Discovering rules for nursery students using apriori algorithm

Mohammad Marufuzzaman¹, D. Gomes², A. A. A. Rupai³, L. M. Sidek⁴

¹Institute of Energy Infrastructure, Universiti Tenaga Nasional, Malaysia ^{2,3}Department of Computer Science, American International University, Bangladesh ⁴Civil Engineering Department, College of Engineering, Universiti Tenaga Nasional, Malaysia

ABSTRACT

Article Info

Article history:

Received Jun 24, 2019 Revised Aug 19, 2019 Accepted Sep 26, 2019

Keywords:

Apriori algorithm Association rules Data analysis Information technology Nursery education Over recent years, there has been a rise in the number of students completing nursery education in Bangladesh. However, in order to achieve a sustainable education goal, the dropout rate in education needs to be reduced. Therefore, this research worked on providing insights that would help to understand the possible causes of dropout from education. Since primary education is the starting point for every student, this research has been conducted on this part of education. The research used data obtained from a European country, Slovenia to use the insights of a developed country. The study was conducted using association rule mining where several mining rules were generated using the Apriori algorithm. The rules obtained had the confidence of 0.95 and support of 0.04. The result showed three major rules of dropping out children in nursery education and eventually helps to ensure higher education for all children.

This is an open access article under the <u>CC BY-SA</u> license.



Corresponding Author:

Mohammad Marufuzzaman, Institute of Energy Infrastructure, Universiti Tenaga Nasional, Jalan IKRAM-UNITEN, Selangor, Malaysia. Email: marufsust@gmail.com

1. INTRODUCTION

United Nations Development Program set 17 Global Goals, which is known as Sustainable Development Goal (SDG). Quality Education for all girls and boys is one of the major goals in that list which asking for ensuring free education until secondary school for all by 2030 [1]. Bangladesh is a developing country and the recent rise in the rate of literacy is constantly accelerating the country's growth and prosperity. According to UNESCO, the literacy rate of the population aged 15 to 24 in 2017 was 92.95% [2]. This is a clear indication of the fact that people are getting aware of the need for education. Students of the current generation are willing to complete their tertiary education, which plays a vital role in the country's development. However, most of the students are not to continue their education until the tertiary level. In other words, it is important to decrease the dropout rates of nursery education. According to the leading newspaper in Bangladesh, the dropout rate was about 18.8% in 2017 [3]. A report by the Bangladesh Bureau of Statistics mentioned that the major reasons for students dropping out from the primary level were lack of financial support, not interested, doing work to support the family and failed examination [4].

In order to decrease the dropout rate, it is important to work on the primary level so that the students can be motivated to continue their education. However, the reasons for the dropout cannot be ignored and any solutions to decrease the dropout rate must address these reasons as well. It is very much clear from the different statistics that we have to work on the primary level of education in order to reduce the dropout

rate. Moreover, recent research shows that educational institutions are in the midst of transformation and becoming more dependent on information technology (IT) to achieve their organizational missions and goals [5]. In fact, artificial intelligence (AI) techniques can be applied in any sector [6-9]. Nowadays researchers use AI techniques as well as optimization algorithms in different fields of research [10-13]. Al the branches of engineering research now required algorithm design and implementation in order to enhance the results [14-18]. Therefore, the implementation of different algorithms or IT related researches would be a good option in the education sector too.

Enrique García et al proposed a model that used data association rule mining to develop web-based courses. These rules were later scored by teachers and experts. The results demonstrate that the proposed architecture constitutes a good starting point for future investigations in order to generalize the results over many course contents [19]. Zanker et. al proposed different strategies to understand user requirements such as knowledge and utility-based methods, collaborative filtering, association rule mining as well as hybrid variants [20]. Masud Ahmed et. al conducted a study to deduce variables that affect the teacher's attitude towards disabled children in the classroom [21]. Data of 738 teachers working on 293 government primary schools were collected. The results indicated that perceived school support for inclusive teaching practices and a range of demographic variables including previous success in teaching students with disabilities and contact with a student with a disability were associated with more positive attitudes of the teachers towards inclusive education. Tissera et al presented an experiment made on an ICT educational institute in Sri Lanka [22]. A series of data mining tasks are applied to find relationships between subjects in the undergraduate syllabus. This provided a considerable amount of insights on the structure of the curriculum syllabus and thus helped to improve the quality of the education. Romero et al conducted a survey on the relevant studies carried out in education [23]. They have described the types of users, types of educational environments and the data they provide. In addition, they have explained in their work the common tasks in the educational environment that have been resolved through data mining techniques. Sun conducted research to understand students learning results [24]. The main idea was to study how the learning processes help the student to do a better result and thus provide a better way for the student to learn. According to this researches, it is being proved that in order to maintain a sustainable education system, the primary dropout needs to be prevented and so the reason for dropping needs to be identified.

