A Review of the Trends and Challenges in Adopting Natural Language Processing Methods for Education Feedback Analysis

THANVEER SHAIK[1], **XIAOHUI TAO**[1], **(Senior Member, IEEE), YAN LI**[1],
CHRISTOPHER DANN[2], **JACQUIE MCDONALD**[3], **PETREA REDMOND**[2],
AND LINDA GALLIGAN[1]

[1]School of Mathematics, Physics and Computing, University of Southern Queensland, Toowoomba, QLD 4350, Australia
[2]School of Education, University of Southern Queensland, Toowoomba, QLD 4350, Australia
[3]Academic Development, University of Southern Queensland, Toowoomba, QLD 4350, Australia

Corresponding author: Thanveer Shaik (thanveer.shaik@usq.edu.au)

ABSTRACT Artificial Intelligence (AI) is a fast-growing area of study that stretching its presence to many business and research domains. Machine learning, deep learning, and natural language processing (NLP) are subsets of AI to tackle different areas of data processing and modelling. This review article presents an overview of AI's impact on education outlining with current opportunities. In the education domain, student feedback data is crucial to uncover the merits and demerits of existing services provided to students. AI can assist in identifying the areas of improvement in educational infrastructure, learning management systems, teaching practices and study environment. NLP techniques play a vital role in analyzing student feedback in textual format. This research focuses on existing NLP methodologies and applications that could be adapted to educational domain applications like sentiment annotations, entity annotations, text summarization, and topic modelling. Trends and challenges in adopting NLP in education were reviewed and explored. Context-based challenges in NLP like sarcasm, domain-specific language, ambiguity, and aspect-based sentiment analysis are explained with existing methodologies to overcome them. Research community approaches to extract the semantic meaning of emoticons and special characters in feedback which conveys user opinion and challenges in adopting NLP in education are explored.

INDEX TERMS Artificial Intelligence, natural language processing, education, deep learning.

I. INTRODUCTION

Artificial Intelligence (AI) is a fast-growing topic with its cognitive human-like intelligence in building decision-making systems. AI can revolutionize education with its capacity for prediction and classification by processing huge amounts of structured data sets such as SQL databases and unstructured datasets such as videos and audios. AI introduces machine learning methodologies to personalize the student learning experience via learning management systems [1], deep learning, and transfer learning to use pre-trained concepts to deal with new similar problems [2], natural language processing (NLP) methods [3] to listen to student feedback, process them and output predictive insights on their opinion towards learning infrastructure. AI can

transform existing educational infrastructures [4] namely online tutoring, learning management systems, curriculum, employment transitions, teacher training, assessments, and research training. The institutional project data are diverse and inclusive of student feedback in textual format class-room recordings in video and audio formats.

Chassignol *et al.* [5] defined AI as an "Artificial Intelligence is that activity devoted to making machines intelligent, and intelligence is that quality that enables an entity to function appropriately and with foresight in its environment". Educational institutions have extensively adopted AI in different forms of service delivery to students [6]. One of the most widely used AI methodologies for student opinion mining is NLP [7]. It plays a key role in interpreting feedback or opinions of end-users. Most institutions in the world invest their time and resources to understand end-users' feedback. NLP can read the feedback in most languages

---

The associate editor coordinating the review of this manuscript and approving it for publication was Yiqi Liu.

without much human intervention and can analyze textual data and unwrap the end-user perception and opinion on a service, product, or human. In recent years, NLP has been applied to review movies, books, gadgets and so on [8]. Topic modelling techniques are part of NLP to read text corpus and can summarize, annotate or categorize text documents. Furthermore, it uses various techniques like part-of-speech (POS) tagging to understand the context of words.

Eggert [9] discussed the opportunities of AI in education. The author proposed an AI method to improve teaching methods by collecting vast amounts of data related to each student's prior knowledge, emotional state, or economic background and adjusting the teaching approach through adaptive learning platforms (ALP). Intelligent tutoring systems (ITS) is one of the ALP components. With automation of repeated tasks would allow teaching staff to design new instructional approach and focus on non-routine work. The other opportunity discussed in that article is to expose students to some AI-driven tools to cope with the future labour world that is highly dependent on technologies and focus on lifelong learning via improved access to Massive Open Online Courses (MOOCs). AI can enhance student's learning experience in MOOCs by identifying areas where personalized guidance is required. Holstein *et al.* [10] also stressed the need for personalized guidance to students in their work on AI-enhanced classrooms. Using participatory speed dating (PSD) [11], the authors found real-time support was needed from the AI system to identify when a student needs a human's help for motivation. Holstein *et al.* [12] also focused on the challenges of involving non-technical stakeholders due to the complexity of learning analytics systems [13]. The authors proposed Konscia, a wearable and real-time awareness tool for teachers working in AI-enhanced K-12 classrooms. In addition, they demonstrated the process of non-technical stakeholders' participation in designing a complex learning analytics system. Alrajhi *et al.* [14] stressed the need to analyse student feedback or comments in MOOC as it would help to understand the student need for intervention from instructors.

Chen *et al.* [6] surveyed the impact of AI on education. The authors discussed the technical aspects of AI in education: assessment of students and schools, grading and evaluating papers and exams, smart schools, personalized intelligent teaching, online and mobile remote education. Their study scope was confined to the application and effects of AI in administration, teaching, and learning. To enable instructors and teachers with effective grading capabilities, an adaptive learning method was used in applications of Knewton and ensured a continuous student improvement in learning [15]. Applications like Grammarly, Ecree, Paper-Rater and Turnitin leverage AI to assist educational institutions and teachers in performing plagiarism checks, typographical and grammatical error checks. The student learning experience is an essential aspect of the education domain. AI enables an adaptive learning system for students based on their

backgrounds to assist in tracking their learning progression and enhance the system to customize the content according to student's needs to leverage a personalized system. A quick interactive system using AI would reduce the gap between students and educational providers and assist in listening to students' opinions and queries.

With the extensive research being conducted in analyzing AI's impact on education [16], [17] and discovering the opportunities in the education domain, educational institutions have focused on building a cognitive intelligent system using AI. In this process, the foremost step is to listen to students' opinions and feedback on existing educational infrastructure, teaching practices, and learning environments. In academic institutions, it is traditional practice to request student feedback to gather students' perception of the teaching team and their learning experience in the course. The student feedback could be in quantitative or qualitative formats, using numerical answers to rate the performance or textual comments to questions [18]. Monitoring and tracking students' feedback manually is a time-consuming and resource-demanding task. NLP can contribute to this task with its annotation and summarization capabilities. This study reviewed NLP methodologies that can contribute to the education domain, and the following research questions were explored:

- What are the existing methodologies being used for NLP?
- What are the generic challenges of using NLP in the education domain?
- What are the current trends of NLP in student feedback analysis?
- How can NLP methodology in other disciplines be adopted to the education domain?

Machine learning and deep learning are part of AI methodologies. Machine learning is a set of algorithms can analyze data, learn and apply. Deep learning techniques holds multi-layer neural network with processing layers to train new concepts and link to previously known concepts. Deep learning enhances NLP with concepts like continuous-bag-of-words and skip-gram model. Convolutional neural networks (CNNs) [19], recurrent neural networks (RNNs), their special cases of long short-term memory (LSTM) and gated recurrent units (GRUs) are different forms of deep learning techniques used in text classification [20], [21]. In this article, existing works using the AI methodologies to analyze text data are explored. Although few research works are not directly related to student feedback, the methods can be adopted to students' feedback analysis.

The contributions of this research are as follows:

- Enhanced understanding of the impact of AI on education with open opportunities in the industry.
- Synthesis of existing NLP methodologies to student user feedback and annotate their views.
- Exploring trends and challenges in NLP that need to be addressed to be adopted to the education domain.

The remainder of the paper is organized into sections. Section II defined feature extraction, feature selection, and topic modelling techniques with other researchers' work. Text evaluation techniques like summarization, knowledge graphs, annotation, existing NLP methodologies being used for NLP are defined. In Section III, challenges in adopting NLP in the education domain are discussed. Section IV presents a discussion on this work. The article concludes with limitations and future work of the study presented in Section V

II. METHODOLOGY

Feature extraction and feature selections are mandatory data preprocessing steps to transform text data into quantitative vector formats before feeding the students' feedback data to traditional machine learning algorithms or machine learning techniques like topic modelling. In this section, existing methods in feature extraction, feature selection, and topic modelling will be discussed.

A. FEATURE EXTRACTION

Feature extraction techniques can be applied to prepare the students' feedback data and transform it for machine learning modelling. For example, in NLP, there are feature extraction techniques like Bag of Words (BoW), Term Frequency (TF)-Inverse Document Frequency (IDF), and Word Embedding [22].

Bag of Words (BoW) [23] is a common feature extraction method that involves a vocabulary of known words and a measure of the presence of known words. The BoW is only concerned with known words in a document. It will not consider the structure or order of words in a document that ignores the context of the words [24]. TF-IDF [25] estimates the importance of each word or term in a document based on their weights [26]. IDF of a word gives how common or rare a word is in a corpus. The closer the value is to zero, the more common a word is in a corpus. TF-IDF is a multiplication of TF and IDF.

Word Embedding [27] is a learned representation of text with similar meaning. It enhances the generalization process and reduces dimensionality. The most common word embedding techniques are Word2Vec, GloVe, Doc2Vec [28] and Bidirectional encoder representations from transformers (BERT) [29]. Word2vec algorithm is built on a neural networks model to learn word associations from a large corpus of text. The trained algorithm can detect synonymous words or even suggest additional words for a partial sentence. Word2Vec generates the number of dimensions for each word in a corpus and then searches at the context level of the occurrence of the words in a sentence. In a vector space, all the words with similar contexts are grouped. A GloVe approach combines the matrix factorization technique and latent semantic analysis (LSA) with a context-based learning in Word2Vec. Doc2Vec is a tool to create vector or numeric representations of documents. BERT is a pre-trained deep bidirectional representations from unlabeled text by jointly

conditioning on both left and right context in all layers. BERT can perform word or sentence embedding to extract vectors from text data. It has an advantage over techniques like Word2Vec where each word has a fixed representation in Word2Vec irrespective of context. BERT can produce word representations dynamically based on words around them [30].

Waykole *et al.* [31] evaluated text classification based on feature extraction techniques such as bag of words, TF-IDF, and Word2Vec. In that study, each of the feature extraction techniques were evaluated with machine learning algorithms, for example logistic regression and random forest classifier with 3-fold stratified cross-validation. The experimentation results showed that Word2Vec was a better feature extraction technique with a random forest classifier was better for the text classification. Similarly, count vectorizer, TF-IDF and Word2Vec techniques were compared using a logistic regression model. Deepa *et al.* [32] proposed an approach to detect the polarity of words from Twitter using three feature extraction techniques count vectorizer, Word2Vec, TF-IDF and two dictionary-based methods of valence aware dictionary and sentiment reasoner (VADER) and SentiWordNet. Feature extraction techniques achieved better accuracy than dictionary-based methods. For example, count vectorizer achieved the highest classification accuracy of 81%. Twitter text is short text analysis which is similar to students' feedback to an open-ended question in an educational institution where feature extraction techniques can be adopted.

TF-IDF feature extraction generates feature vectors with high dimensions in a large text corpus [33]. In the study [33], the TF-IDF extraction technique was evaluated by adding dimensionality reduction techniques, latent semantic analysis (LSA) and linear discriminant analysis (LDA). Using a neural network classifier, the authors compared the classification performance of plain TF-IDF, TF-IDF LSA, and TF-IDF LDA methods on short texts. The research outcome stated that the TF-IDF approach outperformed the other two approaches with larger datasets. With smaller datasets, the TF-IDF and TF-IDF LSA achieved similar accuracy. However, the TF-IDF LDA approach had difficulty accurately classifying the text, as it failed to reduce the noise.

Deep learning techniques were used to evaluate word embedding techniques of Word2Vec and GloVe. In a study by [34], CNNs and RNNs were compared, and ensembles a combination of CNN and LSTM networks and compared. Eight different combinations of deep learning algorithms were implemented. Their comparison results showed that the GloVe system enhanced the performance by about 5-7% compared to Word2Vec. Sangeetha *et al.* [35] proposed a novel approach to analyze and find students' emotions in their feedback. In the study, feedback sentences were processed parallel across a multi-head attention layer with embedding techniques GloVe and contextualized vectors (Cove). The proposed method was tested with dropout rates to improve the accuracy. The authors compared the performance of proposed

methods' with baseline models LSTM, LSTM+ATT, multi-head ATT, and fusion in terms of accuracy 86.27%, 87.49%, 90.03%, and 94.13% respectively.

Zhang *et al.* [36] proposed a fine-tuned BERT model for sentiment analysis of student feedback to courses. In that study, intra domain unsupervised training was performed using the BERT model. To add grammatical constraints to the output of BERT model, a conditional random field layer was introduced. In addition, binding corporate rules — double attention layers were added to target sentiment analysis of the student feedback. Masala *et al.* [37] analyzed student feedback provided to each course and extract important ideas on various components. The authors used the BERT model to extract keywords from student feedback from each course, find contexts for repeated keywords, and group similar contexts. With this approach, 59% of the feedback text was reduced at a cost if mean average error increased to 0.06 while predicting course ratings from student feedback. Wu *et al.* [38] proposed pre-trained word embeddings to automatically create clusters such as homogeneous and heterogeneous student groups based on students' knowledge. Homogeneous groups can assist teachers to provide collective feedback, and heterogeneous groups can support and improve collaborative learning.

Feature extraction methods normally break down students' feedback data into word tokens to prepare the data for semantic and grammatical analysis. Neural network-based BoW, TF-IDF gives the frequency of words in a document, word embedding techniques like Word2Vec, GloVe, Cove, and BERT reduce the dimensionality of a word to group similar contexts. The performances of the feature extraction methods are often compared using machine learning and deep learning methods in [27], [28], [31]–[34], [36]–[39].

B. FEATURE SELECTION

Feature selection is a process of reducing data dimensionality in terms of features. This would maintain or enhance the performance of a machine learning algorithm. The reduction criteria would simplify a model's complexity and consistently maintain accuracy. Considering n features in a dataset, the number of the feature subsets would be 2^n. An increase in the features count would make the modelling infeasible [40], [41]. The stability or robustness of feature subsets was evaluated by grouping similar features or considering all feature subsets, removing the non-contributing features, and the size of the feature subsets. The feature subset evaluation methods are broadly categorized as filters, wrappers, or embedded methods [42], [43].

