I.J. Modern Education and Computer Science, **2026, 3, 17-35**
Published Online on June 8, 2026 by MECS Press (http://www.mecs-press.org/)
DOI: 10.5815/ijmecs.2026.03.02

Leveraging Machine Learning to Predict ICT Proficiency levels among Public School Teachers in Bukidnon

Nathalie Joy G. Casildo*
Department of Information Technology, College of Information Sciences and Computing, Central Mindanao University, Philippines
E-mail: nathaliejoy@cmu.edu.ph
ORCID iD: https://orcid.org/0000-0001-6935-5262
*Corresponding author

Gladys S. Ayunar
Department of Information Technology, College of Information Sciences and Computing, Central Mindanao University, Philippines
E-mail: gladys.ayunar@cmu.edu.ph
ORCID iD: https://orcid.org/0009-0000-8829-8689

Jinky G. Marcelo
Department of Information Technology, College of Information Sciences and Computing, Central Mindanao University, Philippines
E-mail: jinkymarcelo@cmu.edu.ph
ORCID iD: https://orcid.org/0000-0001-7066-6375

Kent Levi A. Bonifacio
Department of Information Technology, College of Information Sciences and Computing, Central Mindanao University, Philippines
E-mail: klab@cmu.edu.ph
ORCID iD: https://orcid.org/0000-0002-3546-3158

Received: January 9, 2025; Revised: June 21, 2025; Accepted: January 13, 2026; Published: June 8, 2026

Abstract: This study utilizes the Digital Competence Framework for Educators (DigCompEdu) and machine learning (ML) techniques to evaluate and predict the ICT proficiency levels of public school teachers in Bukidnon. Analyzing a dataset of 1,275 responses and addressing data imbalances, several classification models were evaluated to identify the most reliable predictor of teacher competence. The findings indicate that the majority of teachers currently operate at the 'Integrator' (B1) level. Key predictors of proficiency include skills in online safety, collaborative learning, and the creative use of digital tools. Among the tested algorithms, Random Forest emerged as the most effective model for accurately classifying teacher skill levels. This research provides a data-driven roadmap for educational policymakers, offering actionable insights for designing targeted professional development programs that foster transformative teaching and improved student outcomes.

Keywords: Class Imbalance, Digcompedu, ICT Proficiency, Machine Learning, Professional Development, Public School Teachers

1. Introduction

Information and Communication Technology (ICT) integration in the classroom has been shown to improve academic achievement and student retention. According to studies, ICT provides interesting and useful teaching resources that increase retention rates in addition to enhancing academic performance [1]. The necessity of accepting and using technology in education effectively is further highlighted by the fact that its use in online learning environments has shown a notable positive influence on student retention [2].

The impact of ICT goes beyond the students; it is essential to teachers' professional growth. Teachers' knowledge, attitudes, and institutional culture must change significantly for technology integration to be effective [3]. According to research, in order to improve training efficacy and teaching outcomes, educators must incorporate ICT into their professional development programs [3,4]. Structured ICT training, monetary investments in technology resources, and ongoing monitoring to guarantee successful implementation are all stressed in practical suggestions [5].

Evaluating teachers' competency levels is essential to achieving meaningful ICT integration in the classroom. Classifying educators' digital competences into six categories, from "Emerging" to "Empowering," frameworks such as the Digital Competence Framework for Educators (DigCompEdu) allow for customized training programs depending on particular needs[6,7]. For example, assessing the ICT proficiency of Bukidnon public school teachers can inform customized interventions aimed at improving their instructional skills and, eventually, student outcomes [8].

Machine learning has developed into a ground-breaking educational tool with predictive skills that can evaluate and analyze a variety of teaching and learning facets. In forecasting academic performance, methods such as Random Forest have shown remarkable accuracy, indicating their potential to tackle challenging educational issues [9]. Predictive models' dependability must be guaranteed, nevertheless, by addressing issues such class imbalance and biases in training data [10,11,12]. SMOTE-ENN is one solution that has demonstrated potential in addressing these problems and improving model fairness and accuracy [12].

The purpose of this study is to forecast the ICT competency levels of Bukidnon public school teachers using machine learning techniques. To do this specifically, we must use machine learning analysis to determine the key variables affecting instructors' ICT competency levels. Furthermore, we will assess how feature selection and oversampling strategies affect machine learning models' predictive accuracy in identifying ICT competence levels. Finally, use the dataset to identify the machine learning model that predicts instructors' ICT competence levels with the best accuracy and dependability.

While previous studies have extensively documented the benefits of ICT in the classroom, a significant gap remains in how teachers' digital competencies are systematically evaluated and predicted for future planning. Most existing research relies on descriptive statistical analyses of survey data, which identifies current skill levels but lacks the predictive power to pinpoint the specific competencies that drive a teacher from one proficiency level to the next. Furthermore, a common challenge in educational technology research is the presence of class imbalance—where the majority of respondents cluster at a single proficiency level—leading to biased models that fail to accurately identify teachers at the highest or lowest ends of the spectrum.

This study addresses these shortcomings by applying a strong machine learning (ML) approach to the DigCompEdu framework. Unlike traditional assessments, this research goes beyond mere description by utilizing feature selection and the SMOTE-ENN oversampling technique to resolve data imbalances that have historically limited the accuracy of educational classification models. This study specifically fills the gap in local educational policy-making within the Bukidnon region by providing a high-precision predictive system that identifies the critical indicators of teacher success.

The purpose of this study is to forecast the ICT competency levels of Bukidnon public school teachers using machine learning. Specifically, we aim to: (1) determine the key variables that most significantly influence teacher proficiency; (2) assess how feature selection and oversampling strategies mitigate data bias to improve predictive accuracy; and (3) identify the optimal machine learning model that provides the most dependable classification for targeted professional development interventions.

2. Literature Review

2.1. The Multi-faceted Impact of ICT in Education

While there is a broad consensus that ICT integration improves academic outcomes, the literature reveals a distinction between subject-specific engagement and broader institutional retention. For instance, while Hussain et al. [1] found that ICT specifically boosted retention in secondary chemistry by making abstract concepts "captivating," Scarpin et al. [2] argued that in online learning environments, the technology itself—rather than the subject matter—is the primary driver of student persistence. This suggests that the effectiveness of ICT is contingent on both the curriculum and the mode of instruction.

Furthermore, a critical tension exists between the logistics of ICT implementation and the psychological readiness of educators. Bouasangthong et al. [5] proposed a roadmap focused on infrastructure, such as financial support and resource rooms. Conversely, Seenivasan [3] argues that such structural investments are insufficient without a fundamental shift in teachers' underlying beliefs and institutional culture. This highlights a gap in current training: while teachers may be provided with tools [5], they often lack the pedagogical shifts required to utilize them with transformation [3].

2.2. Evolution of Competency Frameworks

Selecting an appropriate framework for evaluating teacher proficiency requires a balance between conceptual depth and measurable progression. Several frameworks have been widely used in educational research, yet many lack the structural granularity required for predictive machine learning tasks. For instance, the TPACK (Technological Pedagogical Content Knowledge) framework is highly regarded for describing the intersection of teacher knowledge types; however, it has been criticized for being overly conceptual and lacking a clear, ordinal scale for measuring skill growth [13]. Similarly, the UNESCO ICT Competency Framework for Teachers (ICT-CFT) provides excellent policy-level guidelines, but its broad categories often prove too general for the high-dimensional feature selection needed in classification models [13]. Other models, such as Jisc and TETC, focus primarily on institutional capabilities or teacher educators, making them less applicable to the individual professional growth of public school teachers in rural settings [13].

In contrast, the Digital Competence Framework for Educators (DigCompEdu) [6, 7] was selected for this study because it addresses these limitations through a distinctive six-level developmental hierarchy (A1 to C2). Unlike the more static UNESCO model, DigCompEdu provides a granular roadmap that tracks a teacher's evolution from a "Newcomer" to a "Pioneer." This ordinal structure is uniquely suited for machine learning analysis, as it allows the model to identify specific "pivot skills" that facilitate movement between tiers. While critics note that DigCompEdu relies on self-assessment [7], its focus on 22 specific competences across six pedagogical areas provides the precise data points necessary for robust feature selection and classification. A summarized view of this progression is provided in Table 1.

Table 1. Proficiency level from the developed DigCompEdu [6].

Proficiency Tier	Level	Stage of Development	Core Focus
Substitution	A1–A2	Newcomer / Explorer	Awareness of tools and basic experimentation
Integration	B1–B2	Integrator / Expert	Consistent, innovative, and strategic ICT use
Transformation	C1–C2	Leader / Pioneer	Mentoring others and leading systemic change

Effective professional development must cater to these specific tiers [1,2]. Utilizing a hierarchical framework like DigCompEdu rather than a flat competency model, researchers can offer targeted support and progression pathways rather than a "one-size-fits-all" training program that often fails to move teachers beyond basic integration [7].

2.3. Machine Learning in Education

The landscape of Machine Learning (ML) and Artificial Intelligence (AI) in education has shifted from basic administrative automation to complex predictive modeling aimed at improving pedagogical outcomes. According to a systematic review by Forero-Corba and Bennasar [9], ML applications in the educational sector can be synthesized into three primary thematic pillars:

1. Student Performance and Retention: This includes early detection of academic struggles, predicting dropout risks, and identifying students who require specialized support, such as those with Autism Spectrum Disorders (ASD) [9,11].
2. Pedagogical and Content Enhancement: ML is increasingly used to generate adaptive educational content, motivate learning through mobile platforms, and facilitate critical thinking through computational logic [9].
3. Institutional and Professional Growth: Beyond student metrics, ML serves as a tool for strengthening institutional security, preventing the spread of misinformation on campus networks, and directly improving the digital skills of educators [9].

