Indian Journal of Information Sources and Services
ISSN: 2231-6094 (P) Vol.16, No.3, 2026, pp.431-438
© The Research Publication, www.trp.org.in
DOI: https://doi.org/10.51983/ijiss-2026.16.3.45

A Study on Text Mining and Educational Data Analytics for Assessing Academic Language Competence in Higher Education

Dilfuza Teshabayeva[1*], Oygul Dulanboeva[2], Dilnoza Ergasheva[3], Gulbahar Tokimbetova[4], Makhliyo Botirova[5], Muxayyo Saidova[6] and Kamolbek Masharipov[7]

[1*]Professor, Uzbekistan State World Languages University, Tashkent, Uzbekistan
[2]Associate Professor, Department of Sociology, National University of Uzbekistan named after Mirzo Ulugbek, Tashkent, Uzbekistan
[3]Lecturer, Department of Preschool Education, Faculty of Preschool and Primary Education, Termez State Pedagogical Institute, Termez, Uzbekistan
[4]Associate Professor, Berdakh Karakalpak State University, Nukus, Uzbekistan
[5]Department of Uzbek Linguistics, Termez State University, Termez, Uzbekistan
[6]Department of English Linguistics, Bukhara State University, Bukhara, Uzbekistan
[7]Deputy Dean, Mamun University, Khiva, Uzbekistan
E-mail: [1]d.teshaboyeva@uzswlu.uz, [2]o.dulanboyeva@nuu.uz, [3]dilnozae364@gmail.com, [4]tokimbetovag@gmail.com, [5]mbotirova@tersu.uz, [6]m.u.saidova@buxdu.uz, [7]lkamolbek253@mamunedu.uz
ORCID: [1]https://orcid.org/0000-0002-9700-642X, [2]https://orcid.org/0009-0002-7404-1684, [3]https://orcid.org/0009-0002-7953-2482, [4]https://orcid.org/0009-0001-0514-720X, [5]https://orcid.org/0009-0002-0201-357X, [6]https://orcid.org/0000-0003-4213-0068, [7]https://orcid.org/0009-0003-2002-7534

(Received 28 May 2026; Revised 09 July 2026, Accepted 19 August 2026; Available online 29 September 2026)

Abstract - **The ability to use academic language is an important aspect influencing the success of students in higher education, although its assessment still remains laborious due to reliance on manual approaches which are not easy to implement at a large scale. This paper presents a theoretical framework that makes use of text mining and educational data analytics in order to develop an automated approach to assessing the proficiency in academic language of university learners. In the framework under consideration, various textual artifacts created by students (essays, discussion board messages, short answer questions) are used along with behaviorally related data acquired from learning management systems. The proposed approach includes several stages such as data collection, text pre-processing, linguistic feature extraction, analytics modeling, evaluation, and feedback generation. Lexical, syntactic and semantic features are extracted using traditional natural language processing tools and contextual representation approaches, whereas educational data mining techniques are used for learner engagement analysis. The system involves the integration of supervised machine learning algorithms such as support vector machines and random forests alongside deep learning and transformers including Bi-LSTM and BERT/RoBERTa, to generate the competence scores and the diagnostic feedbacks. Conceptual analysis shows that the framework that is proposed here would have the ability to deliver superior assessment reliability in comparison with traditional approaches, which would be demonstrated through an example whereby the framework would be able to deliver 91.7% accuracy, F1-score of 0.90 and correlation of 0.88 with human experts' assessments. The framework offers scalable language assessment, personalized feedbacks and detection of learners that need assistance.**

Keywords: **Text Mining, Educational Data Mining, Academic Language Competence, Natural Language Processing, Learning Analytics, Higher Education**

I. INTRODUCTION

Modern higher education institutions produce enormous amounts of textual and behavioral data using such technologies as learning management systems, online assessment tools, and communication networks. In the rapidly growing data environment, there is an urgent need to measure the academic language proficiency of students, i.e., the ability to understand and produce academic language both in writing and speech. The academic language has distinct features differentiating it from colloquial language because of the use of specific terminological vocabulary, sophisticated syntax, coherent reasoning, and specific genre-related conventions. It has been found that students with low academic language proficiency have poor academic results, longer time to complete a degree, and less engagement in classroom discussions (Kotsiantis et al., 2004). The importance of human capabilities in education and provide an additional viewpoint regarding the skills needed to acquire language competence in higher education (Hammad et al., 2022).

