IARJSET

ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 8.311 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 12, Issue 5, May 2025

DOI: 10.17148/IARJSET.2025.125349

Multi-Class Adaptive Active Learning for Predicting Student Anxiety

Prasanna G[1], Amrutha E[2], Harshitha M R[3], Jeevitha S R[4], Madhumitha R[5]

Asst professor, Department of Computer Science & Engineering, Maharaja Institute of Technology Mysore, Karnataka, India[1]

Student, Department of Computer Science & Engineering, Maharaja Institute of Technology Mysore, Karnataka, India[2-5]

Abstract: In order to improve early intervention and support systems in educational settings, this study presents a multi-class adaptive active learning framework to predict student anxiety. Because of their static learning processes and lack of labeled data, traditional anxiety prediction models frequently perform poorly. Our method improves model accuracy and robustness by iteratively choosing the most informative data points for labeling using adaptive active learning. The model provides a nuanced understanding of students' mental health by distinguishing between different levels of anxiety through the use of multi-class classification. The effectiveness of the suggested approach is demonstrated by experimental results, which show notable gains in prediction accuracy over baseline models. With its scalable solution for real-time anxiety prediction and contribution to more responsive learning, this study highlights the potential of adaptive active learning in educational data mining.

Keywords: Adaptive Active Learning, Multi-ClassClassification, Student Anxiety Prediction, Educational Data Mining,Machine, Learning in Education,Mental ,Health Assessment,Real-Time Analytics.

I. INTRODUCTION

The need for proactive and successful intervention strategies is highlighted by the growing concern for the mental health of students. Due to their dependence on static learning processes and need for large amounts of labeled data, traditional anxiety prediction models frequently perform poorly. These restrictions make it more difficult to determine students' anxiety levels and postpone the help they need. The suggested system fills this gap by introducing a multi-class adaptive active learning framework that uses cutting-edge machine learning techniques to increase the responsiveness and accuracy of anxiety predictions. The system can operate efficiently with little expert input by dynamically choosing the most instructive data points for labeling. It can also continuously adjust to new student stressors and behavioral patterns.

Real-time updates and a sophisticated categorization of anxiety levels, including none, mild, moderate, and severe, are made possible by this adaptive approach, which provides notable improvements over current models. By incorporating active learning, the training process is optimized and the annotation workload is decreased by ensuring that only the most ambiguous and instructive instances are queried for labeling. By enabling customized interventions, the system's multi-class classification feature enables educational institutions to more accurately and quickly address mental health issues. All things considered, the suggested framework offers a strong, expandable way to improve mental health assistance in learning settings by offering early identification, prompt intervention, and ongoing student needs adaptation.

II. LITERATURE SURVEY

Active learning and adaptive techniques have gained attention for improving educational and mental health prediction systems. These approaches reduce the burden of data labeling while maintaining model accuracy. The following foundational studies provide insights into how multi-class classification, active learning, and adaptive methods can be effectively applied to student anxiety prediction:

Zhou et al. [1]. In order to increase efficiency and lower annotation costs, their work concentrated on clever data selection techniques like uncertainty sampling and query-by-committee. The significance of using active learning in contexts with little labeled data was emphasized by the study. In a variety of educational applications, it established the foundation for sample-efficient learning models that preserve classification accuracy.

DOI: 10.17148/IARJSET.2025.125349

Kumar et al. [2] reviewed adaptive learning techniques used to predict student performance and mental health outcomes. Their work emphasized the limitations of static models and highlighted the ability of adaptive systems to adjust dynamically to new data. They demonstrated that real-time learning adaptation can improve the identification of at-risk students, offering a more personalized and effective intervention mechanism in educational environments.

Singh et al. [3] developed a predictive model combining active learning with multi-class classification to enhance mental health predictions. Their system iteratively refined model accuracy by focusing on the most uncertain and informative data samples. This method significantly improved differentiation across various mental health states, indicating the benefits of hybrid approaches in handling complex psychological assessments.

Miller et al. [4] proposed an adaptive active learning framework for early student intervention. Their model prioritized real-time data relevance to identify early signs of student anxiety and academic decline. The system supported ongoing monitoring, making it suitable for dynamic academic settings. This approach highlighted the potential for intelligent systems to inform counselors and educators in timely support delivery.

