International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING

ISSN:2147-6799 www.ijisae.org Original Research Paper

Analysis of Student's Education Data Based on Data Mining Techniques

Vijayakumar Thota[1], S. Sharanyaa[2], Ayisha Noori V. K.[3], K. R. Shanthy[4], M. Bharathiraja[5]

Submitted: 15/11/2023 **Revised:** 28/12/2023 **Accepted:** 08/01/2024

Abstract: One of the trickiest and most popular research fields in educational data mining (EDM) is student achievement analysis. Scientists get attracted to this issue owing to the fluctuating implications of multiple factors on functionality. This dedication is additionally ignited by the huge consumption of instructional records, especially when it comes to online learning. Although there are numerous EDM surveys in the scholarship, there aren't plenty that solely concentrate on student achievement evaluation and projection. These specialized assessments are small in focus and largely emphasize investigations that seek to find potential predictors or patterns of student achievement. This paper proposes data mining through the required algorithms for the accurate extraction of data for further analysis. A brief overview of the current situation of studies in that field is the goal that this literature review attempts to communicate. We employed a couple of methods for performing a literature review: initially, we employed the primary search engines to identify documents, and then we picked them based on previously established requirements. The info collected from student conversations with learning management systems and assessment tasks were the most important elements in early forecasting. At last, the kind of schooling system identified how early projections could be formed.

Keywords: *Performance prediction, Education data mining, DM techniques*

1. Introduction

Online learning (OL) scenarios are incrementally overtaking traditional educational environments as a consequence of the incorporation of ICT in learning. Students' learning experiences are improved by the OL system, which also minimizes the requirement for their educator to be involved directly. Various DM strategies should be applied when analyzing this data to uncover pertinent content and present it in an approach that renders decision-making simpler and optimizes the efficacy of the learning method. Multiple sources, which comprise enrollment records, course content, education systems, and knowledge regarding the course's timing, achieve diverse kinds of data. Identical to this, an array of other applications that are hosted on the internet used in educational settings, such as educational games, online quizzes and tests, virtual universes, forums and notice boards, interactive multimedia networks, user activity logs, and various other forms of instruction and text, additionally deliver a variety of numerous kinds of data [1]. The expression "data mining" indicates a scheduled, rigorous technique for retrieving patterns and hidden, previously undiscovered, but possibly significant details from an enormous amount of data. Engineers may propose new algorithms and frameworks and health care organizations may generate novel drugs and antibodies through data mining. These researchers argue that data mining delivers novel ideas and directions for scientists' investigation [2]. Even though data mining performs an essential part in business analytics, the field comprises much more than simply information extraction. We begin by determining terms and their implications. Afterwards, we provide our viewpoint on supply chain handling while looking at how academic literature has investigated business analytics, particularly data mining [3]. The most accurate depiction of the method that the data mining action will be utilized for sorting educational data is portrayed in Figure 1.1.

[1]Associate Professor, Business Management,
NSB Business School, Bangalore.
Email: vijayssmba@gmail.com
ORCID: 0000-0002-8738-681X
[2]Assistant Professor, Department of Information Technology,
Panimalar Engineering College, Chennai 600123.
Email: Rnsharanyaa@gmail.com
ORCID: 0000-0001-7119-7718
[3]Assistant professor, Dept of Artificial Intelligence,
Madanappalle Institute of Technology and Science, Kadiri Road,
Angallu Madanapalle, Andhrapradesh - 517325
Email: ayishanoorivk@mits.ac.in
ORCID: 0009-0000-4218-7764
[4]Associate Professor,
ECE Department,
Loyola Institute of Technology, Palanchur, Chennai.
Email: jenefa.christ@gmail.com
ORCID: 0000-0002-0681-2881
[5]Professor, Automobile Engineering,
Bannari Amman Institute of Technology,
Sathyamangalam, Erode.
Email: bharathiraja@bitsathy.ac.in
ORCID: 0000-0003-1021-1840

