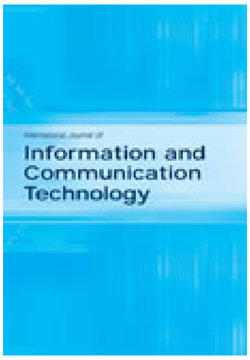


International Journal of Information and Communication Technology

ISSN online: 1741-8070 - ISSN print: 1466-6642
https://www.inderscience.com/ijict

Big data-driven deep mining of online teaching assessment data under affective factor conditions

Ruiting Bai

Article History:

Received:	16 October 2024
Last revised:	01 November 2024
Accepted:	01 November 2024
Published online:	18 December 2024

Big data-driven deep mining of online teaching assessment data under affective factor conditions

Ruiting Bai

Puyang Medical College,
Puyang 457000, China
Email: bairuiting@163.com

Abstract: As online education platforms continue to expand, the effective assessment of teaching quality has become increasingly important. This paper examines the role of emotional factors in evaluating teaching quality within online education and proposes a deep mining approach grounded in big data. We introduce the EduSent-Dig model, which integrates Bi-LSTM and Word2Vec techniques to extract and analyse sentiment-related factors from online assessment data. The model exhibited robust performance on diverse datasets, as indicated by its correct rate, sensitivity, and the F1 metric. Although there are limitations, including dataset bias and model complexity, future research will aim to enhance the model's capabilities in multilingual processing, optimise real-time data analysis, and simplify its structure to improve its overall applicability and effectiveness in enhancing the online education experience.

Keywords: affective factors; big data; online teaching assessment; deep mining.

Reference to this paper should be made as follows: Bai, R. (2024) 'Big data-driven deep mining of online teaching assessment data under affective factor conditions', *Int. J. Information and Communication Technology*, Vol. 25, No. 11, pp.35–51.

Biographical notes: Ruiting Bai received her Master's degree in the Southwest University in June 2012. She is currently an Associate Professor in the Puyang Medical College. Her research interests include machine learning, ideological and political education.

1 Introduction

In 21st century education, technological advancements have sparked innovations. The rise of online platforms and AI technologies has significantly reshaped approaches to education and learning, which have significantly bolstered online education (Hofer et al., 2021). Recently, platforms like Coursera, edX, and MOOCs have allowed students and educators globally to enjoy the convenience and flexibility of online learning. However, the widespread adoption of this teaching approach during the COVID-19 pandemic has introduced new challenges, particularly in assessing educational quality (Chakraborty et al., 2021). The online learning environment differs markedly from traditional classrooms, altering the dynamics of student-teacher interaction and the access and use of learning resources, thereby imposing new demands on teaching evaluation systems.

Analysing emotive elements is one of the several shortcomings of the present online teacher evaluation systems. While neglecting the depth and complexity of student emotional feedback, these instruments usually concentrate on quantitative indications, such test results and student happiness. Actually, students' learning motivation, involvement, and general learning effect are much influenced by their emotional experience. Consequently, it is crucial to create online assessment instruments capable of accurately catching the subtleties of mood. By means of an analysis of the emotive elements, this paper aims to suggest a novel online teaching evaluation technique in order to increase the accuracy and efficacy of the assessment.

As a key indicator of teaching quality, analysing online teaching evaluation data is crucial for optimising instructional methods, enhancing the learning experience, and shaping education policies. Currently, most analyses focus on quantitative metrics like student satisfaction and test scores, which often overlook the depth and complexity of students' emotional feedback. In reality, students' emotional experiences significantly influence their motivation, engagement, and overall learning effectiveness (Lu et al., 2022). Therefore, effectively mining and analysing affective factors in online teaching evaluation data is essential for improving online teaching quality.

With the emergence of the big data era, the education sector has amassed substantial teaching and evaluation data. This information encompasses not only direct student feedback but also rich emotional insights. Extracting valuable information from big data to inform teaching practices poses a pressing challenge in current educational research and practice (Sivarajah et al., 2017). Researchers have made significant strides in this domain. For instance, Baker employed machine learning techniques to analyse students' interactions on an online learning platform, revealing that affective tendencies are closely linked to academic success, with positive emotions boosting motivation and engagement. Additionally, Ganimian and Murnane demonstrated that students' emotional experiences influence academic performance and are directly associated with course appeal and completion rates. These studies provide a theoretical basis for sentiment analysis in online education, highlighting the potential of emotional data to enhance educational quality.

Simultaneously, research in Chinese online education is actively investigating the use of sentiment analysis. Academician Li and Gao (2018) applied deep learning technology to analyse review data from online education platforms, uncovering how students' emotional inclinations towards course content influence their learning experiences. These studies indicate that affective factors not only help educators better understand students' needs but also serve as a crucial foundation for course enhancements and personalised recommendations, ultimately optimising the effectiveness of online education.

While existing research has offered valuable insights into affective factors in online education, several limitations persist. First, most studies have concentrated on analysing affective tendencies, with relatively few addressing the intensity and complexity of emotions, such as specific states like anger, joy, or surprise (Reisenzein, 2000). Second, many studies rely on static datasets and do not incorporate dynamic analyses of real-time data streams. Furthermore, affective analyses are often conducted in isolation from teaching and learning outcomes, lacking a thorough examination of how these emotional factors influence the educational process.

