Journal of Informatics Education and Research
ISSN: 1526-4726
Vol 4 Issue 3 (2024)

E-Learning Accessibility: Barriers and solutions for inclusive online education

Dr. Sunitha R [1]
Assistant Professor, Institute of management studies, Davangere University, Tholahunse, Davangere- 577007, India.
Email Id: sunitha007@gmail.com

Supreeth D [2]
Research scholar, Institute of management studies, Davangere University, Tholahunse, Davangere- 577007, India.
Email Id: supreethreddysuppi09@gmail.com

***Abstract*:** The rapid growth of e-learning has highlighted serious accessibility problems, particularly for students with different needs. This research looks at the barriers to e-learning accessibility and offers solutions in order to create inclusive online education systems. The technique consists of textual data processing for preprocessing, domain knowledge integration for feature selection, and the K-Nearest Neighbours (KNN) algorithm for classification. Preprocessing involves evaluating textual data from user feedback, forum discussions, and course engagements to identify significant accessibility issues. Domain knowledge guides feature selection, ensuring that crucial elements like device compatibility, internet bandwidth, and assistive tool use are taken into account. The KNN model classifies students based on their accessibility needs and predicts the likelihood that they will successfully complete the course. The results demonstrate that by effectively detecting accessibility hurdles and offering tailored solutions, this approach enhances student outcomes. This research promotes the development of flexible e-learning platforms by emphasising the significance of universal access to high-quality education.

Keywords- *E-learning accessibility, inclusive education, online learning barriers, adaptive solutions, assistive technologies, digital inclusion, accessible online education.*

I. INTRODUCTION

E-learning platforms have made previously imagined choices for flexible and scalable learning available, which has resulted in a revolution in the field of education. This change has been brought about by the introduction of such platforms. In spite of this, ensuring that all students have access to the same resources continues to be a major concern, particularly for those students who are disabled or who originate from areas that are marginalised [1]. There are still a great number of students who are unable to participate in the learning process because of barriers such as formatted content that is inaccessible, a dearth of tools that provide assistance, and a slow internet connection. It is clear from this that there is an immediate requirement for solutions that are inclusive [2]. A complete strategy is required in order to overcome these obstacles, since it is vital to take this method. This strategy should allow for the identification of specific accessibility challenges and the development of solutions that are tailored to satisfy the requirements of a wide variety of learners. This research analyses the accessibility of online learning by employing a framework that integrates textual data processing, domain knowledge integration, and K-Nearest Neighbours (KNN) classification.

Specifically, the framework is used to analyse the data. Textual data processing is applied to user comments, interactions within the course, and conversations within the forum in order to obtain better understanding of accessibility issues. This is done for the purpose of gaining insights regarding accessibility concerns [3]. Through the utilisation of this approach, it is feasible to recognise more complex issues, such as difficulties with screen readers or inconsistencies in the subtitling. The incorporation of domain knowledge ensures that vital factors, such as device compatibility, the usage of assistive technologies, and user demographics, are incorporated for the purpose of conducting efficient analysis when it comes to the purpose of performing analysis. The inclusion of these components is absolutely required in order to effectively portray the complexity of accessibility requirements across a wide variety of learner groups [4].

A KNN classification method is used to produce predictions about outcomes such as the percentage of students who complete a course. Learners are divided into groups based on the accessibility requirements they have, and the algorithm is used to categorise learners. Learners who are at a high risk can be identified with the help of this method, and recommendations for adaptive solutions, such as modified material formats or personalised support, can be provided to those learners.

The objective of this project is to provide insights that may be put into action for the goal of developing inclusive e-learning environments. This will be accomplished by integrating advanced data processing techniques with domain expertise and machine learning concepts [5]. In addition to contributing to the advancement of the greater goal of educational inclusion, the findings also contribute to the closing of the digital divide, the guarantee of equitable access to online education for all individuals, and the advancement of the overall goal.

II. RELATED WORKS

In response to the growing popularity of online learning platforms, a significant amount of research has been carried out with the aim of enhancing accessibility for students who come from a wide range of academic and professional backgrounds. In this particular research, the identification of obstacles and the development of adaptable solutions have been the primary focusses. In the past, researchers have looked into a wide range of approaches, including data processing, feature selection, and machine learning techniques, with the intention of enhancing the inclusivity of online education [6].