1) FILTER METHODS

Filter methods rank the key features and select high representative features by setting a threshold [44]. As shown in Figure 1, filter methods rank the features and select them before actual modelling. In addition, this technique filters the low importance features before training a model. The feature
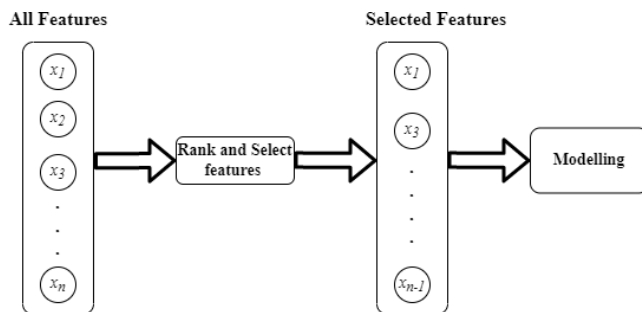

FIGURE 1. Filter methods.

importance technique assesses two measures in ranking the features. The first measure is to check the predictive power of each feature toward the target variable(s). These are called correlation criteria or dependence measures. Mutual information, χ^2 statistic, Markov blank, and minimal-redundancy-maximal-relevancy techniques extract a feature's correlation with a target variable. The second measure in the feature importance technique is redundancy, which assesses the features with redundant information. This detects the redundant features by evaluating relevant measures among the independent variables. An article by Wang [45] presented a redundant feature analysis. Its process is to find the most relevant features in predicting the target variables and use the relevant features to estimate the redundancy in other features.
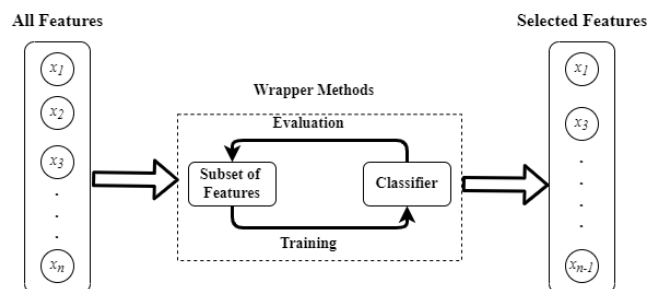

FIGURE 2. Wrapper methods.

2) WRAPPER METHODS

Wrapper methods search for a subset of features using a predefined classifier and then the performance of the subset of features is evaluated using predefined classifiers [44]. In wrapper methods, a machine learning algorithm is used to enhance the feature selection performance. As shown in Figure 2, a subset of features is selected and trained by a classifier with the selected features. Then, the performance of the classifier is evaluated. Sequential forward selection (SFS) is an example of a wrapper method with sequential feature selection methods. It is a greedy search algorithm that extracts an optimal subset of features iteratively based on the classifier performance. Features are selected one-by-one from the pool of all features iteratively.
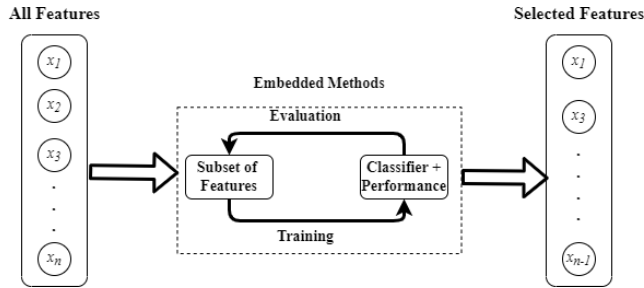
FIGURE 3. Embedded methods.

3) EMBEDDED METHODS

Embedded methods normally combine a filter method and a wrapper method [46]. They overcome the challenges of low accuracy in filter methods and slow computation speed in wrapper methods. Embedded methods analyze the optimum features, contributing to the classifier's accuracy. As shown in Figure 3, embedded methods estimate the performance of each subset of features. One of the most common embedded methods is regularization, which is to reduce the degree of overfitting or variance of a model by adding a penalty against its complexity for L1 regularization methods [47].

Parlar *et al.* [48] proposed a query expansion ranking (QER) method for feature selection. They compared it with other feature selection methods like information gain, Chi-square, document frequency difference, and optimal orthogonal centroid using classifiers like Naïve Bayes-multinomial, a support vector machine, maximum entropy modelling, and a decision tree. The authors tested the model performances on English and Turkish review databases. The proposed feature selection method outperformed the other feature selection methods in the Naïve Bayes multinomial classifier. Similar techniques were adopted to analyze student feedback in teaching evaluation system by Pong-Inwong *et al.* [49]. In the study, filter method was opted for feature selection and number of attributes in the data were reduced to 18 based on Chi-Square value. Three machine learning algorithms ID3, J48 and Naïve Bayes were used for student feedback classification and compared their performance with vote ensemble learning. The voting ensemble learning integrated with Chi-Square feature selection outperformed traditional machine learning algorithms with an accuracy of 87.16%.

Gutiérrez *et al.* [50] proposed social mining model architecture to increase the quality of learning and e-learning based on students' feedback analysis. This approach was focused to enhance teaching techniques and recommend courses for teacher improvement in higher education. As part of feature selection process in the study, random forest importance measure method was used with which weights of each word can be computed and filter them based on higher weights. The selected features were passed to SVM with kernels linear, radial, poly, and random forest classifiers. The machine learning classifiers were trained using k-fold cross validation and SVM model with radial kernel outperformed other models

with an accuracy of 85.17%. Similarly, Soukaina *et al.* [51] proposed an information gain filter method to select most relevant features in students' feedback in an optimized sentiment analysis approach. SVM, random forest, and Naive Bayes classifiers were used in the study and compared their performances before feature selection and after selection. Random forest dominated the other two classifiers before the selection with an accuracy of 81.6% and SVM outperformed with an accuracy of 85.9% after the feature selection.

Feature selection approaches are better with noise resistance and can help to avoid noise or irrelevant data for data modelling. Other modern feature selection methods were proposed and compared with existing methods using machine learning methods in [48], [52]

C. TOPIC MODELING

Topic modelling automatically analyzes a corpus of documents with text data techniques using machine learning techniques and determines cluster words [53], [54]. The technique does not need any training to cluster the words from the corpus. This is an unsupervised machine learning technique [55]. Topic modelling divides a corpus of documents into groups to extract a list of topics covered, and several sets of documents are grouped by the topics they covered. The topic modelling techniques are broadly categorized into probabilistic and non-probabilistic models [56], [57].

1) NON-PROBABILISTIC MODELS

Non-probabilistic models are matrix factorization algebraic approaches. These models came into use with latent semantic analysis (LSA) and Non-negative matrix factorization (NMF) [58]. Both LSA and NMF mechanisms work on BoW approaches. As discussed in Section II-A, BoW converts a corpus into a term-document matrix to extract the frequency of the terms and ignores the order of the terms. LSA is an algebraic method that generates a matrix with words presented in a corpus. It assumes that words that are similar in meaning will occur very close in the text [59]. The technique is based on single value decomposition (SVD) which reduces the number of words while preserving a similar structure. The similarity of the texts will be computed using vector representation and organized into semantic clusters. NMF transforms high dimensional data into low dimensional data with no negative components and clusters simultaneously [60]. This is also called positive matrix factorization (PMF). It is an unsupervised machine learning technique that can extract relevant information without previous insights into the original data.

2) PROBABILISTIC MODELS

Probabilistic models are fully unsupervised approaches that are tweaked to guide in latent dirichlet allocation (LDA) modelling and semi-supervised learning in a probabilistic latent semantic analysis [61]. Probabilistic latent semantic analysis (PLSA) is to detect semantic co-occurrence of words or terms in a corpus [62]. This is built based on the first

statistical model, a model that revealed the semantic co-occurrence in a document term matrix of the corpus. Due to its unsupervised nature, PLSA is capable of determining the number of topics, the probability of a topic and the probability of a document containing the topic. It groups unknown topics of every existing document. LDA is a commonly used technique in topic modelling which is built based on De Finetti's theorem, which states that positively correlated exchangeable observations are conditionally independent relative to some latent variable [63]. It can capture inter and intra document statistical structures on assumptions that a corpus has a pre-defined number of topics and each document in the corpus has a different proportion of the topics. It is a hidden variable model which uncovers hidden patterns in gathered data in a corpus.

An LDA technique based on a topic modelling methodology was selected in mobile learning research to find the topic trends [64]. Out of 50 topics extracted from the LDA, 25 topics were selected and grouped into three dimensions of technology, learning and learners in that mobile learning. Similarly, as part of designing a course structure for virtual reality with augmented reality and mixed or extended reality, the LDA technique was employed in the research study [65]. The study was to understand the motive of learners (students) in joining the course using topic modelling. It revealed that learners had little experience in designing virtual applications. Also, the learners had little experience in a programming language. Designing a massive open online course without understanding learners' engagement would lead to a high dropout rate.

To enhance computer science course teaching materials, Marcal *et al.* [66] proposed an innovative approach for extracting topics from StackOverflow, a question-and-answer website for professional and enthusiastic programmers, to identify unknown or misunderstood topics. Using these topics, to transform the course teaching material, the authors classified the question types into eight categories: debugging, how to, what, is there, possible, looking, advise, and optimal using an SVM. The LDA technique for topic modelling generates five topics for each of the eight types of questions. Based on the keywords in all five topics in each question type, professors or lecturers could compare and enhance their material to fill the gap. Course satisfaction surveys were analyzed to extract student opinions by using the LDA technique [67]. In a study by Cunningham *et al.* [68], nine different topics or aspects were selected using a topic modelling technique. Each student comment was separated into ideas to calculate sentiment and also overall sentiment and satisfaction was considered. The authors visualized one-course feedback over different semesters in terms of aspects like tutorial, lecture, assignment, content, and lecturer.

To analyze international students' needs and perceptions, and grouping them into categories [69], Adriana *et al.* [70] proposed a probabilistic topic model approach using LDA. The authors used a machine learning for language toolkit (MALLET) [71] to run LDA and selected 20 topics based on 59,662 reviews. The topics covered in the research were *language skills, convenient accommodation, weather, academic burdens, interesting courses* and so on. The topics were ordered by their weight in the composition of the whole set of reviews by its importance on what comprises a good university, living expenses, sound teaching, expensive country, and city offerings. As part of the strategic planning of a university to increase student enrollment, knowledge mining on online reviews was performed using ensemble LDA (eLDA) in a study [72]. The authors split the database into training data and held out data where the training data to LDA to extract the probabilistic score of words related to each topic being generated. To avoid inconsistency in the LDA results due to its Collapsed Gibbs Sampling (CGS), multiple LDA models were trained in parallel and stored the results in a database for further sentence labelling. The held-out data were labelled using the trained LDA model. Further, the held-out data were manually annotated with prior knowledge of identified topics in the database. Based on the top five values in each topic, 12 meaningful topics like academic support, diversity, faculty, financial aid, the weather were categorized.

Pyasi *et al.* [73] developed a student feedback analysis tool to extract sentiments and suggestions from students' feedback using sentiment analysis models, NLP techniques, LDA, and visualization techniques. In this study, Textblob [74] and polarity analyzer were used for sentiment analysis and Textblob dominated with Recall is 96.17%, Precision is 67.47% and F-score of 79.30%. Generalized linear models (GLM), SVM, conditional inference tree (CTREE) and decision tree C5.0 classification models were used for suggestion extraction and C5.0 model has better other classifiers with recall is 80.2%, precision is 77.5% and F-Score is 78.1%. For topic modelling, LDA and k-means clustering with cosine similarity scores were compared and LDA model was capable to extract multiple topics on single student comment whereas cosine clusters assigned a single topic to the comment.

Curiskis *et al.* [56], proposed an evaluation of document clustering and topic modelling methods in online social media networks like Twitter and Reddit [75], [76]. The authors used four feature representation techniques Doc2vec, weighted Word2vec, unweighted Word2vec, and TF-IDF on three benchmark databases extracted from Twitter and Reddit API. For document clustering, k-means clustering [77], k-medoids clustering [78], Hierarchical agglomerative clustering [79], non-negative matrix factorization (NMF) techniques were adopted along the LDA topic model to compare the clustering methods. The raw data from the three databases were preprocessed to remove hashtags, punctuations, and stop-words. The word embedding models weighted Word2vec, unweighted Word2vec, and Doc2vec were applied to all three datasets along with k- means clustering. The optimal number of epochs in each approach with its results were compared. All the methods were evaluated using three performance metrics normalized mutual information (NMI) [80], adjusted mutual information (AMI) [81], and adjusted rand index (ARI) measures [82]. With their results,

word embedding models outperformed traditional TF-IDF representations. The research work summarized end-to-end NLP tasks starting from data extraction to methods evaluation including data preparation, document clustering, and topic modelling.

In a study conducted by Patil et al. [54], aspect-level sentiment analysis was proposed to analyze e-commerce site Amazon product reviews [83], [84]. The authors extracted sentiment ratings from the website and categorized them into negative, neutral, and positive based on each product rating. In preprocessing step, tokenization and stemming techniques [85] were implemented on the product reviews or user comments. LDA topic modelling technique and k-means clustering algorithm were used for topic extraction. Three machine learning models logistic regression [86], SVM [87], and Naïve Bayes [88] were developed, one for each sentiment polarity. That model accuracy was calculated to know how the sentiment polarity worked for the textual data. Nine topics from electronic products reviews were extracted using an LDA and k-means clustering. Although that article focused on e-commerce product reviews, the process could be adopted for higher education feedback, where scores can be used to extract sentiment polarity and students comments to extract topics. Similarly, Kastrati et al. [89] proposed weakly supervised framework for aspect-level sentiment analysis and automatically identify sentiment or opinion in MOOC dataset. MOOC related aspects like content, structure, knowledge, skill, experience, assessment, technology, interaction, and general were grouped together to have four aspects for the proposed study. The four aspects are the course that covers both content and structure aspect, the instructor that includes knowledge, skill and experience of the instructor, the assessment, and the technology. Manually annotated dataset collected from Coursera was preprocessed with feature extraction technique TF-IDF, Word2Vec and CNN model was used for the aspect category learning. CNN and LSTM models were used for the aspect polarity assessment task with F1 score of 86.13% for aspect category identification (broader MOOC-related aspects) and 82.10% for aspect sentiment classification.