The main objective of this research is to find how different conditions like parents' financial conditions, number of siblings and so on affect the education of the students enrolled in the nursery. In order to follow the developed country principle, this research used the data of nursery enrolling students in Slovenia. The research wanted to use the insights of a developed country and use them to improve the situation in a developing country like Bangladesh. Considering all the works mentioned above we decided to use association rule mining in order to deduce the relationships between different factors that affected the enrollment of students into a nursery in Slovenia. Later we used this study to propose to provide insights by using association rules that will represent the current enrollment situation of the primary education students taking into consideration some specific features that play an important role in student dropout from the primary level of education.

2. RESEARCH METHOD

The dataset used in this paper is the open Nursery dataset by Vladislav Rajkovic et al. in the UCI repository [25]. The dataset consists of information on Nursery students in Ljubljana, Slovenia. Here due to a huge number of enrollments of nursery students, the nursery dataset was created in order to help in prioritizing specific students during the enrollment process. The main criteria that were used in this process were the financial standing of the student, family structure or a number of family members and profession of parents. Based on the following criteria, the specific students were enrolled in different institutions. The fields and the following probable values of the datasets are shown in Table 1.

| Table 1. Data field and probable values | | |
|---|---|--|
| Field Name | Probable value | |
| parents | usual, pretentious, great_pret | |
| has_nurs | proper, less proper, improper, critical and very critical | |
| form | complete, completed, incomplete, foster | |
| children | 1, 2, 3, more | |
| housing | convenient, less_conv, critical | |
| finance | convenient, inconvenient | |
| social | non-prob, slightly_prob, problematic | |
| health | recommended, priority, not_recom | |

Table 1. Data field and probable values

Discovering rules for nursery students using apriori algorithm (Mohammad Marufuzzaman)

The parent type is defined as *parent*, information of the child has nursery or specific place as *has_nurs*, form of the family as *form*, no of children in the family as *children*, housing condition as *housing*, financial condition of the family as *financial*, social condition of the family as *social* and finally health condition of the family as *health*. The rule mining steps are shown in Figure 1. There were three major steps the pre-processing of the dataset, the tuple conversion of the data and applying the Apriori algorithm. The steps are discussing next.



Figure 1. Rule mining steps

2.1. Pre-processing dataset

The dataset present in the repository was in the data format. That data was first processed using Microsoft Excel to convert to .csv format. The .csv file was first loaded using the Python library pandas. The existing data was in the form of data frames. The data frame then converted into a list using Python Library. The snippet converted each data in a specific column and converted the column values in a list. This gives each row translational value, set of values that can act as a set of items. This allows the association rule mining to be carried out in the Nursery dataset.

2.2. Tuple conversion to lists

The mining process started using a collected separate python class that is transactions. The transaction class provides the functionalities of converting each set of tuples into lists. Evaluating support, confidence and lift from the dataset thus provided. The generated .csv file is first loaded with pandas (python library). Then an empty list known as transactions is initialized.

2.3. Applying the apriori algorithm

The Apriori algorithm discovers the frequent itemsets from a very large database through a series of iterations. In this research, the *.csv* file is traversed for all the 9410 sets of items. The "rules" function from the Apriori class is then used to generate some association rules with minimum support of 0.04, minimum confidence of 0.95, the minimum lift of 3 and minimum rule length of 2.

3. RESULTS AND DISCUSSION

Using the rules above, the selected itemset was then evaluated with respect to support. Here, a pattern of support with respect to itemset is obtained by generating a bar graph as shown in Figure 2. The number of rules generated from the algorithm was originally 12. Here each rule is represented as a frozen set. Here each frozen set represents a list of items. The basic number of rules obtained after implementing the program for a min_support=0.01 and min_confidence=0.2 was 455. Almost all the obtained item sets had poor confidence value. As a result, the number of rules was shortened by improving the min_confidence to 0.95 and min_support for 0.04.

The rules were again plotted with respect to confidence as shown in Figure 3. The bar graph was obtained with respect to confidentiality. The bar graph provides a detailed illustration of item set with respect to confidentiality. Here we see, the confidence of each itemset is close to each other, as a result, it can be concluded that each itemset is viable for our problem at hand.



Figure 2. Selected itemset with respect to support



Figure 3. Selected itemset with respect to confidentiality

From all the itemset at hand, finally, three itemsets were obtained and considered as most viable and important. The following rules were selected from all the rules:

| Proper nursery <i>λ</i> gret_pret ->convenient (housing) | (Rule 1) |
|--|----------|
| Proper nursery <i>λ</i> inconv(finance) ->gret_pret(parents) | (Rule 2) |
| Proper nursery Λ great pret -> priority(health) | (Rule 3) |

Rule 1 provides us that children having proper nursery rooms and having great pretentious parents also have convenient housing. This proves the items pretentious parents, nursery and housing are of prime importance for children's upbringing. Rule 2 describes, children having a proper nursery but having inconvenient financial status also have a pretentious mentality in parents. Rule 3 gives us, having the proper nursery of children and having pretentious parents have a priority health condition.

