Research Paper

Development of a classification model for student programming anxiety levels using logistic regression algorithm

Eduardo R. Yu II[1*], Elmerito D. Pineda[2], Isagani M. Tano[3], Jaime P. Pulumbarit[4], Jonilo C. Mababa[5]
[1*, 2, 3, 4, 5]Graduate Studies Department, La Consolacion University Philippines.

Article Info

Article History:
Received: 07 March 2025
Revised: 24 May 2025
Accepted: 02 July 2025
Published: 09 July 2025

Keywords:
Computer Programming
Anxiety
Machine Learning in Education
Mental Health

ABSTRACT

As programming increasingly becomes a core competency across diverse academic disciplines, mitigating programming anxiety is essential to fostering student engagement, confidence, and academic success in computing-related courses. This study addresses the need for an automated and accurate classification model capable of identifying students currently experiencing programming anxiety. A classification model was developed using the Logistic Regression algorithm, guided by the Cross-Industry Standard Process for Data Mining (CRISP-DM) framework. The methodology involved systematic data preprocessing, feature selection, and the use of the Synthetic Minority Over-sampling Technique (SMOTE) to handle class imbalance. Five supervised classification algorithms were evaluated and compared: Logistic Regression, Support Vector Machine, Naïve Bayes, Random Forest, and Decision Tree. Among these, Logistic Regression produced the best results, achieving an F-measure of 96.77 percent, an accuracy of 97.75 percent, a precision of 96.88 percent, a recall of 96.70 percent, a Cohen's kappa of 0.950, a mean absolute error of 0.0225, a root mean squared error of 0.1464, a relative absolute error of 0.03 percent, and a root relative squared error of 30.71 percent. The resulting model offers practical value for researchers and educators by enabling the automatic detection of programming anxiety, with strong potential for integration into institutional platforms such as learning management systems and academic support services.

Corresponding Author:
Eduardo R. Yu II
Graduate Studies Department, La Consolacion University, Philippines.
Email: eduardoryuii@gmail.com

Journal of Artificial Intelligence, Machine Learning and Neural Network (JAIMLNN)

Copyright © 2025 The Author(s). This is an open access article distributed under the Creative Commons Attribution License, (http://creativecommons.org/licenses/by/4.0/) which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

1. INTRODUCTION

The ability to write computer programs is increasingly recognized as an essential skill across disciplines [1], [2], driving the inclusion of programming subjects in computing-related degree programs [3], [4] and contributing to growing student enrollment in these courses [5]. As this trend continues, educational institutions face the critical challenge of supporting student success in programming subjects, which are foundationl to many computing disciplines. However, programming anxiety has emerged as a significant barrier, characterized by emotional distress and cognitive overload during coding tasks, often resulting in reduced academic performance, avoidance behaviors, and elevated attrition rates [6], [7], [8], [9].

To assess this domain-specific anxiety, the Programming Anxiety Scale (PASc) [10] provides a validated measurement tool. It consists of Likert-scale items categorized under two core dimensions: Classmates Anxiety, which captures peer-related stress and social comparison, and Self-Confidence, which reflects perceived deficiencies in problem-solving, code writing, and comprehension. Despite its reliability, the PASc relies on manual administration and scoring, limiting its applicability for real-time or large-scale educational interventions.

Although machine learning techniques are increasingly employed in educational data mining to classify academic outcomes and identify at-risk students, limited research has explored the development of supervised machine learning models for the automatic detection of current programming anxiety based on available student data.

The primary objective of this study is to develop an automated classification model capable of detecting students' current programming anxiety levels. To achieve this, the study adopts the Cross Industry Standard Process for Data Mining (CRISP-DM) as a structured methodology and evaluates the performance of five supervised classification algorithms: J48 Decision Tree, Random Forest, Support Vector Machine, Logistic Regression, and Naive Bayes. These models are evaluated using multiple performance metrics, including F-measure, accuracy, precision, recall, and Cohen's kappa, as well as error-based indicators such as Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), Relative Absolute Error (RAE), and Root Relative Squared Error (RRSE). This rigorous evaluation ensures the reliability and effectiveness of the proposed classification model in educational settings.

This work contributes to the field of educational data mining by providing a validated and interpretable model for classifying programming anxiety levels. It offers a methodological foundation for future studies in educational machine learning and mental health analytics, with the potential to support timely interventions and enhance student outcomes.

