Business, Management and Economics Engineering
ISSN: 2669-2481 / eISSN: 2669-249X
2023 Volume 21 Issue 01

A MULTICLASS MACHINE LEARNING APPROACH FOR ACADEMIC STUDENT PERFORMANCE PREDICTION

Mr. C. John Paul
Ph.D Research scholar, Bharathiar University, Coimbatore -641046 E-Mail : jpcalling@gmail.com

Dr. K. Geetha
Assistant Professor, Department of Computer Science, Bharathiar University, Coimbatore-641046
E-Mail : geethakab@gmail.com

Abstract: In recent years, there has been an increasing interest in the development of precise and reliable prediction models for evaluating student performance. The prediction of student grades holds great significance in educational institutions as it facilitates the identification of students facing difficulties, improvement of teaching methodologies, and implementation of targeted interventions. It is an essential task in education that empowers educators to recognize academically struggling students and offer them tailored support. Nevertheless, accurately predicting grades can be a challenging task due to the intricate and imbalanced nature of educational datasets. Fortunately, recent advancements in machine learning offer promising solutions to address this challenge, such as the utilization of the Multiclass Student Grade Prediction System employing SMOTE. The Multiclass Student Grade Prediction System incorporates SMOTE, a technique that balances the dataset by increasing the least represented data modules, thereby enhancing the accuracy of the model. This system has been developed through a comprehensive review of relevant literature, a detailed explanation of the methodology employed, the presentation of evaluation results, and a discussion on the implications of these findings for improving academic outcomes in educational institutions. By predicting the final grades of students across multiple classes, the proposed model aims to provide educators with a comprehensive and automated solution, enabling them to make data-driven decisions and enhance academic outcomes. Our generated results demonstrate that the Multiclass Student Grade Prediction System, employing SMOTE, outperforms other traditional methods in predicting student grades across multiple classes. Implementation of this system in educational institutions can significantly assist teachers in identifying academically struggling students, ultimately leading to improved academic outcomes. An Ensemble of Three Classifiers, referred to as ETCs, is introduced as a method for predicting student performance. This approach combines three classifiers: Adaptive Neuro Fuzzy Inference System (ANFIS), Support Vector Machine (SVM) classifier, and Decision Tree (DT). The outcomes of the experiment demonstrate that the proposed method is compared to different algorithms used in classifiers, including DT, Artificial Neural Network (ANN), ANFIS, and SVM.

Keywords: Multiclass prediction model, SMOTE, imbalanced datasets, and grade prediction, Ensemble of Three Classifiers(ETC), Adaptive Neuro Fuzzy Inference System (ANFIS), Support Vector Machine (SVM), Decision Tree (DT).

1. INTRODUCTION:

Accurate prediction of student grades plays a crucial role in the education sector as it allows educators to identify struggling students and implement targeted interventions to improve academic outcomes. Machine learning techniques have emerged as powerful tools for developing prediction models in this field. This research paper introduces a novel multiclass prediction model for student grade prediction that incorporates the Synthetic Minority Over-Sampling Technique (SMOTE) with machine learning algorithms. The primary objective of this model is to overcome the challenges posed by imbalanced class distribution and enhance the overall accuracy and fairness of grade predictions. Predicting grades across multiple classes is a complex task due to various factors, including imbalanced class distribution, missing data, and the existence of multiple classes. Imbalanced class distribution occurs when the grades are unevenly distributed across classes, resulting in biased predictions. Additionally, the presence of missing data further complicates the prediction process. Traditional approaches often struggle to effectively handle these challenges. To address these issues, this research paper proposes a multiclass prediction model that combines machine learning algorithms with the Synthetic Minority Over-Sampling Technique (SMOTE). Machine learning algorithms can leverage large datasets to uncover patterns and make accurate predictions. SMOTE, on the other hand, is a data augmentation technique specifically designed to tackle class imbalance. By generating synthetic samples for the minority class, SMOTE helps balance the class distribution and mitigate bias in the prediction process. To overcome these challenges, this research paper introduces a multiclass prediction model that harnesses the power of machine learning algorithms and integrates the Synthetic Minority Over-Sampling Technique (SMOTE). Machine learning algorithms have demonstrated their effectiveness in uncovering patterns and making accurate predictions when provided with sufficient training data. Conversely, SMOTE is a data augmentation technique explicitly developed to address class imbalance. By generating synthetic samples for the minority class, SMOTE aids in balancing the class distribution and reducing bias in the prediction process. The main objective of this research is to contribute to the existing literature by specifically addressing the challenge of multiclass prediction for student grades using machine learning and SMOTE. By integrating SMOTE with machine learning algorithms, the proposed model aims to improve the accuracy and fairness of grade predictions. Moreover, the model allows educators to identify at-risk students across multiple classes, facilitating targeted interventions tailored to individual needs. The subsequent sections of this paper will delve into the related work, methodology, experimental results, and conclusion. The effectiveness of the proposed multiclass prediction model will be evaluated using appropriate evaluation metrics and compared to existing approaches. This research strives to make a valuable contribution to the field of student grade prediction by providing a comprehensive and robust model that considers the challenges associated with multiclass prediction and addresses class imbalance using the SMOTE technique. This study aims to investigate the effectiveness of ensemble feature selection techniques in generating robust feature selection methods and the impact of combining multiple methods on classification performance. To achieve a better understanding, we employ

Ensemble Swarm based Feature Selection (ESFS) for selecting features/attributes. Subsequently, we utilize three widely-used data mining methods, namely Adaptive Neuro Fuzzy Inference System (ANFIS), Support Vector Machine (SVM) classifier, and Decision Tree (DT), to build an academic performance model. Additionally, an ensemble bagging model is implemented to improve the performance of these classifiers. Finally, the results are evaluated using classification metrics such as precision, recall, F-measure, and accuracy.

