International Journal for Research in Engineering Application & Management (IJREAM)
ISSN : 2454-9150

Forecasting of Deliberate Learners in Education Field using Data Mining Techniques

G.Venugopal, T. D. Ravi kiran
Asst. Professor, Department of IT, P V P Siddhartha Institute of Technology, A.P, India.
D. Syam Sandeep, G. Gnananandh, K. Subha Kiran
IV/IV B.Tech IT Department, P V P Siddhartha Institute of Technology, India.

Abstract Forecasting a deliberate learner in the field of education is not a simple task. This paper focus on identifying the deliberate learners among students community and displaying it by a foretelling data-mining model using classification-based algorithms. The dataset of understudy scholarly records is tried and different arrangement calculations connected, for example, Naïve Bayes, J48 utilizing R, an Open source tool. This paper highlights the importance of forecasting data mining algorithms in the field of education.

Keywords—Classification, forecasting, data mining, Naïve Bayes, J48, Educational Data Mining.

I. INTRODUCTION

Instructive Data Mining (IDM) is the one of the application of Data Mining techniques on informational data. The main objective of IDM is to separate such data and to decide educational research issues[1]. IDM manages growing new strategies to investigate the instructive information, and utilizing Data Mining techniques to better comprehend understudy learning condition. Instructive Data Mining field center around Prediction more much of the time as stand out from make remedy results for future reason. This paper recognizes the components related with understudies whose scholarly execution isn't great and to enhance the nature of instruction by distinguishing moderate students by a prescient information mining model utilizing grouping based calculations so educators can help them separately to enhance their execution Real World informational index from an office information is taken and filtration of wanted potential factors is finished utilizing R an Open Source Tool. The dataset of understudy scholastic records is tried and connected on any of these order calculations, for example, Naïve Bayes, utilizing R an Open source tool. This paper includes the essentialness of Forecasting and Classification based data mining figurings in the field of preparing and moreover demonstrates some promising future lines.

II. DATA ANALYSIS

Understudy's scholarly execution is a pivotal factor in building their future. Scholarly execution of understudy isn't an aftereffect of just a single main factor other than it intensely depends on different elements like individual, financial, mental and other natural factors. The guideline goals of this work are to make data wellspring of insightful elements, Data mining methods to consider understudy execution at graduation level, recognizing confirmation of the direct understudies' execution, conspicuous verification of the significantly influencing perceptive factors on the educational execution of understudies [2][3][4]. In present day's educational system, a student's performance is determined by the internal assessment and end semester examination. Within assessment is finished by the teacher in light of understudy's execution in informative activities, for instance, class test, workshop, assignments, general capacity, interest and lab work. The end semester examination is one that is scored by the understudy in semester examination. Every understudy needs to get least stamps to pass a semester in interior and also end semester examination.

III. CONDITIONAL PROBABILITY

In likelihood hypothesis contingent likelihood is a measure of the likelihood of an occasion given that (by suspicion, assumption, statement or proof) another occasion has happened. On the off chance that the occasion of intrigue is An and the occasion B is known or accepted to have happened, "the contingent likelihood of A given B", or "the likelihood of An under the condition B", is generally composed as P(A|B), or here and there $P_B(A)$.

Given two occasions A and B, from the sigma field of a likelihood space, with P(B) > 0, the contingent likelihood of A given B is characterized as the remainder of the likelihood of the joint of occasions An and B, and the likelihood of B.

Conditional Probability: Bayes' Theorem

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

Bayes' Theorem

Bayes' hypothesis is named after Thomas Bayes, a free thinker English priest who did early work in likelihood and choice hypothesis amid the eighteenth century.

What is Naive Bayes algorithm?

It is a gathering methodology in perspective of Bayes' Theorem with a doubt of flexibility among pointers. In direct terms, a Naive Bayes classifier acknowledge that the closeness of a particular segment in a class is detached to the proximity of some other component. For example, a characteristic item may be believed to be an apple if it is red, round, and around 3 sneaks in separate over. Notwithstanding whether these features depend upon each other or upon the nearness of interchange features, these properties self-governingly add to the probability that this characteristic item is an apple and that is the reason it is known as 'Innocent'. Gullible Bayes display is anything but difficult to manufacture and especially valuable for expansive informational collections. Alongside straightforwardness, Naive Bayes is known to beat even exceptionally modern order strategies.