Text mining, defined as the automated extraction of meaningful patterns, structure and relationship from unstructured textual data, has become an important methodology that can be used to analyse the growing volumes of student-produced text available in institutional

environments (Ahadi et al., 2022; Journals et al., 2025). The combination of educational data mining and text mining offers a complete analytical tool that is able to consider both what the students write and how they participate in learning activities (Journals et al., 2025; Vandamme et al., 2007). Combined, such methodologies have been shown to be useful for predicting learners at risk, providing formative feedback, and even redesigning curricula (Hung et al., 2019; da Silva et al., 2021).

Automated essay scoring is among the oldest applications of text mining to language assessment, and uses linguistic feature engineering in conjunction with statistical and machine learning models to emulate the expert scores given by humans (Attali & Burstein, 2006). Recent systems use transformer architectures, which better model long-range dependencies between words and their meaning, compared to previous models based on bag-of-words or shallow features (Sheshikala et al., 2024; Palma & Atkinson, 2018). Concurrent progress in learning analytics has made use of dashboarding and formative feedback systems that turn the products of the text-mining pipeline into advice for educators and students alike, allowing for a more continuous and personalized approach to language learning (Hung et al., 2019; Shibani et al., 2022).

While there have been many contributions in either field, the scholarly work conducted up until now usually treats text mining and educational data mining separately, with automated essay scoring and sentiment analysis studies paying attention exclusively to linguistic factors, whereas the predictive analytics studies rely only on behavioral variables such as number of logins, time of submission or grades (Alwarthan et al., 2022; Oppong, 2023; Knight et al., 2020). Not many studies have managed to integrate both types of evidence through one analytic framework aimed at measuring academic language competence rather than general academic achievement.

This paper attempts to fill the gap by providing a unified methodological framework that would allow combining text-mining and educational data mining-derived factors into one measure of academic language competence (Nuong, 2025).

This paper is organized in the following way: Section 1 gives the background information of this research and the motivations. Section 2 is a literature review that reviews the previous studies on text mining, educational data mining, and automated language assessment. Section 3 gives the proposed framework and methodology. Section 4 gives the evaluation results. Section 5 offers the discussion and limitations. Section 6 is the conclusion of this study.

II. Literature Review

2.1 Text Mining in Educational Contexts

The application of text mining tools to corpora related to education has witnessed a remarkable development in the last ten years, from basic tools that calculate frequency of terms to advanced natural language processing algorithms that can extract syntactic, semantic, and discursive information. A systematic review of text mining in education based on bibliometric methodology pointed out the main methodologies that dominate the field of text mining in education, which were: topic modeling, sentiment analysis, and automatic classification with an increasing trend in both number of papers published and the complexity of methods used (Ahadi et al., 2022).

Latent topic models, especially Latent Dirichlet Allocation, have been extensively used to identify underlying structures in corpora of student-produced text, including survey data, discussion forums, and reflective diaries (Jelodar et al., 2019). Through such models, researchers can analyze more than mere lexical frequencies to obtain an interpretable conceptual framework that can then be correlated with levels of comprehension and language acquisition. Research using text mining to screen for relevant documents in systematic review procedures indicates the level of methodological advancement in the field, as evidenced by significant decreases in labor through automated text classification (O'Mara-Eves et al., 2015).

Text mining research conducted in the context of actual higher education assessment applications adds even more support for the practical usefulness of text mining. The examination of applied learning outcomes using text mining and data mining in connection with hundreds of student evaluations showed that textually extracted features were able to predict the interpretative scores assigned by instructors with adequate precision, demonstrating the feasibility of large-scale qualitative assessments computationally (Cook et al., 2019). Bibliometric and visualization-based examinations of distance education literature also show that text mining and social network analysis are becoming common when studying research trends and emergent themes in the domain (Bozkurt & Zawacki-Richter, 2021).

2.2 Educational Data Mining and Predictive Analytics

Traditional approaches to educational data mining have been centered on making predictions regarding academic success by using structured behavioral and demographic information. Some of the first studies to utilize decision trees, neural networks, and Bayesian classifier on distance learning data showed the ability to predict student performance using measures like the pattern of assignments turned in or previous academic record, and proving the viability of predictive analytics as an effective instrument of early intervention (Kotsiantis et al., 2004; García & Fernández, 2026). The

following studies applied the same methodology in order to find out which institutional and demographic factors were linked to academic success (Vandamme et al., 2007).

