives us 92% accuracy, and the Kappa statistic is 0.842, precision is 95.746, and recall is 85.94%. More statistics by each class are shown in Table 13.

Recursive partitioning and regression tree algorithm gave us 87.589% accuracy; kappa statistic is 0.759, precision 85.85%, recall 86.24% and detailed statics of rpart shown in Table 14.

We also used a support vector machine algorithm with its polynomial kennel. SVM Poly gave us 89% accuracy, kappa score is 0.7802, precision 89.18% and recall 86.34% and more classification results shown in Table 15.

Finally, we evaluated the k-nearest neighbour's algorithm using the test data, which gave us the lowest performance. KNN gave us 86% accuracy, kappa statistic 0.7289, precision 89.134%, and recall score 75.191%. Statistics of KNN by class are shown in Table 16.

To measure the comparative performances of classification algorithms, we also calculated statistics by classes, as we have three positive, neutral, and negative classes. We used accuracy, precision, recall,

Table 15
Statistics by classes using support vector machine algorithm with its polynomial kernel.

Performance metrics	Negative	Neutral	Positive
Sensitivity	0.84615	0.8092	0.9351
Specificity	0.99491	0.9375	0.8344
Pos Pred value	0.91667	0.8548	0.9041
Neg Pred value	0.98987	0.9153	0.8851
Prevalence	0.06205	0.3126	0.6253
Detection rate	0.05251	0.2530	0.5847
Detection prevalence	0.05728	0.2959	0.6468
Balanced accuracy	0.92053	0.8733	0.8848

Table 16
Statistics by classes using k-nearest neighbours algorithm.

Performance metrics	Negative	Neutral	Positive
Sensitivity	0.50000	0.8473	0.9084
Specificity	1.00000	0.8750	0.8662
Pos Pred value	1.00000	0.7551	0.9189
Neg Pred value	0.96798	0.9265	0.8500
Prevalence	0.06205	0.3126	0.6253
Detection rate	0.03103	0.2649	0.5680
Prevalence	0.03103	0.3508	0.6181
Accuracy	0.75000	0.8612	0.8873

Table 17
Comparative accuracy of models according classes.

ML Models	Positive	Negative	Neutral
rpart	0.777739	0.659082	0.758137
SVM-Linear	0.991090	0.941035	0.981950
SVM-Poly	0.884755	0.920532	0.873330
KNN	0.887320	0.750000	0.861164
Random forest	0.903827	0.884615	0.903189
Neural network	0.939503	0.708994	0.946552

Table 18
Comparative recall of different models according classes.

ML Models	Positive	Negative	Neutral
rpart	0.950382	0.346154	0.564886
SVM-Linear	0.988550	0.884615	0.984733
SVM-Poly	0.935115	0.846154	0.809160
KNN	0.908397	0.500000	0.847328
Random forest	0.992366	0.769231	0.816794
Neural network	0.980916	0.423077	0.931298

Table 19
Comparative precision of models according classes.

ML Models	Positive	Negative	Neutral
rpart	0.800643	0.450000	0.840909
SVM-Linear	0.996154	0.958333	0.955556
SVM-Poly	0.904059	0.916667	0.854839
KNN	0.918919	1.000000	0.755102
Random forest	0.899654	1.000000	0.972727
Neural network	0.941392	0.846154	0.917293

Table 20
Comparative F1-Measure of models according classes.

ML Models	Positive	Negative	Neutral
rpart	0.869110	0.391304	0.675799
SVM-Linear	0.992337	0.920000	0.969925
SVM-Poly	0.919325	0.880000	0.831373
KNN	0.913628	0.666667	0.798561
Random forest	0.943739	0.869565	0.887967
Neural network	0.960748	0.564103	0.924242

and f1-score computational matrices to evaluate models' comparative performance by classes. Comparative accuracy of each machine learning algorithms shown in Table 17 and recall is shown in Table 18 with highlighted colours.

Table 21
Models accuracy with and without feature modification.