The objective of this study is to investigate and apply a data-intensive methodology to thoroughly analyse educational and evaluative datasets from virtual learning environments, with a particular emphasis on extracting and analysing student sentiment factors. By employing advanced text analytics and sentiment analysis techniques, this study aspires to equip educators with an innovative tool for assessing and optimising the online teaching and learning experience. Additionally, it will delve into the intricate relationship between affective factors and teaching effectiveness within online learning environments.

The main innovations and contributions of this work include:

1. Comprehensive analysis of affective factors: this study prioritises students' emotional expressions as the core content of teaching evaluation data, rather than treating them as secondary elements. This approach provides new perspectives and greater depth to online teaching evaluations.
2. Integration of big data technologies: by leveraging big data analytics, this research can process and analyse extensive teaching evaluation datasets, which is a significant advancement over previous sentiment analysis studies. This integration greatly enhances both the breadth and depth of the analysis conducted.
3. Multi-dimensional teaching evaluation data: the proposed model incorporates multiple dimensions and constructs a new framework for analysing online teaching effectiveness from various perspectives, thereby offering a more comprehensive basis for informed educational decision-making.

2 Relevant technologies

2.1 Sentiment analysis techniques

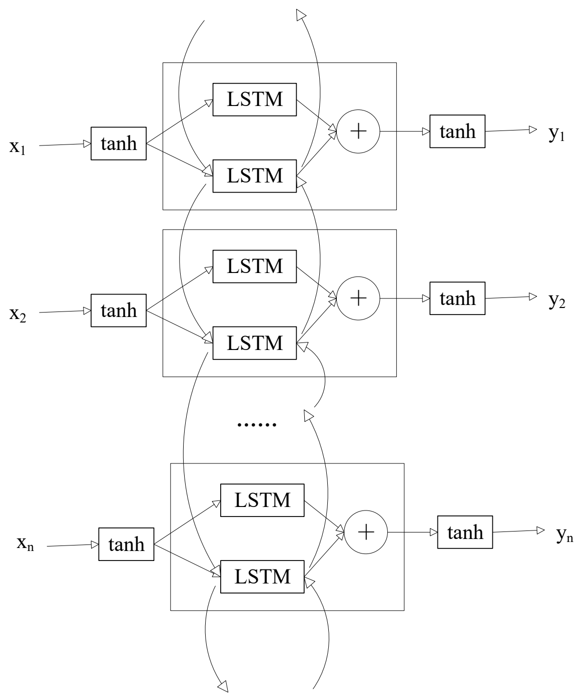
Sentiment analysis, integral to the field of computational linguistics, is vital for discerning consumer emotions and assessing the market's view on various offerings (Gooljar, et al., 2024). In the "Internet Plus" era, platforms such as social media, forums, and customer feedback channels have amassed a vast array of user comments, transforming this data into a valuable resource for gaining insights into public sentiment and enhancing decision-making. As illustrated in Table 1, sentiment analysis transcends academic inquiry, significantly impacting practical domains like business analytics, consumer sentiment evaluation, and societal sentiment tracking (Philander and Zhong, 2016).

Sentiment analysis is a complex process that involves understanding the subtleties and emotional nuances present in textual data. To accomplish this, researchers have developed a range of models, encompassing lexicon-driven strategies, conventional ML algorithms, and advanced deep learning models. Dictionary-based approaches identify emotional tendencies by referencing predefined emotion dictionaries. While these methods are straightforward and efficient, they often struggle with contextual understanding, which can compromise accuracy. Classic machine learning techniques, including the naive Bayes classifier and SVMs, extract text features through feature engineering before performing sentiment classification. Although these methods are effective for smaller datasets, their performance typically declines as data quantities and textual intricacies escalate.

Table 1 Common application scenarios for sentiment analysis

Implementation setting	*Descriptions*
Market feedback synthesis	Assess customer satisfaction regarding the product and subsequently formulate a targeted marketing plan.
Sentiment evaluation	To discern the collective emotional inclination towards trending events, to track the prevailing sentiment, to vigilantly monitor public sentiment promptly and efficiently, and to facilitate relevant authorities with actionable insights.
Movie review evaluation	Gather audience feedback on films or TV shows to enhance narrative structure and appeal to a broader viewership.
Analysis of event forecasts	Forecast pertinent details, such as film box office and election winners, through user comments on an event.
Online education	By analysing the emotional tendency of learners' comments in course reviews, discussion forums, personal home pages, etc., it is possible to understand the learning status of learners and then provide personalised learning solutions.

Figure 1 Network modelling with Bi-LSTM



Neural network-based deep learning approaches, especially, have emerged as a focal point in sentiment analysis research due to their robust nonlinear modelling capabilities and automatic feature extraction. While Word2Vec maps terms into vector space to offer the semantic representation, Bi-LSTM model gathers the context information of text by processing sequence data in two directions. Deep learning models such as CNNs for feature extraction, RNNs for sequence analysis, LSTMs for capturing long-term dependencies, and Bi-LSTMs for enhancing sequence prediction are widely recognised (Yadav et al., 2019). CNNs excel at capturing local features in textual data but often struggle to grasp long-distance dependencies. RNNs and LSTMs are tailored for processing sequential data, adept at detecting patterns across time sequences; however, LSTMs may face challenges such as gradient vanishing or gradient explosion when handling very long sequences. In contrast, Bi-LSTMs enhance sentiment analysis accuracy by effectively capturing contextual information through bidirectional information flow (Naseem et al., 2020).

