Webology, Volume 18, Number 5, 2021
ISSN: 1735-188X
DOI: 10.29121/WEB/V18I5/37

The Application of Data Mining and BI to the Study of and Efforts to Enhance Academic Performance

Rupa Khanna Malhotra[1], Narendra Singh Bohra[2], Pravin P Patil[3], Durgaprasad Gangodkar[4], Dibyahash Bordoloi[5]

[1]Department of Commerce,Graphic Era Deemed to be University, Dehradun, Uttarakhand India, 248002
[2]Department of Management Studies,Graphic Era Deemed to be University, Dehradun, Uttarakhand India, 248002
[3]Department of Mechanical Engineering,Graphic Era Deemed to be University, Dehradun, Uttarakhand India, 248002
[4]Department of Computer Science & Engineering,Graphic Era Deemed to be University, Dehradun, Uttarakhand India, 248002
[5]Head of the Department, Department of Computer Science & Engineering,Graphic Era Hill University, Dehradun, Uttarakhand India, 248002

ABSTRACT

The higher education sector has expressed worry about the high rate of academic failure among university students. The first year of college has the highest attrition rate because students leave because of their academic performance. There has been much discussion and effort put into attempting to determine what may have led to this disappointing showing. Therefore, there are numerous potential applications for the capacity to forecast a student's achievement in higher education. The suggested approach makes use of data mining methods to determine what factors significantly impact and influence undergraduate students' performance. The academic pattern is then analysed using students' demographic and previous academic achievement data. The information is buried in the school database, waiting to be uncovered by data mining tools. Midterm assessments, end-of-semester examinations, abilities, ethics, grasping capacity, extracurricular activities, and other educational data sets may all be mined for this kind of information. Such extraction is facilitated by data categorization algorithms combined with decision trees, and the resulting analysis may be used to generate semantic rules for foretelling students' final grades. Semantic web technologies like ontologies and semantic rules are used by the system to improve the quality of the lessons and exercises that are given to each individual learner. The suggested method inspires a sense of trust among educators and their pupils. As a result, the system intends to analyse the extracted data and mine educational data to generate graphical and statistical results that can aid in the enhancement of student performance and provide instructors with an overall picture of the student's level of proficiency in their chosen field of study.

Webology, Volume 18, Number 5, 2021
ISSN: 1735-188X
DOI: 10.29121/WEB/V18I5/37

Keywords: Data mining, ID3, Naive Bayes, the Perceptron learning rule, and student performance.

INTRODUCTION

An interdisciplinary field of computer science, "data mining" (the analysis phase of "Knowledge Discovery in Databases," or KDD) is the computational process of discovering patterns in large datasets ("big data") using approaches at the crossroads of artificial intelligence, machine learning, statistics, and database systems. The purpose of the data mining process is to discover useful information hidden within a dataset and organise it in a way that can be consumed by the project team. In addition to the initial data collection and processing, this process also includes database and data management, data pre-processing, model and inference considerations, interestingness measures, complexity considerations, post-processing of identified structures, visualisation, and updates.

Data manipulations like this may be used to analyse student performance and find ways to boost their grades. It is implemented using data mining and ANN (Artificial Neural Network) methods. For this function, a data set is constructed, with several pedagogical factors serving as the basis for analysis and decision making. Data is processed via the various methods, and the outcomes are compared. When many results are produced, the best one is compared for efficacy and accuracy before being shown. The inference engine extrapolates and delivers personalised ideas and guidance, and the produced reports are available in statistical, graphical, and textual formats to best communicate the results. [1]

A SUMMARY OF DATA MINING AND THE REASONS WHY IT MAY BE USED FOR THIS TYPE OF ANALYSIS

Extracting Useful Information from Large Data Sets

Many people have issues in the classroom, both students and teachers, whether it be with their grades, course choices, teaching techniques, note-taking, motivation, or testing. The goal of this project is to create a system that uses Data Mining methods to evaluate a student's academic progress. The analysis is used to make projections about the kids' future performance. It aspires to be autonomously deliberative, offering suggestions and advice to both students and instructors. The ultimate goal is to use data mining methods for the purpose of producing such information and drawing such inferences. Also, the suggested system aims to compare and pick the best choice by using data mining methodology and neural networks inside the system's operation. After the findings have been generated and a conclusion has been reached, a graphical/statistical report will be produced to show the results. [2]