2. LITERATURE SURVEY:

Several research studies have been conducted by higher education institutions (HEIs) to utilize machine learning techniques in predicting student grades. These studies explore various factors and collect data from multiple sources to forecast student outcomes. However, the effectiveness of predictive models for unbalanced datasets in the field of education has not been extensively investigated. In one study [12], researchers employed discretization and oversampling techniques, specifically SMOTE, to enhance the accuracy of predicted final grades. They utilized classification techniques like Naive Bayes (NB), Decision Trees (DT), and Neural Networks (NN) to categorize students' final grades into five groups (A, B, C, D, and F). The results indicated that SMOTE combined with NN and NB, as well as equal width binning, outperformed other methods in terms of accuracy. Notably, NB demonstrated superior performance in computational time. Researchers from the College of Minnesota's Computer Science and Engineering (CSE) and Electrical and Computer Engineering (ECE) schools [13] developed a technique for predicting future course grades. They found that Matrix Factorization (MF) and Linear Regression (LinReg) yielded more accurate predictions compared to traditional approaches. Moreover, by utilizing data subsets specific to each course, they were able to improve the predictability of subsequent class grades. In another study [14], 225 datasets of university students were utilized to forecast grades in different courses using algorithms such as MF, collaborative filtering (CF), and Limited Boltzmann Machines (RBM). The results indicated that RBM outperformed CF and MF in modelling tabular data and achieved improved prediction accuracy. Early in the semester, a study [15] aimed to predict students' final grades in introductory courses. The researchers employed WEKA to compare eleven different machine learning algorithms, including Bayes, Function, Lazy (IBK), Rules-Based (RB), and Decision Tree (DT). To address high dimensionality and imbalanced data, they utilized feature selection techniques such as correlation-based and information gain. Additionally, they employed SMOTE to balance the instances across the three classes. Among the algorithms, the Decision Tree classifier (J48) outperformed the others with an accuracy of 88%. In another study [17], three Decision Tree algorithms, namely Random Tree (RT), RepTree, and J48, were employed to forecast student grade achievement. Cross-validation was employed to evaluate the effectiveness of the prediction models, and the findings indicated that RT performed better than other algorithms, achieving an accuracy of 75.188%. Adding more samples and attributes to the dataset could potentially further improve the accuracy of the prediction models. Furthermore, at the University Sultan Zainal Abidin (UniSZA) in Malaysia, a system was developed to predict students' academic success [18]. Cheng et al. [19] proposed a novel approach called Synthetic Feature Selection Approach (SFSA) that integrates Support Vector Machines (SVM) to identify significant patterns and key features influencing students' academic achievement. They conducted an investigation using two databases, namely "Student Profile" and "Tutorship Record," collected from an elementary school in Taiwan. These databases were merged based on students' names to create a combined dataset for analysis. The results demonstrate that the proposed model improves accuracy and facilitates the interpretation of patterns in a hybrid-type dataset of students' academic performance. Almasri et al. [20] presented a comprehensive study with three main objectives. First, they conducted an extensive analysis of selected features and their impact on performance using statistical examination techniques. Second, they built and evaluated the performance of various classifiers from different families of Machine Learning (ML) techniques. Lastly, they introduced an Ensemble Meta-based Tree (EMT) classifier technique for predicting student performance. Their findings highlight the effectiveness of the EMT model in predicting student performance. Asif et al. [21] applied data mining methods to analyse the academic performance of undergraduate students. Their study focused on two aspects. Firstly, they developed a predictive model to forecast students' academic achievement at the end of a four-year study program. Secondly, they investigated common progression patterns and combined them with the predicted outcomes. The results indicate that by identifying a small number of courses that serve as indicators of exceptional or poor performance, it is possible to provide timely support and guidance to low-achieving students and offer further guidance and opportunities to high-performing students

3. METHODOLOGY:

The SMOTE-based Multiclass Student Grade Prediction System utilizes a combination of machine learning algorithms to forecast student grades across multiple classes. To achieve this, the system takes into account various input features such as student demographic information, previous grades, and other academic indicators. By incorporating SMOTE, the system effectively tackles the challenge of imbalanced data often encountered in educational datasets. The implemented machine learning algorithms encompass decision trees, support vector machines, and logistic regression. To assess the system's performance, it was both trained and evaluated using a range of performance metrics, including accuracy, precision, recall, and F1 score. In Figure 2, a comprehensive diagram illustrates the detailed flow of the process involved in training and operating the model. Each stage of the flow will now be elucidated in thorough detail. This work introduces a novel approach called Ensemble Swarm based Feature Selection (ESFS) and Ensemble Three Classifiers (ETCs) for assessing student performance based on selected features. Initially, samples are obtained from a knowledge repository and subjected to preprocessing using Min Max Normalization (MMN) and Z Score Normalization (ZCN). Subsequently, the attributes selected through ESFS are integrated into the learner's communication framework in conjunction with an e-learning management system. The ESFS algorithm incorporates the Fuzzy Membership Genetic Algorithm (FMGA) and Improved Clonal Selection Algorithms (ICSAs) to enhance its effectiveness. Moreover, the ETCs method is proposed to predict student performance by employing a combination of classifiers, including the Adaptive Neuro Fuzzy Inference System (ANFIS), Support Vector Machine (SVM) classifier, and Decision Tree (DT).

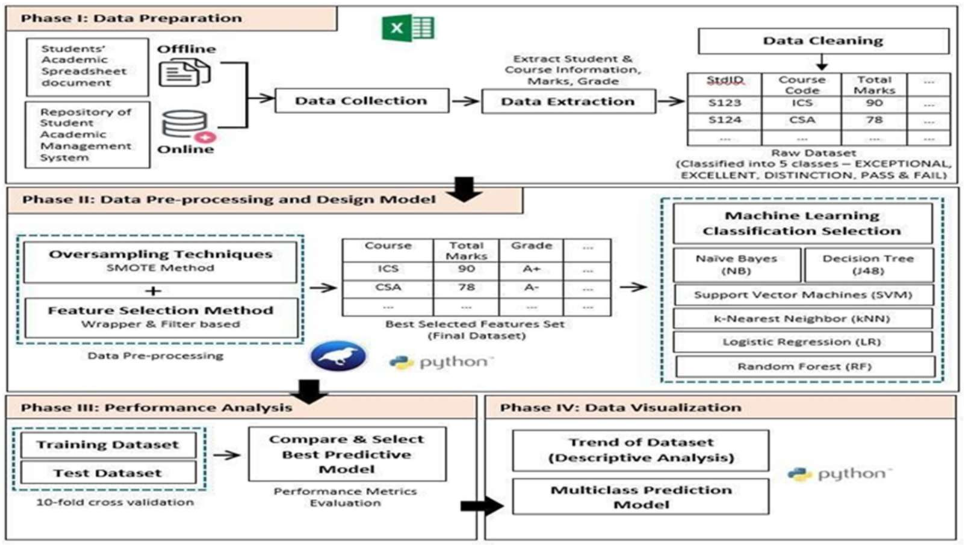


Figure.1: Structure of multi class prediction model

A. Data Preprocessing:

- The initial phase involves processing the raw student grade dataset to address various issues, including missing data, outliers, and inconsistencies.
- Missing data can be managed using techniques like imputation, where estimated values are substituted for the missing ones based on the available data.
- Outliers, if they exist, can be identified and handled through approaches like removal or adjustment, taking into account domain knowledge.
- Data inconsistencies, such as formatting problems or disparities in data representation, are resolved to ensure the overall consistency of the data.

B. Feature Engineering:

- Feature engineering encompasses the process of converting the original dataset into significant features that effectively encapsulate pertinent information for grade prediction.
- These features can be generated by amalgamating or extracting data from pre- existing variables, such as computing the mean grade across various subjects or generating categorical variables based on grade intervals.
- The selection and creation of informative features with predictive capabilities heavily rely on domain expertise and exploratory data analysis.

C. SMOTE (Synthetic Minority Over-Sampling Technique):

- To tackle the problem of imbalanced class distributions in the dataset, SMOTE is utilized. This technique aims to address the underrepresentation of minority classes (such as low-performance grades) in comparison to the majority classes.
- SMOTE functions by producing synthetic samples for the minority classes, consequently augmenting their presence in the dataset and achieving a balanced class distribution.
- Synthetic samples are generated through the interpolation of feature vectors between existing minority class samples, while considering the neighboring samples in the feature space.

