

Advanced Scholastic Analytics by Implementing Data Mining Techniques

Madhav Singh Solanki

Abstract— Over the past few years, the adaption of Edification management System in the sector of education has increased. Mining, clustering of essential data and finding out unique patterns from the field of education to research student's behaviors and performance is widely recognized as Educational Data Mining (EDM) and it is a progressing profession involved with the production of new techniques for discovering distinctive and progressively large-scale information from educational environments and employing those techniques for deeper comprehend learning. It also provides an inherent understanding of teaching and learning processes for the efficient scheduling of education. This paper recommends the use of two information mining techniques in educational data. First, in admittance information, the association rule was implemented to discover some knowledge to support admission schedules. Second, a decision tree was implemented in grades and graduate student job information to estimate work type after graduation. The findings of this research provide an excellent understanding of admittance scheduling and work prediction.

Keywords— educational data mining, classification, association rule mining Education Planning, Clustering, scheduling.

I. INTRODUCTION

In particular, data mining can be described as the continuous handling of big data sets and the identification of concealed facts and important data. Data mining has the aim of understanding data patterns and developing new and operative knowledge. The education industry has seen extreme vertical development leading to an elevation in educational information. Educational information systems or Learning Management System (LMS) can hold enormous amounts of data from numerous locations in different sizes and with different amounts of precision. This information is quite comprehensive and includes a lot of information, for instance, when a teacher received access to educational material, how soon they collaborated with it, how many occasions and in what series they performed with exam assignments and other kinds of assignments, and so on. Therefore, any duration of work in LMS, even a brief one, can generate a big quantity of information for evaluation. Educational Data mining (EDM) is a fresh occurrence in the sector of database information mining and knowledge discovery (KDD) which focuses on the development and discovery of essential knowledge from academic IT structures, like admission schemes, register processes, class leadership structures (models, blackboards, etc.), as well as any other technologies that deal with learners at various stages, from schools, to colleges and universities. Researchers in this area are focused on finding helpful

knowledge to either improve the student management of the schools or to improve the efficiency of teaching. EDM is the method of transmuting pure data collected by instructional technologies into helpful material that can be employed to make educated choices and for finding responses to some research questions. Analyzing data and information about learners to rank people, establish decision making or association rules, for better decision making or improve the efficiency of learners is an exciting area of research that relies primarily on analyzing and comprehension educational data about learners, indicate their academic efficiency and generate particular standards, categories, and predictions. The most common and efficient method of Data mining used to identify and estimate scores is classification.

II. RELATED WORK

Voluminous studies have been conducted to illustrate the significance of "data mining" methods in the education sector. The process of extracting valid and precise data based on the performance and effectiveness in the learning process is demonstrated in [1]. "Data Mining" has also been employed in the area of education methods to evaluate the prospectus and theme of present study subjects, as well as to evaluate learner achievement [2]–[7]. Several studies have been carried out under this suggested topic of research for example [8] used the Naïve Bayes model to forecast pupil output based on thirteen factors. These findings were utilized to construct a prototype that is employed to predefine learners who are at risk of a letdown and enable a program of instruction for counseling. [9] In their studies, the "K implies" procedure utilized a group of eight thousand learners depending on 5 factors (median university entry results of tests/exams, median paper results, seminar records, and interval job records). The findings demonstrated a robust connection between participation and pupil achievement.[10]Claim that information can enhance the teaching scheme, college achievement, and organizational governance by analyzing "Data Mining." [11] Directed an assessment survey; take into consideration the development of the teaching and analysis of exams at the start and finish of the classes. In 2015, [12] described the consumption of the "Neuro-Fuzzy" ranking of student in Electrical Engineering of Malaysian Public University. The analysis presented that the production of the arrangement might conclude the pupil's likelihood of attaining outstanding grade even if the pupil got impotence in some course or subject. [13] Utilized numerous information mining functions to produce qualitative predictive models to

Revised Manuscript Received on 14 September, 2019.

Madhav Singh Solanki, Department of Computer Science and Engineering, Sanskriti University, Uttar Pradesh, India. (E-mail: sanpubip@gmail.com)

assess the grades of learners from academic dataset. The conclusions may encourage the educational organization to conduct information mining tasks on their school information, as well as students to enhance their performance. The conception suggested in [14] [14] on higher education for Big Data Analytics. The suggested architecture is divided into five parts; the data collection device that collects university data from multiple sources (school ticket, personal networking scheme, college information service), the big data leadership data processing and governance scheme, data analysis scheme that processes data systems. The latest survey on the admission criteria of a University of Saudi Arabia has been released [15]. They used algorithms and advanced methods and a model that suits the audience and its factors. Input in prior training courses, entry records and even features describing the requirements of the university were taken into consideration. Some researches demonstrate the utilization of MOODLE by implementing Data Mining [16] explains the various methods for information mining to support the teaching of students on digital systems. The clustering matrix in [17] provide a learning model for students. Some research [18] examined how information is functioned in the school for information to improve the teaching scheme and emphasize understanding. [19] Developed a scheme to recognize facial expressions depending on school student anger or comprehension. [20] Also describes a guide for the use of prediction models for human behavior.