Despite the breadth of these applications, the literature shows a significant concentration on specific algorithmic approaches. As illustrated in Fig.1., there is a clear hierarchy in the techniques favored by researchers [9].

As shown in Fig. 1., Random Forest (RF) and Decision Trees are the most prevalent techniques, while specialized algorithms like the Boruta Algorithm remain underutilized. The dominance of Random Forest is largely attributed to its ensemble nature, which allows it to handle the non-linear relationships common in educational data. For instance, some studies reviewed in [9] achieved a near-perfect 99% prediction accuracy using RF for student performance.

However, a critical contrast must be drawn between model accuracy and model utility. While high-accuracy models are impressive, Barbierato and Gatti [10] argue that "black-box" ensemble methods often lack the interpretability required for teachers to understand why a specific prediction was made. This suggests that in the context of teacher proficiency, a model must not only be accurate—as seen in the high success rates of RF—but also transparent enough to inform actual policy changes and professional development strategies.

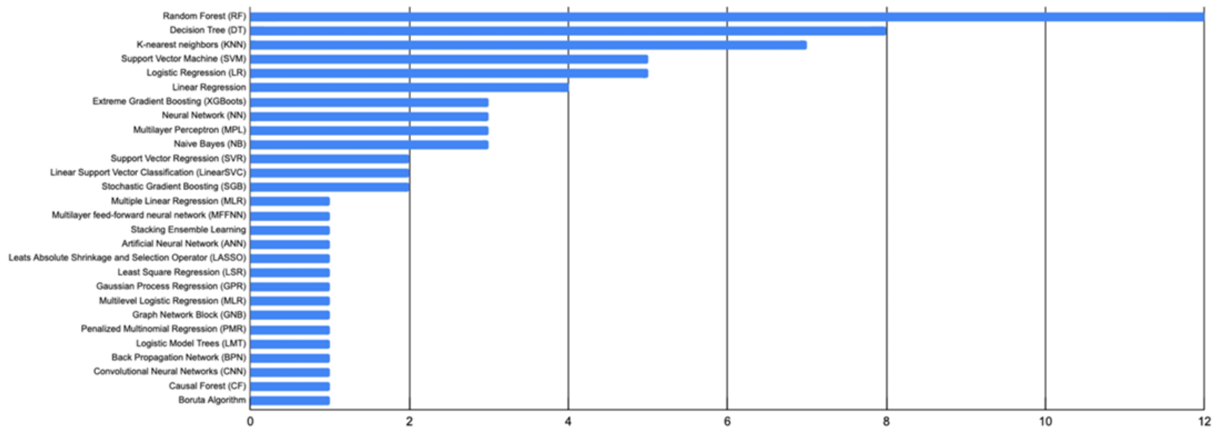

Fig. 1. The most widely used Machine Learning Techniques [9].

2.4. Challenges in Machine Learning

Challenges in machine learning is highlighted in a paper [10], despite ML's strong predictive capabilities, it often lacks causal explanatory power, which is a hallmark of scientific understanding [10]. The paper [10] also discusses the limitations of ML models, such as potential biases in training data, and highlights emerging techniques like reinforcement learning and imitation learning that show promise in mimicking human cognitive skills. [11,14,15] These papers have mentioned common issues in their study which is the class imbalance. To answer this concern a study [12] that we can benchmark which used SMOTE-ENN gives a better result.

3. Methodology

Fig.2. Presents the Research Process of this study. The process begins with the preparation phase, during which the appropriate framework or questionnaire is determined for assessing ICT competency among teachers. Once the framework is established, the data collection phase is conducted to gather relevant information for the study.

Fig. 2. Research Process.

The data pre-processing stage is carried out after data collection in order to arrange the dataset, eliminate discrepancies, and deal with missing or unknown information. An imbalance in the target variable's class distribution was discovered at this phase. Techniques for class distribution balancing were used to remedy this, guaranteeing more equal representation for all classes.

To determine the most important elements affecting teachers' ICT competence levels, feature selection was carried out after the data had been balanced. The efficiency and efficacy of later machine learning applications were improved by this stage, which produced a refined dataset with only the most pertinent attributes.

The prepared dataset was then used to train multiple machine learning models, including Logistic Regression, Support Vector Machines (SVM), Random Forest, and K-Nearest Neighbors (KNN). The performance of the models was evaluated using Accuracy , Precision , Recall, and F1-score in defined as:

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

$$Precision = \frac{TP}{TP+FP} \tag{2}$$

$$Recall = \frac{TP}{TP+FN} \tag{3}$$

$$F1\ score = 2\ x\ \frac{Precision \times Recall}{Precision+Recall} \tag{4}$$

Ultimately, the most accurate predictor of ICT proficiency levels was found to be the model with the best performance indicators. The development of a prediction system that can precisely evaluate instructors' ICT skill levels was based on this chosen model.

3.1. Preparation

This study is based in part on a study [16] on the development of the teacher digital competence validation of DigCompEdu check-in questionnaire in the Andalusian university context. Additionally, [17] has created a Spanish version of the questionnaire. Professional Engagement, Digital Resources, Teaching and Learning, Assessment, Empowering Learners, and Facilitating Learners' Digital Competence were the six competency categories that were the emphasis of the questions as shown in Fig.3.

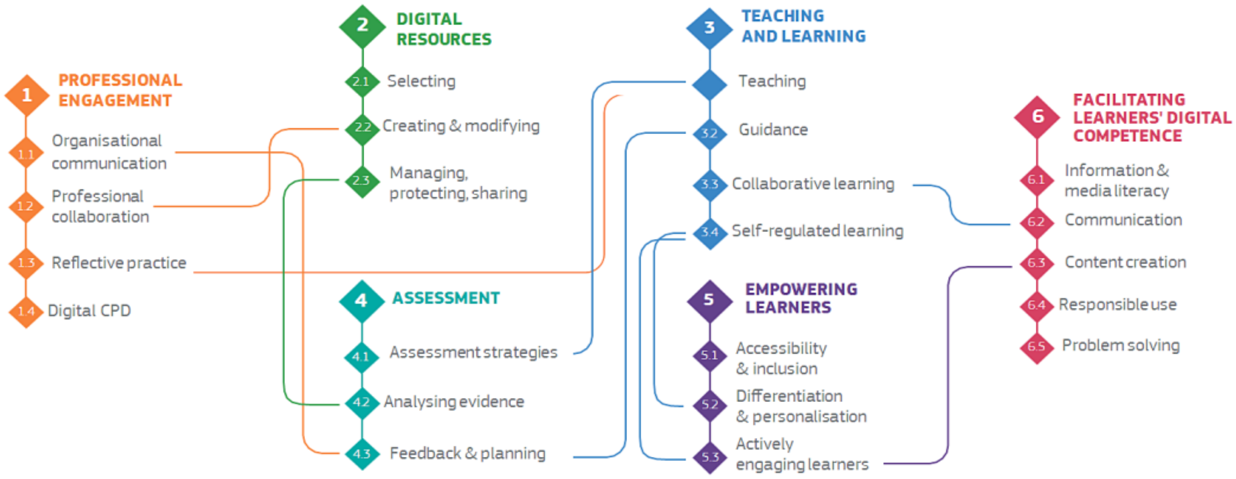

Fig. 3. Six Competence Areas in DigComEdu.

The Framework also defines six proficiency levels:

A1 (Newcomer) which means a person has Limited or no digital skills
A2 (Explorer) with Basic digital knowledge, limited integration into teaching
B1 (Integrator): Regular use of digital tools in teaching.
B2 (Expert): Effective and innovative integration of digital tools.
C1 (Leader): Leads others in using digital tools effectively.
C2 (Pioneer): Contribute to the advancement of digital education at a systemic level.

3.2. Data Collection

Using the DigCompEdu Framework, a total of 1,275 valid responses were gathered from DepEd teachers across 14 municipalities in the province of Bukidnon (Fig.4). The questionnaire focused on six core constructs: Professional Commitment, Digital Resources, Digital Pedagogy, Evaluation and Feedback, Empowering Students, and Facilitating Students' Digital Competence. To ensure the study's findings were representative of the regional teaching population, a stratified voluntary sampling strategy was employed. The DepEd Bukidnon divisions served as the primary strata, and data collection was formally coordinated through administrative channels, where school heads disseminated Google Form links to their respective faculty.

This approach maximized reach across diverse geographic and socio-economic contexts. Coordinating through official channels, the study minimized self-selection bias and ensured a broad cross-section of teachers. The resulting large sample size ($N = 1{,}275$) provides a robust statistical foundation, enhancing the generalizability of the predictive models to the wider regional teacher population.

Based on the framework's scoring system, competency levels were categorized into six tiers according to the score thresholds shown in Table 2.

Table 2. Proficiency Level.