In this study, we leverage the strengths of Bi-LSTM and Word2Vec's word vector representation to develop a new model called EduSent-BiLSTM. The Bi-LSTM architecture captures contextual information from text in both directions, while Word2Vec offers rich semantic insights. This combination enhances the model's capability to efficiently process intricate and ambiguous textual data. Additionally, the Bi-LSTM model addresses the gradient vanishing issue commonly encountered in deep neural network training by incorporating residual connections (Jang et al., 2020), which enhances both training efficiency and model performance, as illustrated in Figure 1.

The EduSent-BiLSTM model's framework consists of several integral elements: initial data preparation, attribute extraction, the central emotion detection unit, the categorisation mechanism, along with the training and assessment protocols.

1 Data pre-processing

Including text cleaning, word splitting and conversion to Word2Vec word vectors to prepare high-quality input data for sentiment analysis (Aoumeur et al., 2023), where V_w indicates the vector representation of term w, with w referring to the term itself in the text.

Word2Vec word vectors:

$$V_W = \text{Word2Vec}(w) \tag{1}$$

2 Feature extraction

Word2Vec-derived vector representations serve as model inputs to discern semantic connections among terms. By training on the corpus, Word2Vec maps semantically proximate words to nearby points in the vector space.

3 Sentiment analysis core module

A bidirectional LSTM (Bi-LSTM) is employed to analyse word vector sequences bidirectionally, thereby extracting the contextual nuances of the text (Airlangga, 2024). At time step t, $\overrightarrow{h_t}$ denotes the forward state and $\overleftarrow{h_t}$ the backward state; the forward input, that is $\overrightarrow{x_t}$, word vectors, for the backward input at t, and the forward hidden state and backward hidden state at $t-1$, respectively, we have.

- Forward LSTM:

$$\overrightarrow{h_t} = \text{LSTM}\left(\overrightarrow{x_t}, \overrightarrow{h_{t-1}}\right) \tag{2}$$

- Backward LSTM:

$$\overleftarrow{h_t} = \text{LSTM}\left(\overleftarrow{x_t}, \overleftarrow{h_{t-1}}\right) \tag{3}$$

Final hidden state:

$$h_t = \left[h_t^f ; h_t^b \right] \quad (4)$$

4 Classifier module

The Bi-LSTM yields are converted to probability distributions of sentiment categories by means of a dense layer followed by a softmax classifier. Here, z is derived from the dense layer's output, W corresponds to the weight matrix, h represents the Bi-LSTM layer's concatenated output of the merged forward and backward hidden states, and b stands for the bias term.

- Full connectivity layer:

$$z = W \cdot h + b \quad (5)$$

- Softmax layer:

$$p(y \mid X) = \frac{e^{z_y}}{\sum_{c=1}^{C} e^{z_c}} \quad (6)$$

In the softmax layer, the likelihood for label y based on X; C is the sentiment category count, which corresponds to the output of the fully-connected layer for the sentiment category z_c. The softmax function transforms the dense layer's output into a probability distribution that reflects the likelihood of various sentiment categories.

5 Model training and evaluation

The model's training is guided by minimising cross-entropy, and its effectiveness is quantified by metrics including accuracy, recall, and the F1 score. Here, L denotes the loss function, and N represents the sample count, y_{ic} is the true sentiment label (one-hot coding) of the i^{th} sample, and p_{ic} denotes the likelihood for sample i to be classified as category c. The cross-entropy loss function quantifies the divergence between the predicted probability distribution and the actual label distribution.

Cross-entropy loss function:

$$L = -\sum_{i=1}^{N} \sum_{c=1}^{C} y_{ic} \log\left(p_{ic}\right) \quad (7)$$

6 Model optimisation

L2 regularisation and dropout are incorporated to enhance the model's generalisation and mitigate overfitting (Demir-Kavuk et al., 2011). Where λ is the regularisation coefficient and W is the model parameter.

L2 regularisation:

$$L_{\text{regularisation}} = \lambda \sum_{w} \|W\|^2 \quad (8)$$

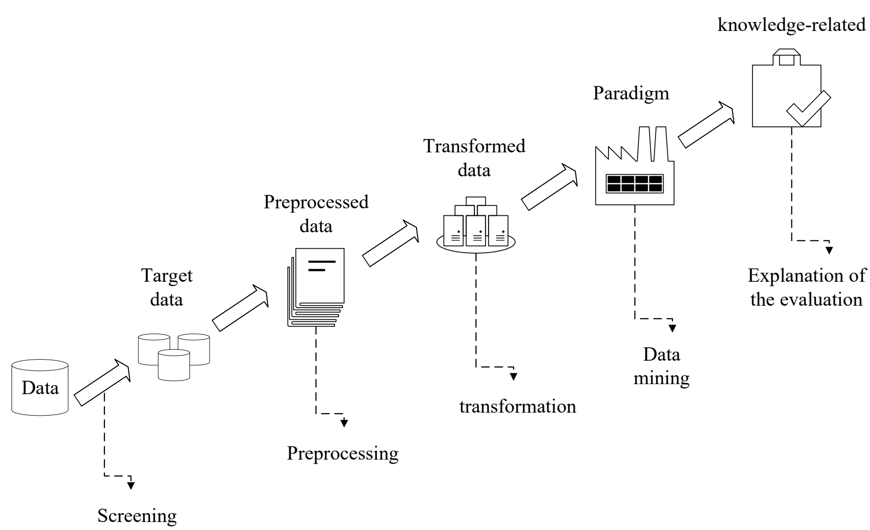
Versus a standard LSTM, the EduSent-Bi-LSTM excels at grasping text's bidirectional nuances, crucial for interpreting complex linguistic expressions like sarcasm and wordplay. Compared to CNN models, Bi-LSTM is more advantageous in dealing with long sequence data because it is able to take into account contextual information from further away. Furthermore, the EduSent-Bi-LSTM model bolsters its generalisation capabilities and guards against overfitting through the implementation of regularisation and dropout methods.