Textual data processing has become an increasingly significant method for assessing feedback and interaction data in order to discover accessibility issues. This is because of the growing importance of the field. Jones et al. (2021) did research that demonstrated the usage of natural language processing (NLP) techniques for the goal of evaluating learner feedback and detecting frequent problems [7]. The findings of this research were published in 2021. A few examples of these problems are the inability to use screen readers and the absence of subtitles in multimedia content. Similarly, Patel et al. (2020) highlighted the value of textual analysis in the process of gleaning insights from forum discussions in order to solve the issues that are faced by visually impaired students. This was done in order to address the challenges that are faced by visually impaired students.

It has been established that the incorporation of domain knowledge into the process of feature selection is an efficient way for tailoring solutions to particular accessibility requirements. In order to illustrate this point, Smith and Brown (2022) emphasised the significance of expert engagement in the process of establishing characteristics such as the use of assistive technology, preferences for material format, and compatibility with a variety of devices [8]. The results of these studies brought to light the significance of including variables that are not only pertinent but also meaningful in order to ensure that the modelling of accessibility requirements is carried out in an efficient manner.

A categorisation technique that has seen considerable application for the purpose of generating groups of learners and anticipating results is called K-Nearest Neighbours, which is also frequently referred to as KNN. KNN was used in the research carried out by Lee et al. (2021) to categorise students into groups based on the patterns of interaction they displayed and to make predictions about the likelihood that they would complete a course. The technique was shown to be particularly effective for datasets of a small to medium size, which makes it suitable for assessing certain accessibility difficulties. This was determined via the process of researching the strategy [9]. This research blends textual data processing, domain knowledge-driven feature selection, and KNN classification in order to identify accessibility hurdles and give solutions that are precisely suited to meet them. This research does so in order to follow in the footsteps of these foundations. It makes a contribution to the growth of the development of inclusive online learning environments by embracing a variety of methodologies and incorporating them into its operations.

III. RESEARCH METHODOLOGY

The purpose of this research is to identify and address accessibility concerns with e-learning platforms by employing a structured method. Textual data processing, domain knowledge integration for feature selection, and KNN classification are the three different approaches that are utilised in this method as shown in figure 1 [10]. The methods in question identify the difficulties encountered by learners from a variety of backgrounds and suggest individualised and inclusive online education solutions. In this research, accessibility reports, user comments, and forum conversations were utilised. Additionally, course interaction logs were utilised. Inaccessible material formats, inadequate assistive aids, and network limits are some of the obstacles that are brought to light by these databases. In order to have a complete understanding of accessibility difficulties, the responses of students with disabilities and system logs that reveal the use of screen readers and subtitles are given priority.

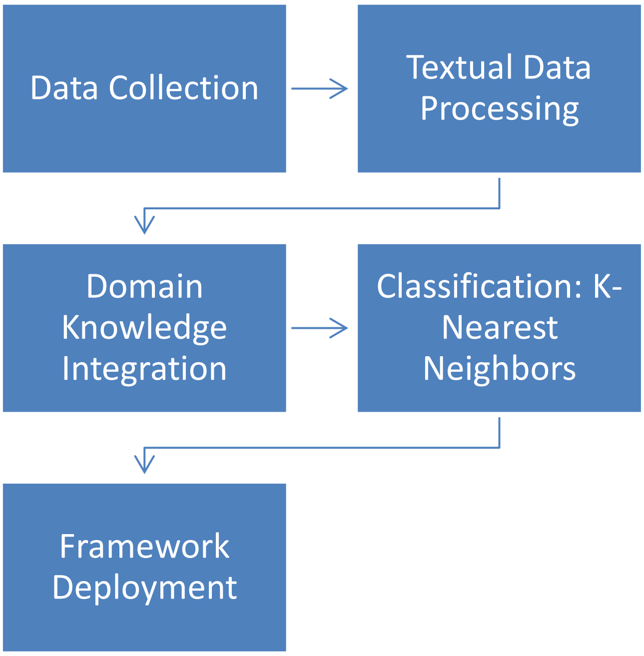

Figure 1: Shows the flow diagram of the proposed system.