A student feedback mining system (SFMS) for analyzing students' feedback was proposed by Gottipati et al. [90] based on agglomerative clustering with cosine similarity. Text analytics model was employed for text evaluation tasks like text categorization, entity extraction, sentiment analysis, document summarization. Ten topics were extracted using agglomerative clustering, and the top 5 to 10 words in each topic were used to label the topic. Logistic regression model was used for sentiment classification with a precision of 80.1%, recall of 86.4% and F-Score of 83.5%. A faculty rating system based on text mining techniques was developed by Krishnaveni et al. [91]. Student feedback was mapped with student database and weights were assigned to student attributes like CGPA, sincerity, attendance, performance, and feedback submit duration. Naive Bayes classifier was used to rate faculty into classes range (1 star—5 star).

Overall, LDA techniques [54], [56], [64]–[68], [70], [72], [75], [76], [83], [84] were the most used topic modelling methodology due to their generative process with the disambiguation of words in each topic and precise alignment of keywords to topics that may closely reflect the original collection.

D. TEXT EVALUATION
In this subsection, NLP applications like text summarization, document categorization, text annotation, and knowledge graphs are discussed.

1) TEXT SUMMARIZATION
There has been an exponential growth in collection of student feedback for evaluation in educational institutions. Consolidating the content and extracting useful resources is a tedious task and would consume massive efforts. The summary of the content would be easier for readers to digest and comprehend. Text summarization technique provides a summary of a student feedback or corpus of the feedback text without losing critical information. Text summarization can be categorized into extractive, abstractive, and hybrid approaches [92].
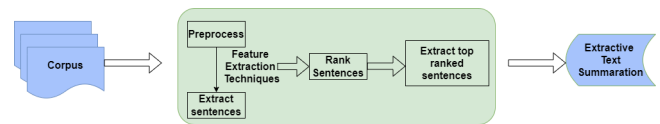

FIGURE 4. Extractive text summarization.

Extractive Text Summarization is a traditional text summarization method. It extracts significant sentences as it is from the document and adds to the summary. As shown in Figure 4, the technique selects a subset of the sentences in an original text using feature extraction techniques like BoW, N-gram, graphs and so on. The extracted sentences are ranked based on their importance. It creates an intermediate representation that highlights the most important information included in the original text [93].
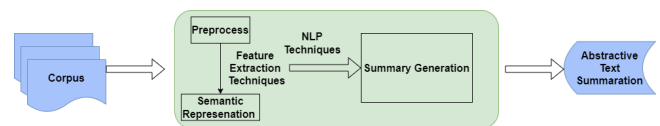

FIGURE 5. Abstractive text summarization.

Abstractive Text Summarization extracts sentences from documents in an intermediate representation and generates a summary of the sentences instead of the original sentences as shown in Figure 5. The technique paraphrases the sentences using NLP techniques and generates a summary that is suitable to human interpretation [94].

Hybrid Text Summarization is an ensemble of extractive and abstractive text summarization as shown in Figure 6. In this mechanism, the top-ranked sentences extracted from
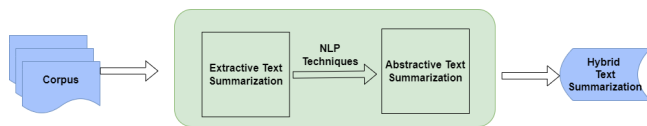
FIGURE 6. Hybrid text summarization.

extractive text summarization are paraphrased using NLP techniques and summarizes the content abstractedly.

Mutlu *et al.* [95] stated that extractive sentiment has an advantage of language independence as it doesnot aim for sentence construction and paraphrasing. As stated earlier, it has intermediate representations where sentence scoring and sentence selection steps were involved. Each sentence is assigned a salience degree and ranked to summarize the important sentences in the original text. Estimating the salience of a sentence in the original text is a classification problem. The authors used a LSTM-NN (neural network) for the sentence selection based on semantic features, synthetic features and ensembled features. Compared the LSTM-NN model with the baseline models of hierarchical attention-based bidirectional gated recurrent unit (Bi-GRU), CNN, Bi-GRU and the newer state-of-the-art models SummaRuNNer [96], BanditSum [97]. The LSTM-NN model outperformed all other four models. Yuxiang *et al.* [98] proposed a reinforced neural extractive text summarization model which optimizes the coherence and importance of summarized information simultaneously. The authors used a CNN model at the word level to extract features and their context. At the sentence level, a Bi-GRU was used to model the context of a sentence. The pre-trained data were fed to a reinforcement learning to compute cross sentence coherence as part of the reward of the proposed reinforced model. The research results showed that the proposed approach could balance cross-sentence coherence and sentence importance.

Statistical methods can be used for the extractive summarization process. Madhuri *et al.* [99] proposed a novel statistical method for extractive summarization. In the study, sentences were tokenized, stopwords removed, parts-of-speech added to each token, and the weights were assigned to each token based on its frequency and the total number of terms in the document. The weighted frequency of the token was calculated, and finally, the sum of the weighted frequency tokens was calculated. The sentences were rearranged in descending order and a summarizer extracts the highest rank sentences and converted them into an audio format. The work was evaluated with human summarized data and the proposed method achieved a higher accuracy.

Fan *et al.* [100] proposed an CourseMIRROR (Mobile In-situ Reflections and Review with Optimized Rubrics) using automatic text summarization techniques to aggregate students' feedback. This approach assisted to extract most significant ones and help students to understand both difficulties and misunderstandings. The authors extracted phrases, grouped them using k-medoids clustering algorithm, and then re-rank the phrases by student coverage. This

approach has generated better results than existing technique, LexRank [101].

Gottipati *et al.* [102] proposed a topic based summarization tool to analyze student online discussion forum in a course and extract topic based summaries. The authors used TextRank Summarizer [103] and LSA Summarizer [104] techniques for text summarization. Three questions were defined to categorize the online forum posts and extracts topics on each of these questions. LSA summary provided additional data recommendations when compared to TextRank summary. Luca *et al.* [105] proposed a methodology to recommend summaries of large teaching documents and these recommendations are customized to student's needs according to the test results conducted at end of lectures. A multiple-choice test was conducted at the end of a lecture to assess the student's level of understanding of different topics. The authors processed the text results and teaching material in parallel and summarized the content with multilingual weighted itemset-based summarizer (MWISum) [106]. Based on the student understanding, teaching material summary was recommended. Similarly, a lecture summarization service was proposed by Miller [107] using BERT model for dynamically sized lecture summarizations. The author used BERT model to generate embeddings for K-means clustering, which is an extractive text summarization approach.

Abstractive text summarization preserves actual information and overall meaning while summarizing the sentences from a corpus with a shorter representation. In a study conducted by Song *et al.* [108], a deep learning-based framework was proposed to construct new sentences based on semantic phrases using an LSTM-CNN model. The semantic phrases were not conventional tokenized sentences, the authors performed a phrase acquisition, phase refinement and phrase combination on the preprocessed database for the phrase extraction. The LSTM-CNN deep learning algorithm was trained using the extracted semantic phrase and set a threshold value to divide the text generation stages into a generating mode and a copy mode. The proposed approach of abstractive text summarization using the LSTM-CNN model outperformed other state-of-the-art systems using a CNN in terms of metric recall-oriented understudy for gisting evaluation (ROUGE-1) [109] of 34.9% (an increase of 4.4% over existing models) and the ROUGE-2 [109] of 17.8% (an increase of 1.6% over existing models).

Asmussen *et al.* [57] developed a smart exploratory literature review where the authors proposed a three-step framework with pre-processing, topic modelling using an LDA technique, and post-processing. In the preprocessing step, articles were loaded to clean the non-value-adding words, to convert words to lowercases, to remove punctuations, special characters, whitespaces, URLs, and emails. The cleaning process differed from domain to domain, as non-value-adding words differed for each domain. Further to this, the number of topics in LDA topic modelling was estimated using a cross-validation technique. Once the number of topics was determined, the LDA model was executed. The outcomes of

the model include the list of articles, a list of probabilities for each article for each topic, and a list of the most frequent words for each topic. In the post-processing step, identified research topics and labelled the topics that are relevant for use in a literature review. The LDA model was evaluated using statistical, semantic, or predictive approaches.

2) DOCUMENT CATEGORIZATION
Document categorization is one form of annotation to annotate a document in a text corpus [110]. It analyzes the content, intent and sentiment within a document and classifies them into predefined labels. Document categorization or text classification analogizes end-to-end entity linking where an entity linking labels individual words or phrases, document categorization annotates an entire text or body of a document with a single label. Sentiment annotation and linguistic annotation are part of document categorization to extract latent semantic and linguistic elements in a document.

Sindhu et al. [111] proposed supervised aspect-based opinion mining of students' feedback for teaching performance evaluation. In this study, six different aspects like teaching pedagogy, behaviour, knowledge, assessment, experience, and general were considered as domain understanding. Student feedback labelled with these aspects and description of each aspect were preprocessed to create academic domain word embeddings to represent words semantically. LSTM model with layer 1 as aspect extraction and layer 2 as opinion orientation was designed, and the model achieved accuracy of 91% accuracy in aspect extraction and 93% in sentiment detection.

Li et al. [19] proposed an integrated hybrid deep learning methodology [112], [113] with a combination of LSTM and CNN models for Chinese text classification [114]. The features from processing serialized information in the LSTM were used along with a convolutional layer to extract more features. A BLSTM-C model was also proposed. The authors used three benchmark datasets in Chinese language with eight categories of articles. A BBC English news dataset with five categories was also tested to compare the experiment with Chinese datasets. All the datasets were preprocessed with word vectors using a Word2vec model and used a maxlen method to denote the maximum length of a sentence. Sentences with shorter lengths were padded with '0' vectors. The authors compared the classification accuracy of a simple LSTM and the proposed BLSTM-C on both English and Chinese datasets. They achieved an accuracy of 91.73%, 94.88% and 91.11%, 96.23% respectively.

3) ENTITY EXTRACTION
To identify named entities, parts of speech and key phrases within a text, an entity annotation technique can be used [115]. Annotators read the text thoroughly to locate the target entities based on predefined labels. The located entities in entity annotation can be connected to larger repositories of data using entity linking. In end-to-end entity linking, preprocess a piece of text for named entity extraction. In entity

disambiguation, extracted named entities will be linked to knowledge databases.

Dess et al. [116] proposed a novel architecture for extracting entities and relations among entities. An existing extractor framework [117] based on a deep learning model and entity detection module was modified and embedded in the proposed architecture to detect six types of entities like task, method, material, metric, other scientific-term, and generic. Seven types of relations like compare, part-of, conjunction, evaluate-for, feature-f, hyponym-of, used-for were defined.

4) KNOWLEDGE GRAPHS
Knowledge graphs can represent information extracted using NLP in an abstract form and integrate the information extracted from multiple data sources. Domain knowledge from knowledge graphs are input into a machine learning model to produce better predictions. A knowledge graph can be served as a data structure which can store information. A combination of human input, automated and semi-automated extracted data can be added to a knowledge graph.
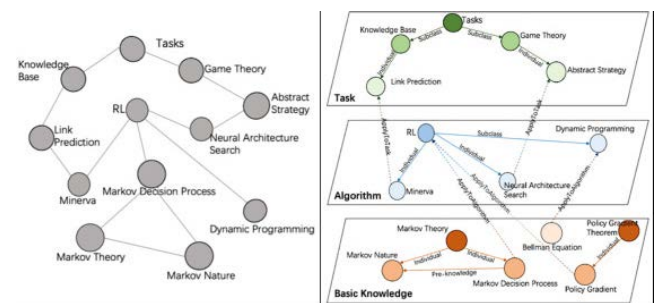

FIGURE 7. Sample knowledge graph [118].

To recommend a well-organized diverse learning path, Shi et al. [118] proposed a learning path recommendation model based on a multidimensional knowledge graph framework. In this framework, learning objects were separately stored in several classes and it also proposed six semantic relationships between the learning objects in the knowledge graph. Figure 7 presents the multidimensional knowledge graph from [118] where dotted lines represent inter-class relationships, solid lines represent intra-class relationships, and the nodes with different colors represent learning objects in different classes. Then a learning path recommendation model was designed to traverse each from the knowledge graph and recommend the best learning path to students. Extracting context data in NLP is critical as it highly influences the model classification [119]. To store the extracted data, knowledge graph is the best approach and can easily map with different objects.

5) SENTIMENT ANNOTATION
One of the most trending annotations in NLP is sentiment annotation which is to label emotion, opinion and sentiment inherent within a text. The label could be a positive sentiment, neutral sentiment, or negative sentiment. It deals with

emotional intelligence quotient in sentiment analysis or opinion mining. In natural language, understanding the context is critical. Without comprehensive understanding, it is difficult to predict the true emotion behind a text message or email. It is much more difficult for machines to mine customer intention in reviews or feedback, especially with sarcasm and humour. Sentiment annotated data are used to train machine learning models and help them to do sentiment analysis or opinion mining [3], [120].

Ibrahim *et al.* [121] proposed a data mining framework to analyze student feedback with classification algorithms like Naive Bayes, SVM, decision tree, and random forest. Sutoyo *et al.* [122] proposed a feedback questionnaire for lecturer evaluation based on student feedback to the questionnaire. Sentiment analysis was performed on the student feedback and classified them into positive or negative sentiment using CNN model. The deep learning model achieved accuracy, precision, recall, and F1-Score of 87.95%, 87%, 78%, and 81%, respectively. Similarly, Kandhro *et al.* [123] proposed LSTM model with predefined word embedding layer for sentiment analysis and achieved an accuracy of 92% and 79% for positive and negative sentiment classification respectively. Using annotation technique, FIT-EBot, a chatbot in a university was proposed in a study conducted by Hien *et al.* [124]. The authors used a NLP technique to extract context and intention in a student query for the chatbot. After analyzing the student's intention and context analysis, three models, namely a pattern-based model [125], a retrieval-based model [126], and a generative model [127] were used to build responses to the student query. While decoding student query, the text messages were classified into 13 predefined topics. A classifier was trained with the manually defined 13 topics based on the results from a survey. The 13 topics defined in the study was course registration, alternative course, prerequisite course, course content, major, course material, scholarship, graduation and others that were used to extract the intent of the student query [128]. After intention, the context of the query was extracted using named entity recognition. For that purpose, a corpus in which each word has been exactly identified with a label was trained using a classifier, so that the model could be used to extract context from the student query. The proposed approach achieved an F1-score of 82.33%, 97.33% for student intent identification and student context extraction respectively. A text annotation was executed to review students' opinion and extracted the sentiment of the opinion in higher education in a study [129]. A MATTER (Model, Annotate, Test, Train, Evaluate, Revise) methodology was implemented as part of the annotation procedure. In that study, student opinion was annotated manually based on a predefined annotation scheme and evaluated using an inter-rater agreement.