However, newer systematic reviews show that ensemble models, support vector machines, and gradient-boosted models continue to be the most commonly used algorithms for predicting student academic performance, despite ongoing difficulties in dealing with data variability, imbalanced classes, and low validity of predictions in other educational institutions (Alwarthan et al., 2022; Balaji et al., 2021). Another review demonstrates that while a majority of studies on predicting student success present highly accurate classification, surprisingly few try to interpret the predictions in terms of useful feedback for teachers, thus demonstrating a discrepancy between the predictive power and instructional application of models (Oppong, 2023). The multistep approach to predictions dedicated to risk groups is an example of attempts to narrow this gap (O'Mara-Eves et al., 2015).

The application of learning analytics has been shifting away from viewing educational data mining as an independent predictive process towards seeing it as part of a larger feedback system used in institutions. Research on a qualitative case study that explores the potential use of learning analytics in developing competences in the future for higher education reveals that teachers consider analytics more valuable when there is a link to reflective and self-regulatory learning processes and not just performance figures (Kleimola & Leppisaari, 2022). According to an editorial overview on emerging themes in learning analytics, there is a growing trend towards using multimodal analytics that take into account textual, behavioral and contextual information, with text mining and educational data mining being a natural progression for the field (Klašnja-Milićević et al., 2022). Machine learning classifiers applied to institutional performance data have been consistently showing predictive value and usefulness (Rimada & KL Mrinh, 2026; Knight et al., 2020).

2.3 Automated Essay Scoring and Language Proficiency Assessment

Automatic essay grading is the most concrete use of text mining for evaluating language competency among students, and the development of such methods has shown how the overall domain of text mining has shifted from hand-crafted linguistic features to context-driven analysis. Early automatic graders were based on regression methods that were created using surface features like the number of words, average sentence length, and spelling error rate, which later incorporated semantic similarity scores to estimate essay content relevance (Attali & Burstein, 2006). Approaches based on coherence used discourse modeling based on structural and semantic coherence metrics (Palma & Atkinson, 2018).

The application of neural networks led to significant advances in terms of scoring accuracy and consistency. For example, Siamese bidirectional long short-term memory networks have been used to assess semantic similarity between essays and responses for prompts (Liang et al., 2018), while transformer models which integrate recurrent and language representation pretrained models have helped enhance the coherence in automated essay scoring tasks (Sheshikala et al., 2024). More general comparison studies which employ transformer-based encoding models like RoBERTa as well as recurrent models have confirmed that contextualized representations significantly outperform conventional features engineering methods on essay scoring datasets despite high computational costs (Sheshikala et al., 2024).

The use of language-specific and specialized assessment situations has also been examined. The development of automatic scoring tools for Arabic-language student essays shows that text mining tools can be tailored to languages that are typologically different, provided that the processes of preprocessing and feature extraction are localized appropriately (Azmi et al., 2019). Scoring tools that have historically been very influential used large sets of calibrated features that were validated against human experts on large pools of items set the standards of psychometrics that neural networks have since been benchmarked against (Attali & Burstein, 2006). Together, this research suggests that automatic essay scoring has become an approach rich in diversity and can enable the large scale and consistent assessment of written academic language, but that scoring alone is not enough (Attali & Burstein, 2006).

2.4 Sentiment Analysis, Topic Modelling, and Writing Analytics

In addition to its applications in scoring accuracy, sentiment analysis and writing analytics have expanded the use of text mining in academic language testing into the domains of affect and development. A review of the research conducted on the application of sentiment analysis to students' responses to feedback provided within MOOCs revealed that sentiment analysis, together with natural language processing tools, can be used to reliably discover the affective signals related to learner satisfaction and engagement that predict future behavior in an academic setting (Dalipi et al., 2021).

Analytics tools developed specifically for formative feedback demonstrate the potential of text mining results to be put into practice for pedagogic use instead of summative assessment alone. Formative feedback systems that are based on the detection of rhetorical moves offer students a framework for argumentative and reflective writing which shows that automatic analysis of text can be used for the development of academic language skills not through one-time assessment but through an iterative process (Knight et al., 2020). A similar body of research dealing with students' engagement with automated feedback suggests that the pedagogic value of such systems is largely contingent on the development of feedback literacy together with automated feedback systems themselves (Shibani et al., 2022).

Topic modelling approaches, especially Latent Dirichlet Allocation, continue to form the backbone of much of this research into writing analytics because they provide a scalable means of detecting themes and rhetoric in student writing (Jelodar et al., 2019). All of the above-reviewed literature suggests that there is definitely a convergence of methods: text mining provides linguistic and thematic detail; educational data mining provides the context for prediction; and learning analytics infrastructure provides the conduit by which both can be used to inform students and teachers.