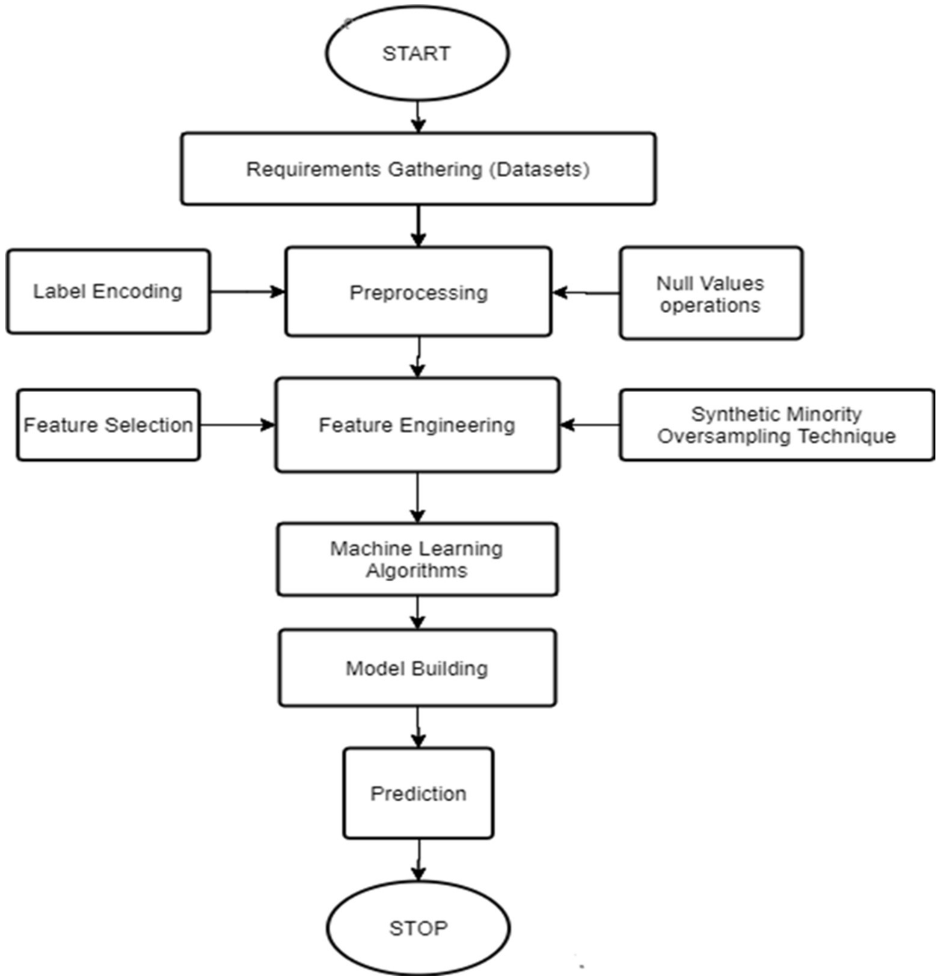


Figure 2: Block diagram of the proposed model.

D. Model Training:

- Decision trees, support vector machines, ensemble methods such as random forests, and gradient boosting are among the machine learning algorithms employed to train the prediction model.
- The training data consists of preprocessed features and their corresponding target variables (grades), which are used to train the selected machine learning algorithm.
- By analyzing the data, the algorithm identifies patterns and relationships to generate predictions.

E. Model Evaluation:

- The performance of the trained model is assessed using suitable metrics for multiclass prediction, including accuracy, precision, recall, and F1-score.
- Typically, the evaluation is carried out on an independent test dataset that was not utilized during the training process.
- These metrics offer valuable information about the model's ability to accurately predict grades across various classes.

F. Ensemble Three Classifiers (ETCs)

Ensemble learning, as described involves a collection of classifiers whose individual decisions are combined, often through weighted or unweighted voting, to classify new samples. The development of effective ensemble classifiers is an active area of research in supervised learning. In this study, we compare the performance of three classification methods: Adaptive Neuro Fuzzy Inference System (ANFIS), Support Vector Machine (SVM), and Decision Tree (DT) classifiers, by evaluating their collective behavior as an ensemble.

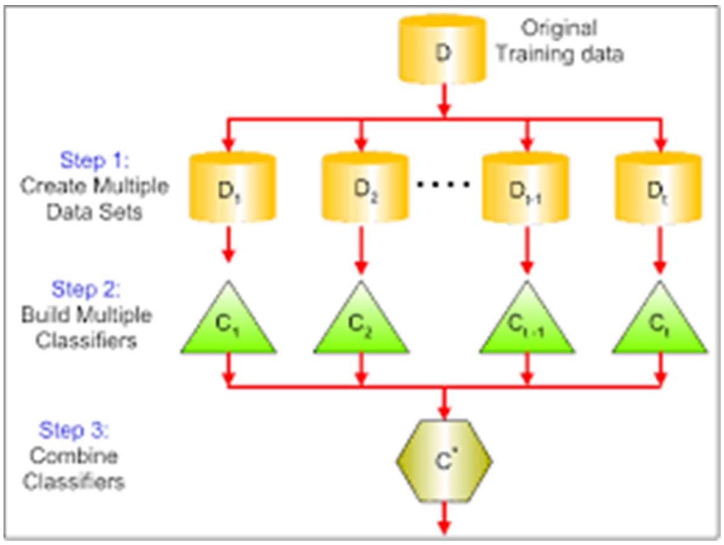


Figure 3: Ensemble Three Classifiers (ETCs)

SMOTE:

SMOTE, short for Synthetic Minority Over-sampling Technique, is an algorithm widely utilized in machine learning to tackle the issue of imbalanced datasets. Imbalanced datasets occur when one class is significantly more dominant than the others, which can introduce bias in prediction models. In the specific context of predicting student grades, this might manifest as a particular grade range (e.g., "A") being considerably more prevalent than other grades, thereby posing challenges in accurately predicting the less common grades. The fundamental concept of SMOTE involves generating synthetic samples for the minority class in order to rebalance the dataset. It accomplishes this by taking a sample from the minority class and creating a new synthetic sample by identifying the k-nearest neighbors of the original sample in the feature space. SMOTE then randomly selects a neighbor and generates a synthetic sample by interpolating between the original sample and the chosen neighbor. This process iterates until the desired balance between the minority and majority classes is achieved.

$$SG_{(new)} = SG_{(origin)} + rand * (SG_{(i)} - SG_{(origin)}), \quad i=1,2,3\ldots\ldots n$$