2. RELATED WORK

Recent literature has extensively explored the application of supervised machine learning algorithms and ensemble models in predicting anxiety and related mental health conditions across various populations. Studies employing psychometric tools like Depression, Anxiety, and Stress Scales (DASS) 21 and 42 have shown high accuracy levels, with [11] reporting perfect classification using RFT. However, the reliance on self-reported data may limit generalizability due to inherent subjectivity. [12] Integrated demographic and occupational features, achieving 82.6% accuracy with CatBoost in anxiety prediction among seafarers. Subsequent studies [13], [14] underscored issues such as class imbalance and model optimization, which remain underexplored in mainstream research. [15], [16] further emphasized the role of context-specific predictors, particularly during the COVID-19 pandemic, where behavioral and environmental stressors significantly influenced model outcomes. Study by [17], used a small dataset and highlighted the value of recall and Cohen's kappa as evaluation metrics, advocating for nuanced model assessment.

This contextualization is also evident in educational settings. [18] adapted CRISP-DM (Cross-Industry Standard Process for Data Mining) to assess mental health among higher education students using World Health Organization Quality of Life (WHOQOL) and DASS-21, with LR and Neural Networks yielding 68–88% accuracy. Social and academic variables emerged as key predictors. [19] Utilized biomedical and demographic data to build interpretable anxiety models for undergraduates, emphasizing transparent feature selection. [20] Extended this approach to children aged 10–15, identifying family income, exposure to violence, and academic performance as significant factors. A meta-analysis by [21] confirmed the effectiveness of SVM and LR while noting Python's dominance and the widespread use of accuracy and precision metrics.

In the Philippine context, few studies have examined ML-based mental health prediction. [22] Applied SVM, DT, and NB to classify academic stress among college students, with SVM achieving 95% accuracy. Key stressors included academic workload and unrealistic expectations, especially during remote learning. [23] Found anxiety prevalence rates of 14% among students and 16% among teachers, identifying stress, fear of COVID-19, and lack of support systems as major predictors.

While these studies demonstrate the potential of machine learning in mental health prediction, they tend to generalize across mental health conditions and often overlook domain-specific constructs such as programming anxiety. Programming anxiety refers to the cognitive, emotional, and behavioral difficulties students experience specifically in learning computer programming. This study addresses this gap by focusing on programming anxiety and applying supervised ML models with context-specific features. By incorporating academic, psychological, and socio-environmental indicators, the research aims to support early identification and targeted interventions within educational settings.

3. METHODOLOGY

This study used descriptive and developmental research to analyze and describe the process of building [24], [25] a classification model with machine learning techniques, following the adapted CRISP-DM methodology, and evaluating the final product. The process was grounded in the supervised machine learning paradigm, wherein the target class guided the training, validation, and evaluation of five classification algorithms.

Cross Industry Standard Process for Data Mining (CRISP-DM)

The CRISP-DM framework offers a structured approach for developing classification models aimed at detecting potential outcomes [26]. The principal phases implemented in this study were as follows:

1. Problem and Data Understanding

This initial phase defined the problem of programming anxiety levels as a classification task and identified relevant data variables. A cross-sectional dataset of 1,732 instances was obtained from computing students at a Philippine public university during the first semester of 2023–2024. The data was gathered by the Computer Studies Department over an estimated period of 4 to 6 weeks, which involved survey distribution across classes, collection of responses, and follow-ups for non-respondents. The dataset summarizes in Table 1 included academic performance, demographics, learning styles, and self-reported scores from a validated Programming Anxiety Scale.

It is important to note that the authors did not administer the survey; rather, the department managed the entire data collection process. While the original Programming Anxiety Score generated ranging from 11 to 55, the actual numerical scores were not provided for security and privacy reasons. Instead, the department internally aggregated the scores and released only the corresponding categories: 606 instances were labeled as low programming anxiety (scores 11–33) and 1,126 as high programming anxiety (scores 34–55). Exploratory data analysis was conducted to assess distributions, detect missing values, and examine correlations, providing a foundation for informed feature selection and model development in later phases.

Table 1. Attributes and Value Provided by the Computer Studies Department.