Competency Level	Proficiency Level Score (out of 88 points)
Novice (A1)	< 20 POINTS
Explorer (A2)	20 TO 33 POINTS
Integrator (B1)	34 TO 49 POINTS
Expert (B2)	50 TO 65 POINTS
Leader (C1)	66 TO 80 POINTS
Pioneer (C2)	> 80 POINTS

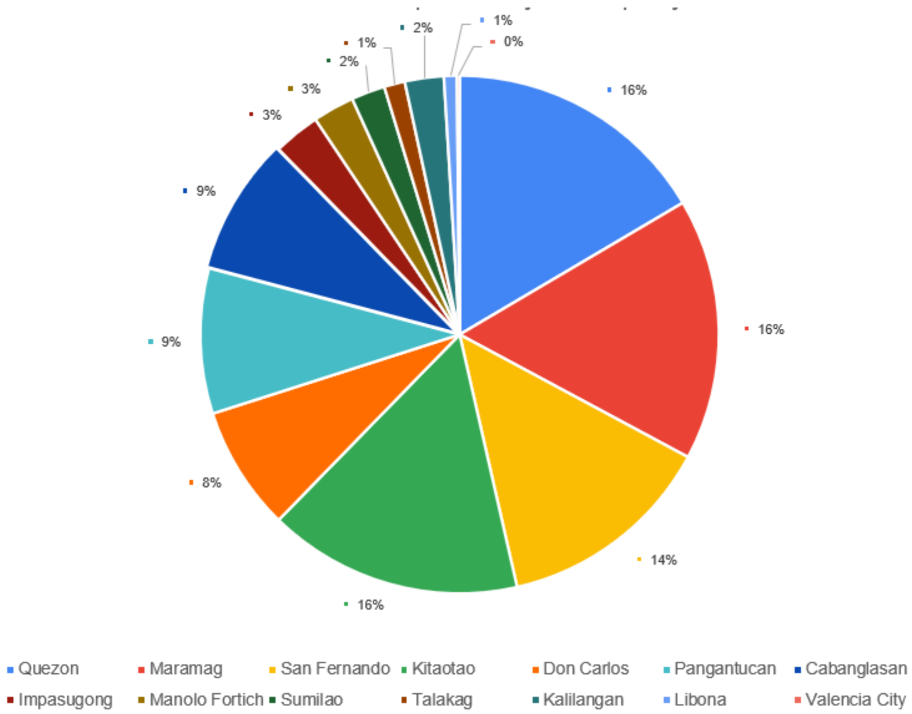

Fig. 4. Teachers Responses per Municipality.

3.2.1. Instrument Validity and Reliability

While internal consistency was measured for the local population, the validity of the DigCompEdu Check-in tool is supported by its rigorous development by the European Commission's Joint Research Centre (JRC). The instrument's validity was established through the following evidence:

- Content Validity: The questionnaire was developed through an extensive multi-stage process involving educational technology experts, stakeholders, and practitioners at a European level. This ensured that the 22 items comprehensively represent the six competence areas defined by the framework [16, 6].
- Construct Validity: Statistical validation in prior research [16, 17] has confirmed that the questionnaire accurately maps to its six theoretical constructs (Professional Engagement, Digital Resources, Digital Pedagogy, Assessment, Empowering Learners, and Facilitating Digital Competence). The alignment of our local results with the hierarchical levels of the framework (A1–C2) further supports the construct's stability across different cultural contexts.
- Face Validity: To ensure the instrument was appropriate for the Bukidnon context, the questionnaire underwent a preliminary review by local educational experts. This step confirmed that the terminology and scenarios presented in the items were relevant and easily understood by public school teachers in the Philippines.

Following the verification of validity, the internal consistency (reliability) of the instrument was tested using Cronbach's alpha (α). This test determines how closely related the 22 items are as a group. The Cronbach's alpha (α) is calculated as follows:

$$\alpha = \frac{k}{k-1}\left(1 - \frac{\sum s_i^2}{s_t^2}\right) \tag{5}$$

Where:

- k is the number of items;
- $\sum s_i^2$ is the sum of the variances; and
- s_t^2 is the variance of total score.

Although the questionnaire has been validated by other institutions, its reliability was further assessed for this specific study using Cronbach's alpha.

3.3. Pre-processing

After we gathered 1, 276 data we were able to clean the data and handle unknown and missing values. We have also removed the columns that could identify the respondents to preserve anonymous respondents. We have added demographic data for this study and results were written in another study [18]. Replacing the questions with item number listed in Table 3.

Table 3. Questions with Item number.

Items	Item Number
I consistently use various digital platforms to enhance connection with students, families, and coworkers. For instance, emails, messaging services like WhatsApp, blogs, and the school website.	Item1
I use ICT technologies to work with my peers inside and outside my educational organization.	Item2
I actively develop my ICT teaching skills	Item3
I participate in online training courses. For example: online courses of administration, MOOCs, webinars.	Item4
I use different internet sites (web pages) and search strategies to find and select a wide range of ICT resources.	Item5
I create my own ICT resources and modify existing ones to adapt them to my needs as a teacher.	Item6
I securely protect sensitive content. For example: exams, grades, personal data	Item7
I carefully consider how, when and why to use ICT technologies in class, to ensure that their added value is exploited.	Item8
I monitor my students' activities and interactions in the online collaborative environments we use.	Item9
When my students work in groups or teams, they use ICT to acquire and document knowledge.	Item10
Use of ICT to allow students to plan, document and assess their learning on their own. For example: self-assessment quizzes, digital portfolio, blogs, forums.	Item11
I use ICT assessment strategies to monitor student progress	
I analyze all available data to identify students who need additional support. "Data" includes: student engagement, performance, grades, attendance, activities, and social interactions in online environments.... "Students in need of additional support" are: those at risk of dropping out of school, underachieving, learning disabilities, specific learning needs, or lacking cross-cutting skills (social, verbal, or study skills).	Item13
I use ICT to provide effective feedback.	Item14
When I propose ICT tasks, I consider and address possible problems such as equal access to digital devices and resources; compatibility issues or low level of digital competence of students.	Item15
I use ICT to provide learners with personalized learning opportunities. For example: assigning different ICT tasks to address individual learning needs, taking into account preferences and interests.	Item16
I use ICT to actively engage students in class.	
I teach students how to assess the reliability of information searched online and to identify erroneous and/or biased information.	Item18
I propose assignments that require students to use ICT media to communicate and collaborate with each other or with an external audience.	
I propose assignments that require students to create ICT content.	
I teach students how to behave safely and responsibly online	Item21
I encourage students to use ICT technologies in creative ways to solve concrete problems. For example, overcoming obstacles or emerging challenges in their learning process.	Item22

Fig. 5. showed the view of the final Dataset consisting of Item Responses, Proficiency Score, Proficiency Level. Each of the columns were renamed with the item number as shown in Fig. 5. It is also shown in this figure the Proficiency Score which were already calculated, with its corresponding Proficiency Level.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X
1	Item 1	Item 2	Item 3	Item 4	Item 5	Item 6	Item 7	Item 8	Item 9	Item 10	Item 11	Item 12	Item 13	Item 14	Item 15	Item 16	Item 17	Item 18	Item 19	Item 20	Item 21	Item 22	Proficiency Score	Proficiency Level
2	4	4	4	4	4	3	5	5	4	4	4	4	5	5	5	5	4	5	4	4	3	5	72	Leader
3	2	1	3	3	1	3	3	2	2	3	1	2	2	3	1	1	3	1	1	1	2	2	21	Explorer
4	4	3	5	2	5	5	3	5	1	1	2	2	3	2	1	1	3	1	1	2	1	1	32	Explorer
5	4	3	5	5	4	4	4	4	4	4	3	3	3	5	5	3	3	4	4	4	5	5	66	Leader
6	1	3	3	3	2	3	5	4	4	3	3	4	2	2	3	3	4	2	1	1	3	2	39	Integrator
7	1	3	4	4	2	3	1	2	5	1	1	2	1	2	1	4	3	1	1	1	1	1	23	Explorer
8	4	4	3	3	4	3	3	4	4	2	3	3	3	3	4	4	4	3	3	4	5	4	55	Expert
9	1	2	1	1	1	1	3	2	1	1	2	2	4	2	1	1	3	4	1	2	3	3	20	Explorer
10	2	2	3	4	3	2	2	2	1	2	2	2	4	2	1	1	2	3	1	2	1	1	23	Explorer
11	3	4	4	3	3	3	3	3	3	3	4	3	3	3	3	4	4	3	1	1	1	1	41	Integrator
12	2	3	3	3	2	2	3	2	1	1	1	2	4	2	1	5	3	1	1	1	1	1	23	Explorer
13	1	3	2	2	2	2	3	2	2	2	3	3	3	3	3	2	3	2	2	2	2	3	30	Explorer
14	4	4	5	5	4	4	5	4	4	3	3	4	4	4	3	4	3	3	3	3	4	3	61	Expert
15	2	2	4	2	2	3	3	2	2	2	2	2	2	2	1	1	3	1	1	2	2	1	22	Explorer
16	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	3	4	4	65	Expert
17	5	1	2	2	1	2	3	4	2	3	3	3	2	2	3	1	2	3	2	2	2	3	31	Explorer
18	5	3	3	2	3	3	5	2	2	3	2	2	3	3	2	2	3	3	2	3	3	1	38	Integrator
19	4	5	4	4	5	4	5	4	3	3	3	5	4	4	4	4	3	3	3	3	4	4	63	Expert
20	2	3	4	1	2	2	3	2	3	2	1	3	3	2	1	1	3	1	2	3	1	3	26	Explorer
21	2	2	4	2	3	3	3	2	2	2	3	2	2	3	3	1	3	2	1	1	1	1	26	Explorer
22	5	4	4	5	4	5	5	4	4	4	5	5	4	5	5	5	5	5	5	5	5	5	81	Pioneer
23	4	4	3	4	4	2	5	3	3	3	1	2	2	2	3	2	2	3	3	3	3	2	41	Integrator
24	4	3	4	4	4	4	5	3	3	2	2	2	3	3	3	3	4	3	3	3	2	3	48	Integrator
25	4	3	4	2	3	3	2	2	2	2	2	2	2	2	1	1	3	1	1	1	1	1	23	Explorer
26	2	3	2	4	2	2	3	2	3	3	3	2	4	2	2	2	3	3	2	2	3	3	35	Integrator
27	3	3	4	2	4	4	3	1	1	3	1	1	2	2	3	5	2	5	2	3	3	3	38	Integrator
28	3	3	5	3	3	3	3	2	3	3	3	3	3	2	1	3	3	3	4	3	2	3	42	Integrator
29	4	4	3	2	2	4	4	3	2	2	2	2	4	2	1	4	3	5	1	1	4	1	38	Integrator
30	4	2	1	3	2	3	4	2	3	5	3	2	2	2	3	3	3	4	4	4	4	3	44	Integrator
31	2	3	4	2	3	2	2	2	2	1	2	3	3	3	3	1	3	2	1	1	1	1	25	Explorer

Fig. 5.View of Final Dataset.