2.2 Data mining and analysis techniques

Data mining entails extracting valuable information and insights from large datasets, integrating fields such as statistics, machine learning, and database technology (Lan et al., 2018). Its core objective is to transform patterns, trends and relationships in a dataset into meaningful information by analysing them to support decision-making and prediction. Data mining employs various analytical techniques, collectively referred to as CRLA (classification, regression, clustering, and association rule analysis), to help researchers and decision-makers uncover crucial insights within the data. In online teaching evaluation, data mining can not only help analyse students' performance and feedback, but also optimise course design and enhance learning outcomes, leading to personalised education and continuous improvement (Hung et al., 2012). Such a comprehensive application helps to improve the quality of teaching and the overall student experience. This work conducted data mining and feature importance evaluation using random forest algorithm. By building several decision trees, random forest can increase classification accuracy and lower over-fitting risk.

As shown in Figure 2, the basic process of data mining includes three stages: data pre-processing, data mining, and result analysis.

Figure 2 General process of data mining



According to these steps, the selection of appropriate algorithms becomes the key to achieve effective data mining. In the scenario of online teaching evaluation, the selection of algorithms needs to take into account factors such as data type, analysis objectives, and model interpretability. Among many algorithms, the random forest algorithm stands out for its excellent performance. Next, we will detail the specific implementation of random forest. Random forest mitigates overfitting by integrating multiple decision trees, effectively handling high-dimensional data. Each tree is trained on a random subset of data, with predictions aggregated by voting. This method enhances classification precision and evaluates feature importance, aiding in the comprehension of pivotal factors in teaching assessment. Therefore, random forest will be the core algorithm of the sentiment-driven online teaching evaluation and data mining model in this project. The following is the working principle and steps of the random forest algorithm.

1 Data preparation.

Collect and clean the data to ensure that the data quality is suitable for model training.

2 Random selection of samples and features.

Multiple subsets are constructed by selecting samples from the training set through putative return sampling (Shi et al., 2019). Each tree is evaluated by randomly selecting k features from p features at split time.

The feature extraction formula is as follows. Here, $TF(t, d)$ is the count of word t in document d, $DF(t)$ is the number of documents that include word t, and N is the total document count.

$$TF - IDF(t, d) = TF(t, d) \times \log \frac{N}{DF(t)} \tag{9}$$

3 Constructing a decision tree.

For each subset, the best splitting feature is selected using information gain or Gini index. Here, $H(D)$ signifies the dataset's entropy, v represents a specific value of attribute A, and $Values(A)$ encompasses all potential values of A.

Gain in information:

$$IG(D, A) = H(D) - \sum_{v \in Values(A)} \frac{|D_v|}{|D|} H(D_v) \tag{10}$$

Entropy formula: here, c denotes a category within the dataset, *Classes* represents all possible categories, and $p(c)$ is the likelihood of category c appearing in dataset D.

$$H(D) = -\sum_{c \in Classes} p(c) \log_2 p(c) \tag{11}$$

4 Generate forest.

Repeat the above steps to construct N decision trees to form a random forest.

5 Prediction.

For new data x, each tree Ti is predicted and the final result is:

$$\hat{y} = \mathrm{mod}\, e\left(\hat{y}_1, \hat{y}_2, \ldots, \hat{y}_N\right) \tag{12}$$

With N representing the decision trees' count, the prediction selects the category that appears most frequently.

6 Feature importance evaluation.

The importance of features can be calculated by summing the gain of each feature, where gain(A, T_j) is the gain of feature A in the j^{th} tree T_j:

$$\text{Importance}(A) = \sum_{j=1}^{N} gain\left(A, T_j\right) \tag{13}$$

The random forest algorithm has become a key tool for sentiment classification due to its powerful performance and flexibility. By effectively handling high-dimensional data and evaluating feature importance, it is able to deeply mine sentiment information in student feedback, thus revealing key factors affecting teaching quality. These advantages make random forest particularly important in building sentiment-driven online teaching evaluation and data mining models. Next, we will explore the construction framework of the model and its applications in detail, showing how the accuracy and personalisation of teaching evaluation can be enhanced by this model.

3 Emotion-driven online teaching evaluation and data mining model EduSent-Dig

3.1 The EduSent-Dig model

The EduSent-Dig model proposed in this study is a deep learning model that integrates sentiment analysis and data mining techniques, aiming to deeply mine sentiment factors from online teaching assessment data and evaluate teaching effectiveness. Bi-LSTM specifically captures contextual information in two directions, forward and backward, therefore offering a complete awareness of text. Word2Vec maps words into vector spaces to offer semantic links between them. This integration helps the model to better grasp the expression of complicated emotions as well as to catch the emotional colours in the text. For instance, Word2Vec offers semantic support at the lexical level whereas Bi-LSTM gathers contextual information to better grasp linguistic events such sarcasm and puns. As shown in Figure 3, the architectural design of the model considers the whole process of data processing, from data pre-processing to feature extraction, to sentiment analysis and final decision output.