Attribute	Description	Values
Gender	Identifies the student's gender.	(Male/Female)
Age	Student's age in complete years.	Numerical
Working Student	Whether the student is employed while studying.	(Yes/No)
Parents with Higher Education	Whether the student's parent completed college or higher education.	(Yes/No)
Family Monthly Income	Whether the family earns eighteen thousand pesos or more monthly.	(Yes/No)
Number of Siblings	Total number of the student's siblings.	Numerical
Course Enrolled	Degree program the student is taking.	(BSIT, BSIS, BSCS, BSEMC)
Current Year Level	Student's year level in college.	Numerical
Previous Semester GPA Remark	Academic remark based on previous GPA.	(Above Average Average, Below Average)
Computer Prog. 1 Final Grade Remark	Academic remark based on the final grade in Programming 1.	(Above Average Average, Below Average)
Computer Prog. 2 Final Grade Remark	Academic remark based on the final grade in Programming 2.	(Above Average Average, Below Average)
Senior High School Track	Whether the student took the ICT or STEM track.	(Yes/No)
Multiple ICT Equipment Access	Whether the student owns more than one ICT device.	(Yes/No)
Uses Mobile Data	Whether mobile data is the student's primary internet source.	(Yes/No)
Multi-modal Learner	Whether the student prefers a multi-modal learning style.	(Yes/No)
Study Hours per Week	Whether the student studies ten hours or more per week on average.	(Yes/No)
Sleep Hours per Night	Whether the student sleeps six hours or more per night on average.	(Yes/No)
In a Relationship	Whether the student is currently in a romantic relationship.	(Yes/No)
Programming Anxiety Level	Student’s level of programming anxiety.	(High/Low)

2. Data Pre-Processing

This phase involved rigorous data preprocessing to prepare the dataset for machine learning. Inaccuracies, missing values, and inconsistencies were addressed using Python libraries such as Scikit-Learn, NumPy, Pandas, and Matplotlib. Categorical variables were transformed via One-Hot Encoding, and Synthetic Minority Oversampling Technique (SMOTE) was used to address class imbalance [27].
Feature selections were followed, identifying nine significant as shown in Table 2. These preprocessing steps ensured data quality, enhanced interpretability, and established a robust foundation for model training and evaluation.

Table 2. Selected Significant Attributes

Feature Selection Technique	Significant Attributes
Random Forest Feature Importances	1. Working Status
Sequential Feature Selection (SFS)	2. Course 3. Current Year Level 4. Prev. Sem GPA 5. Computer Prog. 1 Final Grade Remark 6. Senior High School Track 7. Multiple ICT Equipment Access 8. Multi-modal Learner 9. Sleep Hours per Night

3. Model Engineering

This phase focused on evaluating five classification algorithms selected based on their demonstrated effectiveness in previous anxiety prediction studies. Model development and testing were conducted using Python in a cloud-based environment (Google Colab) with libraries such as scikit-learn, pandas, and NumPy. This platform provided free access to computing resources, enabling efficient data loading, algorithm configuration, and simulations. The computational setup supported rapid experimentation and iterative refinement, ensuring robust algorithm performance in classifying students' programming anxiety levels based on established experimental methodologies.

4. Model Evaluation

The model's performance was evaluated using a separate test dataset and standard metrics: accuracy, precision, recall, F-measure, and Cohen's kappa. These metrics assess overall correctness, positive class identification, sensitivity, balanced performance, and agreement beyond chance, respectively [28], [29]. Complementing classification metrics, error analysis included MAE, RMSE, RAE, and RRSE. These quantify classification deviations, emphasize larger errors, and provide standardized error comparisons. To ensure the robustness of the model's performance, Stratified 10-Fold Cross-Validation was employed. This technique involved splitting the dataset into 10 distinct folds, where each fold served as a validation set exactly once, while the remaining nine folds were used for training. Together, these methods provided a rigorous, multifaceted evaluation of the model's reliability and effectiveness in classifying students' current programming anxiety levels.

5. Model Deployment

The best model was then serialized and saved in a pickle file format (pkl). This format was chosen because it allows the model to be easily loaded and integrated into the web-based service.

4. RESULTS AND DISCUSSION

The following sections present the findings of the study, including the evaluation results of five classification algorithms, the performance of Logistic Regression as the most effective model for predicting students' programming anxiety levels, and its integration into a web-based system.

A) Performance of Classification Models for Student Programming Anxiety Levels

The findings from

Figure 1 and Table 3 show that Logistic Regression outperformed other models, achieving 96.77% F-measure, 97.75% accuracy, 96.88% precision, 96.70% recall, and a Cohen's kappa of 0.950, indicating strong predictive capability. SVM followed closely with 97.69% accuracy and a kappa of 0.949. NB, RF, and DT had lower accuracy, between 94.57% and 94.98%.