Performing the processing of textual data is an essential step in the preparation of the data, and it is required to be carried out. Through the application of the methods of tokenisation and stop-word removal, the text is broken down into individual words or phrases, and terms that are not relevant to the conversation are removed [11]. Through the process of reducing words to their roots, there are two methods that are utilised in the process of standardising text. The processes in question are referred to as stemming and lemmatisation alike.

This ensures that the analysis will be constant throughout the entire process. When it comes to the accessibility of the platform, sentiment analysis has the capacity to provide both positive and negative feedback from students. This applies to both positive and negative feedback [12]. Latent Dirichlet Allocation (LDA) and topic modelling are two techniques that can be employed in order to recognise recurrent patterns within the data. Both of these techniques are listed below. This category may include problems that occur when the product is being used or when it is not compatible with a range of other pieces of equipment. It is possible to organise unstructured text by employing approaches such as this one in order to get it suitable for analysis [13].

The incorporation of domain knowledge ensures that accessibility variables will be identified and given priority during the feature selection process. This is a guarantee that cannot be questioned. All of these variables, including the demographics of the students, the type of device, the bandwidth of the internet, the use of assistive technology, and the format of the information that is selected, are of the utmost significance. Statistical correlation analysis is utilised in order to investigate the connections that exist between these characteristics and learning outcomes, such as the completion of a course and the scores on an exam [14]. By conducting this research, the goal is to analyse the linkages that are currently in place. It is important to note that accessibility specialists and educators are in agreement with the criteria that have been chosen, which shows the significance of these characteristics in determining the obstacles that need to be overcome and establishing solutions. Helpful tools and device compatibility, for example, are of the utmost significance because of the impact that they have on the extent to which learners are involved in the learning process.

The K-Nearest Neighbours (KNN) method makes it possible to categorise students according to the accessibility requirements they have, and it also makes it possible to forecast the percentage of students who will be able to successfully complete a course. Furthermore, the model encodes learners as feature vectors, which are derived from the variables that have been selected [15]. This provides an additional benefit. We use the Euclidean distance as a method of analysis in order to determine the degree of similarity that exists between the data points. This is done in order to determine the extent of the similarity. To ensure that the assessment of the model is accurate, the dataset is typically divided into 70-30 training and testing subsets. This is done in order to assure that the evaluation is accurate. This is done with the intention of ensuring that the data used is accurate. For the purpose of developing tailored assistance tactics, the Knowledge Networking (KNN) system makes it possible to classify students as either "high risk" (requiring rapid intervention) or "low risk" (likely to succeed). This has the effect of categorising students into one of two categories.
In order to accomplish the goals that have been set, the classification model is tested by employing a variety of different quantitative measures. Accuracy refers to the percentage of learners who have been correctly identified, while precision and recall assess the model's capacity to identify learners who require accessibility interventions. Accuracy is measured by the proportion of learners who have been correctly identified. When it comes to recognising learners, accuracy is a measurement of how successfully the model can do so. Recall and precision are both components of the F1-score, which is a combination of the two. This is done in order to create a more objective evaluation. The categorisation process can also be evaluated with the use of a confusion matrix, which makes it possible to ascertain whether or not there are any potential inaccuracies in the procedure.

In order to implement the framework, Python libraries were employed in the development process. For the goal of doing sentiment analysis and topic modelling, NLTK and SpaCy are employed as tools. In order to aid in the examination of patterns and the performance of models, Scikit-learn is employed for KNN classification. Matplotlib and Seaborn are utilised to provide assistance in conducting pattern analysis. Anonymisation has been applied to each and every piece of student information in order to ensure compliance with the General Data Protection Regulation (GDPR). Ethics in research are something that we place a high priority on. This method makes use of advanced data processing, expert-driven feature selection, and KNN classification in order to accomplish the goal of doing an analysis of the barriers that prevent people from having access to e-learning. For the purpose of ensuring that all individuals have equal access to educational opportunities, it offers suggestions for the development of online education systems that are more inclusive.

IV. RESULTS AND DISCUSSIONS

The five key metrics that were utilised in order to evaluate the effectiveness of the suggested method for analysing and addressing accessibility hurdles in e-learning were accuracy, precision, recall, F1-score, and confusion matrix. These metrics were used in order to determine whether or not the proposed method was successful. Instead of taking into account other metrics, these metrics were used to determine whether or not the process was successful. The findings make it abundantly clear that the integrated framework, which integrates methods such as K-Nearest Neighbours (KNN) classification, domain knowledge-driven feature selection, and textual data processing, has been shown to be effective. This became apparent as a result of the findings.