Analysis by Means of Data Mining

Parents and teachers are more worried about their children's academic decline because of the increased attention pupils pay to grades these days. In addition, it is really annoying when students resort to extremely severe methods of therapy on such instances, which might have terrible results. Student life can benefit greatly from the project's concept, which can be used not only by the college staff and dean to generate reports and understand patterns, but also by the student himself, who can use this information to make important decisions like which electives to take and how to best allocate his time and effort.

Webology, Volume 18, Number 5, 2021
ISSN: 1735-188X
DOI: 10.29121/WEB/V18I5/37

The project's central idea may also be widely adopted for use in a variety of college programmes. Classical Techniques, including Statistics, Neighbor-Nets, and Clustering, are just a few examples of algorithmic frameworks that may be used to develop patterns of productive output, each of which can lead to a different inference when applied to the data at hand. Naive Bayesian, ID3 decision making, clustering methods like k means clustering, etc., are only a few examples of these algorithms. The data sets will be processed by these methods, and the optimal solution will be suggested as a consequence. It will also be useful for contrasting the various methods. In order to help the student do better, the suggested system would make judgments based on advise and suggestions.

METHODOLOGIES

The Naive Bayesian

Bayes' theorem is the basis of the Naive Bayesian classifier, which makes no assumptions about the relationship between the predictors. Naive Bayesian models are very helpful for extremely big datasets since they do not need complex iterative parameter estimation. The Naive Bayesian classifier is extensively used despite its lack of complexity because it routinely outperforms more advanced classification algorithms. Naive Bayes classifiers are extremely scalable due to the fact that the number of parameters needed is proportional to the number of features and predictors in a learning issue. Unlike the costly iterative approximation employed by many other kinds of classifiers, maximum-likelihood training may be done simply evaluating a closed-form expression, which requires linear time. Different authors refer to Naive Bayes models by different names in the statistical and computer science field. However, Russell and Norvig point out that "[Naive Bayes] is occasionally dubbed a Bayesian classifier, a rather sloppy use that has inspired actual Bayesians to label it the idiot Bayes model," alluding to the fact that naive Bayes is not (necessarily) a Bayesian approach. Naive Bayes is a straightforward method for developing classifiers, which are models used to ascribe class labels to instances of a problem represented as vectors of feature values, with the class labels being selected at random from a finite set. There isn't just one method for training naive Bayes classifiers; rather, there's a whole family of algorithms that share a fundamental belief: that, given a class variable, a feature's value is invariant to that of any other feature. A crimson, spherical fruit that is around 10 cm in diameter may be identified as an apple. Regardless of any apparent relationships between the characteristics of colour, roundness, and diameter, a naive Bayes classifier treats each feature separately when calculating the probability that this fruit is an apple. Naive Bayes classifiers can be effectively taught in a supervised learning scenario for certain probabilistic models. It is possible to operate with the naive Bayes model without embracing Bayesian probability or utilising any Bayesian techniques, since parameter estimation for naive Bayes models often use the maximum likelihood approach. In spite of their simplistic construction and seemingly simple assumptions, naive Bayes classifiers have proven useful.

performed well in a wide variety of difficult, real-world contexts. An examination of the Bayesian classification issue in 2004 shown that the seemingly improbable success of naive Bayes classifiers had reasonable theoretical underpinnings. Bayes classification is inferior to other methods, such as boosted trees and random forests, according to a thorough assessment of classification algorithms conducted in 2006. Naive Bayes has the benefit of just requiring a minimal quantity of training data to estimate the classification parameters.