The following is the detailed architecture of the model:

1 Data pre-processing module

In the data pre-processing stage, the model first cleans the collected online teaching assessment data to remove invalid or incomplete data records (Amrieh et al., 2016). Subsequently, text segmentation processing is performed to convert continuous text data into discrete word sequences. In addition, the module involves steps such as deactivation word removal, stemming extraction and lexical labelling to improve the quality and efficiency of subsequent feature extraction.

2 Feature extraction module

This module employs Word2Vec to transform pre-processed text into fixed-size vectors, enabling the model to understand word relationships. Word2Vec maps words to vectors in a way that similar words in meaning also cluster closely in vector space, enriching the model's semantic insights.

Word2Vec training formulas include negative sampling formulas and hierarchical softmax formulas.

Negative sampling (NEG) is a technique used for optimisation in Word2Vec training, which improves training efficiency by reducing the number of negative examples to be computed (Qin et al., 2016). Where v is a model parameter, let v_o represent the vector of the target word o, and v_c be the vector of the context word c.

Negative sampling formula:

$$\max_{v} \in \theta\mathbb{E}_{(o,c)} \sim p_{neg}\left[\log\sigma\left(v_o^T v_c\right)\right] \tag{14}$$

Hierarchical softmax (also known as tree softmax) is another technique that improves the efficiency of Word2Vec training by reducing the amount of computation by organising the vocabulary list into a tree. In this context, K is the depth of the tree, v_k represents the vector at node k_k, and u denotes the input vector, often derived from transforming context word vectors.

Hierarchical softmax formula:

$$p\left(wo \middle| w_I\right) = \prod_{k=1}^{K}\left(\sigma\left(v_k^T u\right)\right) \tag{15}$$

3 Sentiment analysis core module

The core module of the sentiment analysis is constructed using Bi-LSTM, which can process word vector sequences from two directions to effectively capture the contextual information of the text. The LSTM processes information in both directions: forward through past contexts and backward through future contexts, thereby providing a comprehensive sentiment analysis of the text data.

4 Classifier module

Data processed by the Bi-LSTM is input into a dense layer, translating the high-level features into sentiment classifications. The softmax layer converts the dense layer's output into sentiment category probabilities for classification.

5 Model training and evaluation module

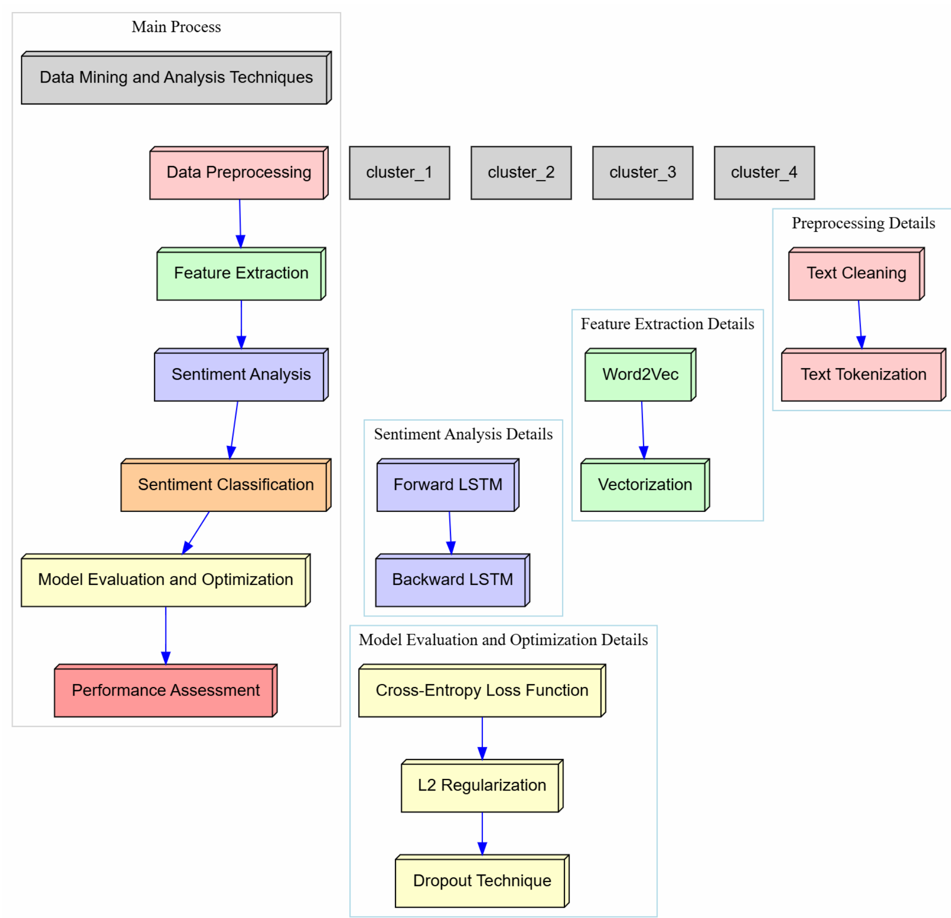
During training, the model minimises cross-entropy to refine network weights, with parameter updates performed via backpropagation. The evaluation phase assesses model performance using metrics including accuracy, precision, recall, and the F1 score.

6 Model optimisation module

To enhance the model's generalisability and guard against overfitting, L2 regularisation and dropout are incorporated into the model. L2 regularisation limits the complexity of the model by penalising excessively large weights, whereas dropout avoids overfitting by stochastically discarding some of the neurons in the network in order to reduce the dependency between parameters.