Each row represents an individual teacher. The dataset includes both numerical variables, item responses, proficiency scores and categorical variable which is proficiency levels. The proficiency scores range across different levels, reflecting a diverse population in terms of digital competence. The "Proficiency Level" column is the target variable for supervised machine learning tasks, where the model predicts proficiency levels based on item responses.

3.4. Machine Learning

The machine learning phase depicted in Fig.6., flowchart of the study's Machine Learning Phase which involves several critical steps designed to prepare the data, train and evaluate models, and fine-tune their performance. Using Google Colaboratory and Python and its libraries this phase becomes possible.

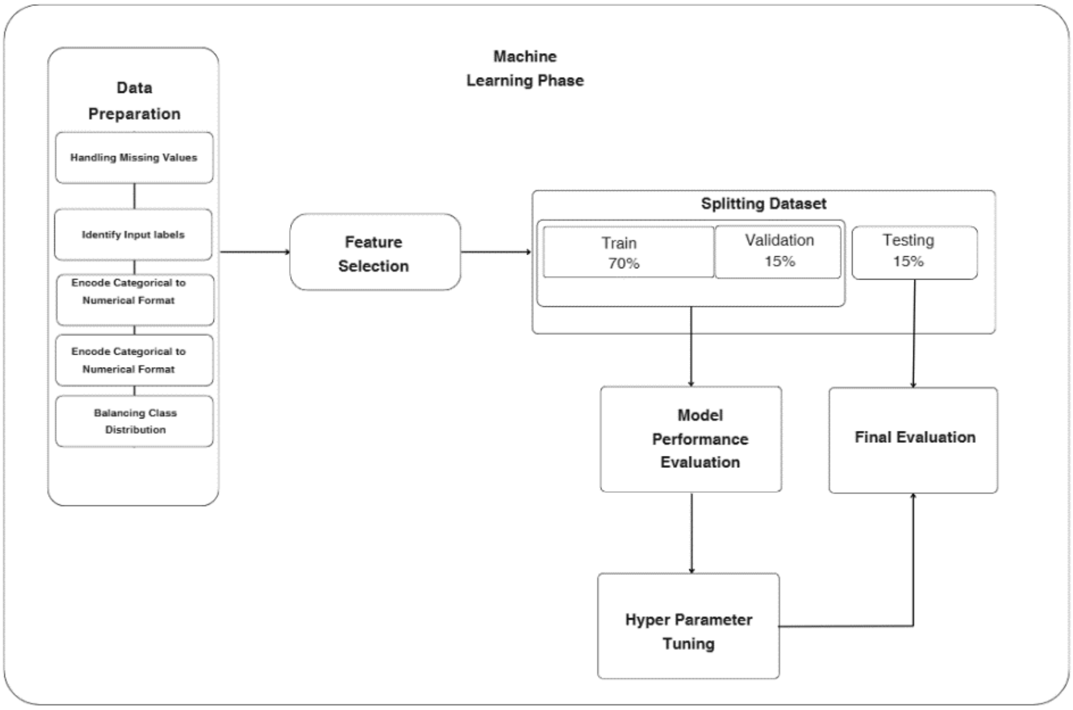

Fig. 6. Flowchart of the study's Machine Learning Phase.

First is the Data Preparation, which ensures the dataset is clean, consistent, and suitable for model training. Missing data points were addressed by applying imputation strategies. To address the class imbalance in the target variable, the SMOTE-ENN technique was employed. Educational datasets frequently suffer from imbalance where respondents cluster around "average" levels (e.g., B1), leaving "Newcomers" (A1) or "Pioneers" (C2) underrepresented. The rationale for choosing SMOTE-ENN is to prevent "majority bias," ensuring the final model accurately identifies teachers at all levels of the spectrum to facilitate more equitable resource distribution.

$$x_{new} = x_i + \lambda(x_{zi} - x_i) \tag{6}$$

Where x_{new} is the synthetic sample, x_i is an instance from the minority class, x_{zi} is one of its k-nearest neighbors, and λi s a random number between 0 and 1. This is followed by ENN, which cleans the overlapping classes by removing instances that differ from their neighbors.

While keeping the most pertinent information, feature selection aids in reducing the dimensionality of the data. The ANOVA F-test (f_classif) was used to evaluate the relationship between each feature and the target variable. The rationale is to isolate "pivot skills"—the specific digital competencies that most significantly drive movement between proficiency tiers. Identifying these features allows educational leaders to focus training on the most impactful areas rather than a broad, less effective approach. The F-statistic is calculated as:

$$F = \frac{MSB}{MSW} = \frac{\sum n_j(\bar{x}_j - \bar{x})^2/(k-1)}{\sum\sum(x_{ij} - \bar{x}_j)^2/(N-k)} \tag{7}$$

The dataset was then split into three subsets: a 70% Training Set to identify patterns, a 15% Validation Set to prevent overfitting, and a 15% Testing Set for final evaluation on unseen data. Training and evaluating the models are the core tasks of this stage. Performance was assessed using a comprehensive suite of metrics including Accuracy, Precision, Recall, and F1-score.

To ensure the models operate at peak efficiency, the best hyperparameter configurations were identified using GridSearchCV. The rationale for this exhaustive optimization is to minimize misclassification errors (false positives). In an educational context, this level of precision ensures that teachers are not assigned to professional development programs that do not match their actual skill levels. The testing set served as the final yardstick to verify the model's practicality and its ability to generalize to fresh, untested data.

Table 4. Results Table for the Machine Learning Phase.

Model	Training Metrics Used	Feature Selection Method	Balancing Technique
Logistic Regression	Classification Report	SelectKBest (f_classif)	SMOTE + Edited Nearest Neighbors
SVM	Classification Report	SelectKBest (f_classif)	SMOTE + Edited Nearest Neighbors
Random Forest	Classification Report	SelectKBest (f_classif)	SMOTE + Edited Nearest Neighbors
Nearest Neighbors	Classification Report	SelectKBest (f_classif)	SMOTE + Edited Nearest Neighbors

3.4.1. Learning Algorithms

Four supervised learning algorithms were evaluated. Logistic Regression was used to model the probability of proficiency levels using the sigmoid function in (8). SVM was implemented to find the maximum-margin hyperplane in (9). For KNN, similarity was determined using Euclidean distance in (10). Then, Random Forest, an ensemble method, utilized Gini Impurity in (11) for node splitting and feature importance ranking.

The sigmoid function for Logistic Regression is defined as:

$$P(y = 1 \mid x) = \frac{1}{1 + e^{-(w^T x + b)}} \tag{8}$$

Where w represents the weights and b is the bias term.

The decision function for the SVM hyperplane is:

$$f(x) = sign(w^T x + b) \tag{9}$$

Where $w^T x + b = 0$ represents the separating hyperplane.

For KNN, the Euclidean distance d between two points x and y in n −dimensional space is:

$$d(x, y) = \sqrt{\sum_{i=1}^{n} (x_i - y_i)^2} \tag{10}$$

The Gini Impurity (I_G) used by Random Forest to determine node splits is calculated as:

$$I_G(p) = 1 - \sum_{i=1}^{C} p_i^2 \tag{11}$$

Where p_i is the probability of an item belonging to class i for C total classes.

3.5. Selecting the Best Model

The selection of the optimal predictive model follows the systematic workflow illustrated in Fig. 7. This process involves methodically iterating through the candidate algorithms and their respective hyperparameter grids. Comparing each model’s performance against established metrics—specifically Accuracy, Precision, Recall, and the F1-score—the most strong configuration is objectively identified. The rationale for this rigorous comparative approach is to ensure that the final prediction system is "policy-ready," providing school administrators with a highly dependable tool that minimizes decision-making errors and maximizes the impact of professional development interventions. This stage serves as the final yardstick to verify that the selected model is not only statistically accurate but also practically reliable for real-world educational planning.