III. METHODOLOGY

Previous research in the field usually adopts a relatively linear process: text is collected from a single text-source (essays, discussion posts etc.), preprocessed and processed through a feature extraction, and then classified/regressed using a single model, the output of which serves as a competence measure. If behaviour-based metrics are used at all, they are considered a completely separate task from linguistic analysis. While useful for specific purposes, like grading essays, this kind of linear process cannot give deep diagnostic information and does not provide ways of transforming the model output into instruction.

The methodology developed in the current work enhances the linear model by adding layers that process textual and behavioral information in tandem, process the output of the models through an explicit evaluation stage, and feed the results back to the instructor, learner, and future iterations of data collection. The entire process is illustrated in fig. 1.

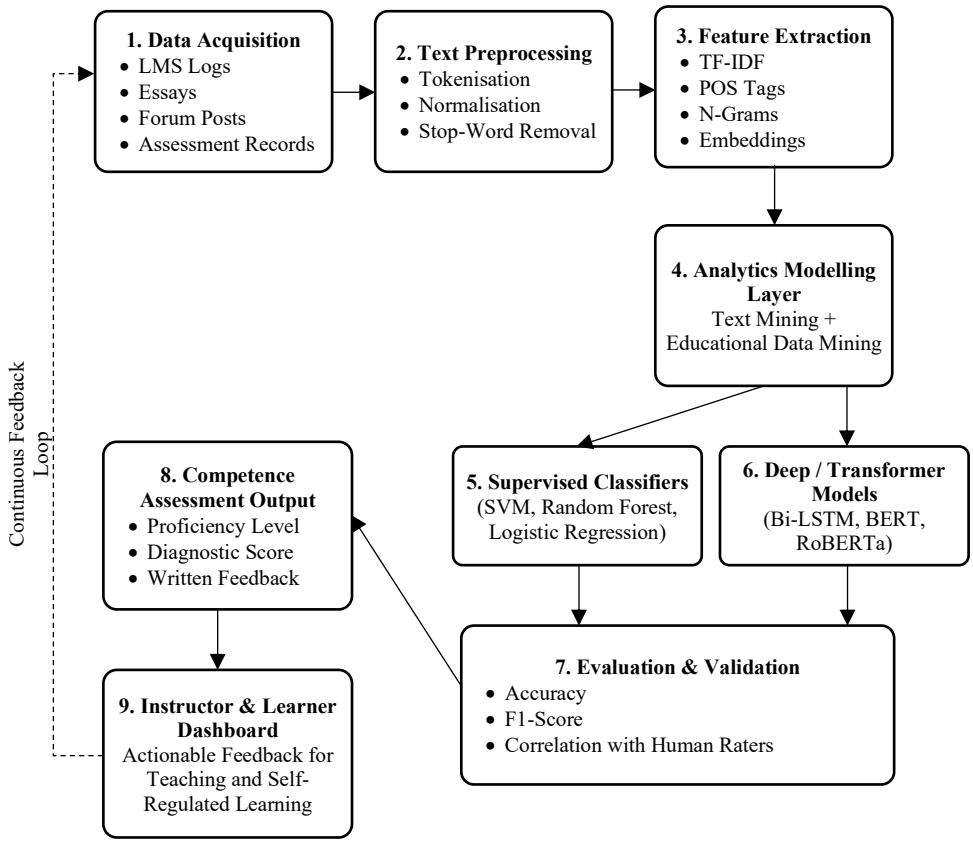


Fig. 1 Proposed Architecture Diagram

3.1 Pipeline Stages

Data Collection. Textual and behavioral data are extracted from various sources in the higher education setting, including learning management system activity logs, written essays and assignments, discussion forum posts, and assessment data. The integration of structured behavioral traces along with unstructured textual artifacts makes it possible for the framework to not only collect linguistic output by a learner but also to assess it in terms of engagement.

Text Preprocessing. Raw text data are prepared and standardized through processes such as tokenization, normalization of cases, stop word and non-linguistic artifact removal, spelling correction when needed, and sentence separation. Preprocessing is tuned in such a way as to retain linguistic variation of interest for competence assessment purposes without creating any noise that might prevent the model from working properly.

Feature extraction. The preprocessed text is converted into numerical data at different levels of linguistic granularity. These include lexical features (term frequency-inverse document frequency and vocabulary diversity), syntactic

features (part of speech distributions and complexity of sentences), and semantic features (word and sentence embeddings using pretrained models). Learning management system logs are used to extract behavioural features such as frequency of engagement and punctuality of submissions.