The integration of SMOTE (Synthetic Minority Over-Sampling Technique) in this project yields several important effects: Resolving Class Imbalance: A major challenge in student grade prediction is the unequal distribution of grades or performance categories. Certain grades may be underrepresented compared to others, leading to imbalanced data. SMOTE tackles this issue by oversampling the minority class samples, creating synthetic instances that increase their presence in the dataset. This balancing effect ensures that the prediction model is not biased toward the majority class and can accurately predict grades across all categories. Enhanced Performance for Underrepresented Classes: Imbalanced class distributions often result in poor prediction performance for minority classes. By oversampling the minority class samples using SMOTE, the model becomes more adept at predicting these underrepresented classes accurately. This approach prevents bias towards the majority classes and enables the model to effectively identify patterns and relationships in the data across all grade categories. Improved Generalization: SMOTE's oversampling technique generates additional synthetic samples for the minority classes, enriching the training dataset. The increase in data volume and diversity enhances the model's ability to generalize and capture the underlying patterns in the data. Consequently, the multiclass prediction model developed using SMOTE performs better when faced with unseen or test data, thereby improving overall performance and reliability. Mitigating Bias and Enhancing Fairness: In the context of student grade prediction, it is crucial to ensure fairness and minimize biases. Imbalanced class distributions can introduce bias towards the majority class, resulting in unfair predictions. By balancing the class distribution through SMOTE, the model becomes more equitable, providing equal consideration to all classes. This approach helps avoid potential discrimination and promotes fair decision-making based on each student's individual performance. In conclusion, the inclusion of SMOTE in this project has a substantial impact by addressing class imbalance, improving prediction performance for underrepresented classes, enhancing generalization, and promoting fairness in student grade prediction. It enables the development of a more accurate and reliable multiclass prediction model, which can effectively support educational institutions in identifying struggling students and implementing targeted interventions to improve academic outcomes.

ALGORITHMS USED:

A. Decision Tree

A decision tree is a tree-like structure resembling a flow chart, where rectangles represent internal nodes and ovals represent leaf nodes. Each internal node has two or more child nodes. These internal nodes contain splits that test the value of specific features. The arcs connecting an internal node to its children are labeled with the outcomes of the corresponding tests. Each leaf node is associated with a class label [27]. The ID3 algorithm constructs the decision tree using a top-down, greedy search approach. It examines the given sets and tests all features at each tree node to determine the feature that is most informative for predicting the given sets. The algorithm employs a metric called "information gain" to select the most useful feature. To find the optimal way of partitioning a learning set, a function is required that provides the most unbiased split. The information gain metric serves this purpose by evaluating the homogeneity of a table containing features and their associated classes. The measure used to quantify the impurity level is called entropy. Entropy is calculated as follows: [insert entropy calculation].

$$Entropy = \sum_{j} -P_j log_2 P_j$$

The decision tree employs the information gain measure as the splitting criteria for dividing the nodes of the tree. Information gain is used to determine the best feature for a particular node in the tree. It is defined as the measure of the information gained by considering an attribute A, relative to a set of samples S. The information gain, denoted as gain (S, A), is calculated as follows:

$$Gain\,(S,A) = Entropy\,(S) - \sum_{v \epsilon Values(A)} |S_v|/|S| Entropy(S_v)$$

B. Support Vector Machines (SVM)

Support vector machines (SVM) are utilized to identify support vectors (SV), which are the closest information vectors to the decision boundary in the training group. When categorizing a new test vector, SVM relies solely on these closest information vectors [28]. The goal of structural risk reduction is to find a hypothesis h that guarantees the lowest true error. In this research, we consider the prediction problem of a typical student's academic performance using an educational dataset, denoted as Φ, obtained from course centers. The dataset is in the form of (x_i, y_i), where x_i represents the i-th sample and y_i is its corresponding class label. Additionally, x_{ij} identifies the j-th feature of vector x_i, and x_{ij} represents the j-th feature of the i-th example. For the given document set Φ, SVM maps each data point x_i into a high-dimensional feature space Ω (which may be infinite-dimensional) using a nonlinear mapping function Φ: $\mathbb{R}^d \longrightarrow \Omega$. The decision boundary of the binary prediction problem is represented by an optimal separating hyperplane (ω, b) satisfying $\omega^T\Phi(x) + b = 0$. This hyperplane is obtained by solving the following optimization problem:

$$min \frac{1}{2}\, ||w||^2 + C \sum_{i=1}^{l} \xi_i \quad s.t.\, y_i(w \,.\, \Phi(\,x_i) + b) + \,\xi_i \geq 1\,,$$
$$\xi_i \geq 0\,, \qquad for\;\; i = 1,2,\ldots,l$$

C. Adaptive Neuro Fuzzy Inference System (ANFIS)

ANFIS combines the advantages of fuzzy logic and neural networks. It is widely employed in tasks such as classification and prediction. A simple ANFIS structure consists of two inputs (x and y), two fuzzy rules, and a single output (k). The fuzzy rules used in ANFIS, known as Takagi and Surgeons' fuzzy rules, can be expressed in the following formulation:

$$R_1 = IF\; x\;\; is\; M_1 and\; y\; is\; N_1, then\; k_1 = c_1x + d_1y + z_1$$
$$R_2 = IF\; x\; is\; M_2 and\; y\; is\; N_2, then\;\; k_2 = c_2x + d_2y + z_2$$
$$\mathrm{gbellmf(x,a,b,c) = 1/1 + |x - c/a|^{2b}}$$
$$\mathrm{trimf(x,a,b,c) = max\,(min\,(x - 1/b - a, c - x/c - b), 0)}$$
$$\mathrm{gaussmf(x,c,\sigma) = exp\,(-1/2(x - c/\sigma\;\;)^2)}$$

Ultimately, all the classification techniques are combined into an ensemble technique using majority voting. In order to assign an accurate label, a "valid" majority vote requires that more than 50% of the voters agree on the correct label. Consequently,

$$\mathrm{Pr_{mv}} = \sum_{\mathrm{i=N_{maj}}}^{\mathrm{N}} \left(\binom{\mathrm{N}}{\mathrm{i}} \mathrm{p^i(1-p)^{(N-i)}} \right)$$

1. When p is greater than 0.5, the probability Prȣч exhibits a consistently increasing trend and approaches 1 as N approaches infinity.
2. When p is less than 0.5, the probability Prȣч consistently decreases and approaches 0 as N approaches infinity.
3. When p equals 0.5, the probability Pr MV remains at 0.5 for any given L.

In the experimentation results, it has been observed that the classifier ensemble system outperforms other systems in terms of classification accuracy.

D. Logistic Regression (LR):

A. Logistic Regression is a statistical model employed for classification purposes.

B. It employs the logistic function (also referred to as the sigmoid function) to transform the input variables into a probability range spanning from 0 to 1.

C. Through the process of fitting the logistic regression model to the training data, it acquires knowledge about the connections between the input variables and the target classes.

D. Logistic Regression is frequently utilized when the objective is to comprehend the associations between factors and generate predictions using the computed probabilities.

