Predicting Students' Grades in a Professional Undergraduate Course Using an Ensemble Model

[1]Mr. Jitendra H. Darji, [2]Dr. Tulsidas V. Nakrani, [3]Dr. Idrish I. Sandhi,
[4]Mr. Rasik D. Patel, [5]Mr. Pyush A. Patel, [6]Ms Prachi D. Raval
[1, 2, 3,4,5,6] Sankalchand Patel University, Department of computer Application (MCA), SPCE, Visnagar-384315, Gujarat, India
[1]mcajayesh8@gmail.com, [2]tvnakrani.mca@spcevng.ac.in, [3]hodmca@spcevng.ac.in, [4]rdpatel.mca@spcevng.ac.in, [5]papatel.mca@spcevng.ac.in, [6]pdraval.mca@spcevng.ac.in

Article Info
Page Number: 9356 - 9370
Publication Issue:
Vol 71 No. 4 (2022)

Abstract

The goal of this study is to come up with a way to predict how well first-year college students in a professional program do in their classes (BCA). Tracking students' academic progress is an important area for ensuring the optimal growth of their analytical and logical skills. Being able to predict a student's academic success in the years immediately after graduation is useful for many different groups, including the government, legislators, and educators. An ensemble model is made for this task using a decision tree, a gradient boost algorithm, and some Naive Bayes techniques. This model gives the most accurate and reliable results. A questionnaire was created to find the factors that affect students' academic, social, behavioral, and demographic performance in school. Then, based on how well each of the three approaches performed, an ensemble model was created. The quality of the outcomes from the suggested ensemble model was evaluated using a 10-fold cross-validation technique. The output of an ensemble model allows for accurate and efficient prediction of student performance, and can help pinpoint students who are at risk of failing or dropping out of school. In order to create the current model, we employ both classification and regression techniques. With the current data set, the model achieves 99.1% accuracy in determining the important factors influencing students' academic success. As the suggested methodology allows for early identification of students who are at danger, it can also offer preventative and remedial strategies to boost students' overall academic performance.

Article History
Article Received: *15 September 2022*
Revised: *25 October 2022*
Accepted: *14 November 2022*
Publication: *21 December 2022*

Keywords:-Higher education institute, Support Vector Machine, Prediction, Educational data mining, Educational, Student performance

---

1. Introduction

Due to HEIs' transformation into complex organizations offering a wider variety of specialized degrees in a multimodal teaching-learning environment, policymakers worldwide are increasingly turning to automated tools for analyzing the massive amounts of data generated by a wide range of educational settings. EDM has proven to be a successful tool for improving education for educators, administrators, and policymakers.

Figure out what influences students' grades. Data mining techniques were used to conduct an analysis of their educational, social, demographic, and behavioral backgrounds. Most significant were the student's previous semester's percentage, their mother's income, a significant family tragedy, and their residence's distance from the university. Kamal and Ahuja (2017)

Romero and Ventura (2010) presented EDM as a new field of multidisciplinary study focused on the creation of tools to investigate information gathered from a classroom setting. Higher education institutions (HEIs) aim to generate high-performing students who will go on to establish successful businesses and develop innovative solutions to pressing global problems.

Performing at a high level. The best way for teachers to help their students succeed is to identify and then optimize the aspects that contribute to those students' successes. Mentors can take prompt action if they are able to identify and address the needs of students who are at danger of falling behind (Kim and Cho, 2017; Marquez- Vera. et al., 2016). The academic, social, and behavioral characteristics of students can be best gleaned from their past records, which may include factors like the students' age, gender, parental education and income, the percentage of students who attended college the previous semester as well as the total number of students who will be attending classes during this semester, the number of hours that will be spent studying outside of the classroom, the impact that the students' social circles will have, as well as the regularity with which they will use alcohol and tobacco.

According to Lu et al. (2018) and Mcquiggan et al. (2007), methods of EDM are frequently utilized to get an early indication of those who are not performing well and are at risk pupils in order to maximize the results. The early identification of students who are considered to be at risk helps higher education institutions to focus greater attention on the performance of these students, which eventually contributes to the achievement of better results.

According to Aluko et al (2018)., monitoring student performance is necessary to produce better results with strong performers. Better outcomes may be obtained through early identification of strong and weak performers.

2. Related work

Wolff et al. (2013) carried out a similar study to determine the academic achievement of first-semester students. Students at risk could be identified and given the necessary help in their first year after graduation because this is regarded as the ideal period for intervention. This program's success percentage is great even from the beginning. Finding underachievers might assist teachers in giving them the necessary guidance. The degree to which teachers are involved in their pupils' academic affairs directly relates to how well they perform. Finding students who are falling behind and assisting them in attending the remedial classes recommended by BH and Suresh et (2018) will result in even greater improvements to the outcomes.

Kamal and Ahuja (2019) argue that a combination of characteristics can boost academic success and retention. This methodology, developed using a variety of data mining techniques, is helpful for estimating a student's likelihood of employment one year after graduation.

Vandamme et al (2007). Had previously collected and utilized the student database sorting students into groups depending on their averages. Contributing factors included performance on in-class tasks, participation in required seminars, and overall grade on final exams. Students were classified as either low-risk (high likelihood of success), medium-risk (chance of success), or high-risk (high potential of failure) based on the results (or dropping out).