Webology, Volume 18, Number 5, 2021
ISSN: 1735-188X
DOI: 10.29121/WEB/V18I5/37

The posterior probability of class (target) given predictor x is denoted by the formula P(c|x) (attribute).
The prior probability of class c is denoted by the notation P(c).
Likelihood, or the probability of a predictor given a class, is denoted by P(x|c).
• P(x) is the prior probability of predictor.
Example: The posterior probability may be computed by first, building a frequency table for each characteristic against the goal. After that, the posterior probability for each class is determined by plugging the data from the likelihood tables into the Naive Bayesian equation.
The predicted class will be the one with the greatest posterior probability. [3]
II. ID3
It is possible to create a decision tree from a dataset using Ross Quinlan's ID3 technique. ID3, a forerunner of C4.5, is often used in the fields of machine learning and natural language processing.

Algorithm
The seed node in the ID3 algorithm is the first recording in the collection. On each pass through the method, it determines the entropy (or information gain) of each unused characteristic in the collection. The property with the highest information gain (lowest entropy) is chosen next. The collection is then partitioned into subgroups based on the given property (for example, age 50, 50 = age 100, age >= 100). The process is repeated over and again for each subset, with each iteration focusing on a new set of unselected qualities. If the set of elements in the subset all belong to the same class (+ or -), then the node is transformed into a leaf and labelled with the class of the instances, and the recursion on the subset will terminate. There are no examples in the subset, which occurs when no example in the parent set was determined to be matching a given value of the specified attribute, and the node is converted into a leaf and labelled with the most common class of the instances in the subset.

for instance, if no exemplars existed when age >= 100. The most frequent category of the child instances is assigned to the newly generated leaf. Each non-terminal node in the decision tree represents the attribute used to partition the data, and each terminal node in a branch represents the class label of the subset that was produced by the method. Summary Determine the entropy of each field in the dataset.

2. Split the collection into subsets using the property for which entropy is smallest (or, equivalently, information gain is largest) (or, equivalently, information gain is maximum)
Create a node in the decision tree where that attribute is included
Using the remaining characteristics, perform a recursive operation on a subset.
[4]

Perceptron Learning Rule
The perceptron is a popular approach in machine learning for supervised learning of binary classifiers, which are functions that can determine whether an input (represented by a vector of numbers) belongs to a certain class. It is a kind of linear classifier, \si.e. a classification technique that generates its predictions based on a linear predictor function mixing a set of weights with the

Webology, Volume 18, Number 5, 2021
ISSN: 1735-188X
DOI: 10.29121/WEB/V18I5/37

feature vector [5]. The method allows for online learning, in that it processes components in the training set one at a time. A neural network, known as the perceptron, capable of categorising patterns into two or more categories is presented. Simple Perceptron for Pattern Classification In this article, we take a NN called a Perceptron into account, which can classify patterns into two or more groups [6]. Training a perceptron requires following a certain learning rule called perceptron learning. The two-class problem will be tackled first, followed by the broader multiclass problem. A single output neuron is sufficient for two-class categorization. In this case, we'll employ bipolar neurons. A layer of N input neurons, a layer with a single output neuron, and no hidden layers would make up the simplest architecture capable of doing the task. The architecture for Hebb learning is identical to what we have seen previously. The output neuron, however, will have a different transfer function: While x is more than, f(x) equals 1, when x is between and, f(x) equals 0, and when x is less than, f(x) equals 1.

SUGGESTIONS

Producing Data Sets

In this lesson, we will construct a massive dataset of students based on a variety of criteria. Different characteristics of a student's academic history form the basis of the data sets (like subjects, attendance, extra- curricular, etc.). The resulting output will be a.csv file that can be opened in Excel [7]. The system will use algorithms on the data set after extracting relevant information from it. Information will be collected from several sources, including final grades, midterm exams, classroom conduct, and other similar activities.

Purifying Data Sets

Removal of errors, mistakes, and other blunders made when compiling a data collection. In order for data to be processed and worked on, such cleansing is required. This produces a unified data set on which further operations may be carried out [8]. Improper algorithm processing is often the result of a tainted data collection. Such outcomes should be avoided, which is why data refinement is essential for effective data processing.