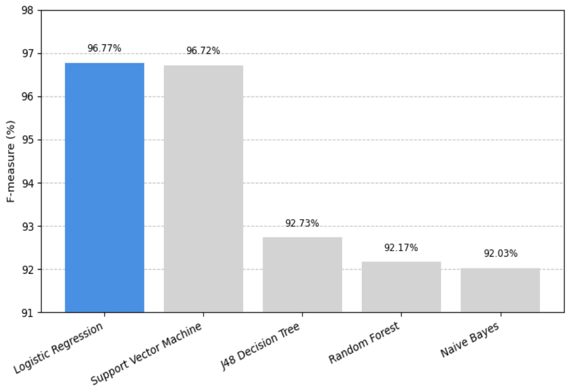

Figure 1. Performance Results of Classification Algorithms Using F-Measure.

Table 3. Performance Comparison of Classification Algorithms for Programming Anxiety Level.

Algorithm	Accuracy (%)	Precision (%)	Recall (%)	Cohen's Kappa (%)
Logistic Regression	97.75	96.88	96.70	95.00
Support Vector Machine	97.69	96.42	97.03	94.90
J48 Decision Tree	94.98	93.97	91.59	88.90
Random Forest	94.57	92.93	91.58	88.00
Naive Bayes	94.57	94.29	90.11	87.90

Further analysis of the error metrics, as shown in Figure 2 and Figure 3, reinforces the performance trends observed among the models. Logistic Regression demonstrated the most consistent and accurate predictions, with the lowest error rates: a MAE of 0.0225, RMSE of 0.1464, RAE of 0.03%, and RRSE of 30.71%. These results indicated minimal deviation and variance, confirming the model's robustness in reducing prediction errors. SVM followed closely, also exhibiting low error values and reliable performance. In contrast, NB, RF, and DT showed higher RMSE and RRSE values, suggesting greater variability and less stability in their predictive outcomes.

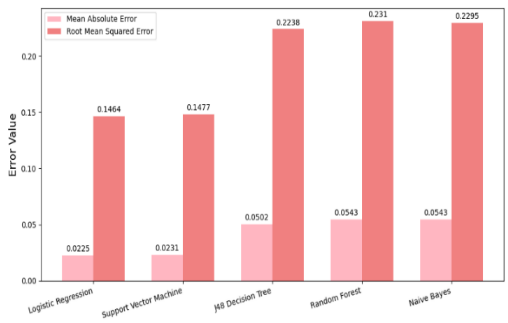

Figure 2. Error Analysis Results of Classification Algorithms Using MAE and RSMSE.

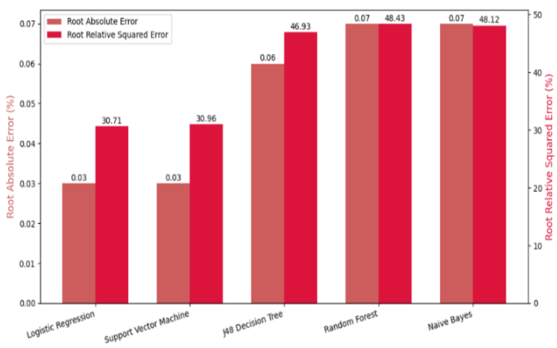

Figure 3. Error Analysis Results of Classification Algorithms Using RAE and RRSE.

B) Logistic Regression as the Best Model for Classifying Student Programming Anxiety Levels

The line chart depicted in Figure 4 illustrates the performance of the Logistic Regression model across ten cross-validation folds, offering a clear visualization of its stability and predictive capability. The results demonstrate consistently high values for both Accuracy and F-measure, with metrics ranging from 97% to 99% across all validation folds. This narrow range reflects minimal performance variation, underscoring the model's reliability when applied to different subsets of unseen data. Furthermore, the integration of the SMOTE to mitigate class imbalance appears to contribute positively to the model's generalizability. Collectively, these findings indicate that the Logistic Regression model exhibits strong and consistent classification performance in predicting students' current programming anxiety levels.

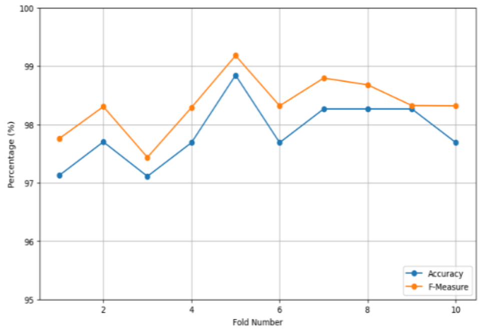

Figure 4. Average Metrics per Cross-Fold for Logistic Regression.