Student behaviour evaluation in the first few years after graduation can be valuable for predicting future results, as suggested by studies such as those conducted by Aluko et al. (2018) and Lu et al. (2018). Both Marquez-Vera et al. (2016) and Alsaffar (2017) employed a classification technique to develop a prediction model for early detection of at-risk kids.

3. Research methodology

3.1 Data preparation

This study analyzed the factors that significantly affected students' academic performance in order to better pinpoint students who were at risk. An exhaustive literature search allowed for the extraction of these variables. Different groups were created for these elements. First category was consistently observed, including demographic characteristics such parental income qualification. The second group included indicators of students' academic success, such as their percentage of classes attended last semester and their current attendance rate. The influence of one's peer group, one's involvement in extracurricular activities, and one's home environment all fall under the third category, which elucidates the societal issues that affect one's academic achievement. Student study habits, including how much time is spent on homework outside of class, how often assignments are turned in, whether or not students smoke or drink, and how often they miss class, were the subject of the fourth heading.

3.2 Data selection

Because of these factors, a survey was developed to determine what factors have the most influence on students' academic performance. A questionnaire was developed using the four groups of impacts on students' academic progress. A five-person team of management, statistics, and data mining experts reviewed the questionnaire's questions and responses. The validation committee suggested using feature selection approaches to reduce there from 79 to 50 questions. A request for information was made to the students of the C.J.Patel College, which is located in Visnagar. This college is affiliated with the Sankalchand Patel University in Chandigarh, which has the title of the oldest university in all of India. On a scale from one to four points, the cumulative grade point average of Panjab University is 3.35, earning it an A grade as of June 25, 2015. There is five-year validity on this mark. In 2015, Times Higher Education ranked it as the best university in India and the 38th best university in Asia. Five hundred and eighty Bachelor of Computer Applications (BCA) students participated during the 2016-2019 academic year. The three-year BCA program is divided into six semesters. Students are tested twice yearly, or every six months. Due to little online participation, we abandoned our Google form in favor of a more traditional survey approach. We distributed the survey during class time, and students filled it out in front of their instructors.

3.3 Characteristics of the data

- There are enough records to do the experiments, and you don't need to split the data into small pieces to do the experiments.
- Our data set does not have any redundant entries. Accordingly, duplicate records have zero effect on classifiers' performance.
- The information is presented in a distinct format (primary data).

3.4 Data pre-processing

The questionnaire was used for offline data collection. The information was first gathered by hand and then entered into a computer system. A spreadsheet in Excel was used for the manual data entry. Since Rapid Miner works with the Excel file format, analyzing the data is a breeze.

3.5 Data cleaning and consolidation

When data was converted to digital form, it was cleaned up and consolidated. The data set was cleaned based on the following three criteria:

1. Errors in formatting were eliminated because all responses were objective in nature and based on a Likert scale. Student input was used for some edits, while other edits, such as changing an employee's status to "employed" when their income is $0, were made manually (via observations). The deciding characteristics were determined through information gathering. Low-importance attributes were eliminated from the database. Upon further inspection, 489 of the initially submitted responses were judged to be valid and full, and were thus included in the analysis.
2. There was less likelihood of missing data because most responses were objective in nature and based on a Likert scale.
3. In those instances in which the absence of data did not create an issue, we accepted the response. For instance, if the respondent did not provide their father's name, we accepted the response; but, if the respondent did not provide the percentage of revenue generated in the previous year, we rejected the response.

Students' information includes personal details as well as information about their academic and social lives. The data set contains 19 demographic factors, such as age, gender, marital status, parental education, employment status, etc. Characteristics in the classroom include grades from the previous term, there are a total of eight possible factors, including but not limited to: grades on coursework, presence at all classes this semester, whether or not a gap year was taken, performance on make-up exams taken during the previous semester, etc. Attitudes are reflected in a student's behavior, and this includes how they complete their assignments (by copying from classmates, for example, or not turning them in on time, or even spending time on self-improvement instead of studying). Learn outside of normal school hours; maintain a regular routine of drinking and tobacco use, etc. In total, there are 17 of them. There are a total of six social factors that have an impact on a student's academic and behavioral outcomes, such as the student's interest in the current course, the amount of time spent with friends, and the prevalence of unhealthy habits. Both demographic and behavioral characteristics were identified as essentially the most formative in terms of a student's academic career. Attributes are tabulated in Table II.

For the sake of training a model that can handle a large amount of data, this survey data is used. The data acquired from students is used to test the established model and demonstrate its usefulness from an Indian viewpoint. For the purpose of doing the analysis, Rapid Miner Studio, version 9.0.2, was utilized.. Feature selection was accomplished via information gathering, and One of the attributes with more weight was chosen since it provided the most accurate distribution. Less-important attributes were thrown out because they provide no useful data. To better understand the research process, please refer to Figure 1.