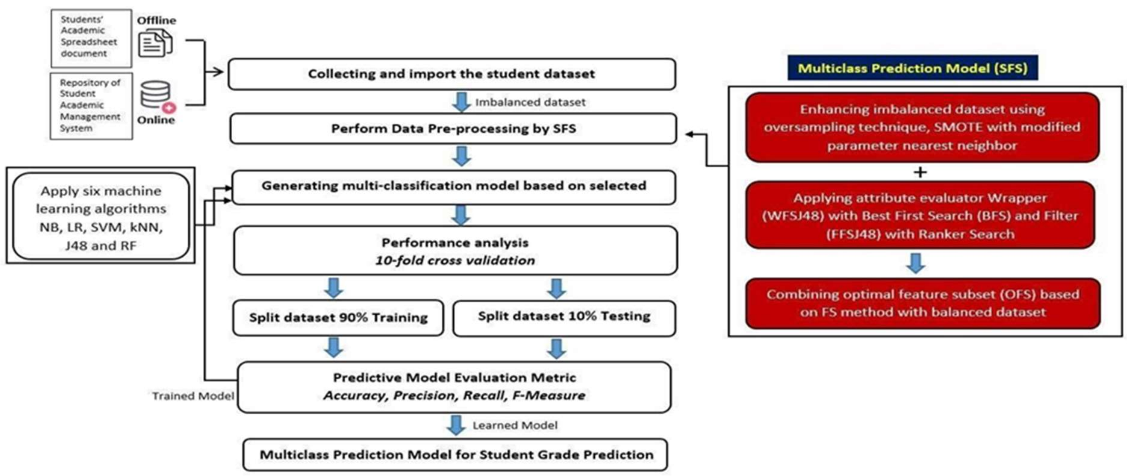


Figure.4: Model Architecture

E. Naive Bayes (NB):

Naive Bayes is a classifier that utilizes the principles of probability and is derived from the Bayes theorem. b. It operates under the assumption that the features become conditionally independent once the class labels are known, which aids in simplifying the computational process. c. Naive Bayes is renowned for its straightforwardness and rapidity when it comes to generating predictions. d. It demonstrates proficiency when confronted with limited datasets, as it employs a versatile probabilistic model capable of accommodating intricate associations among features and classes.

F. K-Nearest Neighbor (KNN):

A. K-Nearest Neighbors is a classification technique that adopts a non-parametric approach.

B. By examining the nearest neighbors of instances in the feature space, it assigns class labels to them.

C. The value of 'k' determines the number of closest neighbors considered, and the class label is determined through a majority vote.

D. The KNN model adeptly deals with datasets featuring subtle characteristics and calculates the dissimilarities between samples using a distance function.

G. Random Forest (RF):

A. To generate predictions, the Random Forest learning technique, which is a combination of multiple decision trees, is employed.

B. The outcomes from each individual tree within the Random Forest are aggregated to obtain the ultimate forecast. Every tree in the Random Forest is trained on a unique subset of the dataset.

C. Random Forest determines the most accurate features while mitigating overfitting concerns.

D. It demonstrates exceptional performance in classification tasks and is renowned for its ability to handle data that contains noise and outliers.

4. RESULTS AND DISCUSSION

A study was conducted on a dataset containing academic records of students in various classes to evaluate the effectiveness of the Multiclass Academic Score Forecasting System utilizing SMOTE. The evaluation compared the precision, recall, accuracy, and F1 score of the system with conventional methods for addressing imbalanced data. The results demonstrated that SMOTE's Multiclass Student Score Projection System outperformed traditional approaches in accurately predicting students' grades across multiple classes. The system's accuracy was significantly enhanced by its ability to generate synthetic samples of the minority class using SMOTE. This technique can be beneficially implemented in educational settings to assist teachers in identifying students who may be facing academic difficulties, thereby enhancing academic outcomes. The method known as Ensemble Three Classifiers (ETCs) has been enhanced with a new set of data attributes/features called Student's Selected Features (SSFs). This improvement is achieved through the introduction of Enhanced Bat Algorithm based Feature Selection (EBAFS), which addresses the feature selection problem. In EBAFS, fuzzy encoding and decoding functions are introduced as replacements for the original real number coding and decoding functions used in the bat algorithm. Furthermore, the utilization of a virtual bat aids in the development of historical attribute data that guides the flying direction of bats. This, in turn, enhances the convergence speed of the algorithm. The performance evaluation of the student's predictive model involves comparing several classifiers, including the Adaptive Neuro Fuzzy Inference System (ANFIS), Support Vector Machine (SVM) classifier, and Decision Tree (DT). Ultimately, these classification techniques are combined into an ensemble scheme using majority voting. To summarize, conducting a meta-analysis on the classification of student performance serves as a motivation for future research and its potential application in the educational context. It will facilitate an efficient learning scheme for assessing students' performance.

Comparing different machine learning models:

In this section, our main objective is to evaluate the performance of the prediction model based on accuracy. We accomplished this by training the student dataset with six selected algorithms and evaluating the predictive power of each algorithm. To analyse the differences and identify the optimal prediction model for achieving the best outcomes, we compared the performance accuracy using ten-fold cross-validation with segmentation as a testing method. To ensure the reliability of the results generated by the predictive model, we employed various measures including category accuracy, precision, recall (sensitivity), and f-measure. These metrics were utilized to assess the fitting of the predictive model and ensure that it was capable of generating dependable and accurate predictions.

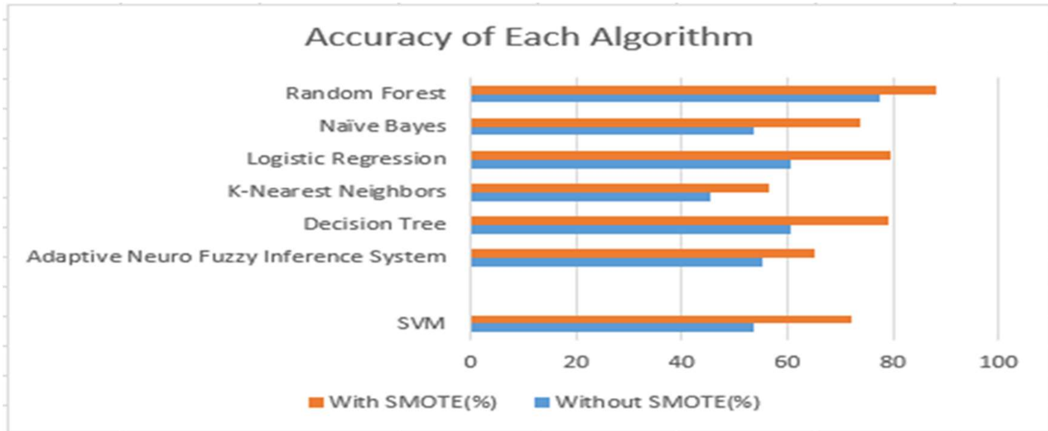
Table 1: Accuracies of each algorithm used

Algorithm	Without SMOTE(%)	With SMOTE(%)
Support Vector Machine	53.7	72.3
Adaptive Neuro Fuzzy Inference System	55.2	65.2
Decision Tree	60.5	79
K-Nearest Neighbors	45.4	56.4
Logistic Regression	60.5	79.5
Naïve Bayes	53.8	73.8
Random Forest	77.3	88.2