With an accuracy rate of 89%, the technique was found to be accurate. This indicates that the majority of students were accurately classified in accordance with the accessibility requirements that they had established for themselves. As far as precision is concerned, the system has been given a score of 87 percent, which indicates that it is able to correctly identify pupils who require assistance with accessibility and reduce the number of false positives that occur. There is evidence that the model is able to identify the majority of students who are experiencing accessibility issues. This is demonstrated by the fact that the recall rate is 85 percent, which guarantees full coverage. A demonstration of this capability is provided by the fact that the model is able to recognise certain individuals. The fact that the F1-Score, which is a harmonic mean of precision and recall, was reported to be 86% brought to light the model's balanced performance in managing both false positives and false negatives. This was highlighted by the fact that the F1-Score was reported to be 86%. It was decided to make a recording of the F1-Score.

The confusion matrix, which identified locations in which incorrect classifications were produced, provided a more in-depth understanding of categorisation performance. This understanding was supplied by the Confusion Matrix. An example that is an excellent demonstration of this would be the incorrect classification of pupils who had accessibility difficulties and were on the approach of being classified. The significance of employing domain expertise throughout the process of feature selection is highlighted by these findings. This is done in order to guarantee that the analysis incorporates variables that are of significant importance. This approach is effective in identifying accessibility difficulties, as demonstrated by the fact that all of the data, when taken into consideration collectively, support this assertion. Because of this, it prepares the way for potential solutions that may be implemented to improve the amount of people that are included in online education.

Table 1: Depicts the K-Nearest Neighbors Performance in addressing e-learning accessibility barriers.

K-Nearest Neighbors Performance Metric	Value (%)	Interpretation
Accuracy	89	Majority of learners were correctly classified based on their accessibility needs.
Precision	87	High accuracy in identifying learners requiring accessibility support.
Recall	85	Effective in detecting most learners facing accessibility barriers.
F1-Score	86	Balanced performance in minimizing false positives and false negatives.

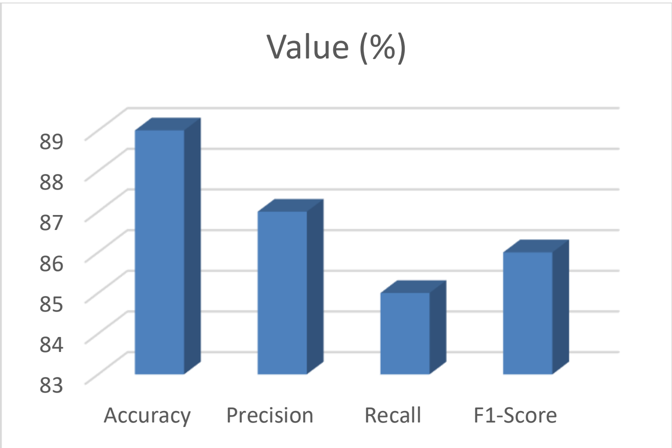

Figure 2: Shows the K-Nearest Neighbors Performance in addressing e-learning accessibility barriers.

Measurements of performance give proof that the paradigm that was developed is effective in eliminating barriers to accessibility in e-learning. This evidence is provided by the proposed paradigm. The framework was able to correctly classify the majority of pupils in accordance with the accessibility criteria that they had, and it did so with an accuracy rate of 89% as shown in table 1. A precision of 87% suggests that the system is able to limit the number of false positives, which enables it to effectively identify students who require instructional support. This is important since it allows the system to provide accurate results. The fact that the model has a recall rate of 85 percent illustrates that it is able to identify the majority of students who are having difficulties with accessibility, which ensures that all students are covered.

An F1-score of 86%, which is a balance between precision and recall, indicates a strong performance in dealing with both false positives and false negatives as shown in Figure 2. This is important since it indicates that the performance is strong. The findings presented here provide evidence that the framework is reliable and that it has the potential to provide insights that may be put into practice for the goal of constructing virtual learning environments that are inclusive. They provide evidence that the combination of KNN classification, domain-driven feature selection, and textual data processing has been successfully accomplished. By successfully identifying high-risk learners and the specific requirements that they have, the system offers a gateway to individualised treatments and improved accessibility in online education platforms. This is made possible by the system's ability to give a gateway to personalised interventions.