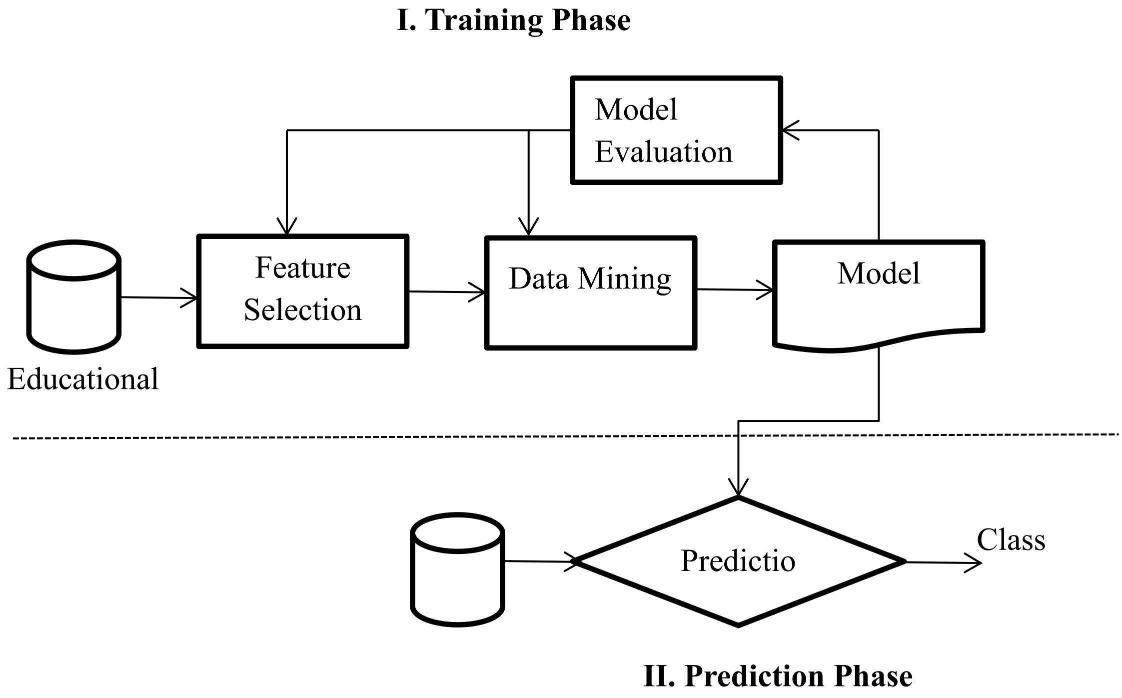

Fig. 1.1. Data mining to sort education performance

The rapid growth of web-based educational platforms in recent years has rendered it vital for us to keep track of enormous quantities of possibly useful information from numerous sources in a wide range of forms and granularities. An immense quantity of data concerning students is collected as well via new educational settings like game-based learning, virtual/enhanced education, mobile/ubiquitous learning, blended learning (BL), etc. Despite the vast amount of highly valuable instructional material produced by all these systems, it is impossible to manually examine it. Thus, the growing need for tools to intelligently assess such types of data stems from the reality that it delivers an extensive amount of academic information that can be analyzed and employed for better comprehension of how kids study [4].

An encompassing summary of each section of this proficient effort is given in this subsequent article. Section 2 renders explicit that the inquiry's primary concentration will be on associated previous studies. The unique attributes of data mining to analyze student's education performance are described in Section 3, with particular significance attached to the facets of the system project, statistical examination, intelligible conceptual framework, and graph-based approach. This third part explores data mining approaches for ensuring the education data analysis with the unique algorithm. A range of unique graphics and charts promoting data mining are shown in Section 4. The conclusion for this approach based on results is available in Section 5.

2. Related Works

Buenaño-Fernández, D., Gil et. al [5] The objectives of knowledge explorers, decision-making, and suggestion creation are all tackled in an extensive range of EDM-related work that conveys plenty of fascinating techniques and strategies. We go through a few of them that have provided data needed for this study in the chapters that follow. Big data methods can be employed in a wide range of ways to facilitate educational analysis, which comprises functionality forecasting, decreased risk detection, the visualization of data, intelligent suggestions, course endorsement, student skill computation, behavior screening, and student grouping and cooperation and in accordance to a study on the utilization of enormous quantities of data in education. The importance of predictive analysis—which concentrates on estimating student behavior, capability, and performance—is pointed out in this article.