A2L approaches	Models	Accuracy %
Baseline	ANN	89.23
	SVM	91.42
After Handling	ANN	92.64
Sparsity	SVM	95.87
After Handling	ANN	93.27
dimensionality	SVM	97.84

We also calculated comparative precision and F1-score using different machine learning algorithms and performances of algorithms shown in Tables 19 and 20 respectively.

Our proposed multi-core parallel processing approach gives us an efficient way to perform the analysis rapidly to attain efficiency from a time perspective. To perform parallel execution, we used the doSNOW package to run our training model similarly, that enhanced and sped up our processing time using multiple machines' multiple logical cores. Using this method, we assigned a specific fold to a particular core to process data simultaneously. We used the "makeCluster" function to create clusters of type sockets to accomplish parallel processing. The overall results are very satisfactory, and it could be a valuable addition for higher educational institutions to evaluate what situational information is floating in students' minds. Our proposed Situational Awareness System using hybrid natural language processing tools and machine learning algorithms, draws a clear image of the whole system. Finally, we also compared our methodology and accuracy with the previously proposed methodologies by different authors. The comparative results are shown in Table 22.

The current study aims to provide a text mining and machine learning-based framework for HEIs to analyse the student's textual recommendations dataset regarding teaching–learning activities. Using the proposed multi-layer A2L approach, academic institutions can analyse and summarize thousands of textual feedback with accurate annotations within no time. The comparative statistical results show that the A2L framework is a rigorous method which performed better than previous studies. The summarized results of A2L can help the authorities understand students' sentiments and opinions about their institution's teaching–learning process. This study provides a theoretical and practical framework that will be a valuable addition to the educational technologies at the HEIs level.

5. Conclusion

This study proposed and successfully implemented an enhanced hybrid multi-layer Aspect2Label (A2L) topic modelling approach to evaluate students' textual feedback and extract meaningful information for Situation Awareness (SA) in the teaching–learning process in HEIs. We proposed and implemented a three-layer approach for this task. We divided A2L into three layers. The first layer is used to extract high-level aspects terms using LDA and co-occurrence analysis. Our proposed novel rule-based A2L algorithm is used to map the aspects terms with relevant feedback. This first layer revealed high-level aspects terms as a teacher, course, and university from the student textual feedback. The second layer of A2L is implemented to extract low-level terms using seed words with LDA and Co-occurrence analysis. Our second proposed algorithm is used to map low-level terms with corresponding high-level aspects terms and feedback. This second layer revealed subject matter knowledge, experience, method, behaviour, and general terms as low-level aspects for the high-level aspect teachers. In contrast, it extracted objective & goals, course content, assessment and manually included the "general" aspect as a low-level aspect for the high-level aspect courses. Similarly, it revealed the environment and policy of the university, and we manually included a "general" aspect as a low-level aspect. The results provide an adequate understanding to analyse students' feedback according to specific aspects, and

Table 22
Result of proposed approach in comparison with prior studies.

References	Methodology	Accuracy in aspects extraction	Sentiment classification accuracy
Shaikh and Doudpotta (2019)	Two-step rule-based strategy using Machine Learning & NLP	84%	90%
Sindhu et al. (2019)	Two-layered LSTM mode	90%	93%
Ozyurt and Akcayol (2021)	Sentence Segment LDA	87%	89%
Gutiérrez et al. (2018)	Machine Learning Algorithms	79%	85%
Proposed A2L Approach	Multi-layer approach using LDA and Co-occurrence analysis	91.3%	97%