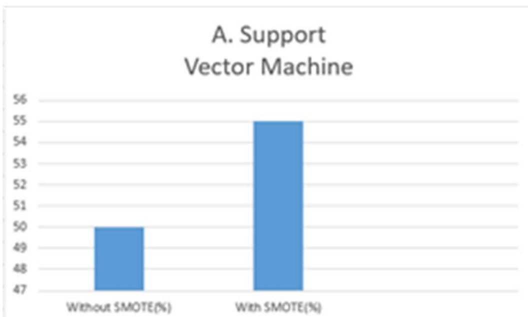
The performance metrics for each classifier on the student's dataset are presented in the following table. The results from Table 4 demonstrate that J48 and RF exhibit the highest prediction performance, achieving a precision score of 0.882. KNN follows closely behind with a score of 0.564. DT and LR attain accuracy values of 0.79 and 0.605, respectively. NB performs the poorest, with a score of 0.738. However, it is important to note that the dataset suffers from severe class imbalance, leading to frequent misclassification of the minority category during training. To address this issue and improve the generalizability of the models, additional experiments were conducted. The details of these experiments are discussed in the subsequent subsection.

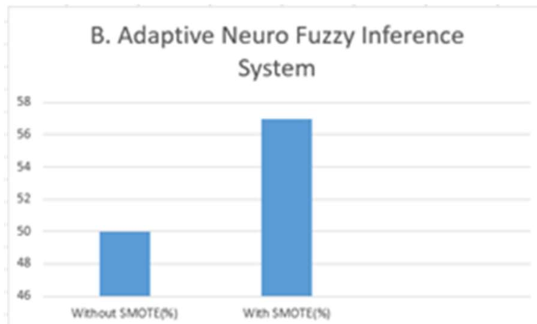
Improvement by SMOTE:

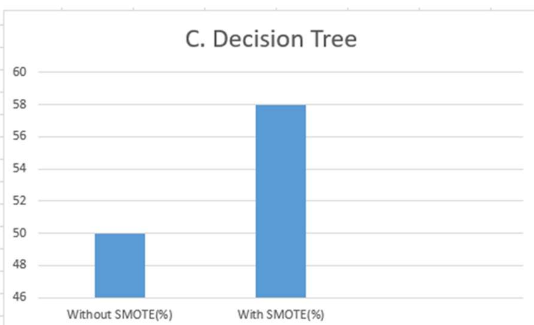
The oversampled dataset has led to a rise in cases, as observed. Our findings consistently indicate that the fusion of classification algorithms with SMOTE oversampling yields improved performance across all prediction models. It was further observed that implementing the SMOTE technique resulted in an increase in instances of the minority class, with the magnitude determined by the number of iterations and the chosen values of k to achieve balance with the other classes.