To elucidate the relationship between the significant predictor variables and the classified outcome, the logistic regression model incorporates the respective coefficients of each feature in the final equation. The probability of a student being classified as having high or low programming anxiety is expressed as:

$$P_{\text{high / low}} = \frac{1}{1 + e^{-(-3.3452 + (-2.7680 \cdot x_1) + (3.2119 \cdot x_2) + \cdots + (3.0043 \cdot x_{22}) + (-2.1269 \cdot x_{23}))}}$$

This equation includes an intercept of −3.3452 and multiple feature coefficients, represented by the ellipsis, which collectively captures the complexity of the model. Each coefficient quantifies the influence of a corresponding feature on the likelihood of a student exhibiting high programming anxiety, thereby highlighting the multifactorial nature of the classification.

The coefficient chart visually represents how each feature's value impacts the Logistic Regression model's classifications, revealing the relative importance of each factor in determining the outcome which provides better understanding of the significance of each feature in classifying students' programming anxiety levels

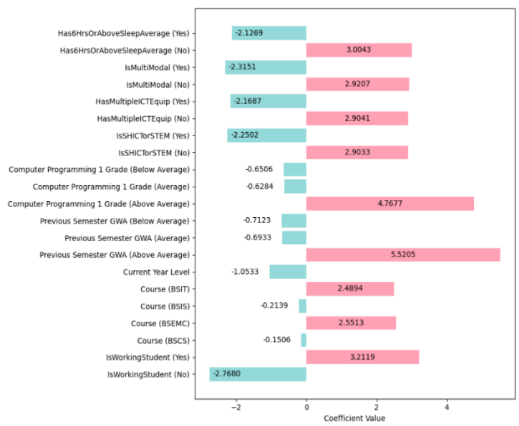

Figure 5. Coefficients Chart for Logistic Regression.

Figure 5 displays the model's coefficients, offering insights into the factors that influence students' programming anxiety levels. Positive coefficients indicate a contribution to higher anxiety, while negative coefficients suggest a mitigating effect on anxiety. To provide a comprehensive view of the model's effectiveness in predicting students' programming anxiety levels, a range of performance metrics is presented. These metrics offer valuable insights into the model's accuracy in classifying anxiety levels and its ability to differentiate between them. By evaluating multiple aspects of the model's performance, a clearer understanding of its predictive capabilities is achieved, enabling a more thorough assessment of its real-world applicability.

Table 4. Results of Area Under the Curve for Logistic Regression Model

Area	Std. Error	Asymptotic Sig.	Asymptotic 95% Confidence Level	
			Lower Bound	Upper Bound
0.98	0.02	0	0.960	0.995

Table 4 shows Area under the Curve (AUC) result of 0.98 with a 95% confidence interval (0.960, 0.995). Additionally, the AUC is significantly different from 0.5 with a p-value of 0, indicating that the logistic regression model classifies the group significantly better than by chance. To determine and evaluate the goodness-of-fit of a logistic regression model, it was tested based on the measures of sensitivity (true positive rate) and 1 - specificity (false positive rate) at various cutoff points, as represented by the receiver operating characteristic (ROC) curve.

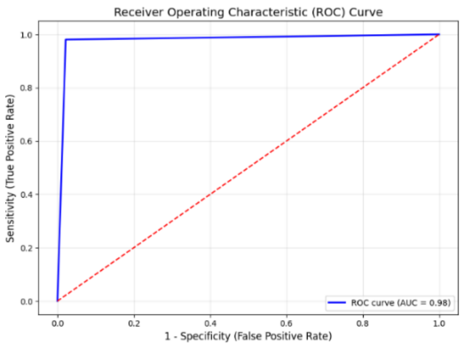

Figure 6. ROC Curve of Logistic Regression Model.

Figure 6 presents the Receiver Operating Characteristic (ROC) curve for the logistic regression model, illustrating its ability to distinguish between different levels of students' programming anxiety. The model achieved an Area Under the Curve (AUC) value of 0.98, indicating excellent discriminative performance. An AUC value close to 1.0 reflects a high level of accuracy in classification, demonstrating the model's strong capability to correctly differentiate between students with varying levels of programming anxiety.

5. CONCLUSION

This study demonstrated the successful development of a high-performing classification model for student programming anxiety levels using the Logistic Regression algorithm. The model achieved 96.77 percent F-measure, 97.75 percent accuracy, 96.88 percent precision, 96.70 percent recall, and a Cohen's kappa of 0.950, indicating strong classification capability. The findings highlight the potential of machine learning in addressing critical challenges in programming education and provide a foundation for future research aimed at supporting timely, data-informed interventions. It is recommended that future studies explore the integration of the developed model into institutional services, such as academic advising systems or learning management platforms, to enable real-time risk detection and the delivery of personalized support to students. Additionally, expanding the dataset in size and diversity and investigating the application of advanced techniques, such as deep learning, are suggested to enhance the model's reliability, generalizability, and practical value [30]. Establishing formal quality assurance procedures as part of future development efforts may further strengthen the robustness of similar models.