III. CHALLENGES

In this section, challenges in implementing NLP techniques in the education domain are discussed.

A. DOMAIN-SPECIFIC LANGUAGE

In order to classify academic dataset or students' feedback, it is required to understand core factors of teaching context [111]. This is considered one of the challenges in implementing NLP in education domain. Considering abundant student feedback being generated from different surveys, questionnaires, and other educational feedback acquiring portals on a course teaching or a learning management system. Without understanding or getting trained on the specific domains, NLP methodologies could not be able to uncover the latent semantic meaning of a text. Nhi *et al.* [130] proposed a domain-specific NLP for students, faculty members, universities in computer science or information technology in higher education sector. The authors extracted tech-related skills using named entity recognition (NER) and built a personalized multi-level course recommendation system. This is a domain-specific NER designed to scrape data like job postings, course descriptions, and MOOCs online courses [131] information from multiple websites and enhance the system with annotated corpus from StackOverflow and GitHub [132]. The annotated StackOverflow data were embedded and split into train, test and validation datasets. The train and test data were fed to a Bi-LSTM and the proposed CSIT-NER models for training, and GitHub data with StackOverflow test dataset was used to evaluate the models. The scraped data were embedded and extracted entities to form a corpus. Pashev *et al.* [133] proposed a methodology to extract entities and their relations using MeaningCloud API and Google Translate API. The authors calculated grades based on the relevance to the topics created by a teacher or auto-generated text from the subject area. Extracting entities or concepts from a huge database available using a data scraping technique and processing them with considerable manual annotation would assist in building corpora for an application domain [134].

B. SARCASM

Decoding sarcasm is critical in NLP tasks like sentiment annotation and opinion analysis. This helps to decipher student opinions and perceptions on course structure and educational infrastructure. In a survey article in [135], automatic sarcasm detection was studied explicitly. The authors surveyed existing traditional sarcasm detection studies and reported the research gap. Sarcasm labels are hidden attributes that need to be predicted by considering conversations and sentences before and after a sarcastic text. The datasets used for sarcasm detection in that research were divided into categories like short text, long text, transcripts, dialogues and miscellaneous. To detect sarcasm, the authors reported three different approaches such as rule-based, statistical, and deep learning approaches. In a rule-based approach, sarcasm can be identified based on key indicators of sarcasm captured as evidence [134], [136], [137]. In a statistical approach to detect sarcasm, punctuations,

sentiment-lexicon-based features, unigrams, word embedding similarity, frequency of the rarest words, sentiment flips and so on were key features to the statistical classifiers [138]–[140]. Traditional machine learning algorithms like SVM [141], logistic regression, decision trees, Naive Bayes, hidden Markov model, and ensemble classification methods were also used in classifying the sarcasm. In deep learning algorithms, RNN models and LSTM methods [142] can be used individually as well as in combination with CNNs [143] for automatic sarcasm detection. The survey article provided a comprehensive understanding of sarcasm detection.

C. AMBIGUITY
Ambiguity in natural languages is common as it depends on context and user perception in reading a text. With challenges in decoding a context, ambiguity in machine learning language processing is more complicated. Ambiguity could be due to the structure, syntactic, or lexical nature of a sentence [144]. In structural ambiguity, a sentence has more than one syntactic structure. In syntactic ambiguity, a grammatical construct error occurs in a sub-part of a sentence that causes grammatical ambiguity in a complete sentence. Lexical ambiguity is due to a word having two different meanings and two words having the same form. Addressing the ambiguity challenge is crucial in analyzing feedback. In a study in [145], word sense disambiguation was addressed by customizing BERT, a language representation model, and selecting the best context-gloss pairs from a group of related pairs [146]. The authors classified the context-gloss pairs into positive and negative sentiment, and example sentences from WordNet 3.0 were combined with the positive and negative gloss pairs. Annotating the combination assisted in creating additional training samples. The proposed BERT model outperformed other existing state-of-art models in terms of F1-score with 77%.

D. EMOTICONS AND SPECIAL CHARACTERS
Emoticons and special characters play a vital role in opinion mining especially students' feedback containing the special symbols to express their emotions. NLP has a challenging phase in processing the emoticons and labelling them with appropriate emotion tags. In a study in 2020 [147], the authors analyzed cross-cultural reactions to the novel coronavirus and detected sentiment polarity and emotion from their tweets and validated them with emoticons. A deep learning model based on LSTM was used in combination with feature extraction methods like GloVe, word embeddings. Six emotions of joy, surprise, sadness, anger, fear, and disgust were validated using different emoticons with their unicodes. Cappallo *et al.* [148] proposed a large dataset with real-world emojis and explained three challenges in emoticons processing. They were emoji processing, emoji anticipation, and query-by-emoji. The authors used two deep learning models, a Bi-LSTM model for text-to-emoji baseline results and a CNN model for image-to- emoji. They then combined the

two algorithms to form a multi-modal method approach for emoticons processing. The work can be adopted in analyzing student opinions and processing the emoticons used in their feedback.

E. ASPECT-BASED SENTIMENT ANALYSIS
Chauhan *et al.* [149] quoted that sentiment analysis tool largely underused in education could not find opinions on different aspects. Most of the research works to process student comments or feedback to classify the positive or negative sentiment using lexicon-based or machine learning methods at document level. Nazir *et al.* [150] conducted a survey on issues and challenges that are related to extraction of different aspects. The study was divided into three topics aspect extraction, aspect sentiment analysis, and sentiment evolution. Each topic was breakdown into sub-categories explicit aspect extraction, implicit aspect extraction, aspect level sentiment analysis, entity level sentiment analysis, multi-word sentiment analysis, recognition of factors in sentiment evolution, and predicting sentiment evolution over social data.

F. DATA IMBALANCE
Data imbalance is one of the most common challenges in AI [151] in which number of samples in one class exceeds the amount in other classes. Considering NLP, a subset of AI, the challenge is inherited. Especially in education domain, it is difficult to acquire of massive labelled data as it requires manual annotation from domain experts. Although the acquired labelled data fed to deep learning algorithms, the classification performance is biased due to data distribution discrepancy [152]. A potential tool to overcome this challenge could transfer learning [153], where a deep learning model trained on a large corpus of student feedback to perform similar tasks on another data source. Other techniques could be sampling techniques [154] to under-sample majority classes or oversample minority classes, which might demand text augmentation tasks [155].

IV. DISCUSSION
According to a Gartner diagram shown in Figure 8 [156], decision intelligence, deep learning, knowledge graphs are at peak point which can be adapted to the education domain to build decision support systems. These areas analyze existing data and streamline the process of data storage. Deep learning methods can be used without much expertise in the application domain and build semantic networks to store interlinked entity data in a domain. It is expected that 70% of the organizations will shift their focus from big to small and wide data by 2025 to provide more context for data analytics [157]. This infers a variety of small structured and unstructured data sources in diverse platforms.

NLP, part of AI, makes it possible to understand human language and listen to their opinions and feedback. Especially, education institutions need to adopt NLP methods to enhance the student learning experience, personalized learning management systems [9], and teacher training. This would help
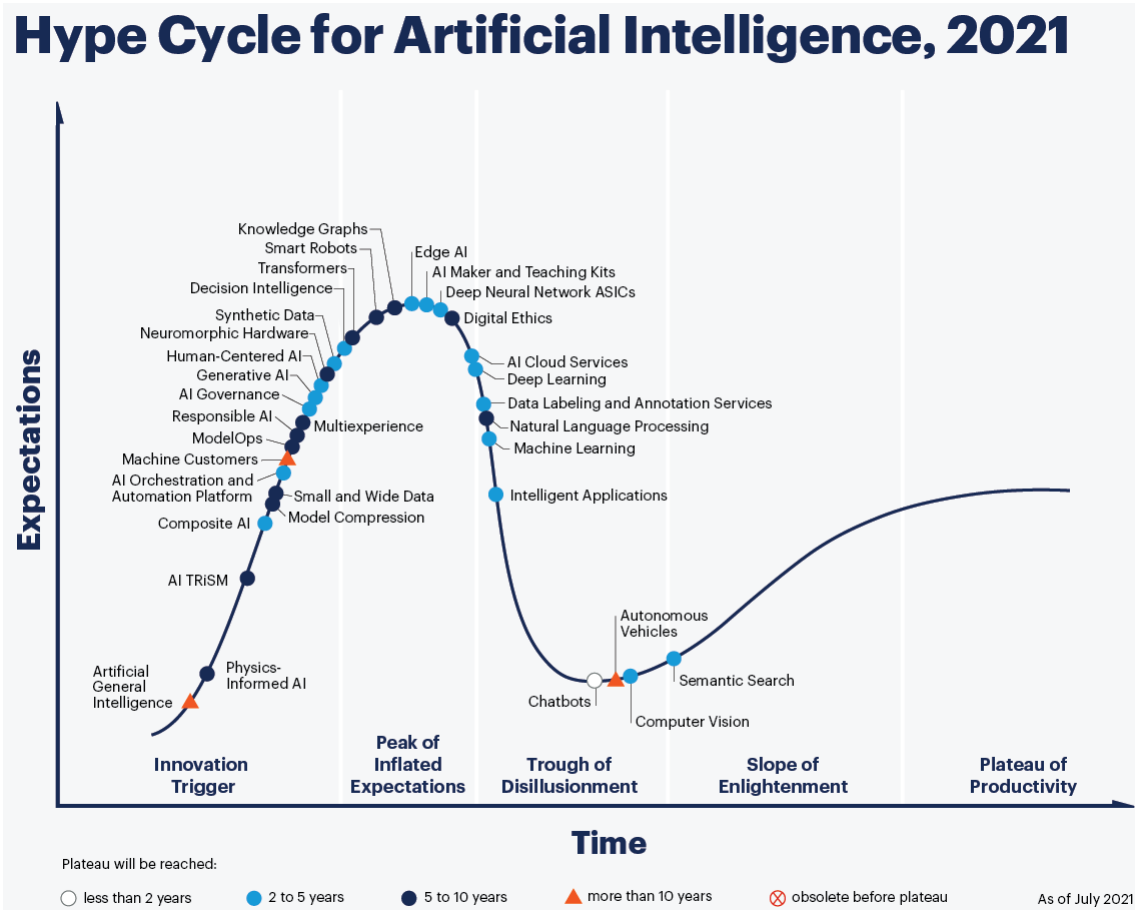
FIGURE 8. Gartner's hype cycle for artificial intelligence 2021 [156].

transform and get students expose to AI- driven tools. There is a wide variety of AI applications in education like smart classrooms with video & audio data annotation [158]–[160], NLP for textual data annotation, classification, summarization, and image processing to detect gestures [161], [162]. In this study, NLP methodologies were focused on and discussed their applications by AI. Although few research articles are related to AI in education directly, their approaches can be adapted to education.

Holmes *et al.* [163] discussed problems and future implications of AI in education. The authors raised two problems of "What we teach, and how we teach it". What we teach refers to what students should learn in the age of AI, and the learning goals should be versatility, relevance, and transferability. To achieve these goals, strategies like emphasize on selective traditional knowledge areas, addition of modern knowledge, interdisciplinary concepts, embedded skills, and meta learning were proposed. Coming to the how-to question, it refers to how AI can enhance and transform education. The authors draw a line between education technology and AI in education. Education technology is to amend the taxonomy and ontology of the field, whereas AI in educations deals with a layered framework of substitution, augmentation,

modification, and redefinition, which includes enhancement and transformation. The authors quoted an assumption where many assume robot teachers teach students with AI in education [164], [165]. Although it could be possible in the future, current research is to transform and evolve the education industry to amplify the student learning experience without considering the enhancement of teaching practices. The authors proposed intelligent tutoring systems which involved a domain model with subject knowledge, a pedagogy model with effective teaching and learning approaches, a learner model for individual student learning.

NLP techniques can be executed on feedback data using different programming languages. Even though there are different languages with pre-built packages to execute NLP methods, Python, Java and R programming languages are widely being used [166]. The factors to be considered in adopting a programming language includes the expertise in the specific programming language, the number of libraries or package tools that could assist in performing NLP tasks [167]. Python [168], with its versatility and simple consistent syntax mirroring human language, can offer a huge number of NLP packages for topic modelling, word embeddings, document classification, sentiment annotations. Java [169] is a

TABLE 1. NLP - Programming packages.

Programming Languages	Packages	Features
Python	spaCy	Part-of-Speech Tagging, Tokenization, Dependency Parsing, Sentence Segmentation, Entity Sentence Recognition, Seamless integration with Deep Learning, Methods for cleaning and normalizing text
	CoreNLP	Lemmatization, Named Entity Recognition, Morphological Tagging, Tokenization, Part-of-Speech tagging
	NLTK	Tokenization, Part-of-Speech Tagging, Stemming
	Gensim	Topic Modelling Algorithms
	PyNLPI	Tokenization, Frequency Lists, Supports formats like FoLiA, timbl, GIZA++, moses
	Pattern	Part-of-Speech Taggers, n-grams, Sentiment Analysis, WordNet, Text Classification, Tokenization
	Polyglot	Tokenization, Language detection, part-of-speech tagging
	TextBlob	Part-of-Speech Tagging, Sentiment Analysis, Classification, Tokenization, n-grams, Word Inflection, WordNet Integration, Language translation and detection powered by Google Translate, Word, and phrase frequencies, Parsing, Spelling correction, Add new models or languages through extensions
	Quepy	Transform natural language questions to queries in a database query language
Java	OpenNLP	Lemmatization, Named Entity Recognition, Morphological Tagging, Tokenization, Part-of-Speech tagging
	StanfordNLP	Lemmatization, Named Entity Recognition, Morphological Tagging, Tokenization, Part-of-Speech tagging
	CogCompNLP	Collects a number of core libraries for NLP for Part-of-Speech Tagging, Lemmatization, Named Entity Recognition, Temporal extractor and normalizer
	MALLET	Statistical NLP with document classification, clustering, topic modelling, information extraction
	NLP4J	Morphological analysis, Data Crawler, Annotator, Parsing, Statistical indexing function
	CORENLP	Derive linguistic annotations for text, Part-of-Speech tagging, Named Entity Recognition, Numeric and Time values, Dependency and Constituency parses, Co-reference, Sentiment, Quote attributions, and relations
R Programming	koRpus	Part-of-Speech tagging, Tree Tagger, functions for automatic language detection, hyphenation, several indices of lexical diversity
	lsa	Latent Semantic Analysis, statistically derive conceptual indices to extract higher order structure.
	OpenNLP	Lemmatization, Named Entity Recognition, Morphological Tagging, Tokenization, Part-of-Speech tagging
	Quanteda	Quantitative analysis of textual data , tokenizers, ngrams, and analyzing keywords
	RWeka	Collection of machine learning algorithms. It deals with data mining, data pre-processing,classification, regression, clustering, association rules, and visualization

platform-independent language with robust architecture that can provide comprehensive text analysis tasks like clustering, tagging, and information extraction using multiple packages. R programming language [170] is popular for its statistical learning, it is also being widely used in NLP tasks. The programming language can handle computationally intensive data analytics and investigate big data applications. Table 1 presents the three programming languages with their NLP packages. It has features for each package and its documentation source.