4. Proposed work

To get there, we applied a number of data mining techniques to the dataset in an effort to identify significant variables associated with students' performance in the classroom. The degrees of precision with which various approaches predicted our outcomes varied. Based on their effectiveness and results, classification techniques such (NB), (DT), (SVM), gradient boost, (RF), clustering, and neural networks (NN) were investigated. Rapid Miner 9.0.1 was used to implement these algorithms, and comparisons were made to determine which provided the most accurate predictions. Based on the accuracy demonstrated in Table III, three methods were chosen. There were 489 valid, complete survey responses, representing 50 attribute values and 1 class variable. A total of 173 students were identified as being at risk following the statistical analysis, with the data set providing 489,51 dimensions. Attribute descriptions are shown in Table 1.

Table-I. Dataset Description

Total instance	489
Total features	50
Target feature	1
At-risk students	173
Dimension	489*51

Table-II. Features Description

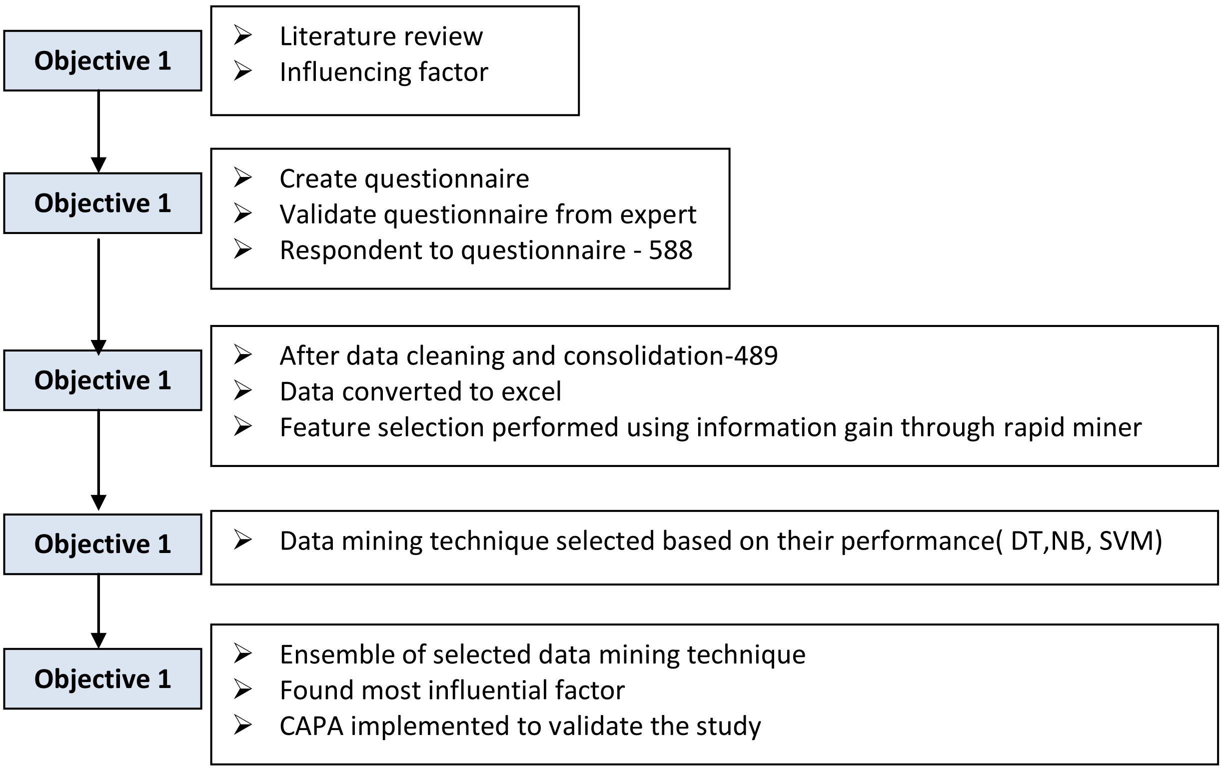
Features	**Number of factor**
The Influence of Demographics	18

Social Factor	7
Educational Consideration	8
Behavioral Factor	16

Table- III. Three algorithms are compared. (NB, SVM, DT)

Model	**Accuracy**
DT	97.2
SVM	99.4
NB	97.0

Figure -1: Flowchart describing Methodology



When these methods were applied to the dataset, SVM produced the highest accuracy (99.4%), followed by DT (97.2%) and NB (97.0%).

4.1 Data mining techniques

4.1.1 Decision tree.

There are many different kinds of machine learning algorithms, but DT is by far the most used. Classification and regression models can be built with its help. When constructing DT, we work from the very top down. The best characteristic is determined for breaking down the class label into its simplest form of child nodes. A DT's nodes, branches, and leaves each stand for an aspect (attribute), an action (rule), and an outcome, respectively. The goal of DT is to create data sets that are noise-free and are simple to categories into discrete nodes. DT does this by building a tree using entropy. Figure 2 shows the influencing factors, and Figure 3 shows the final decision tree model.

4.1.2 Entropy.

Data noise, or entropy, can be thought of as a measure of variety. Entropy is necessary for reducing data noise and quantifying data homogeneity. A higher degree of purity is represented by a lower entropy value. A number of 0 is the minimum value for entropy that can be considered useful. If you calculate entropy, you'll have a sense of how clean something is. Information stands in for the quantity of background knowledge that is roughly expected to be required to decide whether or not a fresh instance needs extra classification to get the desired class label.