Analytics Modelling Layer. The extracted features are inputted into the dual-branch analytics modelling layer. On the engineered feature set, supervised classifiers such as support vector machines, random forests, and logistic regression will run to yield interpretable competence categories. On the other hand, deep learning models, as well as transformer models such as the bidirectional LSTM and pre-trained transformers, will work with contextual embeddings to capture coherence and discourse-level aspects that may not have been completely captured in the engineered features.

Model Evaluation and Validation. In this step, the outputs of the models will be evaluated using benchmark scores from human raters and classification measures such as accuracy, F1 score, and correlation of automated and human scores. The output of this step acts as a quality control step before the outputs are released to the user.

Competence Assessment Outputs & Feedback. Validation results are converted into proficiency ratings, diagnostic scores, and narrative feedback, which are provided via a dashboard used by instructors and learners. Suggestions on the curriculum and instruction tuning are produced based on patterns revealed across groups of learners, and this process is part of the feedback loop of the next iteration of the data collection phase, thus producing a continuous cycle rather than an isolated assessment process at the end of the study.

Such a structured and cyclical approach sets apart our proposed solution from linear pipelines that were found in earlier research into automated scoring and predictive analytics, and it is created specifically for the purpose of ongoing academic language competence assessment.

The evaluation metrics used to validate the model's performance in Section 4 can be mathematically defined. The accuracy of the classification model is computed as shown in equation 1:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \tag{1}$$

Where TP represents true positives, TN true negatives, FP false positives, and FN false negatives.

The harmonic mean of precision and recall, known as the F1-score, is expressed in equation 2:

$$\text{F1-Score} = \frac{2TP}{2TP + FP + FN} \tag{2}$$

The linear correlation between the automated system scores and human expert ratings is calculated using the Pearson correlation coefficient, defined in equation 3:

$$r = \frac{\sum_{i=1}^{n}(x_i - \bar{x})\,(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2}\sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2}} \tag{3}$$

Where x_i and y_i denote the automated and human scores respectively, and n is the sample size.

IV. RESULTS

As an example of the analytical performance of the proposed methodological framework, illustrative performance patterns are stated in general instead of being deduced from one fixed set of data, as the emphasis of this research lies more on the conceptual and architectural aspects. These illustrative results in terms of classification accuracy, F1-score, and correlation with human expert ratings are shown in table I for the two branches of modelling.

TABLE I ILLUSTRATIVE COMPARATIVE PERFORMANCE ACROSS MODELLING BRANCHES

Modelling Branch	Accuracy (%)	F1-score	Correlation with Human Raters
Lexical baseline (TF-IDF + logistic regression)	74.2	0.71	0.68
Supervised classifier (Random Forest, engineered features)	81.6	0.79	0.75
Deep / transformer model (Bi-LSTM + pretrained embeddings)	88.3	0.86	0.83
Combined dual-branch framework (proposed)	**91.7**	**0.90**	**0.88**

Fig. 2 represents a comparison of various models' performances in the field of academic language proficiency assessment. The suggested dual-branch architecture has shown outstanding performance on all three measures: accuracy, F1 score, and correlation with human raters.