In this work, four research questions on NLP in education were explored. For the first research question, the existing methodologies being used for NLP were discussed. Section II discussed in details data preprocessing methods feature extraction and feature selection definitions, types, and research community's work using machine learning and deep learning models. Different approaches of machine learning technique topic modelling were explained and student feedback topic extraction works were explored. Further to this, text evaluation techniques like text summarization, document

TABLE 2. NLP techniques—Research works.

NLP Techniques	Algorithms	Research Works
Methodology		
Feature Extraction	BERT, LSA, LDA, CNN, LSTM, LSTM+ATT	[22]–[39], [171]
Feature Selection	Naive Bayes, Chi-square, SVM, random forest	[40]–[52]
Topic Modelling	LSA, NMF, LDA, PLSA	[18], [53]–[56], [58]–[91], [172]–[174]
Text Evaluation		
Text Summarization	LSTM-NN, CNN, Bi-GRU, TextRank, LSA, BERT, LSTM-CNN	[57], [92]–[109]
Document Categorization	LSTM, CNN, BLSTM-C	[19], [110]–[114]
Entity Extraction	CNN	[115]–[117]
Knowledge Graphs	Multidimensional knowledge graphs	[118], [119]
Sentiment Annotation	CNN, LSTM	[3], [120]–[129]

categorization, text annotation, and knowledge graphs were explained.

The second research question was about the challenges of NLP in the education domain. Generic NLP challenges like domain-specific language, sarcasm, ambiguity, data imbalance are challenging in education to uncover the latent semantic meaning of students' feedback. The research community addressed this challenge using NER, rule-based, statistical, deep learning and BERT modelling. Emoticons and special characters are used to express their sentiment in feedback. To process these special symbols and characters, a multimodal approach was used. Converting emojis to their corresponding unicodes or image processing were used to determine the sentiment. Aspect-based sentiment analysis is more trending challenge of NLP in education domain. This challenge also refers to fine-grained sentiment analysis [3].

The third research question is to address the trends in NLP methodologies that could be adopted in the education domain. Those language processing methods were discussed in detail. AI models need to be trained in a quantitative approach, which will require pre-processing the textual data into vectors using feature extraction and feature selection techniques. In topic modelling techniques, both probabilistic and non-probabilistic models were discussed but the LDA technique from probabilistic models was the most commonly used to extract unsupervised topics from a corpus. Research on the fourth question investigated the community's work from other industry applications with short text analysis was discussed to understand [32], [56] and adapt to the education application domain. In [32], twitter text analysis is short text analysis using VADER and SentiWordNet techniques. This approach can be applied to student feedback analysis, which would be short text in most cases. Query expansion ranking in [48] is feature selection method that can be used in education feedback analysis. Patil *et al.* [54] approach to analyze e-commerce site Amazon product reviews can be
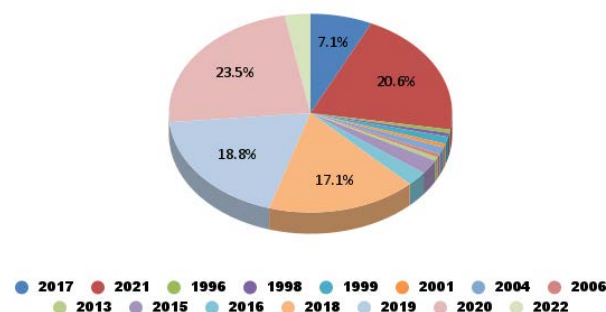

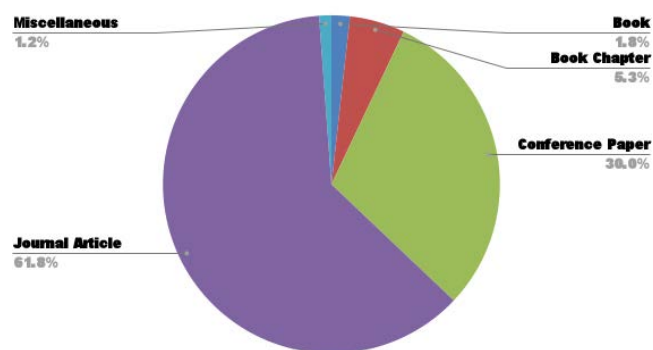**FIGURE 9.** Year-wise references distribution.

FIGURE 10. References distribution.

directly used in higher education feedback analysis to analyze student ratings to a course delivery and extract sentiment aspects from student comments. Emojis are one of the forms' that student use to express their opinion in feedback. The approach in [148] can be used to process emojis in student comments.

V. CONCLUSION

The aim of the study is to explore existing NLP methodologies that can be implemented or adopted in education domain. This assist to understand AI impact on education with open opportunities, synthesize the methods to process student feedback, and annotate their views. The literature review has been performed using Google Scholar covering bibliographic databases such as Wiley, Scopus, Springer, ACM Digital Library, IEEE Xplore, Pub-Med, Science Direct, and Multidisciplinary Digital Publishing Institute (MDPI) and so on. The search results of Google Scholar were manually checked for relevance NLP techniques in student feedback or education applications that can be adopted to the feedback analysis. For example, Twitter data analysis which consists of short text analysis using NLP similar student feedback. As shown in Figure 9, the majority of the references included in this study are from the last 5 years. Also, more than 90% percent of the citation included in this study are journal articles and conference papers. Table 2 presents the NLP techniques explored in this study and corresponding research community works citations.

In this review article, the impact of AI on education was discussed. The scope of introducing AI into educational institutions is detailed based on the opportunities. Limiting the scope of introducing NLP methodologies to education for feedback analysis in this article, existing NLP methodologies were explored. Feature extraction, feature selection and topic modelling methodologies were explained with brief definitions. Further to this, text evaluation techniques text summarization, annotation, and knowledge graphs were reviewed. Each of these applications was defined and existing approaches were discussed. Challenges in adopting NLP methodologies to the education domain were reviewed. The limitation of this research is that this study is confined to AI implementation methodologies with less focus on pedagogy concepts. Data specific challenges like data scarcity and class imbalance were not discussed. This would affect the model learning for deep learning algorithms, which are data hungry. Strategies to interpret deep learning models (black box) were not explored. The future direction of this research would be to explore data challenges while extracting feedback or opinions without affecting privacy.

CONFLICTS OF INTEREST

The authors declare no conflict of interest.

REFERENCES

[1] M. A. Peters, "Deep learning, education and the final stage of automation," *Educ. Philosophy Theory*, vol. 50, nos. 6–7, pp. 549–553, Jul. 2017, doi: 10.1080/00131857.2017.1348928.

[2] X. J. Hunt, I. K. Kabul, and J. Silva, "Transfer learning for education data," in *Proc. ACM SIGKDD Conf.*, vol. 1. Halifax, NS, Canada, 2017, pp. 1–6.

[3] Z. Kastrati, F. Dalipi, A. S. Imran, K. P. Nuci, and M. A. Wani, "Sentiment analysis of students' feedback with NLP and deep learning: A systematic mapping study," *Appl. Sci.*, vol. 11, no. 9, p. 3986, Apr. 2021, doi: 10.3390/app11093986.

[4] *OECD Education Working Papers*, Org. Econ. Co-Oper. Develop., Paris, France, 2021, doi: 10.1787/19939019.

[5] M. Chassignol, A. Khoroshavin, A. Klimova, and A. Bilyatdinova, "Artificial intelligence trends in education: A narrative overview," *Proc. Comput. Sci.*, vol. 136, pp. 16–24, Jan. 2018, doi: 10.1016/j.procs.2018.08.233.

[6] L. Chen, P. Chen, and Z. Lin, "Artificial intelligence in education: A review," *IEEE Access*, vol. 8, pp. 75264–75278, 2020.

[7] M. L. B. Estrada, R. Z. Cabada, R. O. Bustillos, and M. Graff, "Opinion mining and emotion recognition applied to learning environments," *Expert Syst. Appl.*, vol. 150, Jul. 2020, Art. no. 113265, doi: 10.1016/j.eswa.2020.113265.

[8] A. Yadav and D. K. Vishwakarma, "Sentiment analysis using deep learning architectures: A review," *Artif. Intell. Rev.*, vol. 53, no. 6, pp. 4335–4385, Dec. 2019, doi: 10.1007/s10462-019-09794-5.

[9] K. Eggert, "How artificial intelligence will shape universities of tomorrow," in *Proc. Int. Conf.*, 2021, p. 50.

[10] K. Holstein, B. M. McLaren, and V. Aleven, "Designing for complementarity: Teacher and student needs for orchestration support in AI-enhanced classrooms," in *Artificial Intelligence in Education* (Lecture Notes in Computer Science). Cham, Switzerland: Springer, 2019, pp. 157–171, doi: 10.1007/978-3-030-23204-7_14.

[11] A. Bhimdiwala, R. C. Neri, and L. M. Gomez, "Advancing the design and implementation of artificial intelligence in education through continuous improvement," *Int. J. Artif. Intell. Educ.*, vol. 11625, pp. 1–27, Oct. 2021, doi: 10.1007/s40593-021-00278-8.

[12] K. Holstein, B. M. McLaren, and V. Aleven, "Co-designing a real-time classroom orchestration tool to support teacher–AI complementarity," *J. Learn. Anal.*, vol. 6, no. 2, pp. 27–52, Jul. 2019, doi: 10.18608/jla.2019.62.3.

[13] H. Labarthe, V. Luengo, and F. Bouchet, "Analyzing the relationships between learning analytics, educational data mining and ai for education," in *Proc. 14th Int. Conf. Intell. Tutoring Syst. (ITS), Workshop Learn. Anal.*, 2018, pp. 10–19.

[14] L. Alrajhi, A. Alamri, F. D. Pereira, and A. I. Cristea, "Urgency analysis of learners' comments: An automated intervention priority model for MOOC," in *Proc. Int. Conf. Intell. Tutoring Syst.* Cham, Switzerland: Springer, 2021, pp. 148–160.

[15] R. C. Sharma, P. Kawachi, and A. Bozkurt, "The landscape of artificial intelligence in open, online and distance education: Promises and concerns," *Asian J. Distance Educ.*, vol. 14, no. 2, pp. 1–2, 2019.

[16] K. Gulson, A. Murphie, S. Taylor, and S. Sellar, "Education, work and Australian society in an AI world," Gonski Inst. Educ., Univ. New South Wales Sydney, Sydney, NSW, Australia, Tech. Rep., 2018.

[17] F. Pedro, M. Subosa, A. Rivas, and P. Valverde, "Artificial intelligence in education: Challenges and opportunities for sustainable development," Nat. Inst. Educ. Develop., Okahandja, Namibia, Tech. Rep., 2019.

[18] S. Gottipati, V. Shankararaman, and S. Gan, "A conceptual framework for analyzing students' feedback," in *Proc. IEEE Frontiers Educ. Conf. (FIE)*, Oct. 2017, pp. 1–8.

[19] Y. Li, X. Wang, and P. Xu, "Chinese text classification model based on deep learning," *Future Internet*, vol. 10, no. 11, p. 113, Nov. 2018, doi: 10.3390/fi10110113.

[20] S. Ramaswamy and N. DeClerck, "Customer perception analysis using deep learning and NLP," *Proc. Comput. Sci.*, vol. 140, pp. 170–178, Jan. 2018, doi: 10.1016/j.procs.2018.10.326.

[21] S. Prokhorov and V. Safronov, "AI for AI: What NLP techniques help researchers find the right articles on NLP," in *Proc. Int. Conf. Artif. Intell., Appl. Innov. (IC-AIAI)*, Sep. 2019, pp. 765–776, doi: 10.1109/ic-aiai48757.2019.00023.

[22] R. Ahuja, A. Chug, S. Kohli, S. Gupta, and P. Ahuja, "The impact of features extraction on the sentiment analysis," *Proc. Comput. Sci.*, vol. 152, pp. 341–348, Jan. 2019, doi: 10.1016/j.procs.2019.05.008.

[23] A. Balamurali and B. Ananthanarayanan, "Develop a neural model to score bigram of words using bag-of-words model for sentiment analysis," in *Neural Networks for Natural Language Processing*. Pennsylvania, PA, USA: IGI Global, 2020, pp. 122–142, doi: 10.4018/978-1-7998-1159-6.ch008.

[24] Y. Goldberg and G. Hirst, *Neural Network Methods in Natural Language Processing*. San Rafael, CA, USA: Morgan & Claypool, 2017.

[25] I. Arroyo-Fernández, C.-F. Méndez-Cruz, G. Sierra, J.-M. Torres-Moreno, and G. Sidorov, "Unsupervised sentence representations as word information series: Revisiting TF–IDF," *Comput. Speech Lang.*, vol. 56, pp. 107–129, Jul. 2019, doi: 10.1016/j.csl.2019.01.005.

[26] F. M. Shah, F. Haque, R. U. Nur, S. A. Jahan, and Z. Mamud, "A hybridized feature extraction approach to suicidal ideation detection from social media post," in *Proc. IEEE Region 10 Symp. (TENSYMP)*, Jun. 2020, pp. 985–988, doi: 10.1109/TENSYMP50017.2020.9230733.

[27] A. Onan, "Sentiment analysis on product reviews based on weighted word embeddings and deep neural networks," *Concurrency Comput., Pract. Exper.*, vol. 33, no. 23, p. e5909, Jun. 2020, doi: 10.1002/cpe.5909.

[28] S. Amin, M. I. Uddin, M. A. Zeb, A. A. Alarood, M. Mahmoud, and M. H. Alkinani, "Detecting dengue/flu infections based on tweets using LSTM and word embedding," *IEEE Access*, vol. 8, pp. 189054–189068, 2020, doi: 10.1109/ACCESS.2020.3031174.

[29] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of deep bidirectional transformers for language understanding," 2018, *arXiv:1810.04805*.