Khan, A., & Ghosh, S. K. et. al [6] One of the recognized disciplines of instructional studies is the behavior assessment of pupil achievement. The early identification of a student's accomplishment has advantages for both organization and learners. Forecasting of pupil achievement is also helpful in the systematic monitoring of learners by the system of education. Scholars encourage teachers to apply data mining for student achievement inspections for the reason digital information is growing more and more readily available.

Asif, R., Merceron et. al [7] Based on the private, pre-university, and university performance indicators of pupils, a few of the well-known research articles developed algorithms for forecasting students' academic progress. The probes incorporate 10,330 pupils' quantity of data. Twenty attributes—gender, birth year and residence location, country, place, and total score from preceding education, current semester, etc. were used to describe every pupil. The learners were split into five categories employing strategies for data mining, including the decision tree, Naive Bayes, Knearest Neighbors (KNN) algorithms, etc: Excellent, Very Good, Good, Average, or Bad. Having the highest overall precision, the pick-up tree predictor surpassed both k-NN classifier and JRip.

PAPADOGIANNIS, I., POULOPOULOS et. al [8] A variety of academics have lately centered their study on the EDM aspect of student achievement measurement. Several research papers have been independently published, a great number of which pertain to higher learning. Without any hesitation, early identification is essential since educational systems might use it as evidence for their courses and other endeavors. Multiple approaches for data mining and predictive features in similar to biographical and speculative behavioral attributes of students, their online acts, etc.—are implemented in an attempt to predict academic performance. The main objective of this paper is to look back on the studies conducted all around the last five years.

Azwa, A. A., & Fadhilah, A. et. al [9] Institutions of higher learning encounter an enormous challenge in operating an educational setting up to better serve their main clientele, their learners, in all facets. The application of data mining techniques in the field of education is giving the IHL access to new viewpoints that help them make better decisions and find solutions for any problem that arises. Previous study data suggests that SAP-related concerns are the most commonly studied subject. The explanation for this is that every one of the public and private sectors needs competent contenders to fill vacant jobs. Therefore, when trying to categorize SAP and facilitate further action to be executed to raise scores for students, SAP prognosis is also crucial.

Kumar, M., Singh et. al [10] In the context of educational data mining, it is conventional to make estimations on the educational progress of learners. Diverse data mining techniques, such as categorization, aggregating association rule mining, and regression analysis, should be taken into consideration to construct predictive modeling techniques. Practically all research papers employ only one classification approach to predict the educational achievement of students. We are thinking about only decision trees, Naive Bayes, SVM, ANN, K-Nearest Neighbors, SMO, Linear Regression, Random Forest, Random Tree, J48, etc. There are a lot of techniques for classification readily available for projection.

Kabakchieva, D. et. al [11] The key objective of the research project is to demonstrate the enormous scope of data mining applications for educational institutions, especially when it comes to handling the most effective application of data mining techniques and methods for closely examining the acquired historical data. In data mining terms, the assignment's specific objective—classifying university learners based on their pre-university aspects and college academic results—is believed to be an analysis issue that must be finished using the pupil information that is currently accessible. Because these categorization models are generated using details where the desired (or response) variable is known, the activity is included in the category of supervised learning.

Yağcı, M. et. al [12] Two facets of undergraduate students' performance applying DM procedures were the primary focus of the current paper. The first section pertains to assessing the academic achievements of students at the conclusion of a four-year empirical studies program. The following strategy is to examine how youngsters are growing and blend that with outcomes from forecasts. The author detached the students into communities by academic standing: poor achievement as well as elevated achievement. According to the author's research, it's important instructors spend on a particular set of programs that demonstrate extremely good or poor performance. This enables them to nurture high-achieving scholars with prospects and assistance, as well as prompt action cautions and support for failing pupils.

3. Methods and Materials

The schematic diagram of the data mining procedure stages is available in below Figure 3.1. Experimentation through data mining methods will proceed along the unique trajectory that constitutes the sample and data collection, data pre-processing, building precision systems, evaluation, and the final phase is the conclusion.