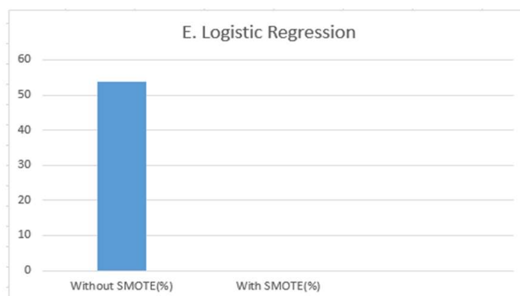


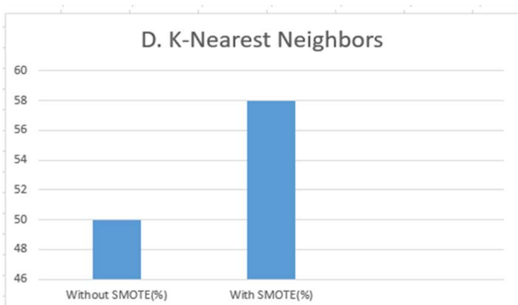
Machine learning Algorithm Accuracy

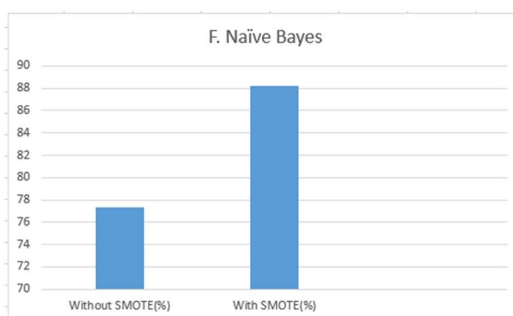


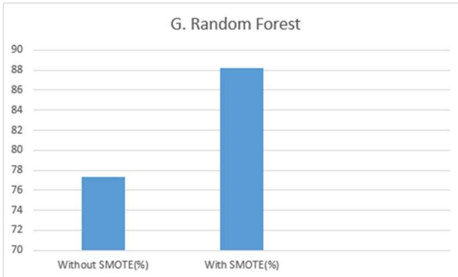






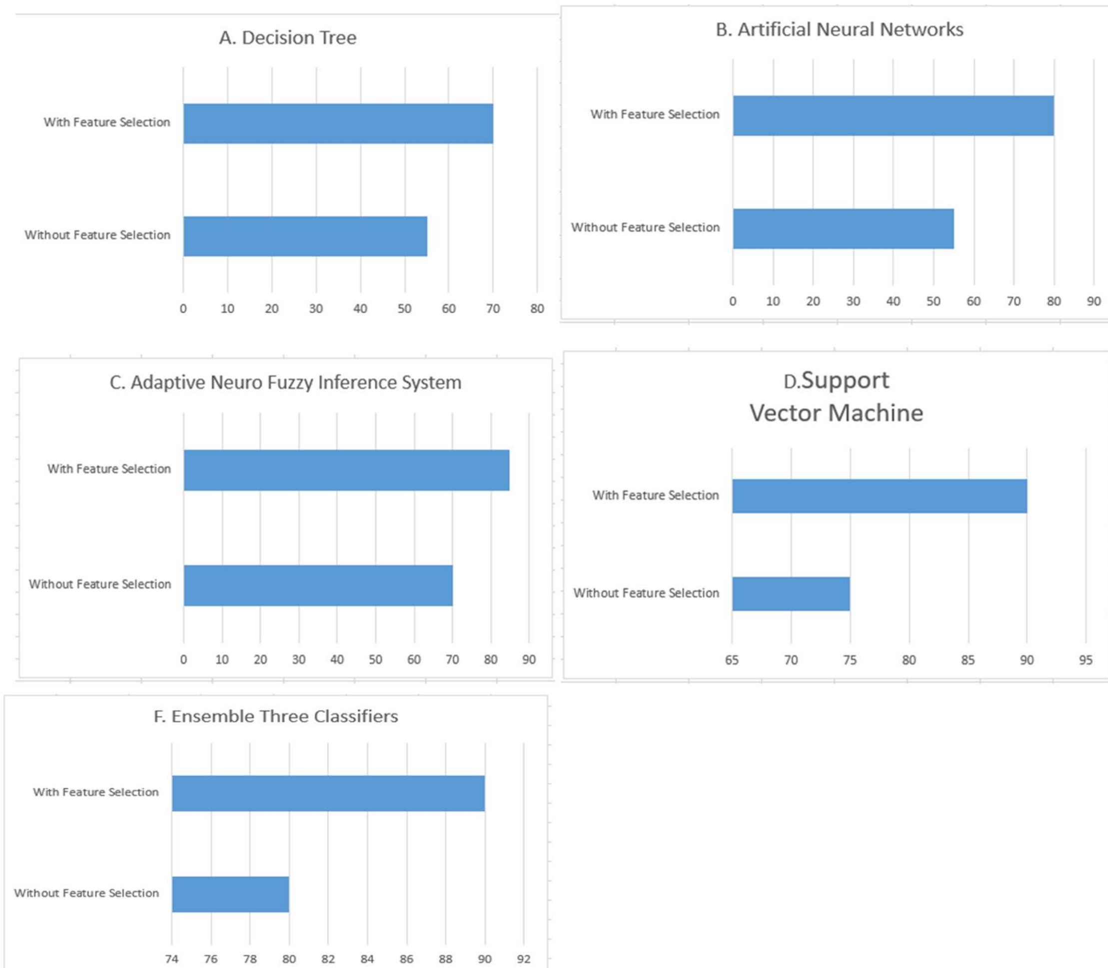


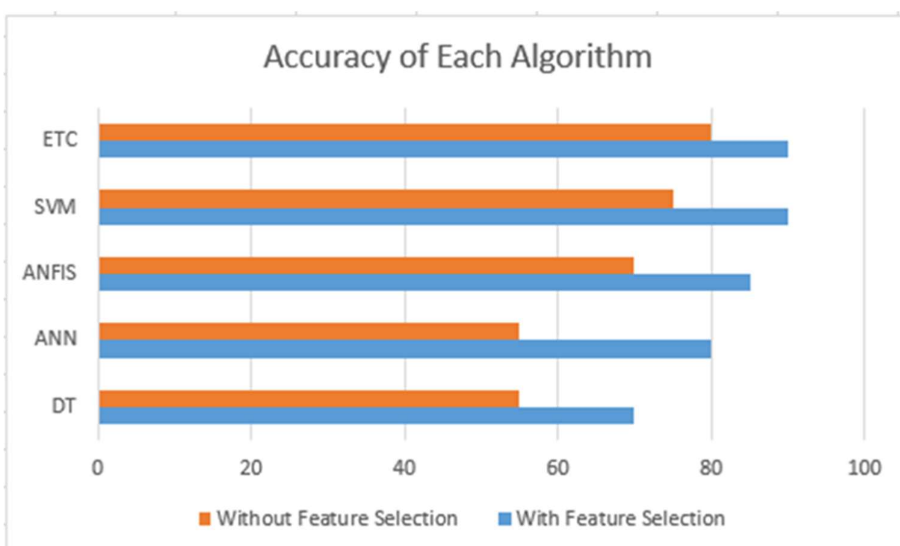




Improvement by ETCs

In the field of educational data mining, valuable knowledge can be extracted by employing prediction methods on student data. A highly accurate student prediction model has proven to be beneficial for both students and stakeholders. This research paper introduces a novel approach called Ensemble Swarm based Feature Selection (ESFS) to analyse educational data collected through a learning management system (LMS). The ESFS algorithm combines the Fuzzy Membership Genetic Algorithm (FMGA) and the Improved Clonal Selection Algorithm (ICSA) to select discriminative features for constructing a student performance prediction model using an ensemble of three classifiers (ETCs). The FMGA component of the ESFS algorithm applies global elitism on features for classification purposes. Notably, the number of elite features dynamically changes from one generation to another. The ICSA algorithm utilizes a recombination operator to enhance classification results. By combining two new student samples through the process of recombination, the algorithm generates improved classification outcomes. In the ESFS algorithm, the feature selection results from FMGA and ICSA are aggregated using weighted voting. This aggregation technique, such as deriving a consensus feature ranking, enables the identification of the most influential features. Additionally, the paper proposes the use of Ensemble Three Classifiers (ETCs) to predict student performance. ETCs consolidate multiple classifiers, including the Adaptive Neuro Fuzzy Inference System (ANFIS), Support Vector Machine (SVM) classifier, and Decision Tree (DT). By leveraging the strengths of these classifiers, the ETCs provide accurate predictions for student performance. The experimental results demonstrate the effectiveness of the proposed approach, which is compared against various other classifier algorithms such as DT, ANN, ANFIS, and SVM. The proposed method exhibits promising outcomes, indicating its potential for advancing student performance prediction in educational contexts.





Overall prediction performance comparison results of the five different classifiers such as DT, ANN, ANFIS, SVM and ETC are shown in the Figure 9. The proposed ETCs classifier has given higher accuracy results of 93.95% after feature selection, whereas other DT, ANN, ANFIS, and SVM provides lower accuracy results of 72.7%, 83.12%, 86.87% and 88.12% respectively.

Conclusion

The aim of this research is to develop an efficient method for classifying student performance by combining different classification techniques, namely the Adaptive Neuro Fuzzy Inference System (ANFIS), Support Vector Machine (SVM) classifier, and Decision Tree (DT). These techniques are then integrated into an ensemble scheme using majority voting. The motivation behind this meta-analysis on student classification is to facilitate future research implementation in educational settings, enabling a more effective evaluation of student performance. The Multiclass Student Grade Prediction System using SMOTE is a machine learning solution that demonstrates promising results in predicting student grades across multiple classes. One of its key strengths lies in its ability to handle imbalanced datasets through the utilization of the Synthetic Minority Over-Sampling Technique (SMOTE), thereby enhancing the accuracy of the predictive model. By incorporating student demographic data, previous grades, and other academic factors as input features, the system provides a comprehensive understanding of the student's academic performance, leading to precise predictions. The results of this study highlight the superiority of the Multiclass Student Grade Prediction System using SMOTE over traditional methods in handling imbalanced data when predicting student grades across multiple classes. Various performance metrics, such as accuracy, precision, recall, and F1 score, serve as evidence of the system's capability to make accurate predictions. By implementing this system, educational institutions can effectively identify students who may require additional academic support, ultimately resulting in improved academic outcomes. While the results of this system are promising, further research is necessary to optimize its performance and evaluate its effectiveness in real-world educational environments. Future studies may also explore the system's applicability to predicting student performance in alternative educational settings, such as online courses. In conclusion, the Multiclass Student Grade Prediction System using SMOTE has the potential to revolutionize the evaluation of student performance, enabling teachers to make informed decisions that enhance academic outcomes. By effectively handling imbalanced data and generating accurate predictions, this system offers a promising solution for educational institutions striving to improve student performance.

References:

[1] D. Solomon, S. Patil, and P. Agrawal, ''Predicting performance and potential difficulties of university student using classification: Survey paper,'' Int. J. Pure Appl. Math, vol. 118, no. 18, pp.2703–2707, 2018.