we achieved 91.3% accuracy in the aspect extraction and annotation process. In the third layer, we extracted sentiments from students' feedback. Our rule-based classification algorithm performs automatic tagging using the lexicon approach. It gave a detailed summary of each aspect, including its high-level and low-level aspects. We also implemented sentiment classification using machine learning algorithms to classify sentiments using the lexicon-based score feature, TF-IDF, Word2Vec and coherence value features to train our models. We performed dimension reduction, including the Single Value Decomposition technique (SVD), which helped reduce extra feature space. We used Variable Stopwords Filtering (VSF) to resolve the sparsity problem and Variable Global Feature Selection Scheme (VGFSS) for dimensionality reduction and solving our dataset's class imbalance. We used five machine learning models for training and testing purposes in the machine learning approach of sentiment analysis. Our experimental results gave us promising results as 97%, 93% accuracy by using Support Vector Machine with the linear kernel (SVM_L) and Artificial Neural Network (ANN) algorithms, respectively. We also applied a multi-core parallel processing technique that reduced our model's training time and made our approach robust and cost-effective. Our proposed SA Aspect-Oriented Sentiment Analysis approach could help higher education institutes to understand student reviews in an appropriate annotated form. These annotated situational information summaries could improve higher educational institutions' decision-making process and their performance in evaluating the teaching–learning process at a large scale.

6. Limitations and future work

The proposed system only deals with comments written in the English language. However, most of the students in Pakistan like to write words in Roman Urdu. Hence, in the future, our proposed approach can further be expanded by processing Roman Urdu to explore the aspects of the text using deep learning algorithms. Usually, students take help from many symbols and emoticons to write feedback about a teacher in online feedback systems. Hence, another aspect is to check the association of emoticons like a happy face, sad face, etc., with sentiment orientation. In future research, we determine each attribute's weight and sentiment analysis, producing better accuracy in sentiment classification.

CRediT authorship contribution statement

Shabir Hussain: Conceptualization, Methodology, Software, Data curation, Writing – original draft, Writing – review & editing, Investigation, Visualization. **Muhammad Ayoub:** Conceptualization, Methodology, Writing – review & editing, Resources, Investigation, Visualization. **Ghulam Jilani:** Data curation, Writing – review & editing, Investigation. **Yang Yu:** Writing - review & editing, Resources, Investigation, Visualization. **Akmal Khan:** Writing - review & editing, Resources, Investigation. **Junaid Abdul Wahid:** Writing - review & editing. **Muhammad Farhan Ali Butt:** Writing - review & editing. **Guangqin Yang:** Supervision, Data curation, Resources, Investigation. **Dietmar P.F. Moller:** Writing - review & editing. **Hou Weiyan:** Supervision, Writing - original draft, Writing - review & editing, Resources, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This research work supported by the National Natural Science Foundation of China [grant number: 92067106] and by the Ministry of Education of the People's Republic of China [grant number: E-GCCRC20200309].

References

Agnihotri, D., Verma, K., & Tripathi, P. (2017). Variable global feature selection scheme for automatic classification of text documents. *Expert Systems with Applications*, *81*, 268–281.

Akhtar, N., Zubair, N., Kumar, A., & Ahmad, T. (2017). Aspect based sentiment oriented summarization of hotel reviews. *Procedia Computer Science*, *115*, 563–571.

Al Amrani, Y., Lazaar, M., & El Kadiri, K. E. (2018). Random forest and support vector machine based hybrid approach to sentiment analysis. *Procedia Computer Science*, *127*, 511–520.

Al Shobaki, M. J., & Abu-Naser, S. S. (2017). The role of the practice of excellence strategies in education to achieve sustainable competitive advantage to institutions of higher education-faculty of engineering and information technology at Al-Azhar university in gaza a model.

Ali, M. Z., Ehsan-ul Haq, A., Rauf, S., Javed, K., & Hussain, S. (2021). Improving hate speech detection of urdu tweets using sentiment analysis. *IEEE Access*.

Altrabsheh, N., Cocea, M., & Fallahkhair, S. (2014). Sentiment analysis: towards a tool for analysing real-time students feedback. In *2014 IEEE 26th international conference on tools with artificial intelligence* (pp. 419–423). IEEE.

Anoop, V., & Asharaf, S. (2018). Aspect-oriented sentiment analysis: A topic modeling-powered approach. *Journal of Intelligent Systems*, *29*(1), 1166–1178.