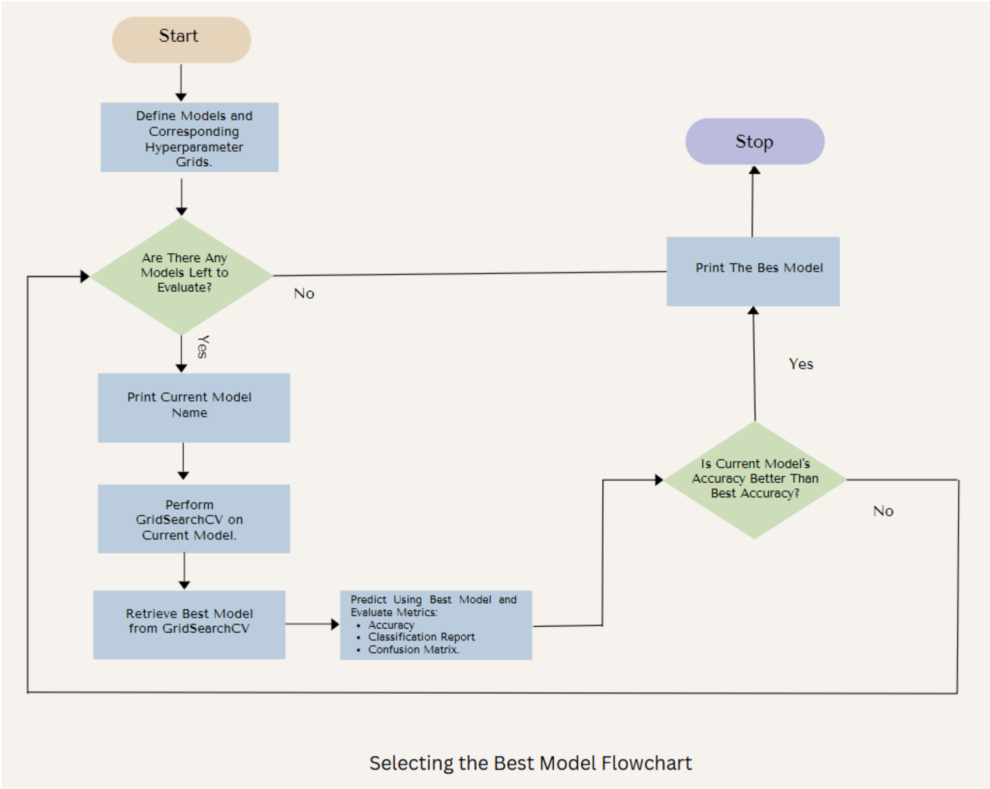

Fig. 7. The study's Flowchart for Selecting the Best Model.

3.6. Predictive System

The flowchart in Fig. 8. illustrates the methodology for using a pre-trained machine learning model to make predictions on new data. The steps outlined, ensure a systematic and accurate application of the model to generate predictions while maintaining consistency with the training pipeline. The process begins with initializing the system for prediction. This ensures that the necessary resources, such as the pre-trained model file and required libraries, are in place.

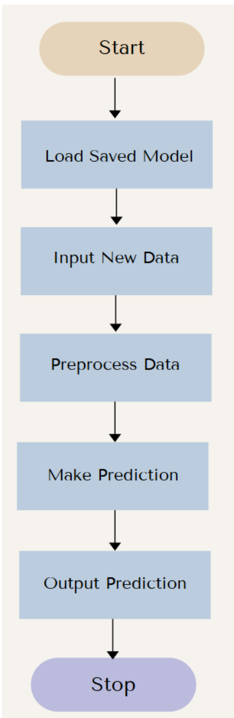

Fig. 8. Flowchart of the Predictive System.

The pre-trained model is loaded using joblib.load(). This step is critical to ensure the model, trained and saved previously, is accessible for making predictions. If the model file is missing or corrupted, appropriate error handling should be implemented.

New data is supplied for prediction. The data format, feature set, and dimensions must match the structure of the data used during model training. This ensures compatibility and avoids runtime errors. Data preprocessing is applied to the input data, including scaling or normalization, using the same techniques and tools (e.g., scaler) utilized during the training phase. This step ensures that the new data aligns with the model's expectations, avoiding discrepancies due to differing scales or feature representations.

The loaded model processes the preprocessed data and generates predictions using the predict() function. This step is the core of the workflow, where the model leverages the learned patterns from the training phase to classify or predict outcomes. The prediction result is displayed or saved. For instance, the predicted class is printed to the console in the code. This step provides the user or system with actionable insights based on the input data. The process concludes after the prediction is output. This signifies the end of the workflow, ensuring no further steps are required unless new data is to be processed.

4. Results

4.1. Cronbach Alpha

Table 5. DigCompEdu Cronbach Alpha Result.

Description	Values
Number of items	22
sum of item variance	21.05
variance of total score	132
cronbach's alpha	0.88

To evaluate the internal consistency of the DigCompEdu instrument, Cronbach's alpha was calculated across the 22 items. The sum of the item variances was 21.05, while the variance of the total score was 132.00. These calculations resulted in a Cronbach's alpha coefficient of 0.88.

4.2. Descriptive Profile of Teacher Proficiency

Based on the proficiency level score in Table 6 the measures of central tendency—mean, mode, and median—derived from the DigCompEdu scores gave us important information on the digital competency levels of Bukidnon's public school teachers after we were able to determine the respondents' proficiency level. These figures provide an overview of the respondents' overall performance and aid in summarizing the score distribution.

Table 6. Measures of Central Tendency.

	Value	Proficiency Level
Mean	43.31137255	Integrator
Mode	44	Integrator
Median	42	Integrator

Figure 9 illustrates a bell-shaped distribution of ICT proficiency scores, with the vast majority of the 1,275 respondents concentrated in the center of the scale, specifically within the score intervals corresponding to the B1 level.

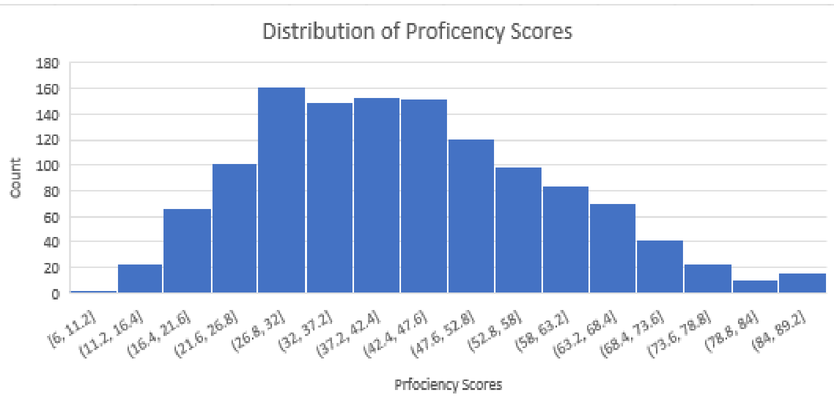

Fig. 9. Distribution of Proficiency Score.

4.3. Impact of Data Balancing (SMOTE-ENN)

After pre-processing we have noticed an imbalanced distribution of class. The class distribution based from the bar chart in Fig.10, the left chart showed unbalanced class distribution. Where certain proficiency levels, such as Integrator (B1) and Explorer (A2), are overrepresented, while others, such as Novice (A1) and Leader (C1), are underrepresented. The largest class is Integrator (B1), followed by Explorer (A2). The smallest class is Novice (A1), with very few examples compared to the dominant classes. The right chart shows a balanced dataset where all proficiency levels now have an equal number of samples after applying oversampling techniques.

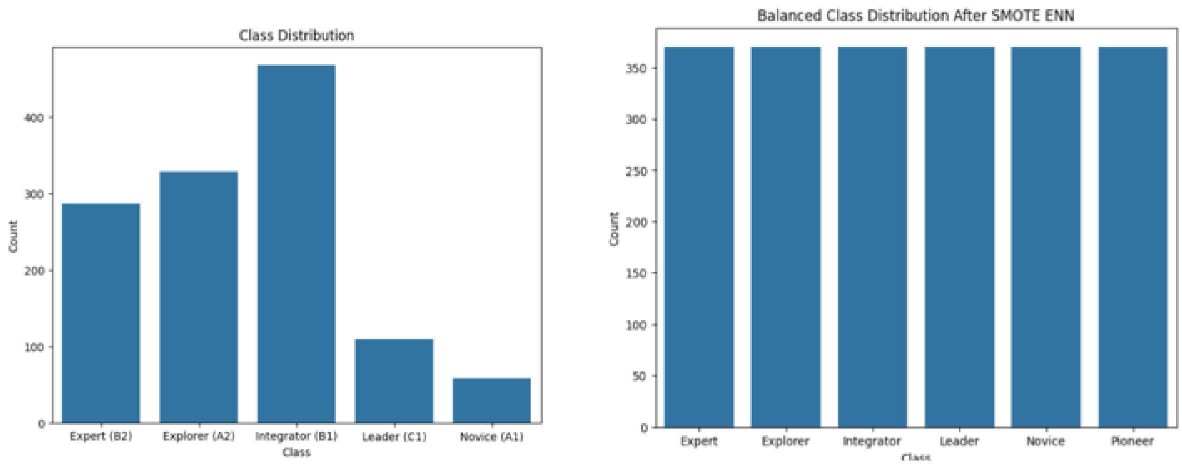

Fig. 10. Unbalanced Class Distribution Versus Balanced Class Distribution.