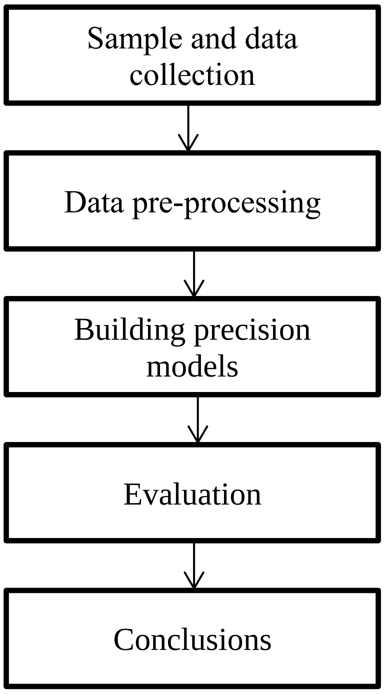

Fig. 3.1. Trial procedure

3.1 Sample and Data Collection

The research in question was undertaken by a vocational and scientific university in Taiwan at the completion of the inaugural term of the educational year 2020–2021. The school assessment and enrollment procedures were utilized for gathering data for the prototypes. Students needed to use the school register technology to fill out a computerized application with their private data when they were first accepted at the institution. Then, as each pupil progressed through their studies, their marks and achievements were entered into the school's evaluation system. In addition, the maintained profile of every pupil at the start of the study incorporated eighteen variables pertaining to their private circumstances and one parameters that reflected the mean score of all the different fields' rankings, which they had succeeded in accomplishing in the initial semester.

3.2 Data pre-processing

We accomplished data cleaning and data harmonization techniques at the data pre-processing stage. Once the 18 variables that made up the output (learning performance) were determined, we sorted category data and corrected missing value instances in the data clean stage. This stage entailed encoding group data and rejecting examples with values that were missing. Equation (1) was employed for normalizing the data in the data purification stage.

$$Y_{mon} = \frac{Y - Y_{min}}{Y_{max} - Y_{min}} \quad (1)$$

where Y_{min}, Y_{max}, and Y_{mon} indicate the minimum, maximum, and normalized values, correspondingly.

3.3 Building prediction models

The experiments included in the research were conducted utilizing 64 GB of RAM and a 3.80 GHz Intel(R) Xeon(R) E-2174G CPU executing Windows operating systems. Reliant to the MLP, RF, and DT computations, four controlled learning models were developed. Python was implemented to create RF prediction models, whereas C5.0 and CART strategies were employed for generating DT prediction models. For all of the models, the investigation was run over five occasions. The average values and the standard error of the arranging results in each case were subsequently utilized as a baseline when assessing the DT and RF computation. The primary goals of several research projects were to find attributes that have a major effect on students' learning performance and analyze and contrast their efficacy in forecasting the academic success of newcomers on the dataset.

In addition, the current investigation incorporates the following three possible types of product information.

- Case 1 is the foundation for the outcome, which is the Excellent, Very Good, Good, Average, and Poor classifications used to calculate the start-up and overall forecasting reliability of the four models.
- Case 2 involves integrating the Very Good, Good, and Average classes from the vast majority output into the Standard class to determine if any of the four approaches can envisage the minority.
- Case 3 narrows the spotlight to the accomplishments of the smaller, the Exceptional and Poor periods.

3.4 Decision Tree

The aforementioned steps were the core of the C5.0 algorithm's testing process for each of the three circumstances in this investigation. Set up decision tree variables, create a preliminary rule tree, create information from testing and training, maintain the tree, and then filter the pruned tree to make it simpler to grasp. Now select the tree that generates the best among all the trees that were established. Then repeat steps 1–6 for ten tests and use the standard deviation and medians of the classification accuracy in ten attempts as a benchmark. Employing a 10-fold cross-validation inquiry, we produced a DT using the C5.0 procedure for every fold in the data set. Ten equal-sized sets were constructed from the information retrieved sets, and each set functioned as the test set in turn. Alongside the test set, we developed DT leveraging nine other sets as our training material. As a consequence, we had ten trees. After picking the tree with the best effectiveness, all of the surviving characteristics were valued identically.