We pick Bi-LSTM as the fundamental emotion analysis module in EduSent-Dig model since, when processing the sequence data, it can efficiently capture the two-way context information of text. The capacity of the LSTM to record information both forward and backward makes it more accurate in recognising the emotional orientation of text than in one-way LSTM.

Figure 3 Architecture of the EduSent-Dig model (see online version for colours)



3.2 *Evaluation indicators*

Since accuracy, recall, and F1 score can entirely represent the model's prediction ability, these indices were chosen in this study as the key evaluation ones. The accuracy rate directly represents the prediction of the correct proportion by the model; the recall rate gauges the model's capacity to find all pertinent instances; and the F1 score aggregates the accuracy rate and recall rate to give a fair performance evaluation. By means of the thorough analysis of these indices, we may more precisely modify and maximise the model parameters to attain the optimum forecasting impact.

1 Accuracy (acc)

Accuracy directly gauges a model's predictive accuracy, indicating the ratio of correct predictions to the overall sample count (Boschetto and Bottini, 2014). Here, true positives (TP) are the instances where the positive outcome is correctly predicted. True negatives (TN) are the instances where the negative outcome is correctly predicted. False positives (FP) are the instances where the negative outcome is incorrectly identified as positive. False negatives (FN) are the instances where the positive outcome is incorrectly identified as negative.

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \tag{16}$$

2 Recall (rec)

Recall assesses the percentage of actual positives that are accurately detected by the model, indicating its effectiveness in finding all relevant instances (Everingham et al., 2015).

$$\text{Recall} = \frac{TP}{TP + FN} \tag{17}$$

3 F1 Score

The F1 score is a harmonic mean of precision and recall, offering a balance between them as a composite performance measure.

$$\text{F1 Score} = 2 \times \frac{Precision \times Recall}{Precision + Recall} \tag{18}$$

The aforementioned evaluation metrics facilitate a thorough analysis of the EduSent-Dig model's effectiveness. The metrics mentioned serve to evaluate the model's performance for sentiment analysis and also guide subsequent refinements. Accuracy offers a measure of the model's overall accuracy in predictions, while recall indicates how many actual positive instances are correctly identified. Additionally, the F1 value combines both accuracy and recall, acting as a key metric for assessing the model's overall effectiveness (Chicco and Jurman, 2020).

Utilising these evaluation metrics, we will select models, fine-tune hyperparameters, and compare models during training to ensure the final model performs optimally in real-world scenarios.

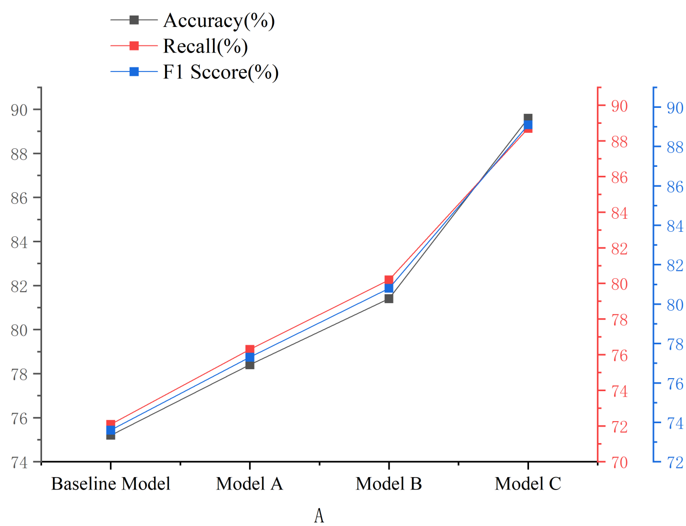
4 Experimental results and analyses

4.1 Ablation experiments

Assessing the efficacy of each component within the EduSent-Dig model, this study conducted ablation experiments (Hoffmann, et al., 2013). These experiments involved systematically removing specific components – such as the core sentiment analysis module and the feature extraction module – from the model one at a time to assess their impact on overall performance.

Through ablation studies, we developed four varied model configurations to evaluate the impact of each component on overall effectiveness. We specifically eliminated one by one the fundamental emotion analysis module and feature extraction module of the model and tracked model performance resulting from these adjustments. For instance, we discovered that the accuracy of the model dropped by 5% and the recall rate dropped by 6% when we deleted Bi-LSTM module, thereby suggesting that Bi-LSTM is essential in obtaining text context information. Likewise, the performance of the model is much lowered when the Word2Vec feature extraction module is eliminated, therefore attesting to the relevance of Word2Vec in offering lexical semantic information. These results direct us in further optimising the model and assist us to grasp the particular contribution of every element to the performance of the model. The baseline leverages traditional machine learning algorithms, such as random forest, without incorporating deep learning methods, serving as our benchmark for comparison. Model A examines the significance of feature extraction by using raw text data directly instead of word vectors generated by Word2Vec. Model B evaluates the importance of the core sentiment analysis module by employing Word2Vec for feature extraction, but replaces the Bi-LSTM classifier with logistic regression. Finally, model C represents our complete model, encompassing all stages of data pre-processing, feature extraction, sentiment analysis, classifier integration, and model training and evaluation.