[30] J. Hou, X. Li, H. Yao, H. Sun, T. Mai, and R. J. I. A. Zhu, "Bert-based Chinese relation extraction for public security," *IEEE Access*, vol. 8, pp. 132367–132375, 2020.

[31] R. N. Waykole and A. D. Thakare, "A review of feature extraction methods for text classification," *Int. J. Adv. Eng. Res. Develop.*, vol. 5, no. 4, pp. 351–354, 2018.

[32] D. Deepa and A. Tamilarasi, "Sentiment analysis using feature extraction and dictionary-based approaches," in *Proc. 3rd Int. Conf. I-SMAC (IoT Social, Mobile, Anal. Cloud) (I-SMAC)*, Dec. 2019, pp. 786–790, doi: 10.1109/i-smac47947.2019.9032456.

[33] R. Dzisevic and D. Sesok, "Text classification using different feature extraction approaches," in *Proc. Open Conf. Electr., Electron. Inf. Sci. (eStream)*, Apr. 2019, pp. 1–4, doi: 10.1109/estream.2019.8732167.

[34] D. Goularas and S. Kamis, "Evaluation of deep learning techniques in sentiment analysis from Twitter data," in *Proc. Int. Conf. Deep Learn. Mach. Learn. Emerg. Appl. (Deep-ML)*, Aug. 2019, pp. 12–17, doi: 10.1109/deep-ml.2019.00011.

[35] K. Sangeetha and D. Prabha, "Sentiment analysis of student feedback using multi-head attention fusion model of word and context embedding for LSTM," *J. Ambient Intell. Humanized Comput.*, vol. 12, no. 3, pp. 4117–4126, Mar. 2020, doi: 10.1007/s12652-020-01791-9.

[36] H. Zhang, J. Dong, L. Min, and P. Bi, "A BERT fine-tuning model for targeted sentiment analysis of Chinese online course reviews," *Int. J. Artif. Intell. Tools*, vol. 29, no. 7, Dec. 2020, Art. no. 2040018.

[37] M. Masala, S. Ruseti, M. Dascalu, and C. Dobre, "Extracting and clustering main ideas from student feedback using language models," in *Proc. Int. Conf. Artif. Intell. Educ.* Cham, Switzerland: Springer, 2021, pp. 282–292.

[38] Y. Wu, J. Nouri, X. Li, R. Weegar, M. Afzaal, and A. Zia, "A word embeddings based clustering approach for collaborative learning group formation," in *Proc. Int. Conf. Artif. Intell. Educ.* Cham, Switzerland: Springer, 2021, pp. 395–400.

[39] D. Wang, J. Su, and H. Yu, "Feature extraction and analysis of natural language processing for deep learning English language," *IEEE Access*, vol. 8, pp. 46335–46345, 2020.

[40] B. Ghojogh, M. N. Samad, S. A. Mashhadi, T. Kapoor, W. Ali, F. Karray, and M. Crowley, "Feature selection and feature extraction in pattern analysis: A literature review," 2019, *arXiv:1905.02845*.

[41] J. Cai, J. Luo, S. Wang, and S. Yang, "Feature selection in machine learning: A new perspective," *Neurocomputing*, vol. 300, pp. 70–79, Jul. 2018, doi: 10.1016/j.neucom.2017.11.077.

[42] G. Kou, P. Yang, Y. Peng, F. Xiao, Y. Chen, and F. E. Alsaadi, "Evaluation of feature selection methods for text classification with small datasets using multiple criteria decision-making methods," *Appl. Soft Comput.*, vol. 86, Jan. 2020, Art. no. 105836, doi: 10.1016/j.asoc.2019.105836.

[43] N. Nikolić, O. Grljević, and A. Kovačević, "Aspect-based sentiment analysis of reviews in the domain of higher education," *Electron. Library*, vol. 38, no. 1, pp. 44–64, Feb. 2020, doi: 10.1108/el-06-2019-0140.

[44] A. Bommert, X. Sun, B. Bischl, J. Rahnenführer, and M. Lang, "Benchmark for filter methods for feature selection in high-dimensional classification data," *Comput. Statist. Data Anal.*, vol. 143, Mar. 2020, Art. no. 106839, doi: 10.1016/j.csda.2019.106839.

[45] M. Wang, X. Tao, and F. Han, "A new method for redundancy analysis in feature selection," in *Proc. 3rd Int. Conf. Algorithms, Comput. Artif. Intell.*, Dec. 2020, pp. 1–5, doi: 10.1145/3446132.3446153.

[46] P. Dhal and C. Azad, "A comprehensive survey on feature selection in the various fields of machine learning," *Int. J. Speech Technol.*, vol. 52, no. 4, pp. 4543–4581, Jul. 2021, doi: 10.1007/s10489-021-02550-9.

[47] H. Osman, M. Ghafari, and O. Nierstrasz, "Automatic feature selection by regularization to improve bug prediction accuracy," in *Proc. IEEE Workshop Mach. Learn. Techn. Softw. Quality Eval. (MaLTeSQuE)*, Feb. 2017, pp. 27–32, doi: 10.1109/MALTESQUE.2017.7882013.

[48] T. Parlar, S. A. Özel, and F. Song, "QER: A new feature selection method for sentiment analysis," *Hum.-Centric Comput. Inf. Sci.*, vol. 8, no. 1, pp. 1–19, Dec. 2018.

[49] C. Pong-Inwong and K. Kaewmak, "Improved sentiment analysis for teaching evaluation using feature selection and voting ensemble learning integration," in *Proc. 2nd IEEE Int. Conf. Comput. Commun. (ICCC)*, Oct. 2016, pp. 1222–1225.

[50] G. Gutiérrez, J. Canul-Reich, A. O. Zezzatti, L. Margain, and J. Ponce, "Mining: Students comments about teacher performance assessment using machine learning algorithms," *Int. J. Combinat. Optim. Problems Informat.*, vol. 9, no. 3, p. 26, 2018.

[51] S. Mihi, B. A. B. Ali, I. E. Bazi, S. Arezki, and N. Laachfoubi, "How digitalization is perceived from Moroccan students: A sentiment analysis study," in *Proc. 4th Int. Conf. Intell. Comput. Data Sci. (ICDS)*, Oct. 2020, pp. 1–7.

[52] W. Tian, J. Li, and H. Li, "A method of feature selection based on Word2Vec in text categorization," in *Proc. 37th Chin. Control Conf. (CCC)*, Jul. 2018, pp. 9452–9455.

[53] H. M. Wallach, "Topic modeling," in *Proc. 23rd Int. Conf. Mach. Learn. (ICML)*, 2006, pp. 977–984, doi: 10.1145/1143844.1143967.

[54] P. P. Patil, S. Phansalkar, and V. V. Kryssanov, "Topic modelling for aspect-level sentiment analysis," in *Proc. 2nd Int. Conf. Data Eng. Commun. Technol.* Singapore: Springer, Oct. 2018, pp. 221–229, doi: 10.1007/978-981-13-1610-4_23.

[55] S. I. Nikolenko, S. Koltcov, and O. Koltsova, "Topic modelling for qualitative studies," *J. Inf. Sci.*, vol. 43, no. 1, pp. 88–102, Jul. 2016, doi: 10.1177/0165551515617393.

[56] S. A. Curiskis, B. Drake, T. R. Osborn, and P. J. Kennedy, "An evaluation of document clustering and topic modelling in two online social networks: Twitter and Reddit," *Inf. Process. Manage.*, vol. 57, no. 2, Mar. 2020, Art. no. 102034, doi: 10.1016/j.ipm.2019.04.002.

[57] C. B. Asmussen and C. Møller, "Smart literature review: A practical topic modelling approach to exploratory literature review," *J. Big Data*, vol. 6, no. 1, pp. 1–18, Oct. 2019, doi: 10.1186/s40537-019-0255-7.

[58] B. V. Barde and A. M. Bainwad, "An overview of topic modeling methods and tools," in *Proc. Int. Conf. Intell. Comput. Control Syst. (ICICCS)*, Jun. 2017, pp. 745–750.

[59] T. K. Landauer, P. W. Foltz, and D. Laham, "An introduction to latent semantic analysis," *Discourse Process.*, vol. 25, nos. 2–3, pp. 259–284, Jan. 1998, doi: 10.1080/01638539809545028.

[60] D. D. Lee and H. S. Seung, "Learning the parts of objects by non-negative matrix factorization," *Nature*, vol. 401, no. 6755, pp. 788–791, Oct. 1999, doi: 10.1038/44565.

[61] S. F. Chen, *Building Probabilistic Models for Natural Language*. Cambridge, MA, USA: Harvard Univ., 1996.

[62] T. Hofmann, "Unsupervised learning by probabilistic latent semantic analysis," *Mach. Learn.*, vol. 42, no. 1, pp. 177–196, Jan. 2001, doi: 10.1023/a:1007617005950.

[63] B. De Finetti, *Theory of Probability: A Critical Introductory Treatment*, vol. 6. Hoboken, NJ, USA: Wiley, 2017.

[64] A. Hamzah, A. F. Hidayatullah, and A. G. Persada, "Discovering trends of mobile learning research using topic modelling approach," *Int. J. Interact. Mobile Technol.*, vol. 14, no. 9, pp. 1–4, Jun. 2020.

[65] D. F. Onah and E. L. Pang, "MOOC design principles: Topic modelling-pyLDAvis visualization & summarisation of learners' engagement," in *Proc. 13th Annu. Int. Conf. Educ. New Learn. Technol.*, 2021, pp. 1–10.

[66] I. Marçal, R. E. Garcia, D. Eler, and R. C. M. Correia, "A strategy to enhance computer science teaching material using topic modelling: Towards overcoming the gap between college and workplace skills," in *Proc. 51st ACM Tech. Symp. Comput. Sci. Educ.*, Feb. 2020, pp. 366–371.

[67] S. Unankard and W. Nadee, "Topic detection for online course feedback using LDA," in *Emerging Technologies for Education*. Cham, Switzerland: Springer, 2020, pp. 133–142, doi: 10.1007/978-3-030-38778-5_16.

[68] S. Cunningham-Nelson, M. Baktashmotlagh, and W. Boles, "Visualizing student opinion through text analysis," *IEEE Trans. Educ.*, vol. 62, no. 4, pp. 305–311, Nov. 2019.

[69] J. K. N. Singh, "Academic resilience among international students: Lived experiences of postgraduate international students in Malaysia," *Asia Pacific Educ. Rev.*, vol. 22, no. 1, pp. 129–138, Nov. 2020, doi: 10.1007/s12564-020-09657-7.

[70] A. Perez-Encinas and J. Rodriguez-Pomeda, "International students' perceptions of their needs when going abroad: Services on demand," *J. Stud. Int. Educ.*, vol. 22, no. 1, pp. 20–36, Feb. 2018.

[71] V. Liermann, "Overview machine learning and deep learning frameworks," in *The Digital Journey of Banking and Insurance*, vol. 3. Cham, Switzerland: Springer, 2021, pp. 187–224, doi: 10.1007/978-3-030-78821-6_12.

[72] S. Srinivas and R. Rajendran, "Topic-based knowledge mining of online student reviews for strategic planning in universities," *Comput. Ind. Eng.*, vol. 128, pp. 974–984, Feb. 2019.

[73] S. Pyasi, S. Gottipati, and V. Shankararaman, "SUFAT—An analytics tool for gaining insights from student feedback comments," in *Proc. IEEE Frontiers Educ. Conf. (FIE)*, Oct. 2018, pp. 1–9.

[74] S. Loria, *Textblob Documentation*, Release 0.15, vol. 2, 2018, p. 269.

[75] M. M. Tadesse, H. Lin, B. Xu, and L. Yang, "Detection of depression-related posts in Reddit social media forum," *IEEE Access*, vol. 7, pp. 44883–44893, 2019.

[76] T. Ruan, Q. Kong, S. K. McBride, A. Sethjiwala, and Q. Lv, "Cross-platform analysis of public responses to the 2019 Ridgecrest earthquake sequence on Twitter and Reddit," *Sci. Rep.*, vol. 12, no. 1, pp. 1–14, Jan. 2022, doi: 10.1038/s41598-022-05359-9.

[77] L. N. Rani, S. Defit, and L. J. Muhammad, "Determination of student subjects in higher education using hybrid data mining method with the K-means algorithm and FP growth," *Int. J. Artif. Intell. Res.*, vol. 5, no. 1, pp. 91–101, Dec. 2021, doi: 10.29099/ijair.v5i1.223.

[78] R. K. Dinata, S. Retno, and N. Hasdyna, "Minimization of the number of iterations in k-medoids clustering with purity algorithm," *Revue d'Intell. Artificielle*, vol. 35, no. 3, pp. 193–199, Jun. 2021, doi: 10.18280/ria.350302.

[79] T.-C. Wang, B. N. Phan, and T. T. T. Nguyen, "Evaluating operation performance in higher education: The case of Vietnam public universities," *Sustainability*, vol. 13, no. 7, p. 4082, Apr. 2021, doi: 10.3390/su13074082.

[80] M. Rahmanian and E. G. Mansoori, "An unsupervised gene selection method based on multivariate normalized mutual information of genes," *Chemometric Intell. Lab. Syst.*, vol. 222, Mar. 2022, Art. no. 104512, doi: 10.1016/j.chemolab.2022.104512.

[81] Y. Hu and K. Dai, "Foreign-born Chinese students learning in China: (Re)shaping intercultural identity in higher education institution," *Int. J. Intercultural Relations*, vol. 80, pp. 89–98, Jan. 2021, doi: 10.1016/j.ijintrel.2020.11.010.

[82] A. D'Ambrosio, S. Amodio, C. Iorio, G. Pandolfo, and R. Siciliano, "Adjusted concordance index: An extensionl of the adjusted Rand index to fuzzy partitions," *J. Classification*, vol. 38, no. 1, pp. 112–128, Jun. 2020, doi: 10.1007/s00357-020-09367-0.

[83] S. Vanaja and M. Belwal, "Aspect-level sentiment analysis on E-commerce data," in *Proc. Int. Conf. Inventive Res. Comput. Appl. (ICIRCA)*, Jul. 2018, pp. 1275–1279.

[84] S. Wassan, X. Chen, T. Shen, M. Waqar, and N. Jhanjhi, "Amazon product sentiment analysis using machine learning techniques," *Revista Argentina de Clínica Psicológica*, vol. 30, no. 1, p. 695, 2021.