The current investigation implemented the CART strategy by Python as a further technique to test, gauge, and interpret the precision of prediction and feature significant choices between C5.0 and CART, after retrieving the outcome of the DT experiment. For every single one of the three situations the CART system was verified using the following technique: Generate testing as well as training data initially. Then it is required to configure the DT settings to start the method. For accurate predictions, analyze the DT using cross-validation, testing, and training. Plotting the outcomes of the Gini importance measure is the subsequent step. Throughout ten experiments, repeat steps 1-4 for the better results. Make use of the standard deviation and mean values of the categorizing performance from the field tests as a benchmark.

3.5 Multilayer Perceptron

A layer of input, an undetectable layer, and a result layer compose the multi-layer architecture of a mixed-learning program (MLP). The input segment acknowledges input, the concealing layer scrutinizes it, and the end result layer produces the model's ultimate product. The subsequent steps make up the MLP experimental process used in the present research. A layer for input, the hidden layer, and a result layer make up the multi-layer framework of a mixed learning program (MLP). The input layer is where the model recognizes input, the layer of concealment assesses it, and the output layer produces the model's outcome. The subsequent steps make up the MLP experimentation process used in the current study.

1. Choose the beginning weight and deviation.
2. Choose the target and training information.
3. Establish the discrepancy concerning the intended and planned results.
4. Adjust the weight and replenish the weight of the entire network.
5. Keep going from step (3) to step (4) until your understanding or alignment is complete [13].

3.6 Data mining Models

Regressions, as well as classification, are dualistic vital DM objectives. Mutually entail supervised learning, in which a prediction is modified to fit an information set of $l \in \{1, \cdots, O\}$ instances each of which plots an input vector $(y_1^l, \cdots, y_J^l)$ to a stipulated target y_l. From the below equation (2) the values that are provided are mentioned clearly. The key difference relates to the outcome depiction, which is independent in the case of sorting and perpetual in the situation of regression. The Root Mean Squared (RMSE) is a commonly employed metric in regression analysis, whereas the Percentage of Correct Classifications (PCC) is frequently employed in assessing predictive models for categorization A good classifier is demonstrated by a high PCC (i.e., very near to 100%) (3), and a regressor should recommend a low global error (i.e., RMSE close to zero) using equation (4). These measurements can be obtained using the equalities:

$$\varphi(j) = \begin{cases} 1, \ if \ z_j = \hat{z}_j \\ 0, \quad else \end{cases} \quad (2)$$

$$PCC = \sum_{j=1}^{M} \varphi(j)/O \times 100(\%) \quad (3)$$

$$RMSE = \sqrt{\sum_{j=1}^{O} (z_j - \hat{z}_j)^2 / O} \quad (4)$$

where the anticipated value for the j-th illustration is indicated by $\hat{z}_j$.

Three oversight methods will be applied when modeling the Mathematics and Portuguese grades in this function:

1. Classification in binary: pass if G3≥10, fail anything else;
2. The Erasmus1 grade conversion method (Table 1) is the fundamentals for the 5-Level groupings;
3. Regression uses the G3 esteem (a measurable output with values varying from 0 to 20).

Table 1. The five level classification system

Country	I (excellent/ very good)	II (good)	III (satisfactory)	IV (sufficient)	V (fail)
Portugal/France	17-21	15-16	13-14	11-12	1-10
Ireland	B	C	D	E	G

For the classification and regression duties, several kinds of DM techniques have been put forward each with distinctive features and objectives. A set of guidelines is symbolized by the Decision Tree (DT), an ordered arrangement that utilizes a hierarchy of levels to differentiate between values. An intuitive set of IF-THEN rules can be drawn from this visual representation. T unpruned DT methods an arrangement known as the Random Forest (RF). The arbitrary feature choice from bootstrap training specimens constitutes the base for each tree, and the RF assessments are made by averaging the T trees' outcomes. When contrasted to just one DT, the radio frequency (RF) signal is more difficult to figure out, but it is nevertheless achievable to offer descriptive data about the importance of the input value. For DM obligations, nonlinear methods like Support Vector Machines (SVM) and Neural Networks (NN) have also been offered; such functions perform more efficiently whenever there is an elevated degree of irregularity. The widely used multilayer perceptron serves as the backbone for the NN model in this research, which has a concealed layer and H concealed nodes. The SVM will use a Gaussian kernel that has a single hyperparameter (β). It needs to be acknowledged that NN and SVM make use of model illustrations that are tricky for human beings to recognize. Furthermore, given that the DT/RF techniques specifically accomplish an internal attribute selection, they are less impacted by insignificant ingredients than NN and SVM [14].