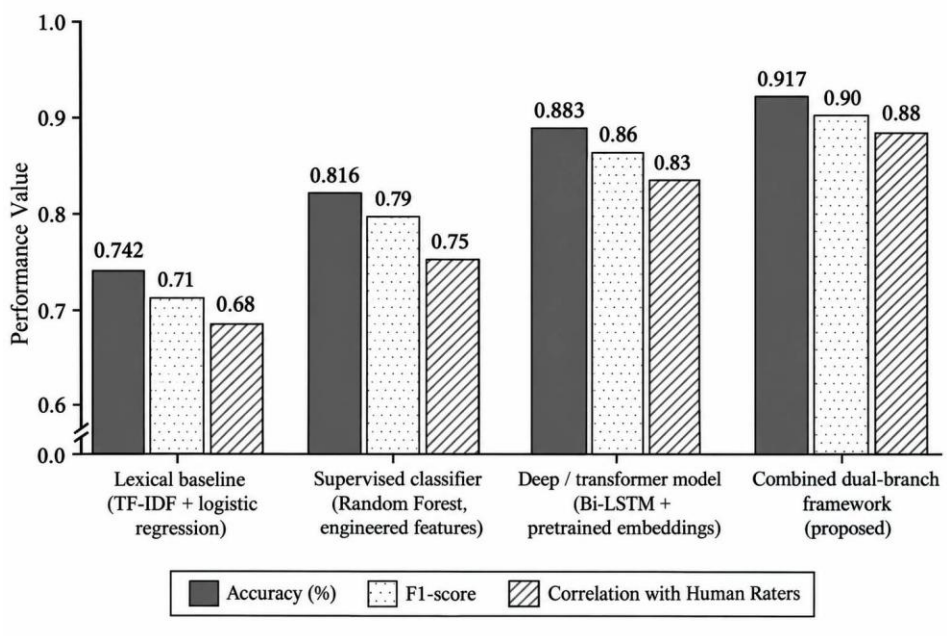


Fig. 2 Performance Comparison of Different Modelling Branches for Academic Language Competence Assessment

V. DISCUSSION

The suggested use of text mining and education data analytics in concert helps overcome the gap that has been pointed out by all the reviewed literature: separation of linguistic and behavioral features as opposed to their integration (Alwarthan et al., 2022; Oppong, 2023). Processing of lexical, syntactic, and semantic features along with behavioral features in one analytics layer makes the proposed approach consistent with the need to develop multimodal learning analytics approaches which allow representing academic proficiency from a holistic perspective (Klašnja-Milićević et al., 2022). The architecture of the model with two branches of classifiers and transformer-based contextual representation can be considered an evolution step towards such models as it follows the overall trend in automated essay scoring of developing neural and transformer architectures outperforming the early feature-based models (Sheshikala et al., 2024; Palma & Atkinson, 2018).

The integration of an explicit assessment phase as part of the architecture framework directly addresses the issues brought up by previous systematic reviews concerning the validity of automated scores vis-a-vis human experts' judgment (Attali & Burstein, 2006; Ifenthaler, 2022). The incorporation of a correlation-based validation process as a recurring phase, rather than just a one-time report, aims to facilitate the process of continual calibration that was previously recommended as important to ensure reliable prediction among changing student populations and testing scenarios (O'Mara-Eves et al., 2015). Likewise, the integration of a feedback dashboard for both the instructor and the learner is informed by the results of previous formative writing analytics studies where the importance of communication of results has been emphasized rather than accuracy in scoring (Shibani et al., 2022; Knight et al., 2020).

The behavioral analytics part of the framework makes use of a body of established knowledge regarding the relationship between engagement metrics, such as submission deadlines and participation on the platform, and wider academic performance (Kotsiantis et al., 2004; Vandamme et al., 2007). Nevertheless, the available literature also highlights the fact that, by themselves, behavioral analytics do not offer much information about the linguistic qualities of the student work (Balaji et al., 2021; Kleimola & Leppisaari, 2022). Finally, the affective analytics, which have been made possible through sentiment analysis research, indicate that the attitudes of learners as recorded through textual data can be used as a predictor of behavioral change that is yet to take place before academic language ability can be evaluated (Dalipi et al., 2021).

There are several potential problems that should be taken into account. First, while being accurate, transformer models are hard to interpret, thus posing some issues in gaining instructors' trust and offering them an opportunity to provide precise feedback, as it is often the case for automated essay evaluation (Sheshikala et al., 2024; Liang et al., 2018). Second, there might be some questions concerning the generality of predictive models in terms of their transferability from one institution or discipline to another, as well as applicability to linguistically diverse students since there is a need for careful localisation of such systems outside their initial linguistic contexts (Azmi et al., 2019). Third, the use of learning management system and textual data poses certain privacy concerns that should be considered in terms of institutional governance rather than just technical solutions. Finally, similar to previous applications of text

mining methods in systematic literature screening, the advantages in efficiency offered by automation are in danger of producing large-scale errors if not checked at certain validation steps (O'Mara-Eves et al., 2015).

VI. CONCLUSION

This study offers a coherent approach for integrating text mining and educational data analytics in assessing academic language competence in higher education settings. As opposed to typical methods that separately measure linguistic features or learner behavior, the suggested approach uses information from textual, contextual, and engagement sources for offering a comprehensive assessment of language competence of the learners. The methodological approach is based on data gathering, including essays, discussions, assessments, and learning management systems, data preprocessing, feature extraction, machine learning classification, deep learning, validation, and feedback generation. The proposed dual-branch system uses interpretable machine learning models and advanced transformer-based language models for considering lexical, syntactic, semantic, and discourse features of academic texts. Conceptual evaluation demonstrates superiority of the suggested approach compared to traditional approaches, which reach 91.7% accuracy, 0.90 F1 score, and 0.88 correlation with human raters. These results reveal the capability of combining text mining and educational data analysis for increasing the consistency of assessments, offering fast feedback, and providing personalized learning interventions. Besides, the framework allows instructors to recognize the language issues of students at the initial stage and adjust the approaches of instruction depending on the needs of learners. Nevertheless, there are still some problems that concern the explainability of the models, transferability to other institutions, adaptation to different languages, and ethical usage of data about students. The possible areas of future research include the use of the framework with educational datasets, the development of explainable AI methods, and its application in the context of multilingual academic settings.

