



A Systematic Review on Digital Literacy using Machine Learning Approaches

Kishore Babu Kalapala¹, Dr. Priti Shaileshbhai Patel²

¹Research scholar, Sarvajani University, Assistant Professor, Metas Adventist College, Surat, India.

²Assistant Professor, Shree Ramkrishna Institute of Computer Education and Applied Sciences, Surat, India.

Email ID: kishoreadventist@gmail.com¹, priti.patel@srki.ac.in²

Abstract

As digital literacy becomes a foundational skill in the digital era, and digital skills become essential in today's world, measuring and understanding digital literacy is more important than ever. This review explores how Machine Learning (ML) is being used to assess digital literacy, moving beyond traditional methods to create more personalized and insightful evaluations. We examined a range of studies that use ML techniques—like decision trees, neural networks, and support vector machines—to analyse digital skills in different educational settings. By identifying where individuals excel and where they may need support, these ML-driven approaches provide a customized view of digital literacy. Our findings show that ML models can not only improve the accuracy of digital literacy assessments but also open up possibilities for adaptive learning experiences, such as intelligent tutoring systems and personalized recommendations. While these approaches offer significant benefits, we also found challenges, particularly around integrating ML into diverse academic fields. This growing intersection between ML and digital literacy research provides new pathways for educators, policymakers, and researchers to promote and support digital skills. Looking ahead, there is a strong need to continue refining ML tools to make digital literacy assessment more inclusive, accurate, and adaptable across different learning contexts.

Keywords: Digital Literacy, Digital Education, Digital Skills, Digital Technologies, Digital Competencies, Media Literacy, Information and Communication Technology (ICT) literacy, 21st Century Digital Skills, Machine Learning, Predictions, Student Knowledge Assessment, Learning outcomes

1. Introduction

As the world becomes increasingly digital, understanding and measuring digital literacy is more important than ever. Recent studies have begun exploring how Machine Learning (ML) techniques can be applied to evaluate digital literacy, offering new insights into how individuals engage with technology and acquire essential digital skills. Rather than relying on traditional assessment methods, these studies use advanced algorithms to analyze patterns in behavior, knowledge, and performance, providing a more refined picture of digital literacy. [1] [2] The research delves into the application of ML techniques such as decision tree classification, neural networks and support vector machines to assess and categorize individuals based on their digital competency levels. By analyzing data from various digital interactions, these models can identify strengths, gaps, and areas where individuals may need additional training or support. This data-driven approach has the potential

to transform how digital literacy is understood and assessed, leading to more personalized and effective learning experiences. [1] [2] These selected papers contribute to the growing field of research at the intersection of Machine Learning (ML) and digital literacy, emphasizing the importance of using technology not only to enhance education but to better understand the digital skills that are essential in today's world. By leveraging Machine Learning, educators and researchers can gain deeper insights into how digital literacy develops, paving the way for more informed strategies to promote these crucial skills. [1]

2. Digital Literacy

Digital literacy is a person's ability to use information and communication technology effectively to access, evaluate, and process information in various digital forms. (Manubey et al., 2022). Digital literacy includes a range of skills that are important for living

in today's digital world. Digital literacy is more than just knowing how to use software or devices; it involves finding, evaluating, creating, and sharing information in different digital formats. These skills are essential for everyone, whether you're a student or a professional, to navigate the digital world effectively and responsibly. It encompasses skills such as understanding how to use digital tools, assessing the credibility of online information, and creating content. [8] The Evolution of Digital Literacy Over Time Digital literacy has grown and evolved alongside the digital age, developing through three key stages:

2.1. Stage 1: The 1990s to Early 2000s

Digital literacy first emerged during this period, focusing mainly on understanding, using, evaluating, and integrating information from digital sources like computers. While technical skills were important, the emphasis was on critical thinking—evaluating and applying digital content thoughtfully was at the heart of digital literacy. [3]

2.2. Stage 2: Early 2000s to 2010s

During this stage, digital literacy became a topic of national interest. In 2006, the European Community's DigEuLit project defined it as a combination of awareness, attitude, and ability to use digital tools in different aspects of life. The focus expanded to not just using digital tools, but also creating new knowledge, producing digital content, and communicating for social progress. [4] Key features of this stage include

- Using digital tools appropriately beyond just computers.
- Recognizing users as creators of digital content, not just consumers.
- Viewing digital literacy as a driver for social and innovative development.