4.1.3 Information gain.

It's also possible to refer to the information gained through mutual interaction as mutual information. It's a measurement of how much of a class's details an attribute reveals. The optimal weight of a separating feature is maximized. The most valuable feature will be chosen for the split. As noise in a data set is minimized, information gain increases in importance. It's calculated by subtracting the entropy of the root from that of the leaves. To put it another way, it aids in the fight against entropy.

Figure-2 influencing factor

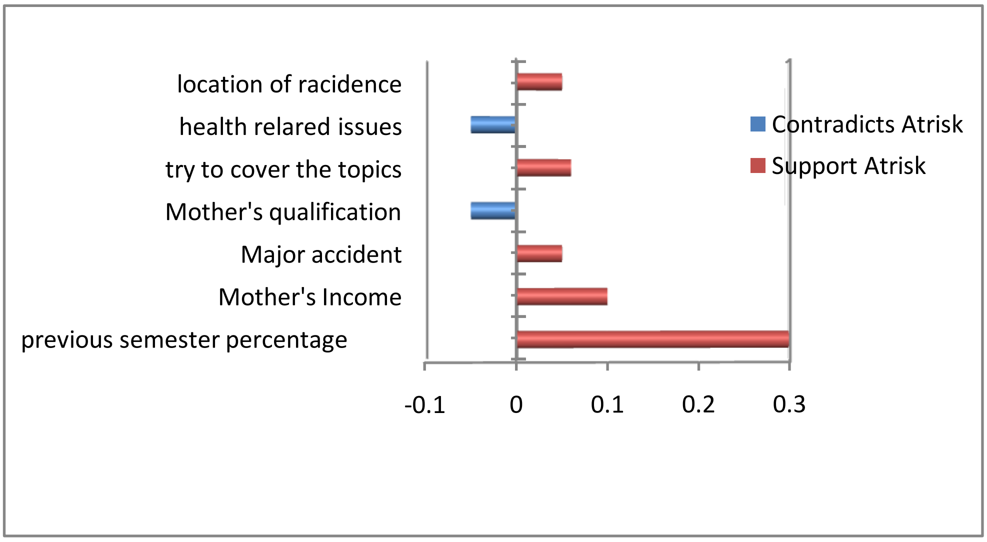
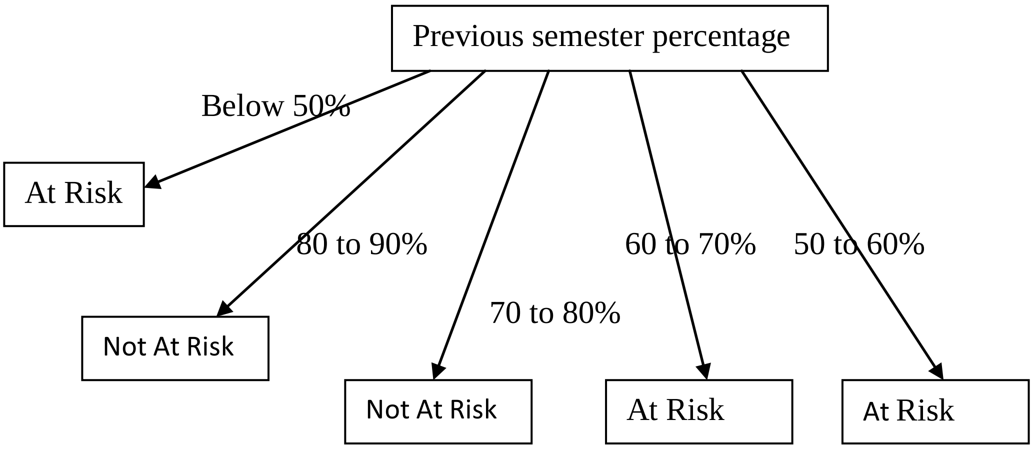