Ardito, C., Di Noia, T., Fasciano, C., Lofú, D., Macchiarulo, N., Mallardi, G., et al. (2020). Towards a situation awareness for ehealth in ageing society. In *AIxAS@ AI* IA* (pp. 40–55).

Arras, L., Montavon, G., Müller, K.-R., & Samek, W. (2017). Explaining recurrent neural network predictions in sentiment analysis. arXiv preprint arXiv:1706.07206.

Asanbe, M., Osofisan, A., & William, F. (2016). Teachers' performance evaluation in higher educational institution using data mining technique. *International Journals of Applied Information System (IJAIS)*, *10*, 10–15.

Aung, K. Z., & Myo, N. N. (2017). Sentiment analysis of students' comment using lexicon based approach. In *2017 IEEE/ACIS 16th international conference on computer and information science (ICIS)* (pp. 149–154). IEEE.

Ayoub, M., Hussain, S., Khan, A., Zahid, M., Wahid, J. A., Zhifang, L., et al. (2022). A predictive machine learning and deep learning approach on agriculture datasets for new moringa oleifera varieties prediction. *Pakistan Journal of Engineering and Technology*, *5*(1), 68–77.

Barrón Estrada, M. L., Zatarain Cabada, R., Oramas Bustillos, R., & Graff, M. (2020). Opinion mining and emotion recognition applied to learning environments. *Expert Systems with Applications*, *150*, Article 113265. http://dx.doi.org/10.1016/j.eswa.2020.113265, URL: https://www.sciencedirect.com/science/article/pii/S0957417420300907.

Bianchi, F., Terragni, S., Hovy, D., Nozza, D., & Fersini, E. (2020). Cross-lingual contextualized topic models with zero-shot learning. arXiv preprint arXiv:2004.07737.

Boring, A. (2017). Gender biases in student evaluations of teaching. *Journal of Public Economics*, *145*, 27–41.

Budiharso, T., & Tarman, B. (2020). Improving quality education through better working conditions of academic institutes.

Buoy, R., Taing, N., & Kor, S. (2021). Joint khmer word segmentation and part-of-speech tagging using deep learning. arXiv preprint arXiv:2103.16801.

Buzydlowski, J. W. (2015). Co-occurrence analysis as a framework for data mining. *Journal of Technology Research*, *6*, 1–19.

Chen, M., Li, Y., & Zhou, X. (2021). CoNet: Co-occurrence neural networks for recommendation. *Future Generation Computer Systems*.

Choi, I.-G., Kwon, J., & Kim, S.-H. (2004). Local feature frequency profile: a method to measure structural similarity in proteins. *Proceedings of the National Academy of Sciences*, *101*(11), 3797–3802.

Constantinou, C., & Wijnen-Meijer, M. (2022). Student evaluations of teaching and the development of a comprehensive measure of teaching effectiveness for medical schools. *BMC Medical Education*, *22*(1), 1–14.

da Costa, T. A. G., Meneguette, R. I., & Ueyama, J. (2022). Providing a greater precision of situational awareness of urban floods through multimodal fusion. *Expert Systems with Applications*, *188*, Article 115923.

Denison, D. G., Mallick, B. K., & Smith, A. F. (1998). A bayesian cart algorithm. *Biometrika*, *85*(2), 363–377.

yu Dong, L., juan Ji, S., jin Zhang, C., Zhang, Q., Chiu, D., qing Qiu, L., et al. (2018). An unsupervised topic-sentiment joint probabilistic model for detecting deceptive reviews. *Expert Systems with Applications*, *114*, 210–223. http://dx.doi.org/10.1016/j.eswa.2018.07.005, URL: https://www.sciencedirect.com/science/article/pii/S0957417418304184.

Drus, Z., & Khalid, H. (2019). Sentiment analysis in social media and its application: Systematic literature review. *Procedia Computer Science*, *161*, 707–714.

