Journal of Neonatal Surgery
ISSN(Online): 2226-0439
Vol. 14, Issue 21s (2025)
https://www.jneonatalsurg.com

OPEN ACCESS

Educational Data Mining: Using NLP For Student Performance Analysis in Nep 2020

Dr. Sulakshana Vispute

Assistant Professor, MCA Department, DES's NMITD, Mumbai University, India.

Email: sulakshana.vispute@despune.org

Cite this paper as: Dr. Sulakshana Vispute, (2025). Educational Data Mining: Using NLP For Student Performance Analysis in Nep 2020. *Journal of Neonatal Surgery*, 14 (21s), 966-970.

ABSTRACT

This study examines the use of Natural Language Processing (NLP) in Educational Data Mining (EDM) to evaluate student performance within the framework of India's National Education Policy (NEP) 2020.It examines how NLP techniques (eg. BERT or GPT,TF-IDF, LDA, etc**.)** can derive meaningful information from textual data, such as learner writing samples, comments, and online forum discussions, to offer a deeper understanding of student learning. The analysis aims to recognize main aspects impacting student achievement and highlight areas where pedagogical interventions can be most effective. The outcomes have important implications for fostering personalized learning experiences, improving overall educational outcomes, and aligning educational practices with the objectives of NEP 2020.

Keywords: *EDM, NLP, NEP 2020, BERT or GPT, TF-IDF, LDA, etc.*

1. INTRODUCTION

EDM is a crucial field for extracting derive meaningful information from educational data, with the potential to enhance teaching strategies and personalize learning experiences. By leveraging data mining techniques to identify patterns and trends in student data, researchers and educators can improve educational outcomes (Society, n.d.). In this situation, NLP plays a vital role by enabling the analysis of non-structured data, such as learner writing samples, feedback, and online discussions, providing a deep understanding of learner performance and learning. The blend of NLP and EDM facilitates a more inclusive assessment of students' skills and knowledge (Lumenalta, 2025). Furthermore, India's National Education Policy (NEP) 2020 advocates for personalized learning pathways and holistic assessment methods to modernize the educational system (National Education Policy, n.d.). Given the policy's emphasis on utilizing technology to enhance education, the convergence of EDM and NLP is particularly relevant in today's educational landscape (Lumenalta, 2025).

1,1 Background on Educational Data Mining

EDM is an cross-disciplinary field that utilizes data mining, machine learning, and statistical procedures to analyse educational data. Its prime goals comprise understanding how students learn, identifying aspects that impact learning outcomes, and developing data-driven strategies to enhance teaching and learning processes. Common EDM techniques include classification (e.g., predicting student performance), regression (e.g., modelling student grades), clustering (e.g., grouping students based on learning styles), and association rule mining (e.g., identifying relationships between student behaviours and outcomes) (Society, n.d.). Over time, EDM has evolved from basic statistical analysis to advanced machine learning approaches, incorporating techniques like deep learning to process complex educational datasets. Effective applications of EDM can be observed in various educational settings, such as predicting student dropout rates, personalizing learning content, and providing early interventions for struggling students (Alshareef, 2020).

1.2 Natural Language Processing (NLP) in Education

NLP offers powerful tools for analysing textual data in educational settings. Student writing samples, forum discussions, feedback forms, and other written materials offer valuable insights into students' understanding, reasoning skills, and learning progress. NLP techniques enable the automatic extraction and analysis of this information, revealing insights that might be challenging to gain through conventional approaches (Thomas, 2025). Key NLP applications in education include text summarization (e.g., generating concise summaries of student writing samples), question answering (e.g., responding to queries based on course materials), and automated essay scoring (e.g., evaluating the quality of student writing). These techniques help educators assess student comprehension more effectively and deliver personalized feedback at scale (Sunar).