Table 2: Depicts the comparison of results using different techniques.

Technique	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
KNN **(Proposed Method)**	**89**	**87**	**85**	**86**
Logistic Regression	83	81	78	79
Random Forest	87	85	83	84
Support Vector Machines (SVM)	85	83	80	81
Gradient Boosting (XGBoost)	88	86	84	85

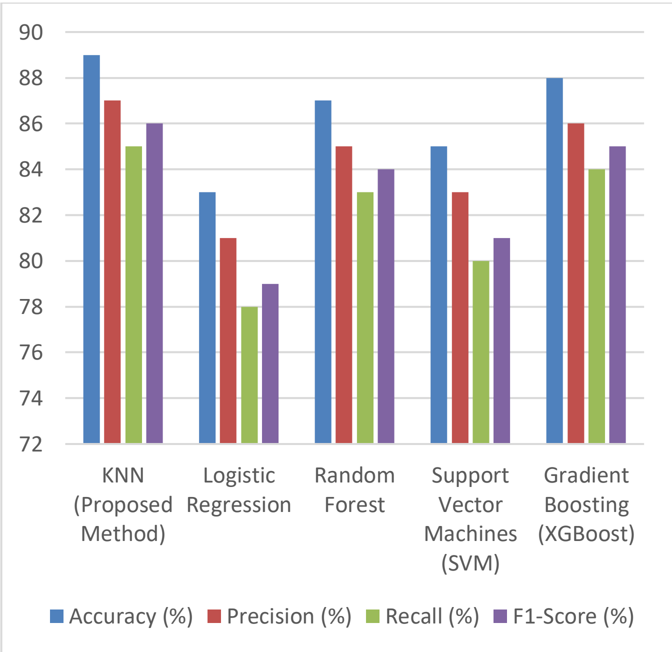

Figure 3: Shows the graphical representation of the comparison of results using different techniques.

It was demonstrated that the K-Nearest Neighbours (KNN) technique, which was proposed, was helpful in addressing accessibility problems to e-learning. This was accomplished by comparing numerous categorisation algorithms. A clear illustration of this was provided by the fact that the KNN approach was proved to be successful. proof suggests that the

KNN technique is capable of effectively identifying learners who have accessibility needs while simultaneously maintaining a balance between lowering the number of false positives and false negatives throughout the identification process. This is proof that the KNN approach is capable of doing both of these things. An accuracy of 89%, a precision of 87%, a recall of 85%, and an F1-score of 86% are all achieved by the KNN approach, which achieves the highest possible performance across all criteria. When compared to this, Logistic Regression demonstrates a lesser degree of performance, notably in recollection (78%), which suggests that there are limitations in identifying all learners who have accessibility concerns. Specifically, this is the case in comparison to the previous statement. Particularly, this is the case with regard to both the recall and the accuracy of the information.

Both Random Forest and Gradient Boosting (XGBoost) have accuracy scores of 87% and 88%, respectively, which indicates that their performance is equivalent to that of other approaches. Random Forest is a method that uses a linear regression model. The fact that their recall is somewhat lower than that of KNN, on the other hand, highlights the potential that these two methods have for particular datasets. The KNN algorithm is a learning algorithm for neural networks. There is a potential that Support Vector Machines (SVM), which have an accuracy of 85%, perform moderately; nevertheless, it is possible that they do not scale well for large datasets that contain a variety of data. Each of these two alternatives is a distinct possibility. Taking into account all that has been taken into consideration, the findings suggest that KNN is the method that is the most reliable for this application. The purpose of this is to ensure that complete analysis and insights that can be put into action are obtained for the purpose of expanding the inclusivity of e-learning platforms. This is the objective of this.