Endsley, M. R. (2017). Toward a theory of situation awareness in dynamic systems. In *Situational awareness* (pp. 9–42). Routledge.

Eng, T. H., Ibrahim, A. F., & Shamsuddin, N. E. (2015). Students' perception: Student feedback online (SuFO) in higher education. *Procedia-Social and Behavioral Sciences*, *167*, 109–116.

Firdaus, R., Asror, I., & Herdiani, A. (2021). Lexicon-based sentiment analysis of Indonesian language student feedback evaluation. *Indonesian Journal on Computing (Indo-JC)*, *6*(1), 1–12.

García-Pablos, A., Cuadros, M., & Rigau, G. (2018). W2VLDA: almost unsupervised system for aspect based sentiment analysis. *Expert Systems with Applications*, *91*, 127–137.

Ge, L., & Moh, T.-S. (2017). Improving text classification with word embedding. In *2017 IEEE international conference on big data (Big data)* (pp. 1796–1805). IEEE.

Goldberg, Y., & Levy, O. (2014). Word2vec explained: deriving Mikolov et al.'s negative-sampling word-embedding method. arXiv preprint arXiv:1402.3722.

Goudjil, M., Koudil, M., Bedda, M., & Ghoggali, N. (2018). A novel active learning method using SVM for text classification. *International Journal of Automation and Computing*, *15*(3), 290–298.

Gutiérrez, G., Canul-Reich, J., Zezzatti, A. O., Margain, L., & Ponce, J. (2018). Mining: Students comments about teacher performance assessment using machine learning algorithms. *International Journal of Combinatorial Optimization Problems and Informatics*, *9*(3), 26–40.

Harshini, G., & Gobi, N. (2020). Student feedback sentiment analysis system for distance education using arm with K-means clustering. *ICTACT Journal on Soft Computing*, *10*(3), 2071–2075.

Holmes, W., Nguyen, Q., Zhang, J., Mavrikis, M., & Rienties, B. (2019). Learning analytics for learning design in online distance learning. *Distance Education*, *40*(3), 309–329.

Hu, M., & Liu, B. (2004). Mining and summarizing customer reviews. In *Proceedings of the tenth ACM SIGKDD international conference on knowledge discovery and data mining* (pp. 168–177).

Hughes, M. C., Hope, G., Weiner, L., McCoy Jr, T. H., Perlis, R. H., Sudderth, E. B., et al. (2018). Semi-supervised prediction-constrained topic models. In *AISTATS* (pp. 1067–1076).

Hunter, S. B., & Springer, M. G. (2022). Critical feedback characteristics, teacher human capital, and early-career teacher performance: A mixed-methods analysis. *Educational Evaluation and Policy Analysis*, Article 01623737211062913.

Hussain, S., Yu, Y., Ayoub, M., Khan, A., Rehman, R., Wahid, J. A., et al. (2021). IoT and deep learning based approach for rapid screening and face mask detection for infection spread control of COVID-19. *Applied Sciences*, *11*(8), 3495.

Jing, L., Ng, M. K., & Huang, J. Z. (2010). Knowledge-based vector space model for text clustering. *Knowledge and Information Systems*, *25*(1), 35–55.

Joachims, T. (1998). Text categorization with support vector machines: Learning with many relevant features. In *European conference on machine learning* (pp. 137–142). Springer.

Jockers, M. (2017). Package 'syuzhet'. URL: https://cran.r-project.org/web/packages/syuzhet.

Jurek, A., Mulvenna, M. D., & Bi, Y. (2015). Improved lexicon-based sentiment analysis for social media analytics. *Security Informatics*, *4*(1), 1–13.

Kandhro, I. A., Wasi, S., Kumar, K., Rind, M., & Ameen, M. (2019). Sentiment analysis of students' comment using long-short term model. *Indian Journal of Science and Technology*, *12*(8), 1–16.