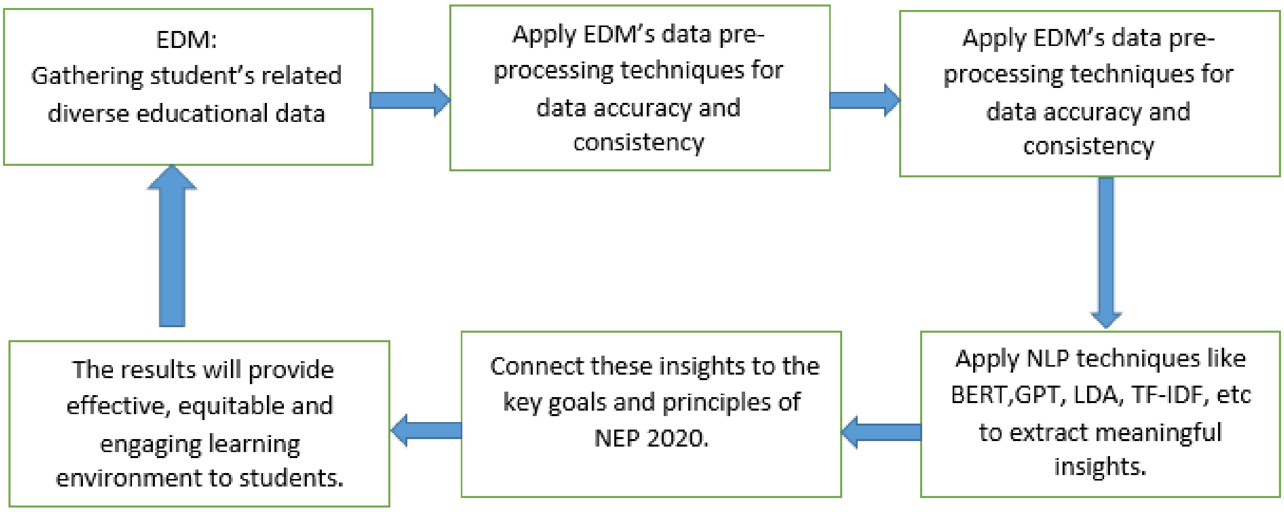
1.3 NEP (The National Education Policy) 2020 and its emphasis on Holistic Assessment:

NEP 2020 marks a ground breaking change in India's education system, advocating a holistic and multidisciplinary approach to learning. A key focus of NEP 2020 is student assessment, promoting a 360-degree evaluation framework that includes self-assessment, peer assessment, and teacher assessment. This comprehensive approach moves beyond traditional exam-based evaluations, offering a more comprehensive assessment of a student's skills and development. Additionally, NEP 2020 emphasizes competency-based education, where outcomes of learning are noticeably defined, and students are assessed on the basis of their skill to demonstrate these competencies.

The policy also highlights the technology's role in enhancing learning outcomes, including the creation of adaptive learning tracks tailored to individual student needs Policy (National Education Policy, n.d.). By integrating technology with innovative assessment methods, NEP 2020 aims to establish a more impactful and equitable education system, aligning seamlessly with the proficiency of Educational Data Mining (EDM) and NLP.

2. METHODOLOGY

This section details the methodology used to assess student performance through NLP methods, aligning with the objectives of the NEP 2020. The methodology involves gathering diverse educational data, preprocessing it to ensure accuracy and consistency, applying NLP techniques to extract meaningful insights, and connecting these insights to the key goals and principles of NEP 2020. This approach provides educators and policymakers with data-driven insights to enhance teaching strategies, personalize learning experiences, and support holistic student development. The process follows a cyclical model of data collection, analysis, interpretation, and implementation, enabling continuous refinement and adaptation to the growing requirements of the education system. Ultimately, this methodology seeks to harness NLP to produce a more effective, equitable, and engaging learning environment consistent with the vision of, NEP 2020 (manager, n.d.). Following cyclic model depicts the same.



3. DATA COLLECTION AND PREPROCESSING

The analysis utilizes a diverse range of educational data sources to acquire in-depth knowledge of student performance. These sources include student-written writing samples and assignments, transcripts of online forum discussions related to course content, teacher feedback, and learning analytics data composed from educational platforms. This data captures both qualitative and quantitative aspects of student learning.

Pre-processing has a key role in in ensuring the quality and relevance of the data for NLP analysis. The process involves several key steps. First, tokenization breaks down the text into parts or tokens. Stemming reduces words to their root form, allowing for the grouping of related terms. Determiner /Stop-word removal sort out common words (e.g., "the," "a," "is") that do not contribute significant meaning. Additionally, methods such as imputation or the removal of irrelevant entries are applied to handle noisy or incomplete data, ensuring data integrity for future analysis (Vidhya, n.d.).