2.3. Stage 3: Post-2010s

As we entered the second decade of the century, technologies like AI, big data, the Internet of Things, and cloud computing rapidly advanced. Digital literacy gained even more importance globally. In 2013, the American Library Association defined a digitally literate person as someone who can discover, understand, create, and communicate

digital information across various formats, effectively manage privacy, and use technology to collaborate and participate in society. [5]. The 2019 DQ Global Standards Report highlighted the idea that digital literacy prepares people to become informed, capable citizens, equipped to use, create, and control technology for human progress. [6] At this stage, digital literacy is characterized by

- A holistic view, linking personal growth with social development.
- A focus on ethics, morality, and legal considerations.
- A global outlook, encouraging a respectful and open-minded approach to the world.

3. Machine Learning

In Machine Learning by Tom M. Mitchell, the concept of machine learning and its types are fundamental to the field. Mitchell defines machine learning as: "A computer program is said to learn from experience E with respect to some class of tasks T and performance measure P , if its performance at tasks in T , as measured by P , improves with experience E . "In core, machine learning is the study of algorithms that improve their performance on tasks by learning from data (experience) without being explicitly programmed.

3.1. Types of Machine Learning

Mitchell outlines several main categories of machine learning, each defined by the nature of the learning task and the type of feedback received from the environment or data.

- **Supervised Learning:** In supervised learning, the algorithm is trained on labeled data. Each training has an input and the corresponding output (label), and the goal is to learn a function that maps inputs to outputs. The algorithm uses this function to make predictions for new, unseen data.
- **Unsupervised Learning:** In unsupervised learning, the algorithm is provided with data that lacks explicit labels or outputs. The goal is to discover underlying patterns or structures within the data, such as clustering or associations.
- **Semi-Supervised Learning:** Semi-supervised learning combines both supervised and

unsupervised learning. The algorithm is trained on a small amount of labeled data and a larger amount of unlabeled data. The goal is to improve learning performance by leveraging both types of data.

- **Reinforcement Learning:** In reinforcement learning, an agent learns to make decisions by interacting with its environment and receiving feedback in the form of rewards or penalties. The algorithm's goal is to maximize cumulative rewards by learning an optimal policy for decision-making.
- **Self-Supervised Learning:** In self-supervised learning, a system generates its own labels from unlabeled data to create a supervised learning task. The method often involves predicting part of the data from another part, allowing the model to learn useful representations without manual labeling.

3.2. Machine Learning Algorithms

In the context of machine learning, algorithms are the core methods that enable computers to learn from data and make predictions or decisions. Depending on the type of learning (supervised, unsupervised, reinforcement, etc.) and the nature of the problem, different algorithms can be applied.

3.2.1. Supervised Learning Algorithms

These algorithms are trained on labeled data, meaning that for each input, the output (target or label) is known.

a) Linear Algorithms

- **Linear Regression:** Used for predicting continuous values. It models the relationship between input features and output using a straight line.
- **Logistic Regression:** Used for binary classification problems. It estimates probabilities using a logistic function to model the relationship between input and output.

b) Tree-Based Algorithms:

- **Decision Tree:** A tree-like structure where internal nodes represent tests on features, and leaves represent class labels or continuous values.
- **Random Forest:** An ensemble of decision trees

that improves accuracy by averaging or voting on the predictions of many trees.