III. PROPOSED WORK

Association rule mining is a technique of Data Mining to find an exciting connection between attributes. There are two significant measures for the association rule mining. The first measure is the technique to search a frequent item set that comprises minimum support and minimum confidence threshold. The second measure is a method to produce instructions from frequent item set. The guidelines produced are assessed by the confidence price. For implementing association rule mining in educational data “Apriori” algorithm is used.

“Apriori” procedure for the proposed system is:

```

ProcedureApriori (T, minSup){
//T is dataset and minSup is minimum support value
Ck: Candidate itemset of size k
Lk: frequent itemset of size k
L1 = {frequent itemsets};
for(k=1; Lk != φ; k++) do begin
Ck+1 = candidates generated from Lk;
for each transaction t in dataset do {
increment the count of all candidates in Ck+1 that are
contained in t
Lk+1 = candidates in Ck+1 with min_sup
}
end
return Uk, Lk
    
```

For the generation of the decision tree, ID3 technique is used, which is a statistical model. ID3 technique is employed to produce the decision tree by utilizing a top-down approach, greedy search by provided sets of training information to train the proposed system at every attribute with node [X]. The root node selection is done by a key provided by the decision tree for recursion. The increase of data can assess how well the selected node can produce the lowest decision tree. The equation gaining data as follows:

$$Gain(S, A) = Entropy(S) - \sum_{v \in Values(A)} \frac{|S_v|}{|S|} Entropy(S_v)$$

Where Value (A) is a set of all possible values for Function A and

S_v is a subset of “S” whose function “A” has value “v”.

The Entropy(S) is an information theory metric that defines how diverse an independent data set is. It is calculated as:

$$Entropy(S) = \sum_{i=1}^c -p_i \log_2 p_i$$

Where pi belongs to category I and pi is the probability of S. The decision tree production method as follows:

```

ProcedureBuildTree(S){
//S = set of not null data
repeat
maxGain ← 0
split A ← null
e ← Entropy(Attributes)
for all Attributes a in S do
gain ← InformationGain(a, e)
if gain > maxGain then
maxGain ← gain
splitA ← a
end if
end for
Partition(S, splitA)
until all partition processed
}
    
```

IV. RESULTS

The proposed system can be divided into parts. The first part discusses how data mining can contribute to the method of admission. The second part shows how the student i.e. the learner work profession can be predicted by data mining techniques. For the first issue, the association rule is used as an instrument to find exciting patterns and linking characteristics in order to solve this issue. The “minsup” and “minconf” thresholds are set as above. Table 1 provides a snapshot of the guidelines after the admission information collection is completed for each region.



Region	Rules	Support / Confidence
North	1. North, Faculty of Medicals, 10% direct admission => Lampang	5% / 80%
	2. North, Faculty of MIS, 10% direct admission => Phayao	4% / 60%
Central	3. Central, School of education, 10% direct admission => Sukhothai	2% / 80%

Region	Rules	Support / Confidence
	4. Central, Faculty of Medicals, 10% direct admission => Phitsanulok	3% / 60%
South East	5. South East, School of education, 10% direct admission => Udon Thani	5% / 80%
	6. South East, Faculty of Medicals, 10% direct admission => Udon Thani	3% / 50%

After investigating the produced association instructions, it has been perceived that:

Rule No. 1 implies that learners from the northern area who are admitted with 10% immediate plan and medical college entry have an assistance price of 5% and a confidence price of 80% in the province of Lampang.

Rule No. 2 means that learners from the North who receive 10% immediate grant and entry to the faculty of MIS have 4% assistance status in the province of Phayao with 60% confidence price.

Rule No.3 implies that learners from key areas who admit 10% of the plan and college entry have 2% assistance importance and 80% certainty compared with the province of Sukhothai.

Rule No.4 implies that learners from key countries who are admitted by 10 percent immediate projects and medical schools have a 3 percent assistance price and a 60 percent confidence price in the province of Phitsanulok.

Rule No.5 implies that learners from the South East and the 10% immediate grant and college entry have an assistance score of 5 percent and 80 points compared with the province of Udon Thani.

Rule No. 6 implies that learners from the South-East area, who are admitted by 10% immediate venture and doctoral admissions, have comparative values of 4% and 60% with the Province of Udon Thani.

The second issue relates to how data mining can forecast the employment of the student. In this suggested system ID3 decision tree is used which is an instrument to forecast the work of a pupil to reply to this query. Important parameters are set to the above setup. The confusion matrix of the decision tree ID3 was created as panel IV after the student grade decision tree was run with 5-fold cross-validation with work dataset.

	True IT	True Non-IT	Class Precision (%)
Prediction IT	45	11	80.35%
Prediction Non-IT	17	33	66.00%
	72.58%	75.00%	

The decision tree of ID3 has been prepared to forecast 78 records out of 106 documents, both favorable and bad. The precision is 73.58 percent of the ID3 decision tree.