4.4. Statistical Feature Ranking

To identify the variables with the most significant predictive power for ICT proficiency levels, an ANOVA F-test was conducted. Table 7 presents the top 10 features ranked by their F-scores, all of which demonstrated high statistical significance with p-values of less than 0.001.

Table 7. Top 10 Features Selected Using SelectKBest for Predicting Teacher Proficiency Levels.

Feature Index	Feature Name	Score	P-value
8	Item9	2225.959	< 0.001
9	Item10	1837.871	< 0.001
10	Item11	2098.685	< 0.001
11	Item12	2336.603	< 0.001
14	Item15	2042.004	< 0.001
17	Item18	1623.884	< 0.001
18	Item19	2257.21	< 0.001
19	Item20	1744.328	< 0.001
20	Item21	2631.907	< 0.001
21	Item22	2434.911	< 0.001

The feature with the highest predictive weight is Item 21 (Teaching Online Safety), which yielded an F-score of 2631.907. This was followed closely by Item 22 (Creative Use of ICT) and Item 12 (Planning Learning), with scores of 2434.911 and 2336.603, respectively. Other highly ranked features included Item 19 (Promoting Collaboration) and Item 9 (Monitoring Students).

The F-scores for the top 10 features ranged from a high of 2631.907 down to 1623.884 (Item 18). These ten features were selected as the primary inputs for the subsequent machine learning models, as they represent the variables with the strongest relationship to the target proficiency levels.

4.5. Comparative Model Performance

Table 8 shows the classification performance of the four models—Logistic Regression, SVM, Random Forest, and K Nearest Neighbors—was evaluated both before and after the application of SMOTE-ENN oversampling.

The Performance ranged from 84% to 89%. Logistic Regression recorded the lowest accuracy at 84%, while SVM and K Nearest Neighbors achieved 85% and 86%, respectively. Random Forest achieved the highest accuracy at 89%. The Random Forest achieved the highest precision at 86%. The other three models—Logistic Regression, SVM, and K Nearest Neighbors—all recorded a macro average precision of 84%. The Random Forest outperformed the other models with a recall of 84%. This was followed by SVM at 83%, K Nearest Neighbors at 82%, and Logistic Regression at 81%.

Table 8. Classification Metrics Summary without and with oversampling SMOTE-ENN.

Model	Metric	Without Oversampling	With Oversampling
Logistic Regression	Accuracy	69%	84%
	Macro Avg Precision	80%	84%
	Macro Avg Recall	64%	81%
	Macro Avg F1-Score	65%	83%
SVM	Accuracy	70%	85%
	Macro Avg Precision	58%	84%
	Macro Avg Recall	65%	83%
	Macro Avg F1-Score	60%	83%
Random Forest	Accuracy	68%	89%
	Macro Avg Precision	75%	86%
	Macro Avg Recall	68%	84%
	Macro Avg F1-Score	70%	85%
K Nearest Neighbors	Accuracy	67%	86%
	Macro Avg Precision	65%	84%
	Macro Avg Recall	66%	82%
	Macro Avg F1-Score	64%	83%

4.6. Error Analysis (Confusion Matrices)

In the confusion matrices presented in Fig.11. and Fig.12., several frequent misclassification patterns were observed across the four models. In Fig.11. with SMOTE-ENN oversampling, the most prominent misclassification occurred with the Expert class, which was frequently mistaken for the Explorer class. This was particularly evident in the Logistic Regression model (31 instances) and the SVM model (33 instances). The Expert class was also frequently misclassified as a Leader, with 10 instances in Logistic Regression, 7 in Nearest Neighbors, and 6 in Random Forest. The Integrator class was occasionally misidentified as either an Expert or a Leader across several models.

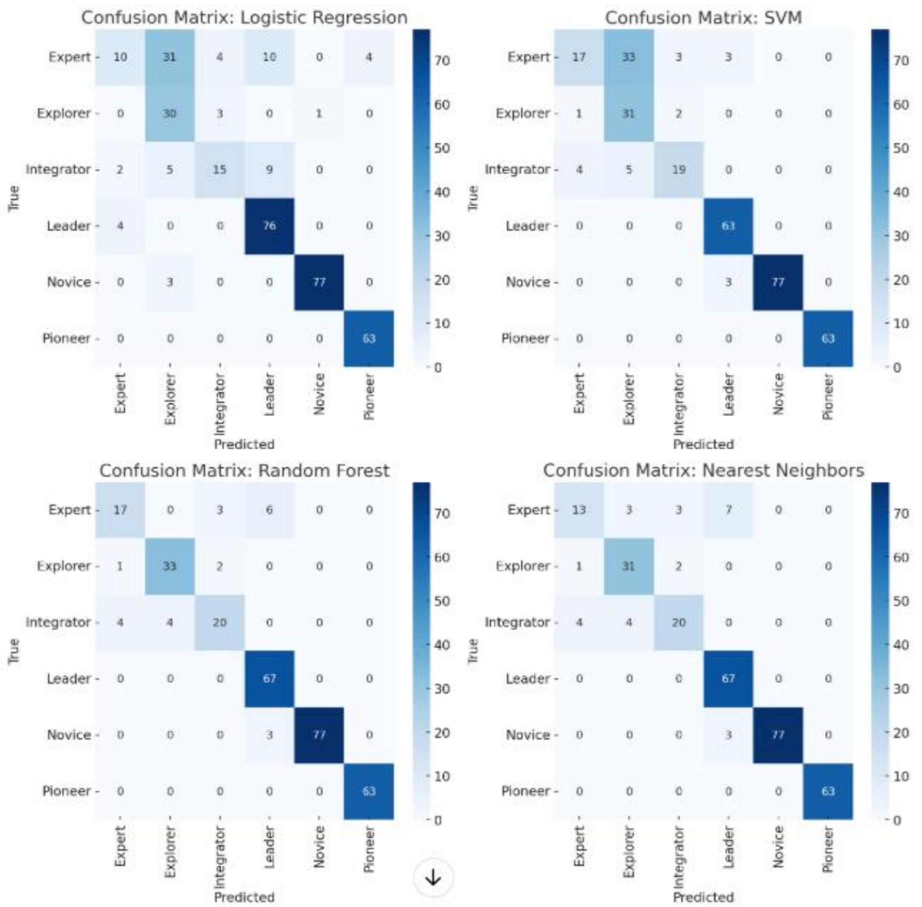

Fig. 11. Confusion Matrix of Algorithm with Oversampling SMOTE-ENN.

In Fig.12. without SMOTE-ENN oversampling, the Explorer class, Class 1 was frequently misclassified as a Novice (Class 4), notably with 8 instances in both Logistic Regression and K-NN. The Expert class, Class 0 was commonly mistaken for an Integrator (Class 2), with 5 instances in Random Forest and 4 in both Logistic Regression and K-NN. In the SVM and K-NN models, several Integrator, Class 2 instances were misclassified as Explorers, Class 1.

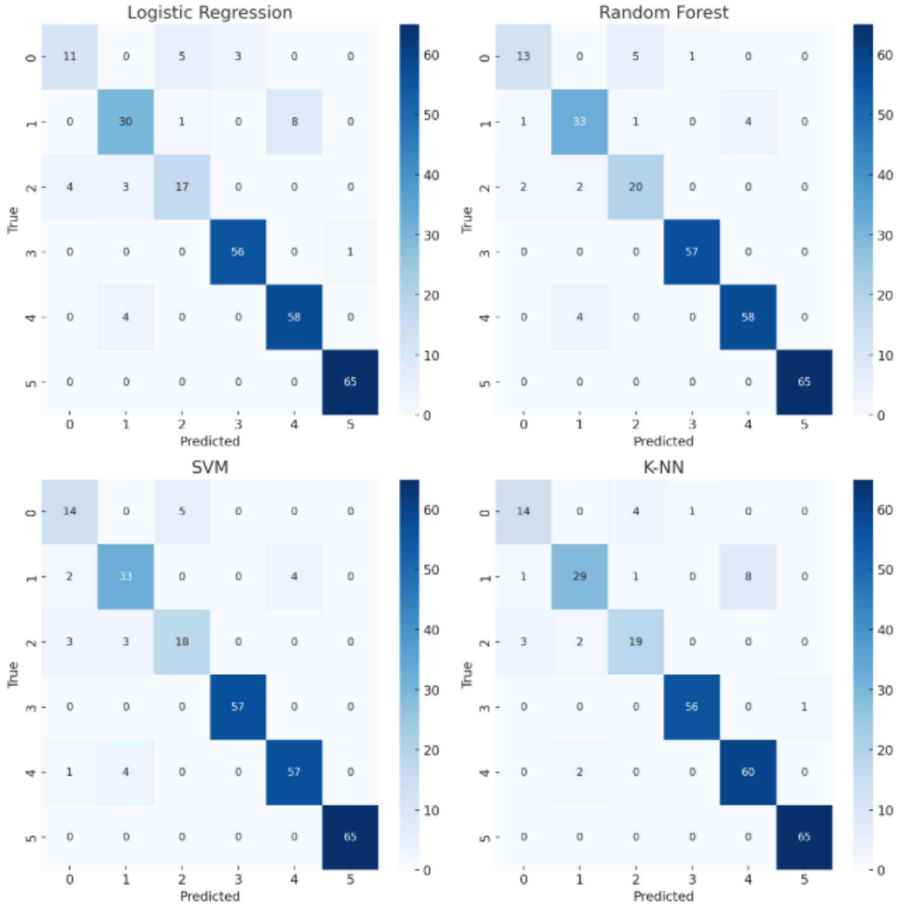

Fig. 12. Confusion Matrix of Algorithm without Oversampling SMOTE-ENN.