Figure 3 Decision Tree



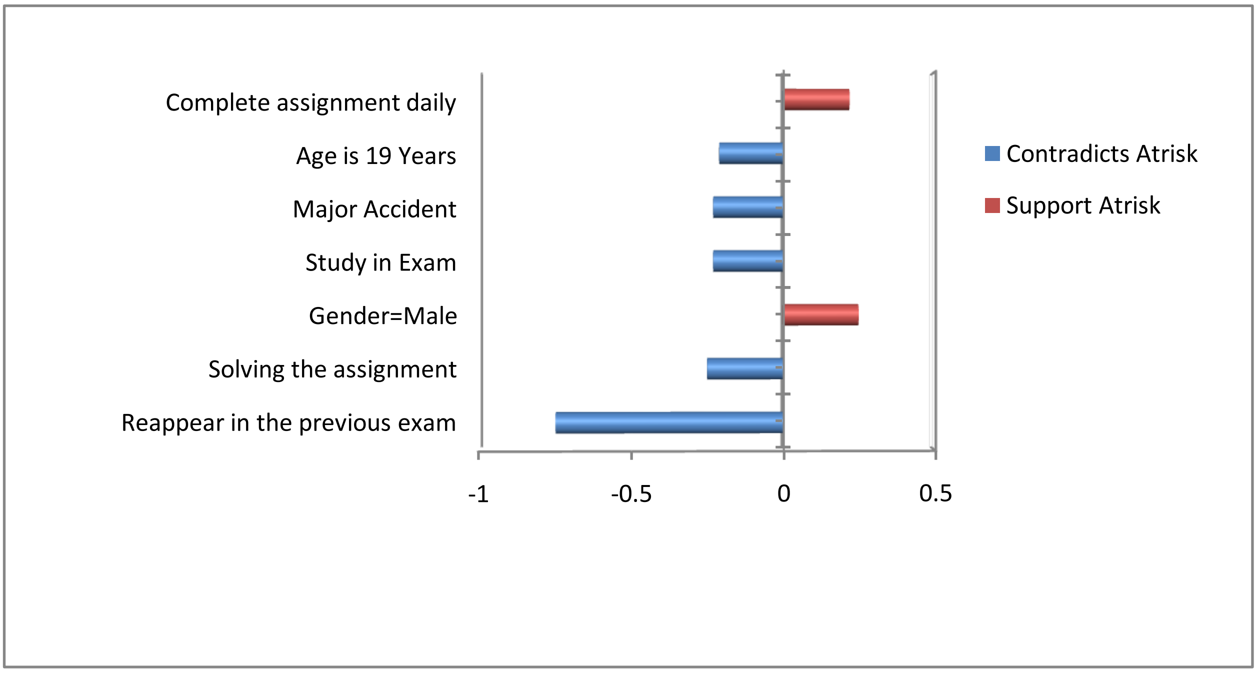
4.1.4 Naïve Bayes

In machine learning algorithms, NB is one of the most effective and widely used algorithms.

The NB algorithm is a probabilistic classifier that operates in a Bayesian setting and generates classifications by applying the highest a posteriori decision rule. The use of conditional probabilities is prevalent in NB. It is predicated on the Bayes theorem, which serves as its foundation. It does this by computing the posterior probability using the prior probability and the conditional probability. It operates with the help of both categorization and numerical qualities.

(Figures 4 and 5).

Figure 4. Important aspects that contribute to their level of academic achievement



4.2 Posterior probability

The likelihood of something happening after the fact, based on the existing data or context. It is the result of reevaluating the baseline probability, or the likelihood that an event will occur in the absence of any new information. The term "posterior probability" refers to an altered likelihood of an event:

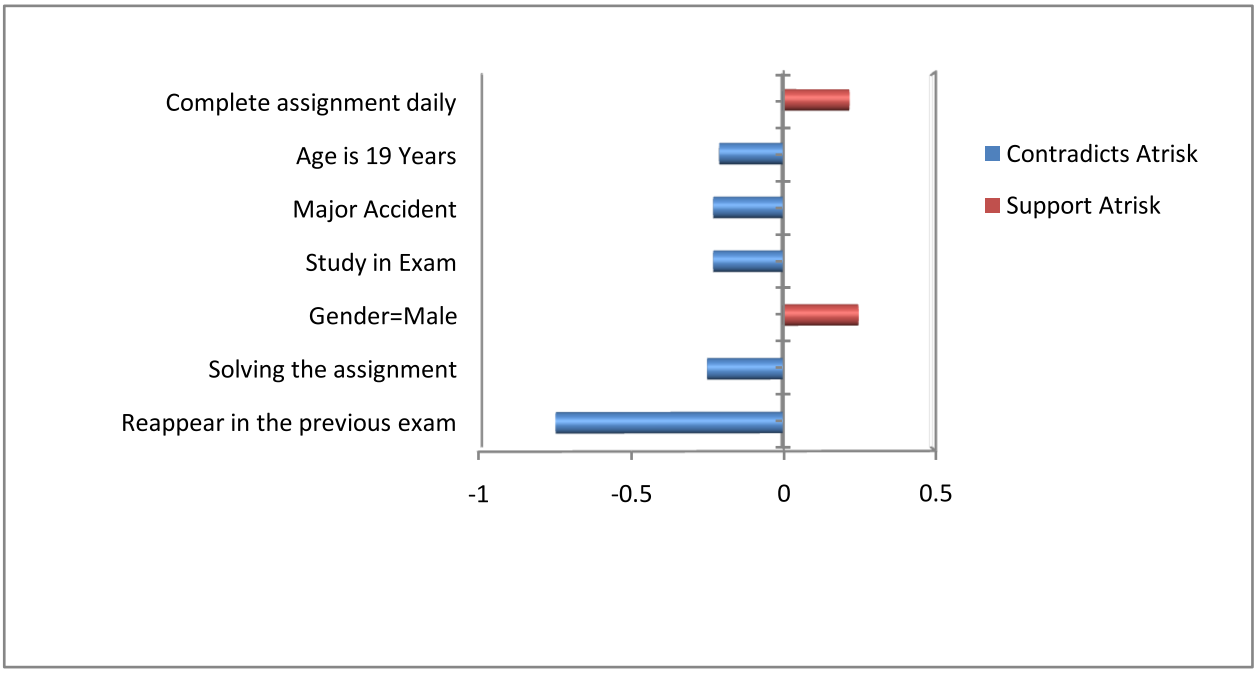
$$\mathrm{P(A|B)} = \frac{P(B|A)P(A)}{P(B)}$$

The calculation yields the conditional probability P(A/B), which is referred to as the posterior probability.