V. CONCLUSIONS

In this paper, two data mining techniques are implemented for discovering essential knowledge from a dataset produced by educational information. The work presented in this paper can be divided into two parts: the first part shows the implementation of a data mining technique to help the admission process. The outcome of this experimentation depicts significant connections among area, admission name, faculty, and province. This outcome aid counselor for the admission process so that they can make strategies for the promotion of their respective educational organization. The second part, present an ID3 decision tree is implemented to estimate job by analyzing learner's core values. The outcome of the second part depicts the subject or course which is essential for the learner for a good career.

REFERENCES

1. B. Kumar and S. Pal, "Mining Educational Data to Analyze Students Performance," *Int. J. Adv. Comput. Sci. Appl.*, 2013.
2. Harwati, A. P. Alfiani, and F. A. Wulandari, "Mapping Student's Performance Based on Data Mining Approach (A Case Study)," *Agric. Agric. Sci. Procedia*, 2015.
3. A. O. Adebayo and M. S. Chaubey, "Data mining classification techniques on the analysis of student's performance," *Dep. Syst. Program.*, 2019.
4. S. N. Bonde and D. K. Kirange, "Survey on Evaluation of Student's Performance in Educational Data Mining," in *Proceedings of the International Conference on Inventive Communication and Computational Technologies, ICICCT 2018*, 2018, pp. 209–213.
5. K. Priya and A. Kumar, "Improving the Student's Performance Using Educational Data Mining," *Int. J. Adv. Netw. Appl.*, vol. 4, no. 4, pp. 1680–1685, 2013.
6. Revathy P., Kalaiarasi P., K. J., and M. D. a., "Data Mining Approach for Suggesting Higher Education Courses Based on Student's Performance," *Int. J. Sci. Technoledge (ISSN 2321 – 919X)*, vol. 3, no. 3, pp. 207–211, 2015.
7. S. Kalaivani, B. Priyadharshini, and B. S. Nalini, "Analyzing Student's Academic Performance Based on Data Mining Approach," *Int. J. Innov. Res. Comput. Sci. Technol.*, vol. 5, no. 1, pp. 194–197, 2017.
8. D. Varun and A. Chadha, "An Empirical Study of the Applications of Data Mining Techniques in Higher Education," *Int. J. Adv. Comput. Sci. Appl.*, 2013.

9. B. M, J. Tomy, U. A, and P. Jacob, "Clustering Student Data to Characterize Performance Patterns," *Int. J. Adv. Comput. Sci. Appl.*, 2013.
10. P. Gulati and A. Sharma, "Educational data mining for improving educational quality," *Int. J. Comput. Sci. Inf. Technol. Secur.*, 2012.
11. M. I. Al-Twijri and A. Y. Noaman, "A New Data Mining Model Adopted for Higher Institutions," in *Procedia Computer Science*, 2015.
12. I. Hidayah, A. E. Permanasari, and N. Ratwastuti, "Student classification for academic performance prediction using neuro fuzzy in a conventional classroom," in *Proceedings - 2013 International Conference on Information Technology and Electrical Engineering: "Intelligent and Green Technologies for Sustainable Development", ICITEE 2013*, 2013.
13. A. Peña-Ayala, "Educational data mining: A survey and a data mining-based analysis of recent works," *Expert Systems with Applications*. 2014.
14. S. Pal, "Mining Educational Data to Analyze Students' Performance," *IJACSA Int. J. Adv. Comput. Sci. Appl.*, vol. 2, no. 6, pp. 63–69, 2011.
15. S. B. Sailesh, K. J. Lu, and M. Al Aali, "Context driven data mining to classify students of higher educational institutions," in *Proceedings of the International Conference on Inventive Computation Technologies, ICICT 2016*, 2017, vol. 2.
16. C. Romero, R. Cerezo, A. Bogarín, and M. Sánchez-Santillán, "Educational process mining: A tutorial and case study using moodle data sets," in *Data Mining And Learning Analytics: Applications in Educational Research*, 2016.
17. K. Kularbphetpong, "Analysis of Students' Behavior Based on Educational Data Mining," *Adv. Intell. Syst. Comput.*, vol. 662, pp. 167–172, 2018.
18. J. Srivastava and D. A. K. Srivastava, "Data Mining in Education Sector: A Review," *Spec. Conf. Issue Natl. Conf. Cloud Comput. Big Data*, 2013.
19. J. F. Grafsgaard, J. B. Wiggins, K. E. Boyer, E. N. Wiebe, and J. C. Lester, "Predicting learning and affect from multimodal data streams in task-oriented tutorial dialogue," *Proc. Seventh Int. Conf. Educ. Data Min.*, 2014.
20. S. J. Lee, Y. Liu, and Z. Popovi, "Learning Individual Behavior in an Educational Game: A Data-Driven Approach," *Proc. 7th Int. Conf. Educ. Data Min.*, 2014.