Bayes hypothesis gives a method for ascertaining back likelihood P(c|x) from P(c), P(x) and P(x|c). Take a gander at the condition beneath:

$$\underbrace{P(c\mid x)}_{\text{Posterior Probability}}=\frac{\overbrace{P(x\mid c)}^{\text{Likelihood}}\,\overbrace{P(c)}^{\text{Class Prior Probability}}}{\underbrace{P(x)}_{\text{Predictor Prior Probability}}}$$

$$P(c\mid X)=P(x_1\mid c)\times P(x_2\mid c)\times\cdots\times P(x_n\mid c)\times P(c)$$

Above,

- P(c|x) is the Posterior Probability of class (c, target) given indicator (x, properties).
- P(c) is the class prior probability.
- P(x|c) is the probability which is the likelihood of indicator given class.
- P(x) is the predictor prior probability.[8]

Dataset used for Training

Student dataset used for training the classifiers specified(Naive Bayes, J48) with the attributes PRSM, CTG, SEMPA, ASSM, GAP, ATTEN, LBW, ESEMM etc., The area esteems for a portion of the factors were characterized for the present examination as takes after:

PRSM – Prior Semester Marks/Grade acquired in B.Tech course. It is part into five class esteems: First means >=60%, Second means >=50% and <60%, Third means >=40% and < half, Fail < 40%.

CTG – Class test grade got. Here in each semester two class tests are coordinated and ordinary of two class test are used to find out session marks. CTG is part into three classes: Poor means < 40%, Average means >= 40% and < =60%, Good means >=60%.

SEMPA – Seminar introduction gained. In each semester course are dealt with to check the execution of understudies. Course execution is surveyed into three classes: Poor means Presentation and correspondence capacity is low, Average means Either presentation is fine or Communication mastery is fine, Good means Both presentation and Communication fitness is fine.

ASSM – Assessment execution. In each semester two assignments are given to understudies by each teacher. Errand execution is disengaged into two classes: Yes implies understudy submitted undertaking, No methods Student not submitted assignment.

GAP - General capacity execution. Like class, in every semester general capacity tests are composed. General Proficiency test is isolated into two classes: Yes implies understudy took part when all is said in done capability, No methods Student not partook all in all capability.

ATTEN – Attendance of Student. Least 70% participation is mandatory to take part in End Semester Examination. Be that as it may, even through in extraordinary cases low participation understudies additionally take an interest in End Semester Examination on veritable reason. Participation is partitioned into three classes: Poor means <60%, Average - > 60% and <80%, Good - >80%.

LBW – Lab Work. Lab work is segregated into two classes: Yes implies understudy completed lab work, No methods understudy not completed lab work.

ESEMM - End semester Marks acquired in B. Tech semester and it is announced as reaction variable. It is part into five class esteems: First means >=60% , Second means >=50% and <60%, Third means >=40% and < 5%, Fail < 40%. [5]

TABLE1. STUDENT RELATED ATTRIBUTES

Variable	Description	Possible Values
CTG	Class Test Grade	{Poor , Average, Good}
SEMPA	Seminar presentation	{Poor , Average, Good}
ASSM	Assignment	{Yes, No}
GAP	General ability	{Yes, No}
ATTEN	Attendance	{Poor , Average, Good}
LBW	Lab Work	{Yes, No}
ESEMM	End Semester Marks	{First >= 60% Second >=50 & <60% Third >=40 & <50% Fail < 40%}

TABLE3. TEST SET USED FOR PREDICTION

S. No.	PSM	CTG	SEM	ASS	GP	ATT	LW	ESM
1	First	Good	Average	No	No	Average	No	First
2	First	Average	Good	No	No	Good	Yes	First
3	Second	Good	Average	Yes	Yes	Average	Yes	First
4	Third	Good	Good	No	No	Good	Yes	Second
5	Third	Average	Average	Yes	Yes	Good	Yes	Second
6	Fail	Poor	Good	No	No	Poor	No	Fail