3.1 NLP Techniques for Performance Analysis

Multiple NLP techniques are applied to extract Findings from the pre-processed educational data.

NLP techniques	Use Case	Example
Sentiment Analysis of Student Feedback	NLP can analyse student feedback on courses, teachers, and assessments to gauge satisfaction and areas of improvement.	Sentiment analysis on student reactions to a broad question about their learning experience in a subject can help identify areas needing more attention (LIU.X, 2022)
Automated Essay Scoring & Summarization	NLP can be used for evaluating descriptive answers, writing samples, and assignments automatically, focusing on coherence, grammar, and subject relevance.	A model like **BERT or GPT** can analyze an essay on "Digital Transformation in India" and assign a score based on predefined rubrics (Shermis, 2013).

Keyword Extraction & Topic Modelling in Assignments	NLP can identify key themes in students' written responses to assess topic comprehension.	Using **Latent Dirichlet Allocation (LDA)**, a system can cluster students' answers on a subject like **Artificial Intelligence**, determining if they understand concepts like machine learning and neural networks (Ghosh, 2020).
Multilingual Text Analysis for Vernacular Learning	Since NEP 2020 promotes **regional languages**, NLP can analyze students' responses across different languages for inclusivity.	A system can translate and analyze Hindi, Marathi, and Tamil answers to measure conceptual understanding uniformly.
Chatbots for Personalized Learning & Doubt Resolution	AI-powered chatbots can provide **instant feedback and support** to students based on their learning progress.	A chatbot powered by **transformers (like GPT-4)** can assist students in explaining difficult topics in Mathematics or Science interactively (Kumar, 2022).
Plagiarism Detection & Academic Integrity	NLP can detect copied content from the web or peers in students' submissions.	Using **TF-IDF or cosine similarity**, an NLP-based tool can flag assignments that have high textual overlap with external sources (Kumar, 2022).
Predictive Analytics for Student Dropout & Performance monitoring	NLP can analyze **student engagement in online forums, learning platforms, and assignments** to predict potential dropouts or underperformance.	If a student frequently posts **negative sentiments ("I don't understand", "This is difficult")**, NLP-based models can trigger early intervention by teachers (Kumar, 2022).

3.2 Linking NLP Insights to NEP 2020 Goals

The insights gained from NLP study can be directly aligned with the objectives of NEP 2020. One of the fundamental main beliefs of NEP 2020, personalized learning, can be enhanced through NLP by identifying individual student learning styles and preferences. As an illustration, sentiment analysis can assess students' attitudes toward different learning methods, helping educators design customized learning pathways (Andersen, 2024).

NLP also supports holistic assessment by analysing learners writing samples and forum discussions, offering a extensive insight of their comprehension and strategic thinking abilities beyond traditional assessments. Additionally, NLP can aid in fostering critical thinking by assessing student arguments and reasoning in written work, pinpointing key competencies and developmental needs. By equipping educators with data-driven perceptions into student learning, NLP enables them to alter instruction to individual needs, contributing to a more operational and equitable educational framework that corresponds with the principles of NEP 2020 (Laura K Allen, 2022).

3.3 Analysis and Results

This section analyzes learner's performance using NLP techniques applied to educational data, in association with the objectives of the NEP 2020. The focus is on identifying key performance indicators, assessing student sentiment, and detecting learning patterns through topic modeling.

3.4 Identifying Key Indicators of Student Performance

NLP offers different tools for retrieving important insights into learner's proficiency beyond traditional measures like grades. Writing quality can be assessed using metrics such as grammatical accuracy, vocabulary diversity, and coherence, with NLP techniques like part-of-speech tagging and readability analysis. Critical thinking skills can be evaluated by examining the structure of arguments, identifying logical fallacies, and assessing the depth of reasoning in student writing. Subject matter understanding can be gathered from the correct use of technical terms, the skill to summarize key concepts, and the application of knowledge to new problems. Engagement levels can be measured by analysing participation in online discussions, the frequency of questions, and the sentiment expressed in student communications. These indicators, derived through NLP, offer complete overview of the student abilities, supporting the personalized learning approaches emphasized in NEP 2020 (Kaddari, 2021).