Utilization of the Data Collection

Data Mining and ANN training methods are used to the data set produced in the preceding phases. When a system undergoes training, it "learns," or gains insight into how to handle a certain kind of data and the specific tasks that should be carried out on that data. Furthermore, it generates novel parameters/attributes upon which to base its evaluations. Information is extracted by using data mining and ANN methods to provide analysis-ready data. The findings from this investigation are then used [9].

• We use Data Mining and Artificial Neural Network methods

After development and refinement of the data sets, data mining methods are used to the created data set to process information and learn via it. Simultaneously, Artificial Neural Network methods are used for both data generation and analysis. In addition to displaying data, these methods create recommendations. In this module, the system processes the input to map the connection between past and future courses in order to infer and draw inferences, training itself to think in the process. These findings are presented in either graphical or textual formats.

Webology, Volume 18, Number 5, 2021
ISSN: 1735-188X
DOI: 10.29121/WEB/V18I5/37

• Making Reports

A report is produced as a result of the system making a decision. Through processing the data, a report is produced. It may be something as easy as helping you choose the right classes, or it could be more involved, like showing you how to do better in key parts of the course [10]. A decision may take the shape of a graph or model representing the students' performance data. Or it may be a statistical graph showing the relationship between all of the pupils. Parents and guardians may use the reports as a basis for discussing their child's progress at home. Both the student and his or her parents may use this information to better understand where their child is succeeding and where they still need work in a certain area of study.

CONCLUSION

Applying this data mining approach will allow for effective analysis and use of real national data. In the future, many applicants may utilise this to choose which area of engineering best suits their interests and needs. Students' information from that field was used in this study. By incorporating association rule mining into the model, new, intriguing hidden patterns in the data may be uncovered. A student's prospects of getting a job after graduating from college may be looked at if he or she gets admitted to that institution and chooses a major in computer technology. Possible future projects may include collecting data on engineering course dropouts and investigating their causes. Similarly, one may look at whether or not a graduate degree improves a person's chances of getting a job. In conclusion, various hidden and important information may be obtained when data mining methods are properly used to a genuine, up-to-date, national-level data base, which can be efficiently utilised by the government and the general public for future policy planning.

REFERENCES

1. A ALTURKI, R. (2016). Measuring and improving student performance in an introductory programming course. Informatics in Education, 15(2), 183-204.
2. Winch, J. K., & Cahn, E. S. (2015). Improving student performance in a management science course with supplemental tutorial videos. Journal of Education for Business, 90(7), 402-409.
3. Feild, J. (2015). Improving Student Performance Using Nudge Analytics. International Educational Data Mining Society.
4. Suchithra, R., Vaidhehi, V., & Iyer, N. E. (2015). Survey of learning analytics based on purpose and techniques for improving student performance. International Journal of Computer Applications, 111(1).
5. Wiegel, V. (2019). Lean in the Classroom: The Powerful Strategy for Improving Student Performance and Developing Efficient Processes. CRC Press.
6. Smith, B. O., Shrader, R., White, D. R., Wooten, J., Dogbey, J., Nath, S., ... & R
7. Mitra, S., & Goldstein, Z. (2015). Designing early detection and intervention techniques via predictive statistical models—A case study on improving student performance in a business statistics course. Communications in Statistics: Case Studies, Data Analysis and Applications, 1(1), 9-21.
8. alah Ahadi, A., KHanmohammadi, M., Masoudinejad, M., & Alirezaie, B. (2015). Improving student performance by proper utilization of daylight in educational environments

Webology, Volume 18, Number 5, 2021
ISSN: 1735-188X
DOI: 10.29121/WEB/V18I5/37

(Case study: IUST School of Architecture). Acta Technica Napocensis: Civil Engineering & Architecture, 59(1).

9. Andolsek, K. M. (2016). Improving the medical student performance evaluation to facilitate resident selection. Academic Medicine, 91(11), 1475-1479.
10. Pabón, O. S., & Villegas, L. M. (2019). Fostering motivation and improving student performance in an introductory programming course: An integrated teaching approach. Revista EIA, 16(31), 65-76.