Academic Performance Prediction of intoxicating Students Using Intelligent Data Mining Techniques

M. Rama[1] | Dr.Ayesha Banu[2] | Zareena Begum[3]

[1]Assistant Professor, Dept. Of Computer science and engineering, Vaagdevi College of Engineering,Bollikunta.
[2]Associate Professor, Dept. of Computer science and engineering (Data Science), Vaagdevi College of Engineering, Bollikunta
[3 890]Assistant Professor, Dept. of Computer science and engineering (Data Science), Vaagdevi College of Engineering, Bollikunta

To Cite this Article
M. Rama, Dr.Ayesha Banu and Zareena Begum. Academic Performance Prediction of intoxicating Students Using Intelligent Data Mining Techniques. *International Journal for Modern Trends in Science and Technology* 2022, 7 pp. 459-462. https://doi.org/10.46501/IJMTST0712082

ABSTRACT

Alcohol consumption by students has become a significant issue these days. Addiction to alcohol results in the poor tutorial performance of scholars. This paper describes few algorithms that facilitate to boost the potency of educational performance of students captivated with alcohol. Within the paper, we tend to area unit victimisation one amongst the favoured Data Mining technique—"Prediction" and checking out the simplest formula among different algorithms. Our project is to investigate the educational excellence of the school professionals by creating use of WEKA toolkit and R Studio. We implement this project by making use of alcohol consumption by student datasets provided by kaggle web site. It is composed of 395 tuples and 33attributes. A classification model is constructed by making use of Naïve Bayes and ID3. Comparison of accuracy is finished between R and WEKA. The prediction is performed so as to seek out out whether or not a student will be promoted or demoted within the next year once previous year marks area unit considered.

Keywords: Data Mining , Prediction ,Naïve Bayes ,ID3 ,WEKA ,*R* studio.

1. INTRODUCTION

A large amount of knowledge is being made from completely different fields a day so as to pull out a legitimate and helpful knowledge that is employed during a decision-making method. The decision-making method is performed by victimization completely different data processing techniques. Various Mining Techniques Classification, Prediction, Clustering and Association. Classification is that the method of composition the information supported similarities. It is a supervised learning technique as we have a tendency to build the model by creating use of coaching data that consists of sophistication labels. Clustering is that the grouping of objects supported the principle of accelerating the intra-cluster distance and decreasing the inter-cluster distance. Prediction is that the method of dig out the knowledge from an enormous quantity of data and helps to predict the end result no matter the past, present, and future events. The prediction data processing technique is employed so as to predict the performance of the scholars. Using Naïve theorem and ID3 algorithms we have a tendency to designed a model and found out the accuracy each in WEKA and R and therefore the comparison between them is performed.

2. RELATED WORK

In [1] they need used totally different data processing technologies to investigate students' performance in the courses. They need used

classification techniques to assess the student's performance. Among all the classification techniques, they need used call tree technique. The data they need used is of scholar's happiness to Yarmouk University of the year 2005 who took C++ course. CRISP-DM a technique is employed to create a classification model. Among the twenty attributes, solely twelve conditional attributes square measure thought-about that affects the performance of the scholars.

In [2], they need created a web-based application creating use of Naïve Bayesian algorithm. The information consists of nineteen attributes like student details, course details, admission details, attending details, etc., from 700 students learning at Amrita Vishwa Vidyapeetham, Mysore. Among all the algorithms they need used, Naïve Bayesian has got the best accuracy. Here, students, tutorial history is taken as input and the output is student's performance on the premise of a semester.

In [3], they need primarily cantered on the information mining techniques that facilitate in learning the educational knowledge primarily in higher learning establishments. This shows however knowledge mining helps in decision-making so as to take care of university name. It predicts the student's performance at the top of their bachelor's degree and located out the students World Health Organization square measure in danger within the early years of their study and provides measured

to improve the standard. They need collected knowledge from 2 different batches of the years 2005–06 and 2006–07 from 214 collegian students happiness to the applied science department at NEDUET, Pakistan. They need used call tree with Gini index, with data gain, with accuracy, Naïve Bayesian, Neural Networks, Random forest.

In [4], the information has been collected from three hundred students of engineering department for all 3 years. The attributes square measure associated with totally different subjects like English, Maths, and artificial language. They need used algorithms like Neural Networks, J48, SOM.

In [5], they need collected 2 completely different sets of knowledge of the scholars happiness to second year and third year of Amrita faculty of Engineering, Bangalore. The dataset constitutes twenty attributes like gender, father education, mother education, etc. Naïve

Bayesian classifier is employed for predicting the student's excellence and according to the performance they need urged a learning vogue for underperformed students.

In [6], they need used alcohol consumption by student's knowledge set age 10–14.They have used SVM technique, call trees, and Naïve Bayes algorithms and that they found out that SVM is a lot of economical than alternative algorithms.