Wang et al. [5] conducted a comparative study on dynamic sampling techniques within active learning frameworks for anxiety prediction. The research evaluated multiple sampling strategies and their impact on model performance. The results showed that intelligent sample selection significantly improved the model's efficiency and accuracy, making the case for targeted data usage in mental health detection.

III. IMPLEMENTATION

With the aid of a clearly defined technology stack and an organized modular approach, the project "Multi-Class Adaptive Active Learning for Predicting Student Anxiety" was implemented. HTML, CSS, and Bootstrap were used in the development of the system's front end to produce a responsive and intuitive user interface. Client-side interactivity was handled by JavaScript. Python with the Flask framework was selected for the back-end due to its adaptability and simplicity in integrating with machine learning components. User input, prediction outcomes, and feedback data were all stored in a MySQL database. The XAMPP server, which offered a simulated environment for executing the application and interacting with the MySQL database, was used to deploy and test the system locally.

There were several modules to the implementation process. Data collection and preprocessing were part of the first module. This module was in charge of importing datasets from trustworthy sources, such as academic transcripts and the findings of psychological surveys. The data was cleaned to deal with missing or null values using libraries like Pandas and NumPy, and LabelEncoder and OneHotEncoder were used to encode categorical variables. To put all attributes on a similar scale, numerical features were normalized. Furthermore, only the most pertinent features for training were kept by using feature selection strategies like correlation matrix analysis.

Initial model training was the main topic of the second module. This entailed using the preprocessed dataset to train baseline machine learning models. The effectiveness of several algorithms in forecasting different degrees of student anxiety was assessed, including Logistic Regression, K-Nearest Neighbors (KNN), Random Forest, Naive Bayes, and XGBoost. The train_test_split function in Scikit-learn was used to divide the dataset into training and testing subsets. Accuracy, precision, recall, and F1-score were among the performance metrics used to train and validate each model. The active learning process began with the model that performed the best.

The adaptive active learning algorithm, which iteratively questioned the most ambiguous data points and retrained the model with freshly labeled samples, was implemented in subsequent modules. This method reduced the quantity of labeled data needed while increasing the model's prediction accuracy. The last module entailed combining all the parts into a comprehensive web-based system that allowed users to enter pertinent student information, get anxiety level predictions, and optionally offer suggestions for model enhancement.

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 8.311 ⁒ Peer-reviewed & Refereed journal ⁒ Vol. 12, Issue 5, May 2025