- **Gradient Boosting Machines (GBM):** A boosting method that builds models sequentially by correcting the errors of the previous models.

c) Support Vector Machines (SVM)

- **SVMs** find the hyperplane that separates the data into classes. It works well for both linear and non-linear classification problems using a kernel trick.

d) Neural Networks

- **Feedforward Neural Networks:** A multi-layered network of neurons that learn to map inputs to outputs. They can model complex, non-linear relationships.
- **Convolutional Neural Networks (CNNs):** A specialized neural network designed for image and video data. CNNs use convolutional layers to capture spatial patterns.
- **Recurrent Neural Networks (RNNs):** A neural network type designed for sequential data, where the output depends on previous inputs, making it useful for time series and natural language processing.

3.2.2. Unsupervised Learning Algorithms

- These algorithms work with unlabeled data, aiming to find patterns, structures, or associations within the data. [16]

a) Clustering Algorithms

- **k-Means Clustering:** Partitions the data into k clusters, where each point belongs to the cluster with the nearest mean.

b) Dimensionality Reduction Algorithms

- **Principal Component Analysis (PCA):** Reduces the number of features in a dataset while retaining as much variance as possible by projecting data into a lower-dimensional space.
- **t-SNE (t-distributed Stochastic Neighbor Embedding):** A visualization technique for reducing high-dimensional data to two or three dimensions, preserving local structure. [13-15]

c) Association Rule Learning

- **Apriori Algorithm:** Used to identify frequent item sets and derive association rules from them, commonly used in market basket analysis.

3.2.3. Semi-Supervised Learning Algorithms

These algorithms leverage both labeled and unlabeled data. A small labeled dataset is combined with a large unlabeled dataset to improve learning.

- Self-Training: The model initially trains on labeled data, then iteratively labels unlabeled data based on its predictions and retrains itself.
- Generative Models: These models learn to generate data similar to the training set, such as using variational autoencoders (VAEs) or generative adversarial networks (GANs).

3.2.4. Reinforcement Learning Algorithms

Reinforcement learning algorithms focus on training an agent that learns to make decisions through trial and error, guided by rewards and penalties.

- Q-Learning: A value-based reinforcement learning algorithm where the agent learns the optimal action-value function by interacting with the environment.
- Deep Q-Networks (DQNs): Combines deep learning with Q-learning to handle complex environments with high-dimensional inputs, such as images.

3.2.5. Ensemble Learning Algorithms

Ensemble methods combine the predictions of multiple models to improve performance and robustness.

- Bagging (Bootstrap Aggregating): Trains multiple models on different subsets of the data and combines their outputs. Random Forest is a popular bagging technique.
- Boosting: Sequentially trains models, each focusing on correcting the mistakes of the previous one. Popular algorithms include Ada Boost and Gradient Boosting. [17]

4. Literature Review

Integration of Machine Learning in Education:

The literature review discusses the integration of Machine Learning (ML) in various facets of education, focusing on techniques and tools used for student performance evaluation and knowledge assessment.

Impact of Digital Literacy Studies: The impact of

digital literacy studies on modern research practices is highlighted, emphasizing the redefinition of skills required for effective scholarly endeavors and the challenges in discerning the authenticity of online information.

Use of AI in Digital Education: Various literature reviews on the use of AI in digital education are identified, presenting methods for recommending systems for online learning & discussing key concepts in educational data mining, learning analytics.

Text Mining Analysis: The text mining analysis using Word Stat software on articles selected for review is described, emphasizing the quantitative analysis of textual data and the objectivity in summarizing studies through text metric analysis.

Systematic Literature Review Methodology: The methodology of conducting a systematic literature review on digital literacy research is outlined, focusing on the search strategy, inclusion criteria, and analysis of selected articles to identify key findings and research gaps. This paper reviews various Machine Learning techniques that examine and analyze the evidence during the investigation process. Each Machine Learning algorithm works on a specific area of digital literacy and digital education based on the features, it overcomes complexity, data volume, time-lining, correlation, consistency, etc. moreover, this study compares Machine Learning algorithms in terms of customary criteria.