4.2.1 Support vector machine.

SVM is the most popular machine learning method and is based on the concept of limiting variables on a decision plane. A decision plane is an alternative to hyper planes for classifying data. It's used to build models for things like categorization, regression, and spotting outliers. Only numeric characteristics can be used with it; however, when a Boolean property is discovered, the Boolean data is immediately translated into binary values of 0 (false) and 1 (true) (True).Therefore, new significant elements that have either good or negative effects on students' academic performance were investigated using SVM processing. Students' academic performance is significantly impacted by things like returning from the previous semester, getting help with assignments, studying for examinations, having a catastrophic accident last semester, etc. Daily task completion, however, had a positive effect on performance. The SVM's accuracy for the results was 99.4% (Figure 6).

Figure 6. Using the Support Vector Machine: Important feature



5. Ensemble method

Combining the best elements of DT, NB, and SVM was the idea behind ensemble. These methods will provide the same accuracy when used on data sets of a similar nature. On the other hand, their accuracy wouldn't change if a new data set was used. It will not decline as a result of a distinct data collection. Ensemble methods are meta-algorithms that combine multiple machine learning techniques into a single predictive model in order to achieve a more consistent result (bagging) and bias (boosting) and enhance predictions (stacking). Combining the average results of multiple models allows for increased precision and generalizability when tackling challenging problems. It has been shown that ensemble techniques, as presented by finally (2014), can improve accuracy by 30% relative to any best single model. Ensemble models can be constructed using any of the three methods (mainly bagging, boosting, and stacking), as stated by Fathian et al (2016). Performance can be improved by using an ensemble model, as demonstrated in Table IV and as stated by Webb and Zheng (2004). On the basis of the classification error of the three processes and the results obtained after applying them to the training data set, an ensemble of three techniques was formed. The creation of the ensemble led to a dramatic increase in precision. Accuracy of 99.4% was achieved by the ensemble, which is much higher than the accuracy attained with the preexisting models. (Figure 7)

Table IV. Error in classifying

Model	**Classification Error (%)**
dt	1.7
svm	1.5
nb	2

Figure 7. The SVM, DT, and NB Ensemble

6. Results and discussion

Although the present models consider a student's family wealth, social circumstances, academic indicators, etc., they usually ignore crucial aspects like peer pressure, self-study habits, time spent studying after classes, the percentage of the prior semester, and significant family tragedies. It is urgently necessary to create novel models in order to evaluate and improve the current educational system. The ensemble model is the most effective tool for examining all the variables required to develop a solid and trustworthy model. The findings of the ensemble model offer efficient and trustworthy forecasts of student performance, in addition to assisting in the identification of students who are at risk of failing their classes or withdrawing from school altogether. Academic achievement from the previous semester has a significant influence on academic performance in the current semester, in addition to other factors (such as a major family tragedy, the amount of money the mother makes, and the distance between the household and the institution). Any serious accidents that has place in the previous year have an impact on academic achievement, in addition to habit-based behavioral factors like alcohol and cigarette usage.

The proposed research aimed to build a model that would aid in detecting underachievers in the years immediately following graduation, to identify the contributing causes to underachievement, and to advise corrective steps, such as additional coursework and mentor support in special cases. Prior studies were reviewed that were relevant to the proposed area of study, and numerous articles were searched through to find relevant data. (Essa and Ayad, 2012; Fathian et al., 2016; Finlay, 2014; Kim and Cho, 2017; Liew et al., 2018; Puyalnithi et al., 2016). (Finlay, 2014). Following the completion of empirical programs to establish which approach was the most successful, an ensemble model was built.

To create the ensemble model, classification methods are used. The model achieves a remarkable 99.4% accuracy using the available data to determine the most important elements in students' academic success. Due to the proposed model's ability to identify potentially struggling students early on, both preventative and remedial approaches for raising students' average grades can be suggested.

The elements that have a good or negative impact on pupils' academic achievement are listed in **Table V**. If a student's percentage from the previous semester was above 60%, that student has a good chance of succeeding this semester; if the percentage drops to 50% or lower, the student is at danger of failing or dropping out. **Table V** displays the total number of students that lied about their percentile ranking. When considering students' motivations, it's helpful to know that if they're determined to succeed academically, they shouldn't have too much trouble achieving those goals over the semester. It is more advantageous for a student to live in a city, as opposed to a smaller town or the countryside, where they would have fewer resources and be less likely to succeed academically. Having more job opportunities available to them is one benefit of living in a city. Last-minute exams have a chilling effect on students' grades in the last weeks of the semester. Students' inability to do well in the current semester is a result of the stress they felt last semester. Students' academic performance might be badly impacted by any major tragedy in the family. Students will be unable to focus on their coursework, increasing the likelihood that they will receive low grades and perform poorly on exams this semester. Students who make poor habits like drinking and smoking a regular part of their lives are doomed to a career of academic failure.