3.4.1 Sentiment Analysis of Student Feedback

Sentiment analysis is essential for understanding student satisfaction and recognizing areas for upgrading in educational programs. By analysing feedback from surveys, course evaluations, and online forums, instructors can derive useful feedback on students' views regarding teaching quality, course content, and the learning environment. Sentiment analysis models can categorize feedback as +ve, -ve, or neutral, offering a quantitative measure of overall satisfaction. However, sentiment analysis in education poses unique challenges. Sarcasm, subjective opinions, and subtle expressions can be difficult for algorithms to interpret accurately. Additionally, the context of the feedback is vital for accurate sentiment classification; a comment that appears negative may actually be constructive criticism intended to enrich the learning journey. To ensure reliable sentiment analysis in academic, it is essential to carefully manage data pre-processing, model selection, and contextual interpretation (Kaddari, 2021).

3.4.2Topic Modeling to Understand Learning Patterns

Topic modelling, especially using Latent Dirichlet Allocation (LDA), can uncover hidden patterns and themes in student writing and discussions, delivering important perspectives into their learning process. By analysing a group of student writing samples, forum posts, or project reports, topic modelling algorithms can detect clusters of related words that represent underlying topics or themes. These topics can highlight student interests, areas of confusion, and preferred learning styles. For instance, topic modelling might reveal a strong interest in environmental sustainability within a group of students or identify a widespread misconception about a specific concept. Faculty can harness these insights to adjust teaching methods, personalize learning materials, and create more engaging educational experiences. Furthermore, topic modelling can be used to detect students struggling with particular concepts, enabling targeted interventions and support (Author: Linda Avila, 2025).

4. DISCUSSION

This section explores the practical implications of using NLP for analyzing student performance, especially with regard to the National Education Policy (NEP) 2020. By examining student text data, such as writing samples and open-ended responses, NLP provides a detailed understanding of their comprehension, critical thinking, and communication skills – aspects often overlooked by traditional assessment methods. Through NLP, instructors can gain actionable insights to adjust their pedagogical techniques and offer targeted support to learners. Moreover, the use of NLP aligns with NEP 2020's focus on holistic and personalized learning, promoting a more student-centered educational environment. The following subsections will examine the benefits, challenges, and future prospects of this transformative approach (Kaddari, 2021).

4.1 Benefits of NLP-Driven Performance Analysis

NLP-driven performance analysis offers several significant benefits in education. First, it enables personalized feedback by automatically pinpointing areas of student proficiency or face challenges, permitting professors to offer customized guidance. This is in contrast to generic feedback, which may not effectively address individual learning needs. Second, NLP can detect struggling students early by recognizing patterns in their writing that suggest difficulties with specific concepts or skills. This initial detection allows for timely interventions, helping students stay on track. Third, analyzing student writing using NLP delivering important perspectives into the effectiveness of teaching strategies. Professors can detect domains where students frequently struggle, indicating a need for adjustments in instructional methods. Lastly, NLP encourages data-driven decision-making in education by offering objective, quantifiable measures of student performance. This data can guide curriculum development, resource allocation, and overall educational planning, contributing to greater and equitable educational outcomes (Author: Linda Avila, 2025).

5. CHALLENGES AND LIMITATIONS

Despite its potential, the use of NLP in education presents several challenges and limitations. One major concern is data privacy, as student writing often contains sensitive personal information. Strong measures must be implemented to maintain the privacy of students and ensure compliance with data protection regulations. Another significant issue is algorithmic bias, as NLP models can reinforce and amplify partialities present in the training data, possibly resulting in biased or unequal treatment of specific student populations. Human interpretation remains crucial, as NLP models generate insights solely from the data used in their training, and educators should apply their critical thinking and expertise to interpret these insights in context. Additionally, capturing subtle aspects of student learning, such as creativity and critical thinking, can be challenging for NLP models, as these potentials are often articulated in complex ways that are challenging to quantify automatically (Yan, 2023).