4. Implementation and Results

Ultimately, the most reliable classifier is incorporated into a straightforward software program to predict pupil achievement, making it simpler for instructors to pinpoint those who are struggling and suggest corrective actions.

4.1 Dataset

Ultimately, the most reliable classifier is incorporated into a straightforward software program to predict pupil achievement, making it simpler for instructors to pinpoint those who are struggling and suggest corrective actions. The information collected in the research set includes to the accomplishments of first-year University students in mathematics, particularly those between the ages of 14 and 15. The private Lyceum "Avgoulea-Linardatou" assembled the data, spanning 279 distinct structures, between 2007 and 2010. The features belong to data concerning the accomplishments of students, which involve grades from oral tests, and final assessments. The set of features provided in Table 2 can be broken down into two primary groups based on how students performed during the first and second semesters, accordingly. In addition, according to the system of groupings applied to instructional evaluation in Greek schools, the pupils were separated into four tiers:

- "Fail" represents a student's score between 0 and 9.
- "Good" suggests an individual's achievement in a hierarchy of 10 to 14.
- A student's academic achievement between 15 and 17 is considered "very good."
- Any achievement of a student between 18 and 20 can be considered "excellent."

Table 2. List of parameters used in our study

Student's attributes of $1^{st}/2^{nd}$ semester	Range values
Oral grade	[0,30]
Score of 1^{st} Test	[0,30]
Score of 2^{nd} Test	[0,30]
Score of final test	[0,30]
Closing rating	[0,30]

The classroom dispersion is seen in Figure 4.1, displaying the shares that are categorized as "Fail" (53 cases), "Good" (76 cases), "Very good" (85 cases), and "Excellent" (65 cases). Two sets of data have been designed with the help of the class distribution and the traits shown in Table 1 since this info is important for a teacher to figure out poor students in the middle of an educational semester.

- $DATA_A$: This incorporates qualities that influence the first semester achievements of the students.
- $DATA_{AB}$: includes attributes connected with the first and second semester accomplishments of the students.

You'll notice that every set of data in our investigation is used to build an individual classifier that distinguishes students who are battling.

Fig. 4.1. Class distribution

Following that, we perform an assortment of tests to find out which learning system best anticipates a student's course (i.e., "Fail," "Good," "Very good," or "Excellent") utilizing their grades from both academic and extracurricular semesters. As a result of this, we picked the most enjoyed and often submitted technique for each machine learning procedure that has been explored. The typical implementation of Bayesian algorithms was the Naive Bayes (NB) approach, which is implemented the most regularly. Presented the state of the class feature, it is a straightforward technique for learning that assumes the assumption that each characteristic does not influence the others. A constructed learning algorithm for producing neural networks, the backpropagation algorithm (BP) with momentum, was an effective representation of ANNs. Since the RIPPER technique is one of the most commonly employed techniques for delivering classification rules, it has been selected as an illustration of rule-learning strategies. The data gain operation is the growth intuition used by RIPPER, which generates rules by continually developing and cutting down. Meanwhile, we utilized the 3NN technique as an instance-based audience, using Euclidean distance as the measurement metric system. The algorithm with C4.5 was the more typical one in our review taken from the decision trees. The C4.5 algorithm chooses what element best breaks down the training examples at every phase of the splits process by utilizing an empirical property called knowledge gain. Table 3 offers an overview of each classifier's accomplishment based on the amount of correctly classified sequences in the datasets that were supplied. It is readily apparent that no algorithm can perpetually exceed all others in terms of achievement. More specifically, BP indicates the greatest proportion of instances correctly identified about dataset $DATA_A$, while 3NN demonstrates the best outcomes in terms of dataset $DATA_{AB}$.

Table 3. Accuracy of classifiers in each dataset

	Datasets	
Classifiers	$\boldsymbol{DATA_A}$**(%)**	$\boldsymbol{DATA_{AB}}$**(%)**
NB	52.6	58.5
BP	59.1	71.3
RIPPER	58.0	68.0
3NN	58.9	68.7
C4.5	57.6	69.1
SMO	58.0	65.2

The accuracy of classifiers is also represented using the following chart in Figure 4.2 which will give clear details regarding the classifier's accuracy.