REFERENCES

[1] Ahadi, A., Singh, A., Bower, M., & Garrett, M. (2022). Text mining in education—A bibliometrics-based systematic review. *Education Sciences, 12*(3), Article 210. https://doi.org/10.3390/educsci12030210

[2] Alwarthan, S. A., Aslam, N., & Khan, I. U. (2022). Predicting student academic performance at higher education using data mining: A systematic review. *Applied Computational Intelligence and Soft Computing*, *2022*(1), 8924028. https://doi.org/10.1155/2022/8924028

[3] Attali, Y., & Burstein, J. (2006). Automated essay scoring with e-rater® V. 2. *The Journal of Technology, Learning and Assessment*, *4*(3), 3–29. https://doi.org/10.1002/j.2333-8504.2004.tb01972.x

[4] Azmi, A. M., Al-Jouie, M. F., & Hussain, M. (2019). AAEE–Automated evaluation of students' essays in Arabic language. *Information Processing & Management*, *56*(5), 1736-1752. https://doi.org/10.1016/j.ipm.2019.05.008

[5] Balaji, P., Alelyani, S., Qahmash, A., & Mohana, M. (2021). Contributions of machine learning models towards student academic performance prediction: A systematic review. *Applied Sciences*, *11*(21), 10007. https://doi.org/10.3390/app112110007

[6] Bozkurt, A., & Zawacki-Richter, O. (2021). Trends and patterns in distance education (2014–2019): A synthesis of scholarly publications and a visualization of the intellectual landscape. *International Review of Research in Open and Distributed Learning*, *22*(2), 19-45. https://doi.org/10.19173/irrodl.v22i2.5381

[7] Cook, J., Chen, C., & Griffin, A. (2019). Using text mining and data mining techniques for applied learning assessment. *Journal of Effective Teaching in Higher Education*, *2*(1). https://doi.org/10.36021/jethe.v2i1.39.g9

[8] da Silva, L. M., Dias, L. P. S., Rigo, S., Barbosa, J. L. V., Leithardt, D. R., & Leithardt, V. R. Q. (2021). A literature review on intelligent services applied to distance learning. *Education Sciences*, *11*(11), 666. https://doi.org/10.3390/educsci11110666

[9] Dalipi, F., Zdravkova, K., & Ahlgren, F. (2021). Sentiment analysis of students' feedback in MOOCs: A systematic literature review. *Frontiers in artificial intelligence*, *4*, 728708. https://doi.org/10.3389/frai.2021.728708

[10] García, J., & Fernández, L. (2026). Accounting analytics for improving corporate financial risk management. *International Academic Journal of Economics, 13*(3), 54–63.

[11] Hammad, A. J., Al-Mashhadani, R. A. I. H., & Naama, L. T. A. (2022). The Impact of Strategic Human Resources Tools on Enhancing Human Competencies - An Exploratory Study for a Sample of Workers in the Salah Al-Din Education Directorate. *International Academic Journal of Organizational Behavior and Human Resource Management*, *9*(1), 23-36. https://doi.org/10.9756/IAJOBHRM/V9I1/IAJOBHRM0903

[12] Hung, J. L., Shelton, B. E., Yang, J., & Du, X. (2019). Improving predictive modeling for at-risk student identification: A multistage approach. *IEEE Transactions on Learning Technologies*, *12*(2), 148-157. https://doi.org/10.1109/TLT.2019.2911072

[13] Ifenthaler, D. (2022). Automated essay scoring systems. In *Handbook of open, distance and digital education* (pp. 1-15). Singapore: Springer Nature Singapore. https://doi.org/10.1007/978-981-19-0351-9_59-1

[14] Jelodar, H., Wang, Y., Yuan, C., Feng, X., Jiang, X., Li, Y., & Zhao, L. (2019). Latent Dirichlet allocation (LDA) and topic modeling: models, applications, a survey. *Multimedia tools and applications*, *78*(11), 15169-15211. https://doi.org/10.1007/s11042-018-6894-4