Table V. An explanation of the concept of influence

Features	**Influence on the Results of**	**Description**

	Performance	
% from the prior semester	Positive	Below 50% (5), 50-60% (164), 60-70% (181), 70-80% (107), 80-90% (32), above 90% (0)
Getting high grades in a semester is simple.	Positive	Yes (479) No (20)
location of the home	Positive	Urban (330), Semi-Urban (81), Rural (78)
return to the prior semester	Negative	No (250), 1(100), 2(106), 3 and above (33)
major accident	Negative	Yes (130), No (359)
Alcohol consumption pattern	Negative	Yes (32), No (457)

After executing the CAPA form provided in **Table VI** with the students, we can verify the elements that had an impact. Students' academic success is considered in the construction of this form. All at-risk pupils must complete out this form. After analyzing the ensemble model findings, a few questions were formulated. Students' answers to these questions, which were incorporated in the CAPA format, will lend credibility to the study's findings. Students' grades can be boosted by following the measures outlined in Table V, which take into account the positive and negative effects of various factors on academic achievement.

Among the most crucial indicators of academic success is a student's percentage from the previous semester; if a student's percentage is below 60%, he or she must schedule an in-person meeting with their mentor and begin CAPA immediately. Their level of dedication to the course is reflected in their attendance. When student attendance drops below 75%, a Comprehensive Attendance and Performance Plan (CAPA) must be submitted along with an explanation for the absences. Students' academic performance is negatively impacted by the time spent commuting to and from school each day; CAPA can assist these students in exploring hostel options. If they choose this path, they will see better outcomes and have better future performance. Prevention strategies that are recommended to students can aid in their future success. Based on the findings of our research, we have developed the following recommendations:

- Students who consistently score below 60% require additional support in order to succeed. Otherwise, CAPA is not required and should not be deployed.
- In general, students are absent from classes if their current semester's attendance is below 75%. This situation calls for the use of a Corrective and Preventive Action plan.
- Students are required to participate in CAPA if they have experienced emotional trauma as a result of a catastrophic family tragedy.
- There should be caution when dealing with the habits of smoking and drinking. It is inappropriate for a mentor to invade someone's privacy in such a direct manner. The Continual Appropriate Action Plan (CAPA) could be useful.
- Students who are having physical or mental health problems require medical assistance. They can improve with the help of CAPA.

- Students that are serious about their education will make up the work they missed. If not, corrective action should be taken.
- The availability of dormitories allows students who live a long way from campus to live on campus and focus on their academics. We must use CAPA.

Table VI. CAPA based on influential factors

% from the prior semester	Above 68% 61-71 Below 61
Attended this semester	Above 73% Below 73%
Has there been a serious incident since the end of the semester?	Yes/No If so, could you elaborate?
Do you regularly consume alcohol or tobacco?	Yes/No (Students are inclined to give incorrect responses because they are self-conscious about getting it wrong. Therefore, teachers should provide feedback based on their unique understanding of each student.
How is your health right now?	Yes/No (If so, please provide further justification for your position.)
When you can't make it to a class during the semester, how do you make up for it? Do you hope to stay in a hostel? (if Required)	

7. Conclusion

Because of their importance to students' academic success, universities and colleges must address aspects like familial problems, prior academic data, and socio-behavioral influences. Multiple aspects of students' academic performance, conduct, demography, and social problems were examined in this article. All of the factors and all of the data acquired during the first few years of a student's education were considered. The kids' academic performance was the subject of an investigation using a specially designed ensemble, which yielded very precise findings. In the future, the ensemble can be fed additional data sets so that students can make more accurate predictions. Since SVM, NB, and DT were the most effective classification methods tested, these three methods were combined to form the suggested and implemented ensemble. Decisions based on ensemble results can be useful to many parties involved. Mentors can use these findings to better prepare early-years remedial strategies to better support their mentees. In this way, the ensemble model will help teachers and tutors monitor students' progress and improve their teaching methods. In conclusion, this framework can point students in the direction of early preventative activities that will have a good impact on their performance taking into consideration the aspects that were taken into account in the research. This group will aid stakeholders, mentors, and most crucially, students, by enhancing learning processes and lowering dropout rates.

Declaration statement on conflict of interest

I, **Jitendra H. Darji** hereby declare that I am the corresponding author of this article. To the best of my knowledge this article contains no material previously published by any other journal. This article has not been submitted for publication nor has it been published elsewhere.