DOI: 10.17148/IARJSET.2025.125349

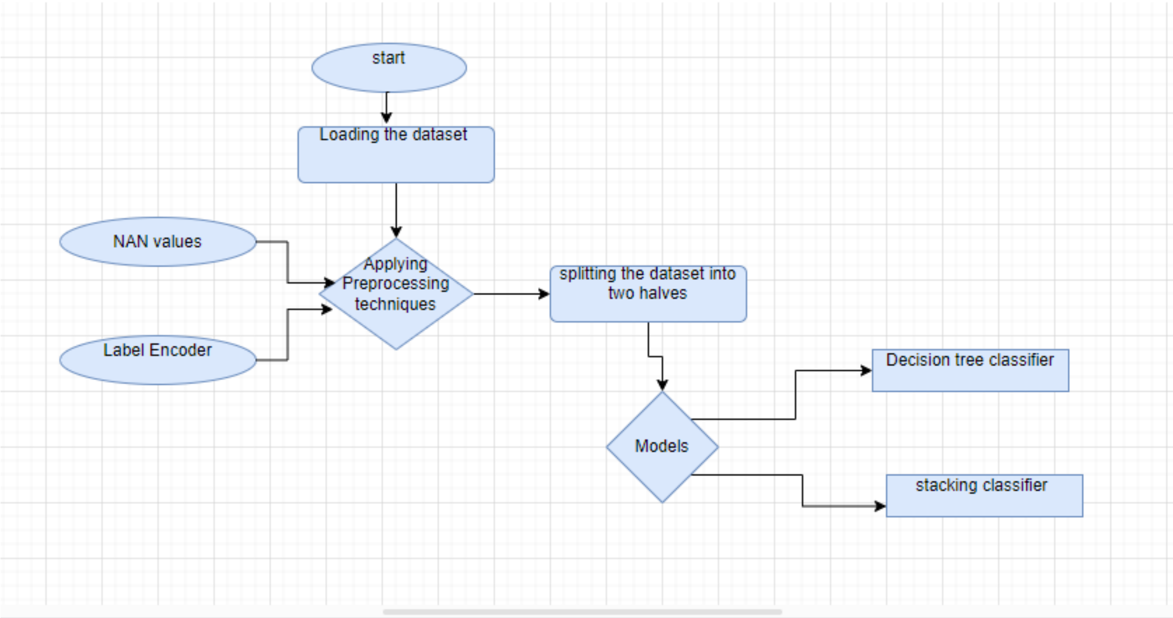

Fig. Implementation for Predicting Student Anxiety

IV. METHODOLOGY

A structured machine learning pipeline is the foundation of the methodology used for this project, which aims to predict sleep disorders from input data. The following details the actions done to meet the project's goals:

Objective 1: Collect and Preprocess Data

- **Method/Procedure**: The dataset was obtained from publicly available datasets or trustworthy medical records. To enhance model performance, preprocessing involved scaling numerical values, handling missing values, and encoding categorical features.

Objective 2: Feature Selection and Engineering

- **Method/Procedure**: To find important predictors like age, BMI, degree of snoring, etc., feature importance techniques (e.g., correlation analysis, statistical significance tests) were used.

Objective 3: Model Selection and Training

- **Method/Procedure**: A variety of models were trained, including SVM, Random Forest, Decision Trees, and Logistic Regression. For final deployment, the model with the highest accuracy, precision, and recall was chosen.

Objective 4: Model Evaluation

- **Method/Procedure**: SVM, Random Forest, Decision Trees, and Logistic Regression were among the models that were trained. The model with the best recall, accuracy, and precision was selected for final deployment.

Objective 5: Deployment and Prediction Interface

- **Method/Procedure**: A web-based interface (such as Flask or Django) was used to deploy the finished model, enabling users to enter symptoms and get real-time predictions.

V. ALGORITHMS

K-Nearest Neighbors: A simple yet powerful classification algorithm, K-Nearest Neighbors (KNN) is employed in a number of machine learning applications. KNN essentially uses the majority class of its 'k' closest neighbors in the feature space to classify data points. The algorithm uses metrics like the Manhattan or Euclidean distance to calculate the distance between the query point and every other point in the training dataset. After that, it determines the 'k' closest neighbors and gives the query point the most prevalent class label among them. Because of its ease of use and versatility in handling different kinds of data without presuming an explicit underlying distribution, KNN is especially beneficial. When used to predict student anxiety, KNN can efficiently classify anxiety levels by using labeled examples and making predictions based on similarity. This improves the model's overall classification accuracy and helps it capture subtle anxiety patterns. With an accuracy of 0.78 in your study, KNN showed promise in distinguishing between different student anxiety levels.

DOI: 10.17148/IARJSET.2025.125349

Logistic Regression: By minimizing a cost function—typically the cross-entropy loss, which quantifies the discrepancy between the actual class labels and the predicted probabilities—the algorithm optimizes the model parameters during training. The model's performance could be enhanced by additional fine-tuning, feature engineering, or the use of more complex algorithms, as the accuracy of 0.60 shows that the Logistic Regression model correctly predicted the anxiety levels 60% of the time.

XGB Classifier: A potent machine learning algorithm built on gradient boosting principles is the Extreme Gradient Boosting Classifier, or XGB Classifier. It sequentially constructs a group of decision trees, each of which aims to fix the mistakes of the one before it. By minimizing the loss function—which quantifies the discrepancy between expected and actual values—this iterative method seeks to enhance the model's performance. To avoid overfitting and improve the model's generalization, XG Boost uses regularization techniques. This algorithm's resilience and capacity to manage intricate data patterns help your research project achieve an accuracy of 0.80 in predicting the anxiety levels of your students. Because of its adaptability, it works well with dynamic and varied educational data.

Naive Bayes: Based on Bayes' theorem, the Naive Bayes classifier is probabilistic and assumes that the features used for classification are conditionally independent given the class label. The algorithm determines the likelihood of a student's anxiety level based on their characteristics (e.g., social interactions, academic performance, etc.) in order to predict their anxiety levels. To do this, it first uses the training data to estimate the prior probabilities of each anxiety class. Then, assuming feature independence, it assesses the probability that each feature belongs to each class. Using Bayes' theorem, the posterior probability of each class is calculated, and the class with the highest posterior probability is assumed to represent the anxiety level of the students. Despite its simplicity, Naive Bayes can be effective for multi-class classification tasks, though in this study, it achieved an accuracy of 0.52, indicating room for improvement compared to other models.