Figure 4 Results of ablation experiments (see online version for colours)



The dataset used for the experiments is sourced from a university's student online course evaluations conducted between 2020 and 2023. We follow a set of pre-processing procedures in the cross-dataset experiment to guarantee data consistency. These processes comprise lexical tagging, text cleaning, word segmentation, stop word elimination, and stem extraction. To remove the effect of data source variances, we specifically pre-processed all datasets using the same technique. For instance, we applied the same stop word list and standardised word segmentation criteria for all datasets. We also standardised the data to help to minimise variations between several data sources. The performance stability of the model on several datasets depends on these pre-processing stages being absolutely important. This data, provided by the university's registrar's office, has been anonymised to ensure data privacy and compliance. The experimental results are illustrated in Figure 4.

The baseline model, which employs the random forest algorithm without utilising deep learning techniques, achieves an accuracy of 75.2%, a recall of 72.1%, and an F1 value of 73.6%, serving as a benchmark for performance comparison. In model A, when we forgo Word2Vec for feature extraction and directly use the raw text data, we observe a slight performance improvement, with accuracy rising to 78.4%, recall increasing to 76.3%, and the F1 value reaching 77.3%. This suggests that using raw text data can yield some performance enhancements, even in the absence of semantic features provided by Word2Vec.

Additionally, model B utilised Word2Vec for feature extraction but replaced the Bi-LSTM classifier with logistic regression. In this configuration, accuracy rose to 81.4%, recall increased to 80.2%, and the F1 value reached 80.8%. These results indicate that Word2Vec effectively extracts semantic features from text, significantly enhancing the accuracy of sentiment analysis.

In the end, model C, integrating all features, delivered the top results: accuracy at 89.6%, recall at 88.7%, and an F1 score of 89.1%. The findings underscore Bi-LSTM's ability to capture contextual information and demonstrate the model's effectiveness for sentiment analysis.

4.2 Integrated experiments

To substantiate the EduSent-Dig model's performance, this research involved applying it to actual teaching evaluation data from various virtual learning platforms.

In the comprehensive experiment, we selected three different online education platforms – Coursera, edX, and MOOCs – and collected course evaluation data from each. The dataset sizes gradually increased, starting with 10,000 reviews for Coursera and reaching 20,000 reviews for MOOCs, enabling an evaluation of the model's effectiveness across a range of dataset sizes. Each dataset underwent the same pre-processing steps, including text cleaning, word segmentation, and conversion to Word2Vec word vectors. We then used this processed data to train and evaluate the EduSent-Dig model. Dataset 1, dataset 2, and dataset 3 correspond to the data from Coursera, edX, and MOOCs, respectively. These datasets were provided by the platforms through public datasets or data-sharing agreements, and all data have been anonymised to protect user privacy. Table 2 presents the comprehensive experimental results.

The experimental data shows that the LSTM-AT model outperforms alternative models in predicting shifts in trading behaviour within the financial market, demonstrating its reliability and validity when applied to financial data. Additionally, this model exhibits the smallest errors across all evaluation metrics, highlighting its enhanced effectiveness.

Table 2 Findings from comprehensive experiment

Dataset	*Accuracy*	*Recall*	*F1 score*
Dataset 1	88.3%	87.2%	87.7%
Dataset 2	90.4%	89.6%	90.0%
Dataset 3	92.1%	91.3%	91.7%

5 Conclusions

We propose a big data approach in this paper to optimise the online teaching experience by deeply mining teaching and evaluation data from online education platforms, with a specific focus on extracting and analysing student sentiment factors. By employing advanced text analytics and sentiment analysis techniques, this study constructs the EduSent-Dig model, which integrates sentiment analysis with data mining methods. This model effectively extracts sentiment factors from online teaching assessment data and evaluates teaching effectiveness.

In this study, we first examine the significance of sentiment factors in online education environments and identify the limitations of current research. We propose integrating sentiment analysis with big data techniques to address these gaps. Subsequently, we provide a detailed overview of the relevant techniques in sentiment analysis and data mining, including deep learning methods like Bi-LSTM for sentiment analysis and the random forest algorithm for data mining.

The EduSent-Dig model we developed integrates Bi-LSTM and Word2Vec to create a novel framework for analysing online teaching effectiveness from various perspectives by fusing multidimensional teaching and evaluation data. Experimental validation across multiple datasets demonstrates high accuracy, recall, and F1 values. Through ablation and comprehensive experiments, we showcase the effectiveness of each model component and evaluate its performance on datasets of varying sizes.

Despite the encouraging findings, this study has its constraints. First, while our dataset encompasses multiple platforms, it may contain geographical or cultural biases that could impact the model's generalisability. Second, the model's relatively complex structure demands significant computational resources, which may restrict its use in resource-constrained settings. Additionally, current research primarily relies on static datasets and lacks the capability for dynamic analysis of real-time data streams. Lastly, although the model effectively analyses affective tendencies, there is limited exploration of affective intensity and complexity, including specific emotional states such as anger, joy, and surprise.

In our future research, we will focus on three key areas. Initially, for global applicability, we will expand the model's capabilities in multilingual and cross-cultural sentiment analysis. This entails gathering and analysing datasets in multiple languages and crafting algorithms that accommodate diverse cultural sentiment expressions, thus enhancing the model's universality and adaptability. Second, given the increasing real-time interactivity of online education platforms, we will explore how to adapt the model for real-time data streams. This includes developing efficient data pre-processing techniques and optimising the model structure to quickly respond to new data, providing educators with immediate feedback and suggestions for pedagogical adjustments. In conclusion, to lessen the model's reliance on computational resources, we will explore optimisation strategies and simplification techniques, such as efficient deep learning models and knowledge distillation, aiming to improve its practical application and deployment.