5. Discussion

5.1. Interpretation of "Pivot Skills"

The feature selection process identified 10 critical "pivot skills" that most significantly influence a teacher’s progression through the DigCompEdu levels. Rather than mere technical rankings, these items reflect the specific pedagogical shifts required for digital transformation in the Bukidnon region.

1. *Digital Responsibility and Safety (Item 21 and Item 18):*

Item 21 (Teaching Online Safety) emerged as the strongest predictor of proficiency. This suggests that a teacher’s ability to manage digital risk is a foundational "gatekeeper" skill. Teachers who prioritize digital citizenship and verify information reliability (Item 18) demonstrate a sophisticated understanding of the digital environment, moving them beyond basic tool usage toward secure and professional classroom management.

2. *Transformative and Creative Pedagogy (Item 22 and Item 15):*

The high significance of Item 22 (Creative Use of ICT) underscores the transition from "Substitution" to "Transformation." High-proficiency teachers do not merely use technology to replace traditional tools; they use it to foster student-led problem-solving. Furthermore, addressing digital access and compatibility (Item 15) indicates that "Expert" level teachers are those who can creatively adapt their digital strategies to meet the diverse socio-economic needs of their students.

3. *Data-Informed Assessment (Item 12 and Item 11):*

Items 12 and 11, concerning Planning and Assessing Learning, rank highly as predictors. This highlights a shift toward data-driven instruction. Teachers who utilize ICT for continuous progress monitoring are better equipped to provide personalized learning pathways. This suggests that moving into the "Expert" (B2) tier is closely tied to a teacher's ability to use digital data to refine their teaching strategies.

4. Collaborative Learning Environments (Item 19, Item 9, and Item 10):

The importance of Promoting Collaboration (Item 19) and Monitoring Online Activities (Items 9 and 10) reflects a move from teacher-centered to student-centered learning. High-proficiency teachers use digital platforms to facilitate peer-to-peer interaction. These items are the primary drivers for teachers moving into the "Leadership" tier (C1), where the educator acts as a facilitator of a digital learning community.

5.2. Addressing the Class Imbalance: Explain why SMOTE-ENN was necessary

The implementation of SMOTE-ENN was a critical step in this research to overcome the inherent limitations of the raw dataset, which was characterized by significant class imbalance. As shown in Figure 12, the models trained without oversampling were susceptible to "majority bias," particularly clustering around the middle-tier ICT proficiency levels (often corresponding to B1-level benchmarks such as "Integrator" or "Leader"). This clustering reflects the reality of many educational environments where the bulk of the population falls into a "competent" middle ground, leaving the extremes—the high-performing "Experts" (C2) and the "Novices" or "Pioneers" (A1)—underrepresented.

Without the intervention of SMOTE-ENN, the algorithms risked becoming "lazy" models. A model is considered "lazy" when it optimizes for overall accuracy by simply predicting the majority class for most cases. While this might result in a high "Accuracy" percentage on paper, it does so by ignoring the nuances of the minority classes. In a real-world educational context, a lazy model is ineffective because it fails to distinguish the specific traits of teachers who either require urgent foundational support or those who possess the advanced skills necessary to lead institutional change.

Applying SMOTE-ENN, the dataset was balanced to ensure that the unique features of the "Expert" and "Novice/Pioneer" classes were as visible to the model as those of the majority. This transformed the system into a "policy-ready" tool. From a management and policy perspective, identifying outliers is more valuable than confirming the status of the average performer.

Specifically:

A1 Teachers (Novice/Pioneer): Accurately identifying this group allows policymakers to design and mandate targeted remedial ICT training, ensuring no teacher is left behind in the digital transition.

C2 Teachers (Expert): Precisely identifying these outliers allows for the selection of "Digital Champions" or mentors who can lead peer-to-peer coaching and drive innovation within their schools.

The enhanced performance demonstrated in Figure 11—where the "Expert" class saw a massive reduction in misclassification—confirms that the model has moved beyond simple statistical probability and toward a strong diagnostic instrument. Mitigating the majority bias, this study provides a framework that is not just mathematically accurate, but practically actionable for educational stakeholders seeking to optimize professional development resources.

5.3. Why Random Forest Succeeded

The superior performance of the Random Forest (RF) model, which achieved 89% accuracy, can be attributed to its ensemble architecture. Unlike Logistic Regression, which assumes a linear relationship between features and the target variable, Random Forest utilizes multiple decision trees to capture the complex, non-linear interactions inherent in educational data.

Teacher proficiency is rarely a linear progression where the addition of one skill leads directly to the next level. Instead, the DigCompEdu framework represents a multidimensional hierarchy where specific combinations of "pivot skills"—such as the simultaneous mastery of digital safety (Item 21) and creative pedagogy (Item 22)—determine a teacher's classification. Random Forest's ability to perform recursive partitioning allows it to identify these specific "decision paths," making it more effective than distance-based models (KNN) or hyperplane-based models (SVM) at distinguishing between closely related tiers like 'Expert' and 'Integrator.

While Random Forest provided the highest accuracy, it is often criticized in educational literature as a "black box" model due to the difficulty in interpreting the logic of hundreds of internal decision trees. As noted by Barbierato and Gatti [10] in the Literature Review, high-accuracy ensemble methods can sometimes lack the transparency required for teachers to understand why a specific prediction was made.

However, this study mitigated the "black box" limitation by integrating a robust Feature Selection phase (ANOVA F-test) before model training. Through identifying the top 10 "pivot skills" (Table 6), the research provides the necessary transparency to move the model from a purely mathematical tool to a practical diagnostic one. Even though the internal weighting of the Random Forest model is complex, the clear identification of specific competencies—such as digital responsibility and collaborative learning—offers educational policymakers a "window" into the model's logic. This ensures that the high accuracy (89%) does not come at the cost of actionable pedagogical insight, providing a balance between predictive power and professional utility.

5.4. Implications for Local Policy (Bukidnon Context)

The clustering of the majority of Bukidnon teachers at the Integrator (B1) level indicates a successful regional baseline of ICT literacy; however, it also reveals a professional "plateau." To advance teachers toward Expert (B2) and Leader (C1) levels, the Department of Education should shift its fiscal and training resources from general software tutorials to the following targeted interventions:

1. Prioritizing Digital Citizenship as a Gatekeeper Skill

Since Item 21 (Teaching Online Safety) emerged as the strongest predictor of proficiency, policy should prioritize "Digital Safety and Ethics" as the primary requirement for advanced certification. In the Bukidnon context, where students are increasingly exposed to unmonitored digital environments, a teacher’s ability to act as a digital guardian is the foundational skill that separates basic users from high-proficiency educators. Training should focus on data privacy, identifying misinformation (Item 18), and secure online classroom management.

2. Incentivizing Transformative and Creative Pedagogy

The high F-score for Item 22 (Creative Use of ICT) underscores a critical policy gap: many teachers use technology merely to replace traditional tools (substitution) rather than to redefine learning tasks. DepEd divisions should launch "Incentivized Innovation Labs" that reward teachers for using digital tools to foster student-led problem-solving. Policies should move beyond counting the number of computers in a school and start measuring the creativity of their application in the curriculum.

3. Addressing the Rural Digital Divide (Inclusive ICT)

The significance of Item 15 (Addressing Digital Access and Compatibility) is particularly relevant for the diverse socio-economic landscape of Bukidnon. High-proficiency teachers in this study were those who could adapt their digital strategies for students with limited internet access or hardware. Regional policy should support the development of "Low-Bandwidth/Offline-First" digital resources, ensuring that ICT integration does not inadvertently marginalize students in more remote municipalities.

4. Shifting to Data-Informed Assessment

The ranking of Items 11 and 12 (Planning and Monitoring Learning) indicates that moving to the "Expert" tier is closely tied to a teacher's ability to use digital data to refine instruction. Local school heads should encourage the use of ICT for continuous progress monitoring and student self-assessment. This shift from "technology as a delivery tool" to "technology as an analytical tool" is essential for the data-driven decision-making required at the Leadership (C1) level.

6. Conclusion

This study successfully leveraged machine learning techniques to address the challenge of evaluating and predicting ICT proficiency levels among public school teachers in Bukidnon. Utilizing the DigCompEdu framework as a structured roadmap for digital competence, the research provides a clear path forward for regional educational development.

6.1. Summary of Contributions

The primary contribution of this research is the development of a high-precision predictive system that identifies the specific drivers of teacher digital growth. Utilizing a dataset of 1,275 responses, this study successfully built a classification model that achieved a near-perfect 89% accuracy using the Random Forest algorithm. A critical methodological contribution was the application of the SMOTE-ENN oversampling technique, which effectively mitigated the "majority bias" inherent in the dataset, ensuring the model could accurately identify teachers at both the highest (Pioneer) and lowest (Novice) ends of the spectrum.