Random Forest: As an ensemble learning technique, the Random Forest algorithm builds several decision trees during training and outputs the mean prediction (regression) or class mode (classification) of each tree separately. A random subset of the data and features is used to train each decision tree in the forest, adding diversity and minimizing overfitting. Each tree casts a vote for a class during the prediction phase, and the class with the most votes becomes the final prediction. This method enables the model to identify intricate relationships and patterns within the data. The Random Forest classifier in your project successfully distinguished between different levels of student anxiety, as evidenced by its accuracy of 0.82. The high accuracy underscores the model's capability to leverage the adaptive active learning framework to iteratively improve predictions and provide valuable insights into student mental health.

Decision Tree Classifier: A flexible tool for managing the multi-class classification of student anxiety levels is the Decision Tree Classifier. The Decision Tree Classifier's internal mechanism builds a tree-like model of decisions by recursively dividing the dataset into subsets according to feature values. The algorithm chooses the feature and matching threshold at each node of the tree that best separates the data based on a criterion such as information gain or Gini impurity. Until the data in each leaf node falls into a single class or satisfies a stopping criterion such as a minimum sample split or maximum tree depth this process keeps going. The decision tree structure enables intuitive understanding of how different features contribute to predictions, and its adaptability in handling various anxiety levels makes it a fitting choice for predicting and interpreting complex patterns in student anxiety data.

Stacking Classifier: In this study, the Stacking Classifier a potent ensemble learning method was employed to forecast students' anxiety levels. In order to enhance overall performance, it combines the predictions from several base models, also known as level-0 models. This method uses the input data to generate predictions from the base models. These predictions are then used as input features for a meta-model (level-1 model), which synthesizes the data to produce the final output. This technique improves predictive accuracy by utilizing the advantages of different classifiers. The Stacking Classifier in our project demonstrated its efficacy in capturing the intricate patterns linked to student anxiety with an impressive accuracy of 0.86. By integrating multiple models' insights, the Stacking Classifier provides a robust solution that enhances prediction accuracy and reliability compared to individual models, aligning with the study's goal of improving early intervention in educational settings.

VI. RESULT DISCUSSION

The application's system module is essential to processing and overseeing the fundamental operations of prediction, model training, and data handling. It starts by storing the user-provided dataset, which serves as the machine learning model's basis. The system preprocesses the dataset by splitting it into subsets for testing, validation, and training during the training phase. Then, using methods like gradient descent, it applies a machine learning algorithm to identify patterns in

DOI: 10.17148/IARJSET.2025.125349

the training data, adjusting its parameters to reduce prediction errors. After training, the model is used to produce predictions using the fresh data that users enter on the Prediction Page. This methodical pipeline guarantees that the model is reliable, effective, and able to produce precise predictions.

From the standpoint of the user, the application provides a safe and easy-to-use interface for communication. By providing verified personal information, users can create accounts through the Registration Page, guaranteeing secure access. To further improve data security, the Login Page allows returning users to authenticate themselves using validated credentials. After logging in, users can interact with the model and evaluate its output to gauge how well the system predicts. This feature provides transparency and fosters application trust by enabling users to comprehend how the model responds to various inputs. All things considered, the smooth and educational experience that supports user engagement and system functionality is made possible by the smooth integration of user and system modules.

VII. CONCLUSION

In order to more precisely and effectively predict student anxiety, our team has created a novel multi-class adaptive active learning framework. This novel method addresses some of the main drawbacks of conventional models, which frequently rely on substantial amounts of labeled data and are not flexible enough to adjust to shifting behaviors. The system reduces labeling effort while greatly enhancing prediction performance and robustness by selectively querying the most informative data points through the use of adaptive active learning.

The model can distinguish between different levels of anxiety, including none, mild, moderate, and severe, thanks to the integration of multi-class classification. Creating individualized interventions and prompt support plans in educational settings requires this finer level of detail. The framework continuously outperformed baseline models in rigorous testing and evaluation, proving its ability to capture dynamic patterns in student behavior.

This work offers a scalable and real-time solution for detecting student anxiety, highlighting the potential of adaptive learning techniques in educational data mining. Our contribution seeks to improve student well-being and foster healthier learning environments by facilitating more responsive, data-driven mental health support.