[85] S. Mystakidis, A. Christopoulos, and N. Pellas, "A systematic mapping review of augmented reality applications to support STEM learning in higher education," *Educ. Inf. Technol.*, vol. 27, no. 2, pp. 1883–1927, Aug. 2021, doi: 10.1007/s10639-021-10682-1.

[86] C. A. Palacios, J. A. Reyes-Suárez, L. A. Bearzotti, V. Leiva, and C. Marchant, "Knowledge discovery for higher education student retention based on data mining: Machine learning algorithms and case study in Chile," *Entropy*, vol. 23, no. 4, p. 485, Apr. 2021, doi: 10.3390/e23040485.

[87] P. D. Gil, S. da Cruz Martins, S. Moro, and J. M. Costa, "A data-driven approach to predict first-year students' academic success in higher education institutions," *Educ. Inf. Technol.*, vol. 26, no. 2, pp. 2165–2190, Oct. 2020, doi: 10.1007/s10639-020-10346-6.

[88] J. G. Perez, P. Bulacan, and E. S. Perez, "Predicting student program completion using Naïve Bayes classification algorithm," *Int. J. Mod. Educ. Comput. Sci.*, vol. 13, no. 3, pp. 57–67, Jun. 2021, doi: 10.5815/ijmecs.2021.03.05.

[89] Z. Kastrati, A. S. Imran, and A. Kurti, "Weakly supervised framework for aspect-based sentiment analysis on students' reviews of MOOCs," *IEEE Access*, vol. 8, pp. 106799–106810, 2020.

[90] S. Gottipati, V. Shankararaman, and J. R. Lin, "Text analytics approach to extract course improvement suggestions from students' feedback," *Res. Pract. Technol. Enhanced Learn.*, vol. 13, no. 1, pp. 1–19, Jun. 2018, doi: 10.1186/s41039-018-0073-0.

[91] K. S. Krishnaveni, R. R. Pai, and V. Iyer, "Faculty rating system based on student feedbacks using sentimental analysis," in *Proc. Int. Conf. Adv. Comput., Commun. Informat. (ICACCI)*, Sep. 2017, pp. 1648–1653.

[92] W. S. El-Kassas, C. R. Salama, A. A. Rafea, and H. K. Mohamed, "Automatic text summarization: A comprehensive survey," *Expert Syst. Appl.*, vol. 165, Mar. 2021, Art. no. 113679, doi: 10.1016/j.eswa.2020.113679.

[93] M. Allahyari, S. Pouriyeh, M. Assefi, S. Safaei, E. D. Trippe, J. B. Gutierrez, and K. Kochut, "Text summarization techniques: A brief survey," 2017, *arXiv:1707.02268*.

[94] L. Abualigah, M. Q. Bashabsheh, H. Alabool, and M. Shehab, "Text summarization: A brief review," in *Recent Advances in NLP: The Case of Arabic Language*. Cham, Switzerland: Springer, Nov. 2019, pp. 1–15, doi: 10.1007/978-3-030-34614-0_1.

[95] B. Mutlu, E. A. Sezer, and M. A. Akcayol, "Candidate sentence selection for extractive text summarization," *Inf. Process. Manage.*, vol. 57, no. 6, Nov. 2020, Art. no. 102359, doi: 10.1016/j.ipm.2020.102359.

[96] A. Sefid, J. Wu, P. Mitra, and L. Giles, "Extractive research slide generation using windowed labeling ranking," 2021, *arXiv:2106.03246*.

[97] M. P. Dhaliwal, R. Kumar, M. Rungta, H. Tiwari, and V. Vala, "On-device extractive text summarization," in *Proc. IEEE 15th Int. Conf. Semantic Comput. (ICSC)*, Jan. 2021, pp. 347–354.

[98] Y. Wu and B. Hu, "Learning to extract coherent summary via deep reinforcement learning," in *Proc. 32nd AAAI Conf. Artif. Intell.*, 2018, pp. 1–8.

[99] J. N. Madhuri and R. G. Kumar, "Extractive text summarization using sentence ranking," in *Proc. Int. Conf. Data Sci. Commun. (IconDSC)*, Mar. 2019, pp. 1–3, doi: 10.1109/IconDSC.2019.8817040.

[100] X. Fan, W. Luo, M. Menekse, D. Litman, and J. Wang, "CourseMIRROR: Enhancing large classroom instructor-student interactions via mobile interfaces and natural language processing," in *Proc. 33rd Annu. ACM Conf. Extended Abstr. Hum. Factors Comput. Syst.*, Apr. 2015, pp. 1473–1478, doi: 10.1145/2702613.2732853.

[101] G. Erkan and D. R. Radev, "LexRank: Graph-based lexical centrality as salience in text summarization," *J. Artif. Intell. Res.*, vol. 22, pp. 457–479, Dec. 2004.

[102] S. Gottipati, V. Shankararaman, and R. Ramesh, "TopicSummary: A tool for analyzing class discussion forums using topic based summarizations," in *Proc. IEEE Frontiers Educ. Conf. (FIE)*, Oct. 2019, pp. 1–9.

[103] R. Mihalcea and P. Tarau, "TextRank: Bringing order into text," in *Proc. Conf. Empirical Methods Natural Lang. Process.*, 2004, pp. 404–411.

[104] O.-M. Foong, S.-P. Yong, and F.-A. Jaid, "Text summarization using latent semantic analysis model in mobile Android platform," in *Proc. 9th Asia Modeling Symp. (AMS)*, Sep. 2015, pp. 35–39.

[105] L. Cagliero, L. Farinetti, and E. Baralis, "Recommending personalized summaries of teaching materials," *IEEE Access*, vol. 7, pp. 22729–22739, 2019.

[106] E. Baralis, L. Cagliero, A. Fiori, and P. Garza, "MWI-Sum: A multilingual summarizer based on frequent weighted itemsets," *ACM Trans. Inf. Syst.*, vol. 34, no. 1, pp. 1–35, Oct. 2015, doi: 10.1145/2809786.

[107] D. Miller, "Leveraging BERT for extractive text summarization on lectures," 2019, *arXiv:1906.04165*.

[108] S. Song, H. Huang, and T. Ruan, "Abstractive text summarization using LSTM-CNN based deep learning," *Multimedia Tools Appl.*, vol. 78, no. 1, pp. 857–875, Jan. 2019.

[109] R. Royan, C. Jayakumaran, and T. Stephan, "Text summarization for automatic grading of descriptive assignments: A hybrid approach," in *Enterprise Digital Transformation: Technology, Tools, and Use Cases*. Boca Raton, FL, USA: CRC Press, 2022, pp. 251–273.

[110] D. Boley, M. Gini, R. Gross, E.-H. S. Han, K. Hastings, G. Karypis, V. Kumar, B. Mobasher, and J. Moore, "Partitioning-based clustering for web document categorization," *Decis. Support Syst.*, vol. 27, no. 3, pp. 329–341, Dec. 1999, doi: 10.1016/s0167-9236(99)00055-x.

[111] I. Sindhu, S. M. Daudpota, K. Badar, M. Bakhtyar, J. Baber, and M. Nurunnabi, "Aspect-based opinion mining on student's feedback for faculty teaching performance evaluation," *IEEE Access*, vol. 7, pp. 108729–108741, 2019, doi: 10.1109/ACCESS.2019.2928872.

[112] S. Minaee, N. Kalchbrenner, E. Cambria, N. Nikzad, M. Chenaghlu, and J. Gao, "Deep learning-based text classification," *ACM Comput. Surv.*, vol. 54, no. 3, pp. 1–40, Apr. 2022, doi: 10.1145/3439726.

[113] P. Lavanya and E. Sasikala, "Deep learning techniques on text classification using natural language processing (NLP) in social healthcare network: A comprehensive survey," in *Proc. 3rd Int. Conf. Signal Process. Commun. (ICPSC)*, May 2021, pp. 603–609.

[114] B. Jang, M. Kim, G. Harerimana, S.-U. Kang, and J. W. Kim, "Bi-LSTM model to increase accuracy in text classification: Combining Word2Vec CNN and attention mechanism," *Appl. Sci.*, vol. 10, no. 17, p. 5841, Aug. 2020, doi: 10.3390/app10175841.

[115] M. Cornolti, P. Ferragina, and M. Ciaramita, "A framework for benchmarking entity-annotation systems," in *Proc. 22nd Int. Conf. World Wide Web (WWW)*. New York, NY, USA: ACM Press, 2013, pp. 249–260, doi: 10.1145/2488388.2488411.

[116] D. Dessì, F. Osborne, D. R. Recupero, D. Buscaldi, and E. Motta, "Generating knowledge graphs by employing natural language processing and machine learning techniques within the scholarly domain," *Future Gener. Comput. Syst.*, vol. 116, pp. 253–264, Mar. 2021, doi: 10.1016/j.future.2020.10.026.

[117] Y. Luan, L. He, M. Ostendorf, and H. Hajishirzi, "Multi-task identification of entities, relations, and coreference for scientific knowledge graph construction," 2018, *arXiv:1808.09602*.

[118] D. Shi, T. Wang, H. Xing, and H. Xu, "A learning path recommendation model based on a multidimensional knowledge graph framework for E-learning," *Knowl.-Based Syst.*, vol. 195, May 2020, Art. no. 105618, doi: 10.1016/j.knosys.2020.105618.

[119] J. Dörpinghaus and A. Stefan, "Knowledge extraction and applications utilizing context data in knowledge graphs," in *Proc. Federated Conf. Comput. Sci. Inf. Syst. (FedCSIS)*, Sep. 2019, pp. 265–272, doi: 10.15439/2019f3.

[120] F. S. Dolianiti, D. Iakovakis, S. B. Dias, S. J. Hadjileontiadou, J. A. Diniz, G. Natsiou, M. Tsitouridou, P. D. Bamidis, and L. J. Hadjileontiadis, "Sentiment analysis on educational datasets: A comparative evaluation of commercial tools," *Educ. J. Univ. Patras UNESCO Chair*, vol. 6, pp. 262–273, Mar. 2019. [Online]. Available: https://pasithee.library.upatras.gr/ejupUNESCOchair/article/view/2987

[121] Z. M. Ibrahim, M. Bader-El-Den, and M. Cocea, "A data mining framework for analyzing students' feedback of assessment," in *Proc. 13th Eur. Conf. Technol. Enhanced Learn. Doctoral Consortium*, 2018, p. 13.

[122] E. Sutoyo, A. Almaarif, and I. T. R. Yanto, "Sentiment analysis of student evaluations of teaching using deep learning approach," in *Proc. Int. Conf. Emerg. Appl. Technol. Ind. 4.0*. Cham, Switzerland: Springer, 2020, pp. 272–281.

[123] I. A. Kandhro, S. Wasi, K. Kumar, M. Rind, and M. Ameen, "Sentiment analysis of students' comment using long-short term model," *Indian J. Sci. Technol.*, vol. 12, no. 8, pp. 1–16, 2019.

[124] H. T. Hien, P.-N. Cuong, L. N. H. Nam, H. L. T. K. Nhung, and L. D. Thang, "Intelligent assistants in higher-education environments: The FIT-EBot, a chatbot for administrative and learning support," in *Proc. 9th Int. Symp. Inf. Commun. Technol. (SoICT)*, 2018, pp. 69–76.

[125] A. Joshi, A. Kunchukuttan, P. Bhattacharyya, and M. J. Carman, "SarcasmBot: An open-source sarcasm-generation module for chatbots," in *Proc. WISDOM Workshop KDD*, 2015, pp. 1–6.

[126] Y. Wu, W. Wu, C. Xing, M. Zhou, and Z. Li, "Sequential matching network: A new architecture for multi-turn response selection in retrieval-based chatbots," 2016, *arXiv:1612.01627*.

[127] M. Patidar, P. Agarwal, L. Vig, and G. Shroff, "Correcting linguistic training bias in an FAQ-bot using LSTM-VAE," in *Proc. DMNLP Workshop ECML-PKDD*, 2017, pp. 1–16.

[128] J. Lee, S. Lee, and S. Lee, "The influence of AI convergence education on students' perception of AI," *J. Korean Assoc. Inf. Educ.*, vol. 25, no. 3, pp. 483–490, Jun. 2021, doi: 10.14352/jkaie.2021.25.3.483.

[129] O. Grljević, Z. Bošnjak, and A. Kovačević, "Opinion mining in higher education: A corpus-based approach," *Enterprise Inf. Syst.*, vol. 16, no. 5, pp. 1–26, May 2022.

[130] N. N. Y. Vo, Q. T. Vu, N. H. Vu, T. A. Vu, B. D. Mach, and G. Xu, "Domain-specific NLP system to support learning path and curriculum design at tech universities," *Comput. Educ., Artif. Intell.*, vol. 3, Jan. 2022, Art. no. 100042, doi: 10.1016/j.caeai.2021.100042.

[131] L. Li, J. Johnson, W. Aarhus, and D. Shah, "Key factors in MOOC pedagogy based on NLP sentiment analysis of learner reviews: What makes a hit," *Comput. Educ.*, vol. 176, Jan. 2022, Art. no. 104354, doi: 10.1016/j.compedu.2021.104354.

[132] J. Tabassum, M. Maddela, W. Xu, and A. Ritter, "Code and named entity recognition in StackOverflow," 2020, *arXiv:2005.01634*.

[133] G. Pashev, S. Gaftandzhieva, and Y. Hopterieva, "Domain specific automated essay scoring using cloud based NLP API," *Int. J. Comput. Sci. Mobile Comput.*, vol. 10, no. 10, pp. 33–39, Oct. 2021, doi: 10.47760/ijcsmc.2021.v10i10.006.

[134] M. Abulaish and A. Kamal, "Self-deprecating sarcasm detection: An amalgamation of rule-based and machine learning approach," in *Proc. IEEE/WIC/ACM Int. Conf. Web Intell. (WI)*, Dec. 2018, pp. 574–579, doi: 10.1109/wi.2018.00-35.

[135] A. Joshi, P. Bhattacharyya, and M. J. Carman, "Automatic sarcasm detection: A survey," *ACM Comput. Surv.*, vol. 50, no. 5, pp. 1–22, 2017.

[136] K. Sundararajan and A. Palanisamy, "Multi-rule based ensemble feature selection model for sarcasm type detection in Twitter," *Comput. Intell. Neurosci.*, vol. 2020, pp. 1–17, Jan. 2020, doi: 10.1155/2020/2860479.