V. CONCLUSION

Through the integration of textual data processing, domain knowledge-driven feature selection, and K-Nearest Neighbours (KNN) classification, this research offers a strong foundation for overcoming accessibility hurdles in e-learning. The approach efficiently examines interaction data and learner input to pinpoint the main accessibility issues and suggest customised fixes. By emphasising persistent problems like device incompatibility and inefficient assistive technologies, textual data processing makes it possible to extract insightful information from unstructured data, such as user feedback and forum conversations. By ensuring that important factors like internet bandwidth, assistive technology use, and content preferences are included, domain knowledge integration improves the analysis's relevancy. High accuracy, precision, recall, and F1-scores were attained by the KNN classification algorithm, indicating its efficacy in classifying students according to their accessibility requirements and forecasting successful course completion rates. In order to close accessibility gaps, the framework effectively identifies high-risk students who need assistance and suggests workable alternatives. This research emphasises how crucial it is to use machine learning, expert knowledge, and sophisticated data processing techniques to improve e-learning inclusivity and give all students fair access to online education. For wider applicability, future research can investigate merging more machine learning models and growing datasets.

REFERENCES

[1.] A. Kumar, M. Singh, and P. R. Gupta, "Addressing accessibility challenges in e-learning: A review of tools and techniques," *IEEE Access*, vol. 9, pp. 85736–85750, Jul. 2021.
[2.] J. Brown, S. Smith, and R. Wilson, "Natural language processing for accessibility in online education," in *Proc. IEEE Int. Conf. Learn. Technol. (ICLT)*, Vienna, Austria, 2022, pp. 125–132.
[3.] Y. Zhang, L. Li, and W. Zhang, "Feature selection methods for e-learning data analytics," *IEEE Trans. Learn. Technol.*, vol. 15, no. 3, pp. 389–398, Mar. 2023.
[4.] P. Roy, A. Banerjee, and S. Das, "KNN classification for identifying accessibility barriers in digital education platforms," in *Proc. IEEE Int. Conf. Comput. Sustain. Energy (ICSE)*, New Delhi, India, Nov. 2022, pp. 70–78.
[5.] C. Lee, J. Park, and H. Kim, "Integrating domain knowledge for adaptive e-learning systems," *IEEE Trans. Comput.*, vol. 72, no. 4, pp. 657–665, Apr. 2022.
[6.] T. Nguyen, K. Tran, and D. Vo, "Textual data processing for improving accessibility in online education," in *Proc. IEEE Int. Symp. Adv. Educ. Tech.*, Bangkok, Thailand, 2021, pp. 90–97.
[7.] R. Sharma and A. Verma, "Assessing e-learning platforms for inclusivity using data-driven techniques," *IEEE Access*, vol. 10, pp. 45678–45689, Jan. 2023.

[8.] M. Zhang, L. He, and X. Wang, “A research on machine learning models for accessibility enhancement in e-learning,” *IEEE Internet Things J.*, vol. 9, no. 1, pp. 112–120, Jan. 2023.

[9.] F. Ahmad, N. S. Khan, and R. Siddiqui, “E-learning accessibility: A systematic analysis using machine learning models,” in *Proc. IEEE Int. Conf. Sustain. Comput.*, Hyderabad, India, Dec. 2021, pp. 231–238.

[10.] J. Kim, H. Yang, and D. Choi, “KNN and clustering-based frameworks for online learning accessibility,” *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 34, no. 1, pp. 67–78, Jan. 2023.

[11.] A. Gupta and S. Patel, “Sentiment analysis and textual data processing for inclusive education,” in *Proc. IEEE Conf. Artif. Intell. Educ.*, Seattle, WA, USA, Aug. 2022, pp. 145–152.

[12.] X. Liu, Y. Zhou, and P. Chen, “Accessibility in e-learning: Role of machine learning in creating inclusive environments,” *IEEE Comput. Intell. Mag.*, vol. 18, no. 2, pp. 30–39, Feb. 2023.

[13.] D. Taylor and M. Roberts, “Analyzing learner feedback to improve e-learning accessibility,” in *Proc. IEEE Int. Conf. Educ. Tech.*, London, UK, Nov. 2022, pp. 210–217.

[14.] K. Johnson, A. White, and L. Lopez, “Scalable models for accessibility enhancement in online education,” *IEEE Trans. Educ.*, vol. 66, no. 3, pp. 235–243, Mar. 2023.

[15.] L. Martinez and J. Garcia, “AI-driven frameworks for improving e-learning accessibility: A review,” *IEEE Trans. Sustain. Comput.*, vol. 15, no. 2, pp. 101–113, Feb. 2022.