VIII. FUTUTRE ENHANCEMENT

Future enhancements of the student anxiety prediction model can significantly improve its effectiveness, adaptability, and ethical application. Integrating multi-modal data—including behavioral patterns, physiological signals like heart rate variability, and academic metrics—will provide a more holistic view of each student's mental health. Personalized intervention strategies can then be developed, tailoring support to individual needs and thereby increasing their impact. Real-time monitoring and feedback systems will enable timely assistance, potentially preventing escalation of anxiety. Additionally, incorporating dynamic learning mechanisms ensures the model evolves with changing student behavior and anxiety trends. Ethical considerations and strong data privacy frameworks will foster trust and ensure responsible data use. Enhancing model explainability will help students and educators understand and trust its outputs. Scalability options should also be explored for deployment across diverse institutions, alongside adaptations for cultural and contextual relevance. Finally, seamless integration with existing student support services will strengthen the overall framework, ensuring students have comprehensive, accessible assistance when needed.

REFERENCES

[1] J. Southworth, K. Migliaccio, J. Glover, J. Glover, D. Reed, C. McCarty, J. Brendemuhl, and A. Thomas, ''Developing a model for AI across the curriculum: Transforming the higher education landscape via innovation in AI literacy,'' Comput. Educ., Artif. Intell., vol. 4, Aug. 2023, Art. no. 100127.
[2] X.-Q. Liu, Y.-X. Guo, and Y. Xu, ''Risk factors and digital interventions for anxiety disorders in college students: Stakeholder perspectives,'' World J. Clin. Cases, vol. 11, no. 7, pp. 1442–1457, Mar. 2023.
[3] M. K. Khaira, R. L. R. Gopal, S. M. Saini, and Z. M. Isa, ''Interventional strategies to reduce test anxiety among nursing students: A systematic review,'' Int. J. Environ. Res. Public Health, vol. 20, no. 2, p. 1233, Jan. 2023.
[4] K. Meeks, A. S. Peak, and A. Dreihaus, ''Depression, anxiety, and stress among students, faculty, and staff,'' J. Amer. College Health, vol. 71, no. 2, pp. 348–354, Feb. 2023.
[5] D. H. Halat, A. Soltani, R. Dalli, L. Alsarraj, and A. Malki, ''Understanding and fostering mental health and well-being among University Faculty: A narrative review,'' J. Clin. Med., vol. 12, no. 13, p. 4425, Jun. 2023.
[6] R. D. Godsil and L. S. Richardson, ''Racial anxiety,'' Iowa Law Rev., vol. 102, p. 2235, Sep. 2016.

DOI: 10.17148/IARJSET.2025.125349

[7] J. Santhosh, D. Dzsotjan, and S. Ishimaru, ''Multimodal assessment of interest levels in reading: Integrating eye-tracking and physiological sensing,'' IEEE Access, vol. 11, pp. 93994–94008, 2023.

[8] R. Qasrawi, S. P. Vicuna Polo, 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 Res., vol. 6, no. 8, Aug. 2022, Art. no. e32736.

[9] P. J. Bota, C. Wang, A. L. N. Fred, and H. P. Da Silva, ''A review, current challenges, and future possibilities on emotion recognition using machine learning and physiological signals,'' IEEE Access, vol. 7, pp. 140990–141020, 2019.

[10] P. Meshram and R. K. Rambola, ''Diagnosis of depression level using multimodal approaches using deep learning techniques with multiple selective features,'' Expert Syst., vol. 40, no. 4, p. e12933, May 2023.

[11] E. Mosqueira-Rey, D. Alonso-Ríos, and A. Baamonde-Lozano, ''Integrating iterative machine teaching and active learning into the machine learning loop,'' Proc. Comput. Sci., vol. 192, pp. 553–562, Jan. 2021.

[12] J. C. Cassady, E. E. Pierson, and J. M. Starling, ''Predicting student depression with measures of general and academic anxieties,'' Frontiers Educ., vol. 4, p. 11, Feb. 2019. [13] H. S. Yi and W. Na, ''How are maths-anxious students identified and what are the key predictors of maths anxiety? Insights gained from Pisa results for Korean adolescents,'' Asia Pacific J. Educ., vol. 40, no. 2, pp. 247–262, Apr. 2020.