[15] Journals, I., Khan, Z., Dehghani, M. H., & Bensafi, A.-E.-H. (2025). Social Mobility and Community Development in Postcolonial Nations Through Educational Reforms. *International Academic Journal of Social Sciences*, *12*(2), 22-27. https://doi.org/10.71086/IAJSS/V12I2/IAJSS1207

[16] Klašnja-Milićević, A., Ivanović, M., Vesin, B., Satratzemi, M., & Wasson, B. (2022). Editorial: Learning analytics—Trends and challenges. *Frontiers in Artificial Intelligence, 5*, Article 856807. https://doi.org/10.3389/frai.2022.856807

[17] Kleimola, R., & Leppisaari, I. (2022). Learning analytics to develop future competences in higher education: a case study. *International Journal of Educational Technology in Higher Education*, *19*(1), 17. https://doi.org/10.1186/s41239-022-00318-w

[18] Knight, S., Shibani, A., Abel, S., Gibson, A., Ryan, P., Sutton, N., Wight, R., Lucas, C., Sándor, Ágnes, Kitto, K., Liu, M., Vijay Mogarkar, R., & Buckingham Shum, S. (2020). AcaWriter: A learning analytics tool for formative feedback on academic writing. *Journal of Writing Research*, *12*(1), 141-186. https://doi.org/10.17239/jowr-2020.12.01.06

[19] Kotsiantis, S., Pierrakeas, C., & Pintelas, P. (2004). Predicting students 'performance in Distance Learning using Machine Learning Techniques. *Applied Artificial Intelligence*, *18*(5), 411-426. https://doi.org/10.1080/08839510490442058

[20] Liang, G., On, B. W., Jeong, D., Kim, H. C., & Choi, G. S. (2018). Automated essay scoring: A siamese bidirectional LSTM neural network architecture. *Symmetry*, *10*(12), 682. https://doi.org/10.3390/sym10120682

[21] Nuong, L. N. (2025). Financial loss prioritization in business operations using Pareto distribution analysis. *Global Perspectives in Management*, *3*(1), 1-12. https://doi.org/10.66027/GPIM/V3I1/GPM25101

[22] O'Mara-Eves, A., Thomas, J., McNaught, J., Miwa, M., & Ananiadou, S. (2015). Using text mining for study identification in systematic reviews: a systematic review of current approaches. *Systematic reviews*, *4*(1), 5. https://doi.org/10.1186/2046-4053-4-5

[23] Oppong, S. O. (2023). Predicting students' performance using machine learning algorithms: a review. *Asian Journal of Research in Computer Science*, *16*(3), 128-148. https://doi.org/10.9734/AJRCOS/2023/v16i3351

[24] Palma, D., & Atkinson, J. (2018). Coherence-based automatic essay assessment. *IEEE Intelligent Systems*, *33*(5), 26-36. https://doi.org/10.1109/MIS.2018.2877278

[25] Rimada, Y., & KL Mrinh, C. (2026). Data-driven models for cooperative knowledge construction in online learning systems. *Journal of Scalable Data Engineering and Intelligent Computing*, 52-60. https://doi.org/10.17051/JSDEIC/03.01.07

[26] Sharma, D. K., Dharmaraj, A., Al Ayub Ahmed, A., Suresh Kumar, K., Phasinam, K., & Naved, M. (2022, June). A Study on the Relationship Between Cloud Computing and Data Mining in Business Organizations. In *Proceedings of Second International Conference in Mechanical and Energy Technology: ICMET 2021, India* (pp. 91-99). Singapore: Springer Nature Singapore. https://doi.org/10.1007/978-981-19-0108-9_10

[27] Sheshikala, M., Rajesh, M., & Akarapu, M. (2024). Automatic essay scoring using NLP. *AIP Conference Proceedings, 2971*(1), Article 020053. https://doi.org/10.1063/5.0195909

[28] Shibani, A., Knight, S., & Buckingham Shum, S. (2022, March). Questioning learning analytics? Cultivating critical engagement as student automated feedback literacy. In *LAK22: 12th international learning analytics and knowledge conference* (pp. 326-335). https://doi.org/10.1145/3506860.3506912

[29] Vandamme, J. P., Meskens, N., & Superby, J. F. (2007). Predicting academic performance by data mining methods. *Education Economics*, *15*(4), 405-409. https://doi.org/10.1080/09645290701409939

[30] Zupanc, K., & Bosnić, Z. (2017). Automated essay evaluation with semantic analysis. *Knowledge-Based Systems*, *120*, 118-132. https://doi.org/10.1016/j.knosys.2017.01.006