5. Objectives of The Study

- To be acquainted with Machine Learning currently being applied to evaluate digital literacy [19]
- To understand the primary Machine Learning techniques and tools used to analyze digital literacy.
- To know the key benefits of using Machine Learning techniques in promoting digital literacy.
- To uncover Challenges and Limitations arise when using Machine Learning tools or techniques to examine digital literacy.

To fulfill the objective of the study, we conducted a systematic review of the literature using Machine

Learning techniques.

5.1. Machine Learning Applications in Evaluating Digital Literacy

- Machine Learning techniques are utilized to build classification models for student knowledge assessment in e-learning systems, employing classifiers like SVM, LR, RF, DT, GBM, GNB, and MLP to classify students into different performance classes. [2]
- Artificial neural network and support vector machine algorithms are commonly used in digital education to predict student performance, dropout rates, and adapt learning styles, showcasing the versatility and effectiveness of these algorithms in educational settings. [10]
- The intersection of Machine Learning and digital literacy research highlights the importance of understanding how emerging technologies like artificial intelligence influence the acquisition and utilization of digital competencies in diverse scholarly domains, emphasizing the need for scholars to effectively navigate cutting-edge tools. [9]
- Machine Learning algorithms, such as gradient boosting machines (GBM), are employed to predict student academic performance, enabling educators to anticipate behavior, extract hidden patterns, and apply information in educational scenarios, ultimately aiding in the assessment and appraisal of student learning and performance. [2]

5.2. Primary Machine Learning Techniques and Tools for Analyzing Digital Literacy

- Machine Learning techniques such as Support Vector Machines (SVM), Logistic Regression (LR), Random Forest (RF), Decision Trees (DT), Gradient Boosting Machines (GBM), Gaussian Naive Bayes (GNB), and Multilayer Perceptron (MLP) are commonly used to build classification models for student knowledge assessment in e-learning systems.
- Artificial Neural Networks and Support Vector Machine algorithms are prevalent in digital education for predicting student performance, dropout rates, and adapting learning styles,

showcasing their effectiveness in educational settings.

- The intersection of Machine Learning and digital literacy research emphasizes the importance of understanding how emerging technologies like artificial intelligence influence the acquisition and utilization of digital competencies in diverse scholarly domains, highlighting the need for scholars to navigate cutting-edge tools effectively.
- Gradient Boosting Machines (GBM) are utilized in predicting student academic performance, enabling educators to anticipate behavior, extract hidden patterns, and apply information in educational scenarios, aiding in the assessment and appraisal of student learning and performance. [9]

5.3. Key Benefits of Using Machine Learning Techniques in Promoting Digital Literacy

Enhanced Personalization: Machine Learning algorithms can tailor educational content and learning experiences to individual students' needs, preferences, and learning styles, promoting personalized learning and engagement. **Predictive Analytics:** Machine Learning enables the prediction of student performance, behavior, and learning outcomes, allowing educators to intervene early, provide targeted support, and enhance overall learning effectiveness.

Adaptive Learning Paths: By analyzing student data and behavior, Machine Learning can suggest adaptive learning paths, resources, and activities to optimize learning outcomes and address individual learning gaps.

Efficient Resource Allocation: Machine Learning can help educational institutions allocate resources effectively by identifying areas where additional support or intervention is needed, optimizing resource utilization and improving overall educational outcomes. [10]

5.4. Challenges and Limitations of Using Machine Learning in Examining Digital Literacy

Data Privacy Concerns: Machine Learning tools may require access to sensitive student data, raising

concerns about data privacy, security, and ethical considerations in educational settings.

Algorithm Bias: Machine Learning algorithms can exhibit bias based on the data they are trained on, potentially perpetuating existing inequalities or misconceptions related to digital literacy.