In [7], classification of knowledge mining techniques illustrates few techniques to classify data together with their applications to health care. IF-THEN prediction rules are one among them that could be a common technique in data processing. This theme gift discovered data at highest level of abstraction.

In [8], they need used datasets associated with education and that they have used numerous data mining techniques to predict and valuate their performance.

In [9], they need analyzed the faculty student performance for Villupuram district. They have used bunch technique and k-means bunch formula. They have used mathematician mixture model so as to boost the accuracy.

In [10], the authors have collected student data from B. J school and analyzed them victimisation K-means bunch. the info set used for this analysis was obtained by measuring the B.C.A students from B. J College.

In [11], vital relationship between variables from an oversized knowledge set is analyzed. The authors propose a mechanism which may be utilized by the academics to look at the academic growth of their students.

In [12, 13], they need great deal of knowledge from medical business that consists of attributes like concerning the patient together with the main points, diagnosis, and medications. Using this knowledge we have a tendency to train the model and realize a pattern that helps in prediction.

3. PROPOSED SYSTEM

Dataset: the info is collected from the net that consists of thirty four attributes. The dataset is obtained from kaggle web site. Name of the college, gender, age, qualification of parents, occupation of oldsters, weekly

study time, net access, alcohol consumption levels on operating day similarly as weekends, health condition and grades are a number of the most important attributes listed.
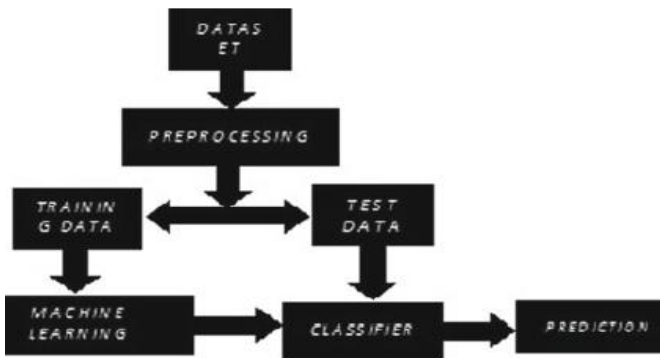

Figure 1: Proposed Model

Preprocessing: when grouping the info we want to preprocess it. The pre processing includes four steps. They are

Data Cleaning: during this step, we tend to take away the rip-roaring knowledge and fill the missing knowledge.

Data Integration: haircare totally different sorts of knowledge into single type.

Data Transformation: during this step knowledge of various types is formed into one form.

Data Reduction: this system is applied to get reduced illustration of information.

Training Data: We split and think about the primary three hundred tuples as coaching knowledge. The coaching data consists of sophistication labels. exploitation this train knowledge we tend to build a classifier.

Machine Learning: so as to create a classifier we tend to use machine learning techniques like prediction. In prediction, we've several data processing algorithms. exploitation these algorithms we tend to build a classifier.

Classifiers: we've enforced Naïve theorem and ID3 algorithmic rule in R and
WEKA to create a classifier.

Naïve Bayesian

Input: dataset

Output: confusion matrix and predicted class labels

Do

For each value of the class label (Ci,Cj) find probability

For each attribute belonging to the class label (either Ci or Cj) find probability

Compare probabilities of each attribute of different class labels

If p(Ci) > p(Cj)

Class label will be Ci else Cj

P(Ci|X) = P(X|Ci)*P(Ci)/P(X)

Confusion matrix: It is a tool for finding the accuracy.

ID3

Input: dataset

Output: confusion matrix and predicted class labels

Do

Calculate information gain of all the attribute

The attribute with highest information gain value will be taken as the root node

and according to the outcomes the tree will be further extended till all the leaf

node becomes the class labels.

Test Data

The left out dataset is taken as test data and this test data is given as input to the classifier and the prediction is performed (Fig.2).

Figure 2 : Test Data

Prediction: it's a mining technique to predict the end result of the information tuple. We perform prediction by training the classifier with trained knowledge that consists of sophistication labels so the check knowledge while not class label is given to the model and sophistication label is foretold. This category label indicates the performance of the student within the next school year i.e., yes = pass, no = fail.

4. RESULT ANALYSIS:

Table 1 describes the comparison between accuracy in R and Weka Tool for all the attributes.

Table 1: Comparision of Accuracy for ALL Attributes using R and WEKA

Algorithm	R (%)	WEKA (%)
Naïve Bayes	96.8	95.95
ID3	94.9	92

Table 2 showss the comparison between accuracy in R and WEKA for the attributes that effects student's performance.