Kastrati, Z., Dalipi, F., Imran, A. S., Pireva Nuci, K., & Wani, M. A. (2021). Sentiment analysis of students' feedback with NLP and deep learning: A systematic mapping study. *Applied Sciences*, *11*(9), 3986.

Khan, A., & Ghosh, S. K. (2021). Student performance analysis and prediction in classroom learning: A review of educational data mining studies. *Education and Information Technologies*, *26*(1), 205–240.

Koopmanschap, R., Hoogendoorn, M., & Roessingh, J. J. (2015). Tailoring a cognitive model for situation awareness using machine learning. *Applied Intelligence: The International Journal of Artificial Intelligence, Neural Networks, and Complex Problem-Solving Technologies*, *42*(1), 36–48.

Kumar, S., Gahalawat, M., Roy, P. P., Dogra, D. P., & Kim, B.-G. (2020). Exploring impact of age and gender on sentiment analysis using machine learning. *Electronics*, *9*(2), 374.

Kumar, S., Kar, A. K., & Ilavarasan, P. V. (2021). Applications of text mining in services management: A systematic literature review. *International Journal of Information Management Data Insights*, *1*(1), Article 100008.

Kumar, S., Yadava, M., & Roy, P. P. (2019). Fusion of EEG response and sentiment analysis of products review to predict customer satisfaction. *Information Fusion*, *52*, 41–52.

Lagerstrom, R., Arzhaeva, Y., Szul, P., Obst, O., Power, R., Robinson, B., et al. (2016). Image classification to support emergency situation awareness. *Frontiers in Robotics and AI*, *3*, 54.

Larkey, L. S., Ballesteros, L., & Connell, M. E. (2002). Improving stemming for Arabic information retrieval: light stemming and co-occurrence analysis. In *Proceedings of the 25th annual international ACM SIGIR conference on research and development in information retrieval* (pp. 275–282).

Lee Chang, A., Dym, A. A., Venegas-Borsellino, C., Bangar, M., Kazzi, M., Lisenenkov, D., et al. (2017). Comparison between simulation-based training and lecture-based education in teaching situation awareness. A randomized controlled study. *Annals of the American Thoracic Society*, *14*(4), 529–535.

Levitz, R. N. (2018). *Cost of recruiting an undergraduate student report*. Cedar Rapids, Iowa: Ruffalo Noel Levitz.

Levy, O., & Goldberg, Y. (2014). Dependency-based word embeddings. In *Proceedings of the 52nd annual meeting of the association for computational linguistics (Volume 2: Short papers)* (pp. 302–308).

Lin, Q., Zhu, Y., Zhang, S., Shi, P., Guo, Q., & Niu, Z. (2019). Lexical based automated teaching evaluation via students' short reviews. *Computer Applications in Engineering Education*, *27*(1), 194–205.

Lisena, P., Harrando, I., Kandakji, O., & Troncy, R. (2020). ToModAPI: A topic modeling API to train, use and compare topic models. In *Proceedings of second workshop for NLP open source software (NLP-OSS)* (pp. 132–140).

Liu, Z., Lv, X., Liu, K., & Shi, S. (2010). Study on SVM compared with the other text classification methods. In *2010 second international workshop on education technology and computer science, Vol. 1* (pp. 219–222). IEEE.

Liu, Y., Wang, Y., & Zhang, J. (2012). New machine learning algorithm: Random forest. In *International conference on information computing and applications* (pp. 246–252). Springer.

Mantyla, M. V., Claes, M., & Farooq, U. (2018). Measuring LDA topic stability from clusters of replicated runs. In *Proceedings of the 12th ACM/IEEE international symposium on empirical software engineering and measurement* (pp. 1–4).

McKenna, L., Missen, K., Cooper, S., Bogossian, F., Bucknall, T., & Cant, R. (2014). Situation awareness in undergraduate nursing students managing simulated patient deterioration. *Nurse Education Today*, *34*(6), e27–e31.

Mehanna, Y. S., & Mahmuddin, M. (2021). The effect of pre-processing techniques on the accuracy of sentiment analysis using bag-of-concepts text representation. *SN Computer Science*, *2*(4), 1–13.