Interpretability Issues: The complexity of some Machine Learning models may hinder their interpretability, making it challenging for educators and researchers to understand how decisions are made or insights are derived.

Resource Intensiveness: Implementing Machine Learning techniques in educational settings may require significant computational resources, expertise, and infrastructure, posing challenges for institutions with limited resources.

Generalization Challenges: Machine Learning models developed for specific contexts or datasets may struggle to generalize to diverse student populations or educational environments, limiting their applicability and effectiveness. [9]

5.5. Applications of Machine Learning

In today's world, during the Fourth Industrial Revolution (4IR), machine learning is gaining popularity in many different fields because it can learn from past data and make smart decisions. For a machine learning model to work well, both the quality of the data and the performance of the algorithms are important. This study has shown that no single measure can fully describe how well a classifier works, and there isn't one classifier that can meet all the criteria. [11] [12]

5.6. Summary of Research Gap

Digital Literacy Disparities: The research highlights the disparities in access to digital resources, technological infrastructure, and digital skill development, particularly in underserved communities. [18]

Policy Implications: There is a gap in policy initiatives addressing differences in digital literacy and developing comprehensive digital competencies among scholars. [20-21]

Algorithm Bias Concerns: Challenges related to algorithm bias in Machine Learning tools examining digital literacy may pose limitations in promoting equitable learning outcomes.

Data Privacy Issues: Concerns about data privacy and security when using Machine Learning techniques to examine digital literacy present challenges that need to be addressed in educational settings.

6. Result and Discussion

- The comparative assessment of digital literacy findings across various academic settings provides valuable insights into the diverse manifestations and implications of digital competencies, shedding light on disciplinary variations in digital skill acquisition and utilization.
- The study conducted experiments using accurate dataset from an e-learning system to build and assess various classification models for student knowledge assessment. Various classifiers like SVM, LR, RF, DT, GBM, GNB, and MLP were employed to classify students into different classes, aiming to provide a comprehensive evaluation of student performance.
- There has been a growing interest in the study of digital literacy and education, as evidenced by the increasing number of publications over the years. This trend reflects the evolving landscape of digital literacy research and its relevance in scholarly pursuits. [18]
- Past paper references have explored methodologies like Machine Learning based on smart LMS for online learning and focused on algorithms that improve with use, indicating a shift towards leveraging ML techniques in educational settings.
- The systematic revision process involved filtration based on inclusion-exclusion criteria, refining search strings related to artificial intelligence, deep learning, and machine learning, and evaluating performance based on recall and precision metrics.
- Literature review studies have explored into recommendation systems for smart learning management systems, learning management systems overview, and the strengths of using artificial intelligence in education, highlighting different aspects of digital education research.

- The analysis of publications over the years underscores the increasing importance of digital literacy and learning in research, with a significant rise in the number of articles published on the topic. This growth reflects the expanding scope and relevance of digital literacy across various fields of study.

Conclusion

The conclusion emphasizes the importance of understanding disciplinary variations in digital literacy to tailor initiatives that meet the unique needs of diverse academic domains. Researchers highlight the significance of improving reporting, design, and research behavior to maintain the credibility and dependability of findings. Intelligent tutoring systems utilizing Machine Learning models like Bayesian nets, content-based recommenders, and collaborative filtering algorithms have shown promise in assisting students with personalized learning resources and feedback. The study highlights the increasing interest in the application of Machine Learning and deep learning in digital education, with a focus on themes like intelligent tutors, performance prediction, and adaptive learning styles. Researchers note a growing trend in using artificial neural networks and support vector machines across various classification and regression problems in digital education, indicating their versatility and effectiveness. The systematic review underscores the prevalence of experimental studies in the field, showcasing a keen interest in comparing algorithmic results for tasks such as student performance prediction and dropout analysis.