References

1. Adejo, O.W. and Connolly, T. (2018), “Predicting student academic performance using multi-model heterogeneous ensemble approach”, Journal of Applied Research in Higher Education, Vol. 10 No. 1, pp. 61-75.
2. Finlay, S. (2014), Predictive Analytics, Data Mining and Big Data Myths, Misconceptions and Methods, Palgrave MacMillan Publisher, New York, NY, p. 89.
3. Aluko, R.O., Daniel, E.I., Oshodi, O., Aigbavboa, C.O. and Abisuga, A.O. (2018), “Towards reliable prediction of academic performance of architecture students using data mining techniques”, Journal of Engineering, Design and Technology, Vol. 16 No. 3, pp. 385-397.
4. Huang, S. and Fang, N. (2013), “Predicting student academic performance in an engineering dynamics course: a comparison of four types of predictive mathematical models”, Computers and Education, Vol. 61, pp. 133-145.
5. BH, H.M. and Suresh, L. (2018), “Data mining in higher education system and the quality of faculty affecting students academic performance: a systematic review”, International Journal of Innovations and Advancement in Computer Science, Vol. 7 No. 3, pp. 2347-8616. ISSN.
6. Kamal, P. and Ahuja, S. (2017), “A review on prediction of academic performance of student’s at-risk using data mining techniques”, Journal on Today’s Technologies, Vol. 5 No. 1, pp. 30-39, doi: 10.15415/jotitt.2017.51002.
7. Kamal, P. and Ahuja, S. (2019), “Academic performance prediction using data mining techniques: Identification of influential factors effecting the academic performance in undergrad professional course”, in Harmony Search and Nature Inspired Optimization Algorithms, Springer, pp. 835-843, available at: https://doi.org/10.1007/978-981-13-0761-4_79.
8. Essa, A. and Ayad, H. (2012), “Student success system: risk analytics and data visualization using ensembles of predictive models”, in Proceedings of the 2nd International Conference on Learning Analytics and Knowledge, ACM, pp. 158-161.
9. Kim, K.J. and Cho, S.B. (2017), “Ensemble Bayesian networks evolved with speciation for high-performance prediction in data mining”, Soft Computing, Vol. 21 No. 4, pp. 1065-1080.
10. Liew, J., Cao, Q., Hughes, J.N. and Deutz, M.H. (2018), “Academic resilience despite early academic adversity: a three-wave longitudinal study on regulation-related resiliency, interpersonal relationships, and achievement in first to third grade”, Early Education and Development, Vol. 29 No. 5, pp. 762-779.
11. Fathian, M., Hoseinpoor, Y. and Minaei-Bidgoli, B. (2016), “Offering a hybrid approach of data mining to predict the customer churn based on bagging and boosting methods”, Kybernetes, Vol. 45 No. 5, pp. 732-743.

12. McQuiggan, S.W., Lee, S. and Lester, J.C. (2007), "Early prediction of student frustration", in International Conference on Affective Computing and Intelligent Interaction, Springer, Berlin, Heidelberg, pp. 698-709.
13. Márquez-Vera, C., Cano, A., Romero, C., Noaman, A.Y.M., Mousa Fardoun, H. and Ventura, S. (2016), "Early dropout prediction using data mining: a case study with high school students", Expert Systems, Vol. 33 No. 1, pp. 107-124.
14. Lu, O.H., Huang, A.Y., Huang, J.C., Lin, A.J., Ogata, H. and Yang, S.J. (2018), "Applying learning analytics for the early prediction of students' academic performance in blended learning", Journal of Educational Technology and Society, Vol. 21 No. 2, pp. 220-232.
15. Alsaffar, A.H. (2017), "Empirical study on the effect of using synthetic attributes on classification algorithms", International Journal of Intelligent Computing and Cybernetics, Vol. 10 No. 2, pp. 111-129.
16. Timande, S., & Dhabliya, D. (2019). Designing multi-cloud server for scalable and secure sharing over web. International Journal of Psychosocial Rehabilitation, 23(5), 835-841. doi:10.37200/IJPR/V23I5/PR190698
17. Umbarkar, A. M., Sherie, N. P., Agrawal, S. A., Kharche, P. P., & Dhabliya, D. (2021). Robust design of optimal location analysis for piezoelectric sensor in a cantilever beam. Materials Today: Proceedings, doi:10.1016/j.matpr.2020.12.1058
18. Ramaswami, M. and Bhaskaran, R. (2010), "A CHAID based performance prediction model in educational data mining", International Journal of Computer Science, Vol. 7 No. 1, pp. 10-18.
19. Romero, C., Ventura, S., Espejo, P.G. and Hervás, C. (2008), "Data mining algorithms to classify students", in R. de Baker, T. Barnes, J. Beck (Eds), Proceedings of the 1st International Conference on Educational Data Mining, pp. 8-17, available at: www.educationaldatamining.org/ EDM2008/uploads/proc/1_Romero_3.pdf
20. Puyalnithi, T., Madhu Viswanatham, V. and Singh, A. (2016), "Comparison of performance of various data classification algorithms with ensemble methods using RAPIDMINER", International Journal, Vol. 6 No. 5.
21. Webb, G.I. and Zheng, Z. (2004), "Multistrategy ensemble learning: reducing error by combining ensemble learning techniques", IEEE Transactions on Knowledge and Data Engineering, Vol. 16 No. 8, pp. 980-991.
22. Wolff, A., Zdrahal, Z., Nikolov, A. and Pantucek, M. (2013), "Improving retention: predicting at-risk students by analysing clicking behaviour in a virtual learning environment", in Proceedings of the Third International Conference on Learning Analytics and Knowledge, ACM, pp. 145-149.
23. Vandamme, J.P., Meskens, N. and Superby, J.F. (2007), "Predicting academic performance by data mining methods", Education Economics, Vol. 15 No. 4, pp. 405-419.