Mendon, S., Dutta, P., Behl, A., & Lessmann, S. (2021). A hybrid approach of machine learning and lexicons to sentiment analysis: enhanced insights from twitter data of natural disasters. *Information Systems Frontiers*, 1–24.

Nanda, G., Douglas, K. A., Waller, D. R., Merzdorf, H. E., & Goldwasser, D. (2021). Analyzing large collections of open-ended feedback from MOOC learners using LDA topic modeling and qualitative analysis. *IEEE Transactions on Learning Technologies*.

Nawaz, R., Sun, Q., Shardlow, M., Kontonatsios, G., Aljohani, N. R., Visvizi, A., et al. (2022). Leveraging AI and machine learning for national student survey: Actionable insights from textual feedback to enhance quality of teaching and learning in UK's higher education. *Applied Sciences*, *12*(1), 514.

Nikolić, N., Grljević, O., & Kovačević, A. (2020). Aspect-based sentiment analysis of reviews in the domain of higher education. *The Electronic Library*.

Özgür, A., Cetin, B., & Bingol, H. (2008). Co-occurrence network of reuters news. *International Journal of Modern Physics C*, *19*(05), 689–702.

Ozyurt, B., & Akcayol, M. A. (2021). A new topic modeling based approach for aspect extraction in aspect based sentiment analysis: SS-LDA. *Expert Systems with Applications*, *168*, Article 114231.

Prasad, G. (2021). Identification of bloom's taxonomy level for the given question paper using NLP tokenization technique. *Turkish Journal of Computer and Mathematics Education (TURCOMAT)*, *12*(13), 1872–1875.

Prasanna, P. L., & Rao, D. R. (2018). Text classification using artificial neural networks. *International Journal of Engineering & Technology*, *7*(1.1), 603–606.

Rajput, Q., Haider, S., & Ghani, S. (2016). Lexicon-based sentiment analysis of teachers' evaluation. *Applied Computational Intelligence and Soft Computing*, *2016*.

Ray, P., & Chakrabarti, A. (2017). Twitter sentiment analysis for product review using lexicon method. In *2017 international conference on data management, analytics and innovation (ICDMAI)* (pp. 211–216). IEEE.

Robertson, S. (2004). Understanding inverse document frequency: on theoretical arguments for IDF. *Journal of Documentation*.

Rudniy, A., & Elliot, N. (2016). Collaborative review in writing analytics: N-gram analysis of instructor and student comments. In *EDM (Workshops)*.

Salles, T., Gonçalves, M., Rodrigues, V., & Rocha, L. (2018). Improving random forests by neighborhood projection for effective text classification. *Information Systems*, *77*, 1–21.

Sangeetha, K., & Prabha, D. (2020). 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*, 1–10.

Shaikh, S., & Doudpotta, S. M. (2019). Aspects based opinion mining for teacher and course evaluation. *Sukkur IBA Journal of Computing and Mathematical Sciences*, *3*(1), 34–43.

Sievert, C., & Shirley, K. (2014). LDAvis: A method for visualizing and interpreting topics. In *Proceedings of the workshop on interactive language learning, visualization, and interfaces* (pp. 63–70).

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

Sozer, E. M., Zeybekoglu, Z., & Kaya, M. (2019). Using mid-semester course evaluation as a feedback tool for improving learning and teaching in higher education. *Assessment & Evaluation in Higher Education*.

Taboada, M., Brooke, J., Tofiloski, M., Voll, K., & Stede, M. (2011). Lexicon-based methods for sentiment analysis. *Computational Linguistics*, *37*(2), 267–307.

Tan, S. (2005). Neighbor-weighted k-nearest neighbor for unbalanced text corpus. *Expert Systems with Applications*, *28*(4), 667–671.

To, W., & Tang, M. N. (2019). Computer-based course evaluation: An extended technology acceptance model. *Educational Studies*, *45*(2), 131–144.