6.2. Actionable Recommendations

To translate these findings into actionable policy, a strategic roadmap for the Department of Education (DepEd) is recommended, focusing on three primary pillars of professional development. First, the Department should transition from generalized computer training toward targeted "pivot skill" programs that prioritize the specific predictors identified in this study—Online Safety, Creative Use of ICT, and Planning Learning—as mastering these areas offers the most efficient path for elevating teachers from the ‘Integrator’ to ‘Expert’ level. Second, this progression should be supported by a formal mentorship network that utilizes the study’s predictive model to identify "Digital Champions" (C1 and C2 level teachers) who can provide localized, peer-to-peer coaching to bridge the regional "Pioneer Gap." Finally, by institutionalizing these predictive tools within national teacher evaluation systems, DepEd can establish data-driven certification pathways. This shift would replace one-size-fits-all curricula with customized professional development modules tailored to each teacher's specific proficiency gaps, ensuring a more precise and effective digital transformation across the teaching force

6.3. Limitations and Future Work

While this study provides a strong predictive tool, it is not without limitations. A primary limitation is that the data relies on self-reported survey responses, which may be subject to social desirability bias or overestimation of one's own skills. Future research should look to integrate objective performance tasks—such as digital portfolio reviews or classroom observations—to validate self-reported scores. Additionally, longitudinal studies are recommended to track how teachers move between proficiency tiers over several years. This would allow researchers to measure the long-term impact of specific training interventions and further refine the predictive power of the model as teacher skills evolve.

All the Declarations and Statements

Author Contributions Statement

Nathalie Joy G. Casildo– Conceptualization, Methodology, Software, Formal Analysis, Investigation.
Gladys S. Ayunar– Investigation (Data Acquisition), Resources (Institutional Coordination).
Jinky G. Marcelo – Conceptualization (Theoretical Framework), Validation.
Kent Levi A. Bonifacio – Investigation (Data Acquisition), Resources (Institutional Coordination).

All authors have read and agreed to the published version of the manuscript.

Conflict of Interest Statement

The authors declare no conflicts of interest.

Funding Declaration

This research was supported by Central Mindanao University "R-0355" "Information and Computing Assessment and Solutions".

Data Availability Statement

The data that support the findings of this study are available on request from the corresponding author. The data are not publicly available to protect the privacy of the research participants.

Ethical Declarations

This study followed the ethical standards of Central Mindanao University and was formally approved by the DepEd Bukidnon divisions. Informed consent was obtained from all 1,275 participants prior to data collection. To ensure participant protection, all data were anonymized and processed to maintain complete respondent confidentiality.

Acknowledgement

We sincerely thank the experts for their professional evaluation and valuable recommendations, which have contributed to improving the quality of the experiment and the reliability of its results. We would like to thank Central Mindanao University for funding the research under the project R-0355.

Declaration of Generative AI in Scholarly Writing

During the preparation of this manuscript, the authors used Google AI Studio to improve the language and readability of the text. After using this tool, the authors reviewed and edited the content as needed and take full responsibility for the accuracy and integrity of the information contained in the work. No AI tools were used for data analysis or drawing insights from the research data.

Abbreviations

None.

Appendix

None.

References

[1] Hussain, I., Suleman, Q., & Shafique, F. (2017). Effects of Information and Communication Technology (ICT) on Students' Academic Achievement and Retention in Chemistry at Secondary Level. Journal of Education and Educational Development, 4(1), 73-93.

[2] Scarpin, J. E., Mondini, V. E. D., & Scarpin, M. R. S. (2018). Technology Acceptance Factors and Student Retention in Online Courses. E-Journal of Business Education and Scholarship of Teaching, 12(3), 44-68.

[3] Seenivasan, R. (2024). ICT in Education: A Critical Literature Review and Its Implications. International Journal of Finance, Insurance and Risk Management, 14(1), 12-27.

[4] Mohiman, A. U., Ullah, N., & Zafar, J. M. (2024). Identification of Information and Communication Technologies'(ICTs) Needs regarding Professional Development and Modern Teaching Methods for Madarassa Teachers. Pakistan Journal of Humanities and Social Sciences, 12(2), 1855-1862.

[5] Bouasangthong, V., Phonekeo, S., Soukhavong, S., Thalungsy, K., Phongphanit, T., Vathana, P., ... & Phengphilavong, C. (2024). An Investigation into the Conditions of ICT Application at the Teacher Education Institutions. *Indonesian Journal of Educational Research and Technology*, *4*(1), 89-104.

[6] Redecker, C., & Punie, Y. (2017). Digital competence of educators. *Edited by Yves Punie*.

[7] Johnson, F.; Schmit, J.; Schneider, C.; Rossa, H.; Müller, L. Evaluating the Effectiveness of an Extracurricular Teacher Education Training Program for DigCompEdu Competences. Educ. Sci. 2024, 14, 1390. https://doi.org/10.3390/educsci14121390

[8] Minamatov, Y. E. O. G. L., & Nasirdinova, M. H. Q. (2022). Application of ICT in education and teaching technologies. *Scientific progress*, *3*(4), 738-740.

[9] Forero-Corba, W., & Bennasar, F. N. (2024). Techniques and applications of Machine Learning and Artificial Intelligence in education: a systematic review. RIED-Revista Iberoamericana de Educación a Distancia, 27(1).

[10] Barbierato, E., & Gatti, A. (2024). The challenges of machine learning: A critical review. Electronics, 13(2), 416.

[11] Sarker, S., Paul, M. K., Thasin, S. T. H., & Hasan, M. A. M. (2024). Analyzing students' academic performance using educational data mining. Computers and Education: Artificial Intelligence, 7, 100263.

[12] Pratama, I., Prasetyaningrum, P. T., Chandra, A. Y., & Suria, O. (2024). Measuring Resampling Methods on Imbalanced Educational Dataset's Classification Performance. Register: Jurnal Ilmiah Teknologi Sistem Informasi, 10(1), 1-11.

[13] Förster, K., & Grafe, S. (2021, March). ICT-related Educational Competencies of Teacher Educators from an Intercultural Perspective. A Systematic Analysis of Competency Frameworks. In Society for Information Technology & Teacher Education International Conference (pp. 1586-1595). Association for the Advancement of Computing in Education (AACE).

[14] Chachoui, Y., Azizi, N., Hotte, R., & Bensebaa, T. (2024). Enhancing algorithmic assessment in education: Equi-fused-data-based SMOTE for balanced learning. Computers and Education: Artificial Intelligence, 6, 100222.

[15] Goswami, S., & Singh, A. K. (2024). A literature survey on various aspect of class imbalance problem in data mining. Multimedia Tools and Applications, 1-26.

[16] Cabero-Almenara, J., Gutiérrez-Castillo, J. J., Palacios-Rodríguez, A., & Barroso-Osuna, J. (2020). Development of the teacher digital competence validation of DigCompEdu check-in questionnaire in the university context of Andalusia (Spain). Sustainability, 12(15), 6094.

[17] Cabero-Almenara, J.; Palacios-Rodríguez, A. Digital Competence Framework for Educators «DigCompEdu». Translationandadaptationof«DigCompEduCheck-In»questionnaire. EDMETIC2020,9,213–234.

[18] Ayunar, G. S., Casildo, N. J. G., Talandron-Felipe, M. M. P., Bonifacio, K. L. A., Marcelo, J. G., & Sebuguero, F. S. (2024, November 25–29). Determinants of ICT competency among public school teachers in Bukidnon. Paper presented at the 32nd International Conference on Computers in Education (ICCE 2024), Manila, Philippines.

Authors' Profiles

Nathalie Joy G. Casildo holds a Master of Information Technology degree. She earned her graduate degree from Central Mindanao University, Philippines. She serves as a Faculty Researcher at Central Mindanao University. Her professional work focuses on applying machine learning and data analytics to real-world challenges, particularly in the fields of ICT education and operational optimization. Ms. Casildo's technical expertise includes predictive analytics, intelligent systems, and cloud technologies. Her current research interests include data-driven solutions for educational improvement, information security, and AI-driven applications.

Gladys S. Ayunar holds a Master of Information Systems from the University of the Philippines Open University. She is a Faculty Member at Central Mindanao University (CMU). Her professional background includes extensive expertise in ICT instruction, academic research, and the development of instructional materials. Ms. Ayunar has co-authored various instructional materials and presented research on technology-enhanced learning and decision support systems. Her current research interests include data science, educational technology, and leveraging machine learning for academic interventions.

Jinky G. Marcelo is currently pursuing Ph.D. studies in Computer Science at De La Salle University, Philippines. She previously earned her degrees in information technology and computer science from Central Mindanao University. She is a faculty member at Central Mindanao University. Her professional publications span diverse areas such as camouflaged image detection, vision-based detection systems, and intelligent tutoring systems. Mrs. Marcelo specializes in object detection models, generative adversarial networks, and educational technology. She is an active member of professional organizations, including the Asia Pacific Neural Network Society and the Computing Society of the Philippines.

Kent Levi A. Bonifacio holds a Master of Information Systems from the University of the Philippines Open University. He also holds multiple professional certifications, including Bloomberg Financial Markets Professional and Advanced Technical Analysis. He is a Faculty Member at Central Mindanao University (CMU), where he leads the College of Information Sciences and Computing. He formerly served as a College Dean at the same institution. Mr. Bonifacio's research focuses on educational technology, programming education, and computer-assisted learning. He has authored several publications examining narrative-centered digital games and programming pedagogy.

How to cite this paper

Nathalie Joy G. Casildo, Gladys S. Ayunar, Jinky G. Marcelo, Kent Levi A. Bonifacio, "Leveraging Machine Learning to Predict ICT Proficiency levels among Public School Teachers in Bukidnon", International Journal of Modern Education and Computer Science(IJMECS), Vol.18, No.3, pp. 17-35, 2026. DOI:10.5815/ijmecs.2026.03.02