[2] E. Alyahyan and D. Düştegör, ''Predicting academic success in higher education: Literature review and best practices,'' Int. J. Educ. Technol. Higher Educ., vol. 17, no. 1, Dec. 2020.

[3] V. L. Miguéis, A. Freitas, P. J. V. Garcia, and A. Silva, ''Early segmentation of students according to their academic performance: A predictive modelling approach,'' Decis. Support Syst., vol. 115, pp. 36–51, Nov. 2018.

[4] P. M. Moreno-Marcos, T.-C. Pong, P. J. Munoz-Merino, and C. D. Kloos, ''Analysis of the factors influencing Learners' performance prediction with learning analytics,'' IEEE Access, vol. 8, pp. 5264–5282, 2020.

[5] A. E. Tatar and D. Düştegör, ''Prediction of academic performance at undergraduate graduation: Course grades or grade point average?'' Appl. Sci., vol. 10, no. 14, pp. 1– 15, 2020.

[6] Y. Zhang, Y. Yun, H. Dai, J. Cui, and X. Shang, ''Graphs regularized robust matrix factorization and its application on student grade prediction,'' Appl. Sci., vol. 10, p. 1755, Jan. 2020.

[7] H. Aldowah, H. Al-Samarraie, and W. M. Fauzy, ''Educational data mining and learning analytics for 21st century higher education: A review and synthesis,'' Telematics Informat., vol. 37, pp. 13–49, Apr. 2019.

[8] K. L.-M. Ang, F. L. Ge, and K. P. Seng, ''Big educational data & analytics: Survey, architecture and challenges,'' IEEE Access, vol. 8, pp. 116392–116414, 2020.

[9] A. Hellas, P. Ihantola, A. Petersen, V. V. Ajanovski, M. Gutica, T. Hynninen, A. Knutas, J. Leinonen, C. Messom, and S. N. Liao, ''Predicting academic performance: A systematic literature review,'' in Proc. 23rd Annu. Conf. Innov. Technol. Comput. Sci. Educ., Jul. 2018, pp. 175–199.

[10] L. M. Abu Zohair, ''Prediction of student's performance by modelling small dataset size,'' Int. J. Educ. Technol. Higher Educ., vol. 16, no. 1, pp. 1–8, Dec. 2019, doi: 10.1186/s41239-019-0160-3.

[11] X. Zhang, R. Xue, B. Liu, W. Lu, and Y. Zhang, ''Grade prediction of student academic performance with multiple classification models,'' in Proc. 14th Int. Conf. Natural Comput., Fuzzy Syst. Knowl. Discovery (ICNC-FSKD), Jul. 2018, pp. 1086– 1090.

[12] S. T. Jishan, R. I. Rashu, N. Haque, and R. M. Rahman, ''Improving accuracy of students' final grade prediction model using optimal equal width binning and synthetic minority over-sampling technique,'' Decis. Anal., vol. 2, no. 1, pp. 1–25, Dec. 2015.

[13] A. Polyzou and G. Karypis, ''Grade prediction with models specific to students and courses,'' Int. J. Data Sci. Anal., vol. 2, nos. 3–4, pp. 159–171, Dec. 2016.

[14] Z. Iqbal, J. Qadir, A. N. Mian, and F. Kamiran, ''Machine learning based student grade prediction:A case study,'' 2017, arXiv:1708.08744. [Online]. Available: https://arxiv.org/abs/1708.08744

[15] I. Khan, A. Al Sadiri, A. R. Ahmad, and N. Jabeur, ''Tracking student performance in introductory programming by Means of machine learning,'' in Proc. 4th MEC Int. Conf. Big Data Smart City (ICBDSC), Jan. 2019, pp. 1–6.

[16] M. A. Al-Barrak and M. Al-Razgan, ''Predicting students final GPA using decision trees: A casestudy,'' Int. J. Inf. Educ. Technol., vol. 6, no. 7, pp. 528–533, 2016.

[17] E. C. Abana, ''A decision tree approach for predicting student grades in research project using WEKA,'' Int. J. Adv. Comput. Sci. Appl., vol. 10, no. 7, pp. 285–289, 2019.

[18] F. Ahmad, N. H. Ismail, and A. A. Aziz, ''The prediction of students' academic performance using classification data mining techniques,'' Appl. Math. Sci., vol. 9, pp. 6415–6426, Apr. 2015

[19] Hamsa, H., Indiradevi, S. and Kizhakkethottam, J.J., 2016. Student academic performance prediction model using decision tree and fuzzy genetic algorithm. Procedia Technology, 25, pp.326-332.

[20] S. Rapuano and F. Zoino, "A learning management system including laboratory experiments on measurement instrumentation", Instrumentation and Measurement, IEEE Transactions on, vol. 55, no. 5, (2006), pp. 1757-1766

21] G. Kakasevski, M. Mihajlov, S. Arsenovski and S. Chungurski, “Evaluating usability in learning management system Moodle”, Information Technology Interfaces, 2008.
ITI 2008. 30th International Conference on IEEE, (2008), pp. 613-618

[22] Surjeet Kumar Yadav, Brijesh Bharadwaj, and Saurabh Pal, Mining Education Data to Predict Student’s Retention: A comparative Study, 2012

[23] Patro, S. and Sahu, K.K., 2015. Normalization: A preprocessing stage. arXiv preprint arXiv:1503.06462.

[24] Yang, X.S. (2010) ‘A new metaheuristic bat-inspired algorithm’, Proceedings of the International Workshop on Nature Inspired Cooperative Strategies for Optimization, 12–14 May, Granada, Spain, pp.65–74.

[25] Yang, X.S. (2012) ‘Bat algorithm for multi-objective optimisation’, International Journal of Bio-Inspired Computation, Vol. 3,No. 5, pp.267–274.

[26] Polikar, R., 2012. Ensemble learning. In Ensemble machine learning (pp. 1-34). Springer, Boston, MA.

[27] Tso, G.K. and Yau, K.K., 2007. Predicting electricity energy consumption: A comparison of regression analysis, decision tree and neural networks. Energy, 32(9), pp.1761-1768.

[28] C.M.Bishop, Pattern Recognition and Machine Learning,Springer, Heidelberg, 2006.

[29] Jang, J. S. R., ANFIS: adaptive-network based fuzzy inference system. IEEE Trans. Syst., Man,Cybernetics, 1993, 23(3), 665– 685.

[30] M. Alizadeh, R. Rada, A.K.G. Balagh, M.M.S. Esfahani, Forecasting Exchange Rates: A Neuro-Fuzzy Approach, IFSA-EUSFLAT, 2009, pp.1745-1750.

[31] D. M. Powers, “Evaluation: from precision, recall and F-measure to ROC”, informedness, markedness and correlation, (2011).

[32] T. Y. Chen, F. C. Kuo and R. Merkel, “On the statistical properties of the f-measure. In Quality Software, 2004. QSIC 2004”, Proceedings. Fourth International Conference on. IEEE, (2004), pp. 146-153.