Acknowledgments

We sincerely thank all who contributed to the success of this study. We acknowledge the support of participating institutions and individuals in data collection, and the encouragement of our peers, colleagues, and families. Their time, effort, and motivation were invaluable throughout this research.

Funding Information

This study did not receive external funding support. However, a manuscript grant from the National Research Council of the Philippines has been approved. All other expenses were personally covered by the authors.

Ethical Approval

The researchers obtained local approvals to ensure ethical and institutional compliance. Data collection followed the Philippine Data Privacy Act of 2012, with strict measures to protect participant privacy.

Data Availability

The data supporting the results of this study are available from the corresponding author upon reasonable request and subject to institutional approval and compliance with the Philippine Data Privacy Act of 2012.

Authors Contributions Statement

Name of Author	C	M	So	Va	Fo	I	R	D	O	E	Vi	Su	P	Fu
Eduardo R. Yu II	✓	✓	✓		✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Elmerito D. Pineda	✓	✓								✓		✓	✓	
Isagani M. Tano	✓	✓		✓	✓					✓	✓		✓	
Jaime P. Pulumbarit	✓			✓	✓					✓	✓		✓	
Jonilo C. Mababa	✓	✓	✓	✓	✓					✓	✓		✓	

C : **C**onceptualization
M : **M**ethodology
So : **So**ftware
Va : **Va**lidation
Fo : **Fo**rmal analysis
I : **I**nvestigation
R : **R**esources
D : **D**ata Curation
O : Writing - **O**riginal Draft
E : Writing - Review & **E**diting
Vi : **Vi**sualization
Su : **Su**pervision
P : **P**roject administration
Fu : **Fu**nding acquisition

Conflict of Interest

The authors declare no conflicts of interest.

Informed Consent

All participants were informed of the study's purpose, procedures, and their right to voluntary participation.

REFERENCES

[1] R. Yazdanian, R. L. Davis, X. Guo, F. Lim, P. Dillenbourg, and M.-Y. Kan, 'On the radar: Predicting near-future surges in skills' hiring demand to provide early warning to educators', Computers and Education: Artificial Intelligence, vol. 3, p. 100043, Jan. 2022. doi.org/10.1016/j.caeai.2021.100043

[2] K. M. N. Rebuta, I. M. P. Cabaron, R. J. C. Pucong, J. M. C. Bisquera, R. T. Llerado, and M. V. M. Buladaco, 'Relationship of programming skills and perceived value of learning programming among Information Technology education students in Davao Del Sur', Int. J. Res. Innov. Soc. Sci., vol. 6, no. 6, pp. 882–887, 2022. doi.org/10.47772/ijriss.2022.6633

[3] Philippine Commission on Higher Education, CMO No. 25, Series of 2015: Policies, standards, and guidelines for the Bachelor of Science in Information Technology program, 2015. [Online]. Available: https://ched.gov.ph/wp-content/uploads/2017/10/CMO-no.-25-s.-2015.pdf

[4] Commission on Higher Education, "CHED Memorandum Order No. 24, series of 2015: Policies, Standards and Guidelines for the Bachelor of Library and Information Science (BLIS) Program," Oct. 12, 2015. [Online]. Available: https://ched.gov.ph/wp-content/uploads/2017/10/CMO-no.-24-s.-2015.pdf

[5] L. Hu, 'Programming and 21st century skill development in K-12 schools: A multidimensional meta-analysis', Journal of Computer Assisted Learning, vol. 40, no. 2, pp. 610–636, Nov. 2023. doi.org/10.1111/jcal.12904

[6] J. Zheng, M. Duffy, and G. Zhu, 'Predictors of university students' intentions to enroll in computer programming courses: a mixed-method investigation', Discover Education, vol. 3, no. 1, Sep. 2024. doi.org/10.1007/s44217-024-00232-5

[7] C. N. P. Olipas, R. F. Leona, A. C. A. Villegas, A. I. Cunanan Jr., and C. L. P. Javate, 'The academic performance and the computer programming anxiety of BSIT students: A basis for instructional strategy improvement', Int. J. Adv. Eng. Manag. Sci., vol. 7, no. 6, pp. 125–129, 2021. dx.doi.org/10.22161/ijaems.76.15

[8] Ahmed, 'Student performance prediction using machine learning algorithms', Applied Computational Intelligence and Soft Computing, 2024. doi.org/10.1155/2024/4067721