6. CONCLUSION

This exploration of applying NLP to learners performance analysis, particularly within the outline of India's National Education Policy (NEP) 2020, highlights its transformative potential. NLP techniques offer valuable tools for extracting meaningful insights from educational data, enabling a more profound insight into the student learning patterns and challenges. By analyzing textual data such as writing samples , feedback, and classroom discussions, NLP can provide personalized feedback, identify students at risk, and customize educational interventions. The integration of NLP aligns directly with NEP 2020's focus on personalized learning and holistic development, paving the way for a more equitable and effective education system. Ultimately, embracing NLP in education can enhance student outcomes and create a more enriching learning experience for all.

Further research ought to emphasize on developing more innovative NLP models that can more precisely convey the complexities of student learning. This includes incorporating contextual factors, such as student background and learning history, into these models. Combining NLP with other educational technologies, like learning management systems (LMS), can offer seamless, Individualized instruction for students. It is also essential to explore the ethical implications of using AI in education to ensure these technologies are used responsibly and equitably, addressing concerns like bias, fairness, and transparency. Lastly, developing NLP-powered tools for personalized learning, such as automated essay scoring systems and intelligent tutoring systems, can encourage learners to take ownership of their education and maximize their growth.

REFERENCES

[1] Alshareef, F. (2020). Educational Data Mining Applications and Techniques. *International Journal of Advanced Computer Science and Applications(IJACSA)*.

[2] Andersen, G. (2024). *The Role of Natural Language Processing in Improving Student Feedback and Evaluation Systems.* Retrieved from Molsstud: https://moldstud.com/articles/p-the-role-of-natural-language-processing-in-improving-student-feedback-and-evaluation-systems

[3] Author: Linda Avila, D. S. (2025). Using NLP to Evaluate the Effectiveness of Teaching Methods Date. *Data to insights*.

[4] Ghosh, S. S. (2020). Using BERT for Automated Assessment of Student Writing in Indian Education System.

[5] Kaddari, Z. (2021). *Natural Language Processing: Challenges and Future Directions.*

[6] Kumar, R. &. (2022). AI-Powered Chatbots in Education: A Study on Student Interaction and Learning Enhancement. *IEEE international conference*.

[7] Laura K Allen, S. C. (2022). Chapter 5: Natural Language Processing as a Tool for Learning Analytics -Towards a Multi-Dimensional View of the Learning Process. In L. K. Allen, *Handbook of Learning Analytics.*

[8] LIU.X. (2022). Sentiment Analysis in Student Feedback for Learning Enhancement. *Journal of learning analytics*.

[9] Lumenalta. (2025, Jan 27). *Natural Language processing*. Retrieved from 5 examples of natural language processing in education.

[10] manager, C. c. (n.d.). *Holistic Progress Report Card as per NEP: A Comprehensive Shift in Student Evaluation*. Retrieved from https://www.cyberschoolmanager.com/blog/holistic-progress-report-card-as-per-nep-a-comprehensive-shift-in-student-evaluation/

[11] *National Education Policy*. (n.d.). Retrieved from https://dsel.education.gov.in/search?q=national+education+policy#gsc.tab=0&gsc.q=national%20education%20policy&gsc.page=1

[12] Shermis, M. D. (2013). *Handbook of Automated Essay Scoring: Theory, Practice, and Applications.* Routledge.

[13] Society, I. E. (n.d.). *educational data mining*. Retrieved from https://educationaldatamining.org/?utm_source=chatgpt.com

[14] Sunar, A. S. (n.d.). Natural Language Processing of Student's Feedback to Instructors: A Systematic Review. *IEEE Transactions on Learning Technologies PP(99):1-14*.

[15] Thomas, J. (2025). Assessing Student Engagement and Performance Through NLP-Driven Text Analysis Author.

[16] Vidhya, a. (n.d.). *Text Preprocessing in NLP with Python Codes*. Retrieved from https://www.analyticsvidhya.com/blog/2021/06/text-preprocessing-in-nlp-with-python-codes/

[17] Yan, L. (2023). Practical and ethical challenges of large language models in education: A systematic scoping review. *British Journal of educational technology*.