Through these efforts, we aim to enhance the EduSent-Dig model, making it more powerful, flexible, and easier to deploy. This will enable it to better serve the online education sector and deliver more effective teaching and learning experiences for both educators and students.

Acknowledgements

This work is supported by the Henan Medical Education Research Project: Ethical Issues and Practice in the Application of Human Assisted Reproductive Technology (No. WJLX2023224).

References

Airlangga, G. (2024) 'Advancing fake news detection: a comparative study of RNN, LSTM, and bidirectional LSTM Architectures', *Jurnal Teknik Informatika CIT Medicom*, Vol. 16, No. 1, pp.13–23.

Amrieh, E.A., Hamtini, T. and Aljarah, I. (2016) 'Mining educational data to predict student's academic performance using ensemble methods', *International Journal of Database Theory and Application*, Vol. 9, No. 8, pp.119–136.

Aoumeur, N.E., Li, Z. and Alshari, E.M. (2023) 'Improving the polarity of text through word2vec embedding for primary classical Arabic sentiment analysis', *Neural Processing Letters*, Vol. 55, No. 3, pp.2249–2264.

Boschetto, A. and Bottini, L. (2014) 'Accuracy prediction in fused deposition modeling', *The International Journal of Advanced Manufacturing Technology*, Vol. 73, pp.913–928.

Chakraborty, P., Mittal, P., Gupta, M.S. et al. (2021) 'Opinion of students on online education during the COVID-19 pandemic', *Human Behavior and Emerging Technologies*, Vol. 3, No. 3, pp.357–365.

Chicco, D. and Jurman, G. (2020) 'The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation', *BMC Genomics*, Vol. 21, pp.1–13.

Demir-Kavuk, O., Kamada, M., Akutsu, T. et al. (2011) 'Prediction using step-wise L1, L2 regularization and feature selection for small data sets with large number of features', *BMC Bioinformatics*, Vol. 12, pp.1–10.

Everingham, M., Eslami, S.A., Van Gool, L. et al. (2015) 'The Pascal visual object classes challenge: a retrospective', *International Journal of Computer Vision*, Vol. 111, pp.98–136.

Gooljar, V., Issa, T., Hardin-Ramanan, S. et al. (2024) 'Sentiment-based predictive models for online purchases in the era of marketing 5.0: a systematic review', *Journal of Big Data*, Vol. 11, No. 1, p.107.

Hofer, S.I., Nistor, N. and Scheibenzuber, C. (2021) 'Online teaching and learning in higher education: lessons learned in crisis situations', *Computers in Human Behavior*, Vol. 121, p.106789.

Hoffmann, R., Rempp, H., Erhard, L. et al. (2013) 'Comparison of four microwave ablation devices: an experimental study in ex vivo bovine liver', *Radiology*, Vol. 268, No. 1, pp.89–97.

Hung, J-L., Hsu, Y-C. and Rice, K. (2012) 'Integrating data mining in program evaluation of K-12 online education', *Journal of Educational Technology & Society*, Vol. 15, No. 3, pp.27–41.

Jang, B., Kim, M., Harerimana, G. et al. (2020) 'Bi-LSTM model to increase accuracy in text classification: combining Word2vec CNN and attention mechanism', *Applied Sciences*, Vol. 10, No. 17, p.5841.

Lan, K., Wang, D-T., Fong, S. et al. (2018) 'A survey of data mining and deep learning in bioinformatics', *Journal of Medical Systems*, Vol. 42, pp.1–20.

Li, D. and Gao, H. (2018) 'A hardware platform framework for an intelligent vehicle based on a driving brain', *Engineering*, Vol. 4, No. 4, pp.464–470.

Lu, G., Xie, K. and Liu, Q. (2022) 'What influences student situational engagement in smart classrooms: perception of the learning environment and students' motivation', *British Journal of Educational Technology*, Vol. 53, No. 6, pp.1665–1687.

Naseem, U., Razzak, I., Musial, K. et al. (2020) 'Transformer based deep intelligent contextual embedding for twitter sentiment analysis', *Future Generation Computer Systems*, Vol. 113, pp.58–69.

Philander, K. and Zhong, Y. (2016) 'Twitter sentiment analysis: capturing sentiment from integrated resort tweets', *International Journal of Hospitality Management*, Vol. 55, pp.16–24.

Qin, P., Xu, W. and Guo, J. (2016) 'A novel negative sampling based on TFIDF for learning word representation', *Neurocomputing*, Vol. 177, pp.257–265.

Reisenzein, R. (2000) 'Exploring the strength of association between the components of emotion syndromes: the case of surprise', *Cognition & Emotion*, Vol. 14, No. 1, pp.1–38.

Shi, L., Westerhuis, J.A., Rosén, J. et al. (2019) 'Variable selection and validation in multivariate modelling', *Bioinformatics*, Vol. 35, No. 6, pp.972–980.

Sivarajah, U., Kamal, M.M., Irani, Z. et al. (2017) 'Critical analysis of big data challenges and analytical methods', *Journal of Business Research*, Vol. 70, pp.263–286.

Yadav, S., Ekbal, A., Saha, S. et al. (2019) 'Feature assisted stacked attentive shortest dependency path based Bi-LSTM model for protein-protein interaction', *Knowledge-based Systems*, Vol. 166, pp.18–29.