Table 2: Comparision of Accuracy FEW Attributes using R and WEKA

Algorithm	R (%)	WEKA (%)
Naive Bayes	94.7	87
ID3	100	100

Figure 3 represents performance of the attributes and quality of the attributes. Each bar chart within the on top of figure indicates distribution of every attributes and every color indicates completely different categories. The attributes used period, study time, gout, G1, G2, G3, Dalc, Walc, performance.
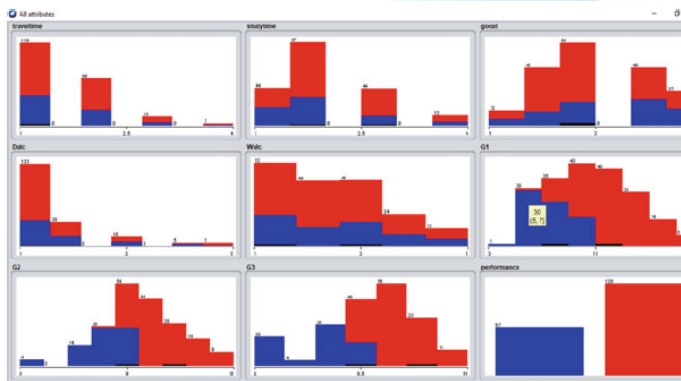

Figure 3: Visualization of Attributes

5. CONCLUSION:

Using Naive theorem and ID3 classifier we tend to got the best accuracy than several other algorithms. The attributes that have gotten highest priority got less accuracy once we have used ID3 algorithmic program than the opposite attributes once thought of. Once all the attributes square measure thought of naïve Naive Bayes once enforced in R has a lot of accuracy than in rail. The accuracy obtained in R and rail for ID3 once all the attributes square measure thought of is same because the accuracy obtained in R and rail for ID3 once

just some of the attributes square measure thought of. From this, we can infer that the time needed to implement the classifier for under a number of attributes is a smaller amount than the time once all attributes are thought of. The category label is foretold and this means that whether or not a student is pass or fail by considering previous year's marks.

REFERENCES

[1] Al-Radaideh, Q., Al-Shawakfa, E., Al-Najjar, M.I.: Mining student data using decision trees.The Int. Arab J. Inf. Technol.—IAJIT (2006)

[2] Devasia, T., Vinushree T.P., Hegde, V.: Prediction of students performance using educational data mining. In: International Conference on Data Mining and Advanced Computing (SAPIENCE), pp. 91–95 (2016)

[3] Asif, R., Hina, S., Haque, S.I.: Predicting student academic performance using data mining methods. Int J Comput. Sci. Netw. Secur. (IJCSNS) 17(5), 187–191 (2017)

[4] Ramesh, V., Parkavi, P., Yasodha, P.: Performance analysis of data mining techniques for placement chance prediction. Int. J. Sci. Eng. Res. 2, 2229–5518 (2011)

[5] Krishna, K.S., Sasikala T.: Prognostication of students performance and suggesting suitable learning style for under performing students. In: International Conference on Computational Systems and Information Technology for Sustainable Solutions (CSITSS—2018), December 2018.

[6] Fabio, M.P., Roberto, M.O., Ubaldo, M.P., Jorge, D.M., Alexis, D.L.H.M., Harold, C.N.: Designing A Method for Alcohol Consumption Prediction Based on Clustering and Support Vector Machines. Res. J. Appl. Sci., Eng. Technol. 14, 146–154

[7] Sreevidya B., Rajesh M., Sasikala T.: Performance analysis of various anonymization techniques for privacy preservation of sensitive data. In: Hemanth J., Fernando X., Lafata P., Baig Z. (Eds.) International Conference on Intelligent Data Communication Technologies and Internet of Things (ICICI) 2018. ICICI 2018. Lecture Notes on Data Engineering and Communications Technologies, vol 26. Springer (2019)

[8] Krishnaiah,V.,Narsimha, G., Subhash Chandra, N.:Diagnosis of lung cancer prediction system using data mining classification techniques. Int. J. Comput. Sci. Inf. Technol. 4, 39–45 (2013).

[9] Shelke, N.: A survey of data mining approaches in performance analysis and evaluation. Int. J. Adv. Res. Comput. Sci. Softw. Eng. (2015)

[10] Jyothi, J.K. Venkatalakshmi, K.: Intellectual performance analysis of students by using data mining techniques. Int. J. Innov. Res. Sci. Eng. Technol. 3, (2014)

[11] Sreevidya, B.: An enhanced and productive technique for privacy preserving mining of association rules from horizontal distributed database. Int. J. Appl. Eng. Res. (2015)

[12] Bhise, R.: Importance of data mining in higher education system. IOSR J. Hum. Soc. Sci. 18–21 (2013).

[13] Sumitha Thankachan, Suchithra, Data mining warehousing algorithms and its application in medical science. IJCSMC, 6 (2010).