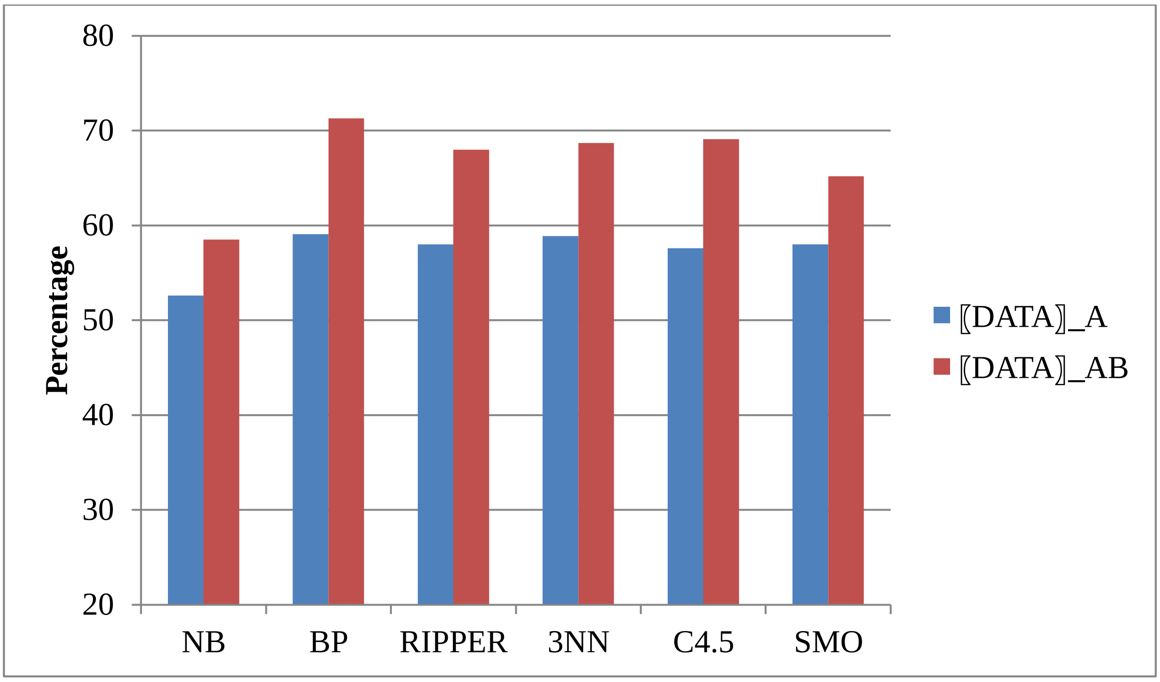

Fig. 4.2. Histogram for Classifier's accuracy

Our primary objective is to develop more specific and truthful classification happenings, so we employ voting, arranging things, and assessing methodology to amalgamate the anticipated outcomes of the various procedures on the datasets that are laid out. The next sentence indicates the function of the strategies listed in Table 4's first column:

- BestCV is an abbreviation for a superior classifier preference procedure.
- Voting implies a relatively easy voting method that integrates the estimates of the different techniques demonstrated in Table 3.
- Stacking is an acronym for stacking procedure, which uses MLR as a meta-level algorithm and the identical base classifiers as election.
- Grading is a shortened form of grading procedure that utilizes the situation primary classifier 10NN as the meta-level classifier and the same fundamental classifiers as voting.
- Voting∗ implies a relatively easy voting process comprised of SMO, BP, 3NN, and RIPPER as base classifiers.
- The word "stacking" corresponds to a stacking methodology that involves the use of MLR as a meta-level classifier and the same starting point classifiers as Voting.
- Grading∗ refers to an evaluating procedure that uses 10NN as the example's base classifier and Voting∗ as the framework for the meta-level algorithm.