TABLE 2. DATASET USED FOR TRAINING

S. No.	PSM	CTG	SEM	ASS	GP	ATT	LW	ESM
1	First	Good	Good	Yes	Yes	Good	Yes	First
2	First	Good	Average	Yes	No	Good	Yes	First
3	First	Good	Average	No	No	Average	No	First
4	First	Average	Good	No	No	Good	Yes	First
5	First	Average	Average	No	Yes	Good	Yes	First
6	First	Poor	Average	No	No	Average	Yes	First
7	First	Poor	Average	No	No	Poor	Yes	Second
8	First	Average	Poor	Yes	Yes	Average	No	First
9	First	Poor	Poor	No	No	Poor	No	Third
10	First	Average	Average	Yes	Yes	Good	No	First
11	Second	Good	Good	Yes	Yes	Good	Yes	First
12	Second	Good	Average	Yes	Yes	Good	Yes	First
13	Second	Good	Average	Yes	No	Good	No	First
14	Second	Average	Good	Yes	Yes	Good	No	First
15	Second	Good	Average	Yes	Yes	Average	Yes	First
16	Second	Good	Average	Yes	Yes	Poor	Yes	Second
17	Second	Average	Average	Yes	Yes	Good	Yes	Second
18	Second	Average	Average	Yes	Yes	Poor	Yes	Second
19	Second	Poor	Average	No	Yes	Good	Yes	Second
20	Second	Average	Poor	Yes	No	Average	Yes	Second
21	Second	Poor	Average	No	Yes	Poor	No	Third
22	Second	Poor	Poor	Yes	Yes	Average	Yes	Third
23	Second	Poor	Poor	No	Yes	Average	Yes	Third
24	Second	Poor	Poor	Yes	Yes	Good	Yes	Second
25	Second	Poor	Poor	Yes	Yes	Poor	Yes	Third
26	Second	Poor	Poor	No	No	Poor	Yes	Fail
27	Third	Good	Good	Yes	Yes	Good	Yes	First
28	Third	Average	Good	Yes	Yes	Good	Yes	Second
29	Third	Good	Average	Yes	Yes	Good	Yes	Second
30	Third	Good	Good	Yes	Yes	Average	Yes	Second
31	Third	Good	Good	No	No	Good	Yes	Second
32	Third	Average	Average	Yes	Yes	Good	Yes	Second
33	Third	Average	Average	No	Yes	Average	Yes	Third
34	Third	Average	Good	No	No	Good	Yes	Third
35	Third	Good	Average	No	Yes	Average	Yes	Third
36	Third	Average	Poor	No	No	Average	Yes	Third
37	Third	Poor	Average	Yes	No	Average	Yes	Third
38	Third	Poor	Average	No	Yes	Poor	Yes	Fail
39	Third	Average	Average	No	Yes	Poor	Yes	Third
40	Third	Poor	Poor	No	No	Good	No	Third
41	Third	Poor	Poor	No	Yes	Poor	Yes	Fail
42	Third	Poor	Poor	No	No	Poor	No	Fail
43	Fail	Good	Good	Yes	Yes	Good	Yes	Second
44	Fail	Good	Good	Yes	Yes	Average	Yes	Second
45	Fail	Average	Good	Yes	Yes	Average	Yes	Third
46	Fail	Poor	Poor	Yes	Yes	Average	No	Fail
47	Fail	Good	Poor	No	Yes	Poor	Yes	Fail
48	Fail	Poor	Poor	No	No	Poor	Yes	Fail
49	Fail	Average	Average	Yes	Yes	Good	Yes	Second
50	Fail	Poor	Good	No	No	Poor	No	Fail

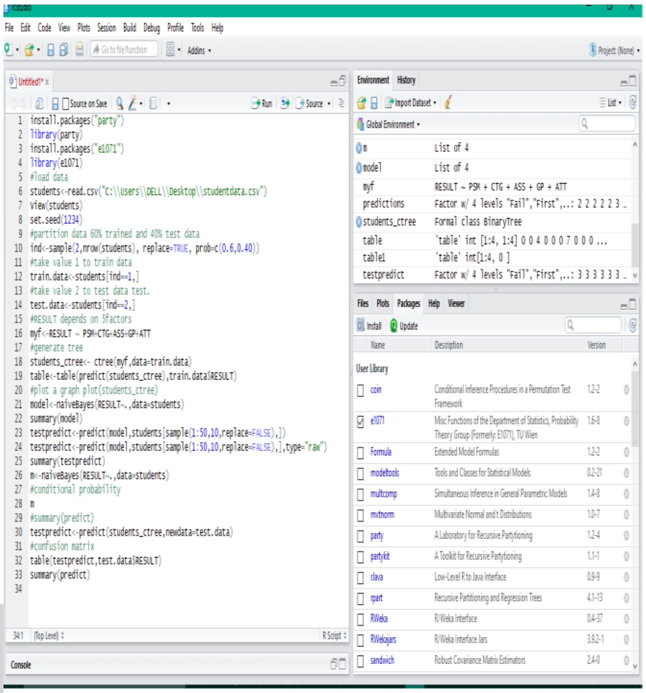

Figure1. Naïve Bayes Code in R in predicting slow learners

Figure1. contains Naïve bayes code in R and predict the class labeled attribute manually by doing the analysis of train data. Next, we generate the confusion matrix of a class labeled attribute which shows student divisions.[9]