Tondeur, J., Scherer, R., Siddiq, F., & Baran, E. (2020). Enhancing pre-service teachers' technological pedagogical content knowledge (TPACK): A mixed-method study. *Educational Technology Research and Development*, *68*(1), 319–343.

Tseng, M.-L., Tran, T. P. T., Ha, H. M., Bui, T.-D., & Lim, M. K. (2021). Sustainable industrial and operation engineering trends and challenges Toward Industry 4.0: A data driven analysis. *Journal of Industrial and Production Engineering*, *38*(8), 581–598.

Uysal, A. K. (2016). An improved global feature selection scheme for text classification. *Expert Systems with Applications*, *43*, 82–92.

Van, V. D., Thai, T., & Nghiem, M.-Q. (2017). Combining convolution and recursive neural networks for sentiment analysis. In *Proceedings of the eighth international symposium on information and communication technology* (pp. 151–158).

Vidaurre, C., Kawanabe, M., von Bünau, P., Blankertz, B., & Müller, K.-R. (2010). Toward unsupervised adaptation of LDA for brain–computer interfaces. *IEEE Transactions on Biomedical Engineering*, *58*(3), 587–597.

Vorontsov, K., Frei, O., Apishev, M., Romov, P., & Dudarenko, M. (2015). Bigartm: Open source library for regularized multimodal topic modeling of large collections. In *International conference on analysis of images, social networks and texts* (pp. 370–381). Springer.

Wahid, J. A., Shi, L., Gao, Y., Yang, B., Tao, Y., Wei, L., et al. (2021). Topic2features: a novel framework to classify noisy and sparse textual data using LDA topic distributions. *PeerJ Computer Science*, *7*, Article e677.

Wahid, J. A., Shi, L., Gao, Y., Yang, B., Wei, L., Tao, Y., et al. (2022). Topic2Labels: A framework to annotate and classify the social media data through LDA topics and deep learning models for crisis response. *Expert Systems with Applications*, *195*, Article 116562.

Wang, C.-H., & Lien, C.-Y. (2019). Combining design science with data analytics to forecast user intention to adopt customer relationship management systems. *Journal of Industrial and Production Engineering*, *36*(4), 193–204.

Wilbur, W. J., & Kim, W. (2009). The ineffectiveness of within-document term frequency in text classification. *Information Retrieval*, *12*(5), 509–525.

Wook, M., Razali, N. A. M., Ramli, S., Wahab, N. A., Hasbullah, N. A., Zainudin, N. M., et al. (2019). Opinion mining technique for developing student feedback analysis system using lexicon-based approach (OMFeedback). *Education and Information Technologies*, 1–12.

Yu, Y., Qin, X., Hussain, S., Hou, W., & Weis, T. (2022). Pedestrian counting based on piezoelectric vibration sensor. *Applied Sciences*, *12*(4), 1920.

Yu, S.-O., et al. (2016). Using students' feedback to evaluate teachers' effectiveness.

Zainuddin, N., & Selamat, A. (2014). Sentiment analysis using support vector machine. In *2014 international conference on computer, communications, and control technology (I4CT)* (pp. 333–337). IEEE.

Zhang, W., Xu, H., & Wan, W. (2012). Weakness finder: Find product weakness from Chinese reviews by using aspects based sentiment analysis. *Expert Systems with Applications*, *39*(11), 10283–10291.

Zhou, R., Niu, L., & Yang, H. (2021). Unsupervised feature selection for attributed graphs. *Expert Systems with Applications*, *168*, Article 114402. http://dx.doi.org/10.1016/j.eswa.2020.114402, URL: https://www.sciencedirect.com/science/article/pii/S0957417420310733.

Zou, Y., Gui, T., Zhang, Q., & Huang, X.-J. (2018). A lexicon-based supervised attention model for neural sentiment analysis. In *Proceedings of the 27th international conference on computational linguistics* (pp. 868–877).