Table 4. Precision of Accuracy for each dataset

	Datasets	
Classifiers	$\boldsymbol{DATA_A}$**(%)**	$\boldsymbol{DATA_{AB}}$**(%)**
Best CV	60.9	71.4
Voting	59.2	87.1
Stacking	57.7	69.6
Grading	58.8	72.8
Voting*	61.7	91.4
Stacking*	58.8	72.8

By Table 4's comprehension, the Voting approach exceeds the other combinations concerning accuracy, with VotingTM confirming the best effectiveness throughout the board for both datasets [15].

5. Conclusion

Utilizing machine learning and data mining mathematical methods for estimation is an effective instrument that assists instructors in recognizing students who are bound to perform badly early on and is an invaluable start in their strategies for intervention. We built a case study about the first-year University final mathematics examinations, as well as a simple-to-utilize decision support tool for foreseeing students' performance in this project. To predict students' learning effectiveness, we used family background elements in the present research, which can be acquired at the start of the freshman season. As soon as newcomers start school, we may employ the existing models to estimate their academic performance. With the help of this DM approach, it is straightforward to examine the classes that learners acquire for academic as well as extracurricular activities. Decision Trees (DT), Random Forests (RF), Neural Networks (NN), and Support Vector Machines (SVM) are the four DM methods that were evaluated. Three independent DM goals (5-level classification and regression) were also examined. Diverse input options (such as incorporating or removing prior grades) were explored as well. The findings gathered show that if the results for the first and/or second school time frame are accessible, a highly accurate prediction can be accomplished. This helps the research findings that prior spectacles have an imperative guidance on achievement for pupils. To increase the student databases, we are also going to broaden our tests to other educational institutions and classroom years. As a consequence of employing this data mining technique, the investigation of student education data is exact and clear.

Reference

[1] Shrestha, S., & Pokharel, M. (2020). Data Mining Applications Used in Education Sector. *Journal of Education and Research*, *10*(2), 27-51.

[2] Huang, Y., Liu, H., & Pan, J. (2021). Identification of data mining research frontier based on conference papers. *International Journal of Crowd Science*, *5*(2), 143-153.

[3] Olson, D. L. (2020). A review of supply chain data mining publications. *Journal of Supply Chain Management Science*, *1*(1-2), 15-26.

[4] Romero, C., & Ventura, S. (2020). Educational data mining and learning analytics: an updated survey. WIREs Data Mining and Knowledge Discovery 10 (3): e1355.

[5] Buenaño-Fernández, D., Gil, D., & Luján-Mora, S. (2019). Application of machine learning in predicting performance for computer engineering students: A case study. *Sustainability*, *11*(10), 2833.

[6] Khan, A., & Ghosh, S. K. (2018). Data mining based analysis to explore the effect of teaching on student performance. *Education and Information Technologies*, *23*, 1677-1697.

[7] Asif, R., Merceron, A., Ali, S. A., & Haider, N. G. (2017). Analyzing undergraduate students' performance using educational data mining. *Computers & education*, *113*, 177-194.

[8] PAPADOGIANNIS, I., POULOPOULOS, V., & WALLACE, M. (2020). A Critical Review of Data Mining for Education: What has been done, what has been learnt and what remains to be seen. *International Journal of Educational Research Review*, *5*(4), 353-372.

[9] Azwa, A. A., & Fadhilah, A. (2014). First semester computer science students' academic performances analysis by using data mining classification algorithms.

[10] Kumar, M., Singh, A. J., & Handa, D. (2017). Literature survey on student's performance prediction in education using data mining techniques. *International Journal of Education and Management Engineering*, *7*(6), 40-49.

[11] Kabakchieva, D. (2013). Predicting student performance by using data mining methods for classification. *Cybernetics and information technologies*, *13*(1), 61-72.

[12] Yağcı, M. (2022). Educational data mining: prediction of students' academic performance using machine learning algorithms. *Smart Learning Environments*, *9*(1), 11.

[13] Huynh-Cam, T. T., Chen, L. S., & Le, H. (2021). Using decision trees and random forest algorithms to predict and determine factors contributing to first-year university students' learning performance. *Algorithms*, *14*(11), 318.

[14] Cortez, P., & Silva, A. M. G. (2008). Using data mining to predict secondary school student performance.

[15] Livieris, I., Mikropoulos, T., & Pintelas, P. (2016). A decision support system for predicting students' performance. *Themes in Science and Technology Education*, *9*(1), 43-57.