RESULT

○ table(testpredict,test.data$RESULT)

testpredict	Fail	First	Second	Third
Fail	0	0	0	0
First	0	0	0	0
Second	4	7	4	4
Third	0	0	0	0

Figure2. Confusion Matrix to prediction result

We have taken test data upon which we have applied Naïve Bayes. Figure2. shows the predicted result from the test data

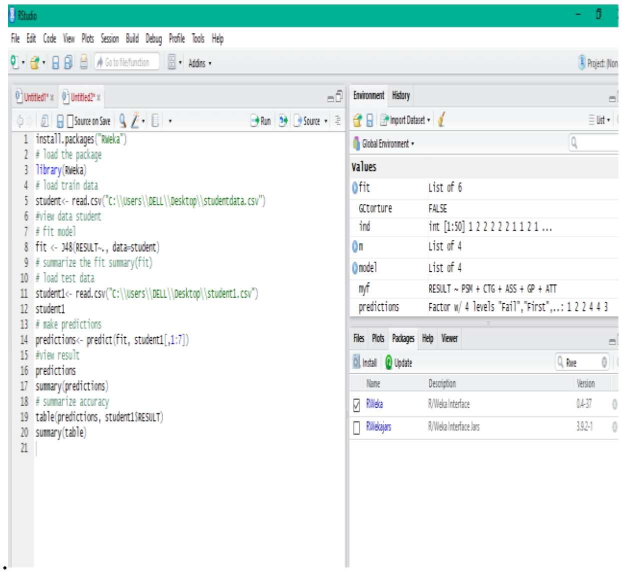

Figure3. J48 code in R in predicting slow learners

With the same dataset used for training in naïve bayes, we ar e performing prediction using J48 algorithm in R. Figure3. S hows the algorithm for prediction J48 in R.

RESULT

o summary(predictions)

```
Fail  First Second  Third
   1      2      1      2
```

Figure4. Confusion Matrix to prediction result

Dataset used for testing in Bayes, used here for the same purpose. Figure4. Predicts the slow learners using J48 algorithm.

IV. CONCLUSION

The classification task done on student dataset to predict the students grade on the basis of previous dataset. We have many techniques for classification, the decision tree method, naïve bayes and J48 are used here, given 70% accuracy with a misclassification of 30%. With the attributes specified in the dataset, are gathered from past database to anticipate the execution toward the finish of the semester. This investigation will help to the both parties(Students & teachers) to enhance the grade of the understudy.

REFERENCES

[1] Cristobal Romero (2010), “Educational Data Mining: A Review of the State-of-the-Art”, IEEE Transactions on systems , man and cybernetics- Part C: Applications and Reviews vol. 40 issue 6, pp 601 – 618.

[2] Zaïane, O. (2001),” Web usage mining for a better web-based learning environment”, Proceedings Of Conference on Advanced Technology For Education, 60-64.

[3] Merceron, A., Yacef, K. (2003),” A web-based tutoring tool with mining facilities to improve learning and teaching”. Proceedings of the 11th International Conference on Artificial Intelligence in Education, 201–208

[4] Romero, C., Ventura, S., de Bra, P., & Castro, C. (2003),” Discovering prediction rules in aha! Courses”. Proceedings of the International Conference on User Modelling, 25–34.

[5] Brijesh Kumar Baradwaj, Saurabh Pal, (2011), “Mining Educational Data to Analyze Students‟ Performance”,
(IJACSA), Vol. 2, No. 6,63-69.

[6] https://www.sciencedirect.com/science/article/pii/.../pdf ?md5...pid=1-s2.0...1

[7] https://www.r-bloggers.com/understanding-naive-bayes-classifier-using-r/

[8] https://www.analyticsvidhya.com/blog/2017/09/naive-bayes-explained/

[9] Ratnam Dodda , G.Venugopal , R.V.S Naga Raju(2017) “Impact Of Weather Condition On Agriculture At Specific Area”, IJAASE, Volume 5 Issue 10 June 2017, 2320-6144.