[9] C. Connolly, E. Murphy, and S. Moore, 'Programming Anxiety Amongst Computing Students—A key in the retention debate? ', IEEE Transactions on Education, vol. 52, no. 1, pp. 52–56, Aug. 2008. doi.org/10.1109/te.2008.917193

[10] Yildirim, O. G., & Özdener, N. (2022), 'Development and validation of the Programming Anxiety Scale', International Journal of Computer Science Education in Schools, 5(3), 17–34, 2022. doi.org/10.21585/ijcses.v5i3.140

[11] P. Kumar, S. Garg, and A. Garg, 'Assessment of anxiety, depression and stress using machine learning models', Procedia Computer Science, vol. 171, pp. 1989–1998, 2020. doi.org/10.1016/j.procs.2020.04.213

[12] A. Sau and I. Bhakta, 'Screening of anxiety and depression among seafarers using machine learning technology', Informatics in Medicine Unlocked, vol. 16, p. 100228, 2019. doi.org/10.1016/j.imu.2019.100228

[13] A. Sau and I. Bhakta, 'Predicting anxiety and depression in elderly patients using machine learning technology', Healthcare Technology Letters, vol. 4, no. 6, pp. 238–243, Nov. 2017. doi.org/10.1049/htl.2016.0096

[14] A. Priya, S. Garg, and N. P. Tigga, 'Predicting anxiety, depression and stress in modern life using machine learning algorithms', Procedia Computer Science, vol. 167, pp. 1258–1267, 2020. doi.org/10.1016/j.procs.2020.03.442

[15] J. D. Elhai, H. Yang, D. McKay, G. J. G. Asmundson, and C. Montag, 'Modeling anxiety and fear of COVID-19 using machine learning in a sample of Chinese adults: associations with psychopathology, sociodemographic, and exposure variables', Anxiety, Stress, & Coping, vol. 34, no. 2, pp. 130–144, 2021. doi.org/10.1080/10615806.2021.1878158

[16] Albagmi, F. M., Alansari, A., Shawan, D. S. A., AlNujaidi, H., & Olatunji, S. O., 'Prediction of generalized anxiety levels during the Covid-19 pandemic: A machine learning-based modeling approach', Informatics in Medicine Unlocked, 28, 100854, 2022. doi.org/10.1016/j.imu.2022.100854

[17] S. A. Farooq, O. Konda, A. Kunwar, and N. Rajeev, 'Anxiety prediction and analysis - A machine learning based approach', 4th International Conference for Emerging Technology (INCET), 2023. doi.org/10.1109/incet57972.2023.10170115