References

- [1]. Munir, H., Vogel, B., & Jacobsson, A. (2022). Artificial intelligence and machine learning approaches in digital education: A systematic revision. *Information*, 13(4), 203.
- [2]. Alruwais, N., & Zakariah, M. (2023). Evaluating student knowledge assessment using machine learning techniques. *Sustainability*, 15(7), 6229.
- [3]. Gilster, P., Glistler, P.: *Digital Literacy*. Wiley Computer Pub, New York (1997)
- [4]. Martin, A., Grudziecki, J.: *DigEuLit: Concepts and tools for digital literacy development*. *Innov. Teach. Learn. Inform. Comput. Sci.* 5, 249–267 (2015)
- [5]. Cordell, R.M.: Information literacy and digital literacy: Competing or complementary? *Commun. Inf. Lit.* 7, 14 (2013)
- [6]. Rahman, T., Amalia, A., Aziz, Z.: From digital literacy to digital intelligence. In: 4th International Conference on Sustainable Innovation 2020–Social, Humanity, and Education (ICoSIHESS 2020), pp. 154–159 (2021)
- [7]. Mitchell, T. M. (1997). *Machine learning* (Vol. 1). McGraw-hill New York.
- [8]. Alakrash, H. M., & Abdul Razak, N. (2021). Technology-based language learning: Investigation of digital technology and digital literacy. *Sustainability*, 13(21), 12304.
- [9]. Susanty, L. (2024). Critical Analysis of the Research on Digital Literacy. *Sinergi International Journal of Education*, 2(1), 12–25.
- [10]. Munir, H., Vogel, B., & Jacobsson, A. (2022). Artificial intelligence and machine learning approaches in digital education: A systematic revision. *Information*, 13(4), 203.
- [11]. Sarker, I. H. (2021). Machine learning: Algorithms, real-world applications and research directions. *SN computer science*, 2(3), 160.
- [12]. Mohamed, A. E. (2017). Comparative study of four supervised machine learning techniques for classification. *International Journal of Applied*, 7(2), 1–15.
- [13]. Wu, D. (2024, June). Digital Literacy: Evolution, Evaluation and Enhancement. In *International Conference on Blended Learning* (pp. 62–74). Singapore: Springer Nature Singapore.
- [14]. Tinmaz, H., Lee, Y. T., Fanea-Ivanovici, M., & Baber, H. (2022). A systematic review on digital literacy. *Smart Learning Environments*, 9(1), 21.
- [15]. Kaddoura, S., Popescu, D. E., & Hemanth, J. D. (2022). A systematic review on machine learning models for online learning and



examination systems. PeerJ Computer Science, 8, e986.

- [16]. Joseph, G. V., Athira, P., Thomas, M. A., Jose, D., Roy, T. V., & Prasad, M. (2024). Impact of Digital Literacy, Use of AI Tools and Peer Collaboration on AI Assisted Learning: Perceptions of the University Students. *Digital Education Review*, 45, 43-49.
- [17]. Reddy, D. R. O., Reddy, R. N., Radha, M., & Vani, S. (2017). A Review of Machine Learning Approaches in Data Sensitive Real-World Applications '. *Journal of advanced research in dynamical and control systems*, 9(3), 165-171.
- [18]. Sassirekha, M. S., & Vijayalakshmi, S. (2022). Predicting the academic progression in student's standpoint using machine learning. *Automatika: časopis za automatiku, mjerenje, elektroniku, računarstvo i komunikacije*, 63(4), 605-617.
- [19]. Fakhrudin, A. (2023). Digital Literacy Analysis of Primary School Students. *KnE Social Sciences*, 13-22.
- [20]. Murray, M. C., Pérez, J., & Fluker, J. (2022). Digital Literacy in the Core: The Emerging Higher Education Landscape. *Issues in Informing Science and Information Technology*, 19, 001-013.
- [21]. Audrin, C., & Audrin, B. (2022). Key factors in digital literacy in learning and education: a systematic literature review using text mining. *Education and Information Technologies*, 27(6), 7395-74