[137] K. Gaanoun and I. Benelallam, "Sarcasm and sentiment detection in Arabic language a hybrid approach combining embeddings and rule-based features," in *Proc. 6th Arabic Natural Lang. Process. Workshop*. Kyiv, Ukraine: Assoc. Comput. Linguistics, Apr. 2021, pp. 351–356. [Online]. Available: https://aclanthology.org/2021.wanlp-1.45

[138] R. Gupta, J. Kumar, and H. Agrawal, "A statistical approach for sarcasm detection using Twitter data," in *Proc. 4th Int. Conf. Intell. Comput. Control Syst. (ICICCS)*, May 2020, pp. 633–638.

[139] A. Khatri, "Sarcasm detection in tweets with BERT and GloVe embeddings," in *Proc. 2nd Workshop Figurative Lang. Process.* Stroudsburg, PA, USA: Assoc. Comput. Linguistics, 2020, pp. 56–60. [Online]. Available: https://aclanthology.org/2020.figlang-1.7

[140] K. Parmar, N. Limbasiya, and M. Dhamecha, "Feature based composite approach for sarcasm detection using MapReduce," in *Proc. 2nd Int. Conf. Comput. Methodologies Commun. (ICCMC)*, Feb. 2018, pp. 587–591.

[141] A. Garg and N. Duhan, "Sarcasm detection on Twitter data using support vector machine," *ICTACT J. Soft Comput.*, vol. 10, no. 4, pp. 2165–2170, 2020.

[142] A. Kumar, V. T. Narapareddy, V. A. Srikanth, A. Malapati, and L. B. M. Neti, "Sarcasm detection using multi-head attention based bidirectional LSTM," *IEEE Access*, vol. 8, pp. 6388–6397, 2020.

[143] D. Das and A. J. Clark, "Sarcasm detection on Flickr using a CNN," in *Proc. Int. Conf. Comput. Big Data (ICCBD)*. New York, NY, USA: Assoc. Comput. Machinery, 2018, pp. 56–61, doi: 10.1145/3277104.3277118.

[144] S. Jusoh, "A study on NLP applications and ambiguity problems," *J. Theor. Appl. Inf. Technol.*, vol. 96, no. 6, pp. 1–14, 2018.

[145] B. P. Yap, A. Koh, and E. S. Chng, "Adapting BERT for word sense disambiguation with gloss selection objective and example sentences," 2020, *arXiv:2009.11795*.

[146] L. Huang, C. Sun, X. Qiu, and X. Huang, "GlossBERT: BERT for word sense disambiguation with gloss knowledge," in *Proc. Conf. Empirical Methods Natural Lang. Process. 9th Int. Joint Conf. Natural Lang. Process. (EMNLP-IJCNLP)*. Hong Kong: Assoc. Comput. Linguistics, 2019, pp. 3509–3514. [Online]. Available: https://aclanthology.org/D19-1355

[147] A. S. Imran, S. M. Daudpota, Z. Kastrati, and R. Batra, "Cross-cultural polarity and emotion detection using sentiment analysis and deep learning on COVID-19 related tweets," *IEEE Access*, vol. 8, pp. 181074–181090, 2020.

[148] S. Cappallo, S. Svetlichnaya, P. Garrigues, T. Mensink, and C. G. M. Snoek, "New modality: Emoji challenges in prediction, anticipation, and retrieval," *IEEE Trans. Multimedia*, vol. 21, no. 2, pp. 402–415, Feb. 2019.

[149] G. S. Chauhan, P. Agrawal, and Y. K. Meena, "Aspect-based sentiment analysis of students' feedback to improve teaching–process," in *Information and Communication Technology for Intelligent Systems*. Singapore: Springer, Dec. 2018, pp. 259–266, doi: 10.1007/978-981-13-1747-7_25.

[150] A. Nazir, Y. Rao, L. Wu, and L. Sun, "Issues and challenges of aspect-based sentiment analysis: A comprehensive survey," *IEEE Trans. Affect. Comput.*, early access, Jan. 30, 2020, doi: 10.1109/TAFFC.2020.2970399.

[151] F. Thabtah, S. Hammoud, F. Kamalov, and A. Gonsalves, "Data imbalance in classification: Experimental evaluation," *Inf. Sci.*, vol. 513, pp. 429–441, Mar. 2020, doi: 10.1016/j.ins.2019.11.004.

[152] L. Guo, Y. Lei, S. Xing, T. Yan, and N. Li, "Deep convolutional transfer learning network: A new method for intelligent fault diagnosis of machines with unlabeled data," *IEEE Trans. Ind. Electron.*, vol. 66, no. 9, pp. 7316–7325, Sep. 2019.

[153] S. Ruder, M. Peters, S. Swayamdipta, and T. Wolf, "Transfer learning in natural language processing," in *Proc. Conf. North Amer. Chapter Assoc. Comput. Linguistics, Tuts.*, Minneapolis, MN, USA: Assoc. Comput. Linguistics, Jun. 2019, pp. 15–18. [Online]. Available: https://aclanthology.org/N19-5004

[154] S. Shaikh, S. M. Daudpota, A. S. Imran, and Z. Kastrati, "Towards improved classification accuracy on highly imbalanced text dataset using deep neural language models," *Appl. Sci.*, vol. 11, no. 2, p. 869, Jan. 2021, doi: 10.3390/app11020869.

[155] C. Shorten, T. M. Khoshgoftaar, and B. Furht, "Text data augmentation for deep learning," *J. Big Data*, vol. 8, no. 1, pp. 1–34, Jul. 2021, doi: 10.1186/s40537-021-00492-0.

[156] *Gartner Identifies Four Trends Driving Near-Term Artificial Intelligence Innovation*. Accessed: Mar. 2022. [Online]. Available: https://www.gartner.com/en/newsroom/press-releases/ 2021-09-07-gartner-identifies-four-trends-driving-near-term-artificial-intelligence-innovation

[157] N. Almazmomi, A. Ilmudeen, and A. A. Qaffas, "The impact of business analytics capability on data-driven culture and exploration: Achieving a competitive advantage," *Benchmarking, Int. J.*, vol. 29, no. 4, pp. 1264–1283, Aug. 2021, doi: 10.1108/bij-01-2021-0021.

[158] Y. Kim, T. Soyata, and R. F. Behnagh, "Towards emotionally aware AI smart classroom: Current issues and directions for engineering and education," *IEEE Access*, vol. 6, pp. 5308–5331, 2018.

[159] D. Gerritsen, J. Zimmerman, and A. Ogan, "Towards a framework for smart classrooms that teach instructors to teach," in *Proc. Int. Conf. Learn. Sci.*, vol. 3, 2018, pp. 1–4.

[160] G. Steinbauer, M. Kandlhofer, T. Chklovski, F. Heintz, and S. Koenig, "A differentiated discussion about AI education K-12," *KI-Künstliche Intelligenz*, vol. 35, no. 2, pp. 131–137, May 2021, doi: 10.1007/s13218-021-00724-8.

[161] M. J. Timms, "Letting artificial intelligence in education out of the box: Educational cobots and smart classrooms," *Int. J. Artif. Intell. Educ.*, vol. 26, no. 2, pp. 701–712, Jan. 2016, doi: 10.1007/s40593-016-0095-y.

[162] L. Chen and D. Gerritsen, "Building interpretable descriptors for student posture analysis in a physical classroom," in *Proc. 22nd Int. Conf. Artif. Intell. Educ. (AIED)*, 2021, pp. 1–5.

[163] W. Holmes, M. Bialik, and C. Fadel, *Artificial Intelligence in Education*. Boston, MA, USA: Center Curriculum Redesign, 2019, pp. 1–35.

[164] C. V. Felix, "The role of the teacher and AI in education," in *Innovations in Higher Education Teaching and Learning*. Bingley, U.K.: Emerald Publishing Ltd., Nov. 2020, pp. 33–48, doi: 10.1108/s2055-364120200000033003.

[165] N. Selwyn, *Should Robots Replace Teachers?: AI and the Future of Education*. Hoboken, NJ, USA: Wiley, 2019.

[166] R. Egger, "Software and tools," in *Applied Data Science in Tourism*. Cham, Switzerland: Springer, 2022, pp. 547–588, doi: 10.1007/978-3-030-88389-8_26.

[167] J. J. Thomas, V. Suresh, M. Anas, S. Sajeev, and K. S. Sunil, "Programming with natural languages: A survey," in *Computer Networks and Inventive Communication Technologies*. Singapore: Springer, Sep. 2021, pp. 767–779, doi: 10.1007/978-981-16-3728-5_57.

[168] D. Sarkar, *Text Analytics With Python: A Practitioner's Guide to Natural Language Processing*. Bangalore, India: Springer, 2019.

[169] R. M. Reese and A. Bhatia, *Natural Language Processing With Java: Techniques for Building Machine Learning and Neural Network Models for NLP*. Birmingham, U.K.: Packt Publishing Ltd, 2018.

[170] M. L. Jockers and R. Thalken, *Text Analysis With R*. Cham, Switzerland: Springer, 2020.

[171] T. Joseph, S. A. Kalaiselvan, S. U. Aswathy, R. Radhakrishnan, and A. R. Shamna, "A multimodal biometric authentication scheme based on feature fusion for improving security in cloud environment," *J. Ambient Intell. Humanized Comput.*, vol. 12, no. 6, pp. 6141–6149, Jun. 2020, doi: 10.1007/s12652-020-02184-8.

[172] Y. Wang, X. Liu, and S. Shi, "Deep neural solver for math word problems," in *Proc. Conf. Empirical Methods Natural Lang. Process.*, 2017, pp. 845–854.

[173] M. Fowler, B. Chen, S. Azad, M. West, and C. Zilles, "Autograding 'explain in plain English' questions using NLP," in *Proc. 52nd ACM Tech. Symp. Comput. Sci. Educ.*, Mar. 2021, pp. 1163–1169.

[174] M. Zhang, Z. Wang, R. Baraniuk, and A. Lan, "Math operation embeddings for open-ended solution analysis and feedback," 2021, *arXiv:2104.12047*.

THANVEER SHAIK received the master's degree in applied data science from the University of Southern Queensland, Australia, where he is currently pursuing the Ph.D. degree. His research interests include cognitive computing, biometrics, and NLP with expertise in artificial intelligence (AI), machine learning, and predictive analysis.
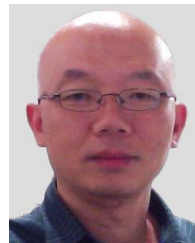
XIAOHUI TAO (Senior Member, IEEE) received the Ph.D. degree from the Queensland University of Technology, Brisbane, QLD, Australia. He is currently an Associate Professor (computing) at the School of Mathematics, Physics and Computing (SoMPC), University of Southern Queensland (USQ), Australia. His research interests include data analytics, machine learning, knowledge engineering, information retrieval, and health informatics. His research outcomes have been published on many top-tier journals, such as IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING (TKDE), *Information Processing and Management* (IPM), *Knowledge-Based Systems* (KBS), *Expert Systems with Applications* (ESWA), and *Physical Review Letters* (PRL), and conferences, such as IJCAI, ICDE, CIKM, PAKDD, and WISE. He is a Senior Member of ACM. He is also an Active Researcher in AI. He received the ARC DP Grant, from 2022 to 2024, and the Australian Endeavour Research Fellow, from 2015 to 2016. He was awarded with the Research Performance Award and the Discipline Research Performance Improvement Award from SoMPC, USQ, among many others. He has been active in professional services. He was an Editor or a Guest Editor of many journals, including *Information Fusion* (INFFUS) and *World Wide Web: Internet and Web Information Systems Journal* (WWWJ). He has served as the PC Chair for WI, in 2017 and 2018, WI-IAT, in 2021, and BESC, in 2018 and 2021.

YAN LI is currently a Professor in computer science with the School of Mathematics, Physics and Computing, University of Southern Queensland, Australia. Her research interests include artificial intelligence, big data analytics, signal and image processing, biomedical engineering, and computer networking technologies and security.

CHRISTOPHER DANN is an Inclusive Goal Orientated Leader whose purpose is to make a positive impact on the educational experiences of learners 'globally.' He is currently a Senior Lecturer in curriculum and pedagogy (technologies) with the School of Teacher Education, University of Southern Queensland. His current research interests include machine learning and artificial intelligence on the teaching and learning process from the perspective of teachers and students across educational context.

JACQUIE MCDONALD is currently an Honorary Associate Professor at the University of Southern Queensland (USQ), Australia, and a Higher Education Community of Practice (CoP) Consultant. She previously worked for over 26 years as a Learning and Teaching Designer at USQ designing online and distance learning courses and programs. Since 2006, she has facilitated, researched, and coached the implementation of inter/national higher education CoPs and led a number of institutional and national fellowships and grants. Her recent publications co-edited books *Communities of Practice: Facilitating Social Learning in Higher Education* (Springer, 2017), *Implementing Communities of Practice in Higher Education: Dreamers and Schemers* (Springer, 2021), and *Sustaining Communities of Practice with Early Career Teachers: Supporting Early Career Teachers* in Australian and International Primary and Secondary Schools. Her publications and research interests on social learning spaces, including communities of practice, designing for online learning and teaching, and educational professional development. She is a member of the Australian Learning and Teaching Fellows Alumni. She has won university and national awards, citations, and best paper awards at international conferences. She is on the Editorial Board of the *Journal on Excellence in College Teaching*.

PETREA REDMOND received the Diploma degree in teaching, bachelor's degree in business, master's degree in education, and the Ph.D. degree in educational technology from the University of Southern Queensland, in 2011.

She worked in high schools and universities in Australia and Canada. She is currently the Associate Head of the School (Education), Research. She has over 80 publications, including two edited books. Her current teaching and research interests include AI and education, telepresence robots, online engagement, and digital pedagogies.

Prof. Redmond is currently a member of the Executive of the Australasian Society for Computers in Learning in Tertiary Education (ASCILITE) and the Society for Information Technology and Teacher Education (SITE). She has received national and international awards for publications and research programs, including best paper awards at international conferences: short listed for U.K. Association for Learning Technology (ALT) of the Year Team Award, the USQ Excellence Award for Advancing Student Success, the USQ Excellence Award for Online Learning Innovation, the Queensland Government Our Women, Our State Award, and the Australasian Society for Computers in Learning in Tertiary Education (ASCILITE) Fellow Award. She has been an Editor of the *Australasian Journal of Educational Technology* (AJET).

LINDA GALLIGAN received the Ph.D. degree in mathematics education from QUT. She is currently an Experienced Associate Professor with a demonstrated history of working in the higher education industry. She has skilled, within the context of mathematics, lecturing, educational technology, educational research, and applied linguistics.

• • •