[18] S. Mutalib, 'Mental health prediction models using machine learning in higher education institutions', Turkish Journal of Computer and Mathematics Education, vol. 12, no. 5, pp. 1782–1792, 2021. doi.org/10.17762/turcomat.v12i5.2181
[19] M. D. Nemesure, M. V. Heinz, R. Huang, and N. C. Jacobson, 'Predictive modeling of depression and anxiety using electronic health records and a novel machine learning approach with artificial intelligence', Scientific Reports, vol. 11, no. 1, 2021. doi.org/10.1038/s41598-021-81368-4
[20] R. Qasrawi, S. VicunaPolo, D. A. Al-Halawa, S. Hallaq, and Z. Abdeen, 'Assessment and prediction of depression and anxiety risk factors in schoolchildren: Machine learning techniques performance analysis', JMIR Formative Research, vol. 6, no. 8, e32736, 2022. doi.org/10.2196/32736
[21] A. D. Vergaray, N. S. Ríos, J. I. Necochea-Chamorro, K. Z. Ramos, and Y. Del Rosario Vásquez Valencia, 'Systematic review of machine learning techniques to predict anxiety and stress in college students', Informatics in Medicine Unlocked, vol. 43, p. 101391, 2023. doi.org/10.1016/j.imu.2023.101391
[22] Geronimo, S. M., Hernandez, A. A., Abisado, M. B., Rodriguez, R. L., Nova, A. C., Caluya, S. S., & Blancaflor, E. B., 'Understanding Perceived Academic Stress among Filipino Students during COVID-19 using Machine Learning', SIGITE '23: Proceedings of the 24th Annual Conference on Information Technology Education, 4, 54–59, 2023. doi.org/10.1145/3585059.3611412
[23] N. B. Mendoza, R. B. King, and J. Y. Haw, 'The mental health and well-being of students and teachers during the COVID-19 pandemic: combining classical statistics and machine learning approaches', Educational Psychology, vol. 43, no. 5, pp. 430–451, 2023. doi.org/10.1080/01443410.2023.2226846
[24] Ibrahim, A., 'Definition, purpose, and procedure of developmental research: An analytical review', Asian Research Journal of Arts & Social Sciences, 1(6), 1–6, 2016. doi.org/10.9734/arjass/2016/30478
[25] M. Vale, 'Descriptive research design and its myriad uses', Elsevier Author Services - Articles, Dec. 27, 2023. [Online]. Available: https://scientific-publishing.webshop.elsevier.com/research-process/descriptive-research-design-and-its-myriad-uses/. [Accessed: Apr. 12, 2025].
[26] C. Schröer, F. Kruse, and J. C. M. Gómez, 'A systematic literature review on applying CRISP-DM process model', Procedia Computer Science, vol. 181, pp. 526–534, Jan. 2021. doi.org/10.1016/j.procs.2021.01.199
[27] Alija, S., Beqiri, E., Gaafar, A. S., & Hamoud, A. K., 'Predicting students' performance using supervised machine learning based on imbalanced dataset and wrapper feature selection', Informatica, 47(1), 2023. doi.org/10.31449/inf.v47i1.4519
[28] W. L. Ku and H. Min, 'Evaluating machine learning stability in predicting depression and anxiety amidst subjective response errors', Healthcare, vol. 12, no. 6, p. 625, 2024. doi.org/10.3390/healthcare12060625
[29] Bendebane, L., Laboudi, Z., Saighi, A., Al-Tarawneh, H., Ouannas, A., & Grassi, G., 'A Multi-Class Deep Learning Approach for Early Detection of Depressive and Anxiety Disorders Using Twitter Data', Algorithms, 16(12), 543, 2023. doi.org/10.3390/a16120543
[30] Tian, X., Zhu, L., Zhang, M., Wang, S., Lu, Y., Xu, X., Jia, W., Zheng, Y., & Song, S. ,'Social anxiety prediction based on ERP features: A deep learning approach', Journal of Affective Disorders, 367, 545–553, 2024. doi.org/10.1016/j.jad.2024.09.006

How to Cite: Eduardo R. Yu II, Elmerito D. Pineda, Isagani M. Tano, Jaime P. Pulumbarit, Jonilo C. Mababa. (2025). Development of a classification model for student programming anxiety levels using logistic regression algorithm. Journal of Artificial Intelligence, Machine Learning and Neural Network (JAIMLNN), 5(2), 1–12. https://doi.org/10.55529/jaimlnn.52.1.12

BIOGRAPHY OF AUTHORS

	**Eduardo R. Yu II** Software Engineer with a decade of industry experience and a part-time faculty member at the University of Caloocan City and Arellano University. He holds a Doctor of Information Technology from La Consolacion University Philippines, a Master of Information Technology from Philippine Christian University, and a Bachelor of Science in Information Technology from City of Malabon University. Email: eduardoryuii@gmail.com/ eduardo.yuii@email.lcup.edu.ph
	**Elmerito D. Pineda** Former Vice President for Academic Affairs at the City of Malabon University. He holds a Doctor of Business Administration from the International Business Academy of Switzerland, has academically completed the Doctor in Information Technology program at AMA Computer University, and holds a Bachelor of Science in Mechanical Engineering from Mapúa University. Email: elmerito.pineda@email.lcup.edu.ph
	**Isagani M. Tano** Dean of the College of Computer Studies at Quezon City University. He holds a Doctor in Information Technology from AMA Computer University and a PhD in Education from La Consolacion University Philippines. He also holds a Master in Information Technology from Rizal Technological University and is currently pursuing a Master in Computer Science at AMA University and a Doctorate in Public Administration at the University of Luzon. Email: isagani.tano@email.lcup.edu.ph
	**Jaime P. Pulumbarit** Professor VI at Bulacan State University, specializing in Information Technology and Education. He holds a Doctor in Information Technology from La Consolacion University and a Master of Information Technology from De La Salle University. Recognized in the AD Scientific Index, his work focuses on the integration of technology in education and data-driven research. Email: jaime.pulumbarit@email.lcup.edu.ph
	**Jonilo C. Mababa** Program Chair of the IT Department at Angeles University Foundation and President of the Philippine Society of Information Technology Educators – Central Luzon for the 2022–2025 term. He holds a Doctorate in Information Technology and is pursuing a Ph.D. in Education. He published research on digital security, big data analytics, and IoT-based systems, contributing to advancements in both educational practices and information technology. Email: jonilo.mababa@email.lcup.edu.ph