The results also provide us with an indication that conditions from the three rules, main ly convenient housing, convenient finance and health priority are the most important factors for the proper upbringing of children. This strong statement is validated by the research results as well. The association rules in this research can later be used to provide much more depth in students, how to manage the environment around them, more insights about the present condition of the students and their educational conditions. Moreover, in the case of education in Bangladesh, the research can be extended into more association rules that can provide us the enrollment condition of students in primary education and obtain the dropout rates.

By doing this research, the specific features that affect student dropouts from primary education are identified. The rules associated with our research are just a few that affect the present condition in the education sector. The research shows the way to discover more factors by extending among students, mainly through studying the different demography, different cultures and finding common ground.

4. CONCLUSION

The rules generated from this research hints that parenting a child in developing countries needs several factors such as proper nursery rooms for children, pretentious behavior in parents and health prioritizing. Sustainable education can be delivered in a developing country like Bangladesh by ensuring these facilities. In this research, it is being proved that convenient or proper housing for children is achieved if a separate nursery for children can be maintained. Moreover, the research shows that even with proper nursery and inconvenient finance great pretentious behaviors in parents exist. Sometimes money is not necessary to carry out the proper upbringing of children. Finally, the researchers predict that with proper nursery and with great pretentious behavior in parents priority health status in children can be achieved which is a positive sign for child health as well as achieve a sustainable education goal.

ACKNOWLEDGEMENTS

The authors would like to express sincere gratitude to Universiti Tenaga Nasional and American International University, Bangladesh for supporting this research project.

REFERENCES

- [1] S. Fredman, J. Kuosmanen, and M. Campbell, "Transformative equality: Making the sustainable development goals work for women," *Ethics & International Affairs*, vol. 30, no. 2, pp. 177-187, 2016.
- [2] Z. Hossain, "Status of secondary school libraries and librarians in Bangladesh," *IFLA Journal*, vol. 45, no. 2, pp. 157-167, 2019.
- [3] Nurul Kabir, "18.8pc dropout in primary education," United News of Bangladesh, [Online], Available: http://www.newagebd.net/article/32556/188pc-dropout-in-primary-education.
- [4] M. Shilpi, S. Hasnayen, T. Ilahi, M. Parvin, K. Sultana, "Education Scenario in Bangladesh: Gender perspective," Bangladesh Bureau of Statistics, UCEP and Diakonia Bangladesh, 2017.
- [5] B. A. Ajayi and H. Hussin, "Conceptualizing Information Technology Governance Model for Higher Education: An Absorptive Capacity Approach," *Bulletin of Electrical Engineering and Informatics (BEEI)*, vol. 7, no. 1, pp. 117-124, 2018.
- [6] M. Marufuzzaman et al, "A time series based sequence prediction algorithm to detect activities of daily living in smart home," *Methods of information in medicine*, vol. 54, no. 3, pp. 262-270, 2015.
- [7] M. Marufuzzaman et al, "A Location Based Sequence Prediction Algorithm for Determining Next Activity in Smart Home," *Journal of Engineering Science & Technology Review*, vol. 10, no. 2, pp. 161-165, 2017.
- [8] I. Matazi et al., "Development of Intelligent Multi-agents System for Collaborative e-learning Support," *Bulletin of Electrical Engineering and Informatics (BEEI)*, vol. 7, no. 2, pp.294-305. 2018.
- [9] M. B. I. Reaz and M. Marufuzzaman, "Pattern matching and reinforcement learning to predict the user next action of smart home device usage," Acta Technica Corviniensis-Bulletin of Engineering, vol. 6, no. 3, pp. 37-40, 2013.
- [10] M. A. Aziz and I. H. Ninggal, "Scalable workflow scheduling algorithm for minimizing makespan and failure probability," *Bulletin of Electrical Engineering and Informatics (BEEI)*, vol. 8, no. 1, pp. 283-290, 2019.
- [11] M. S. Hossain et al., "Passive congregation theory for particle swarm optimization (PSO): An application in reservoir system operation," *International Journal of Engineering and Technology (UAE)*, vol. 7, no. 4, pp. 383-387, 2018.
- [12] T. M. Shin et al., "A comparative study of PSO, GSA and SCA in parameters optimization of surface grinding process," *Bulletin of Electrical Engineering and Informatics (BEEI)*, vol. 8, no. 3, pp. 1117-1127, 2019.
- [13] A. A. Firdaus, O. Penangsang, and A. Soeprijanto, "Distribution Network Reconfiguration Using Binary Particle Swarm Optimization to Minimize Losses and Decrease Voltage Stability Index," *Bulletin of Electrical Engineering and Informatics (BEEI)*, vol. 7, no. 4, pp. 514-521, 2018.
- [14] M. Marufuzzaman and M. B. I. Reaz, "Hardware simulation of pattern matching and reinforcement learning to predict the user next action of smart home device usage," *World Applied Sciences Journal*, vol. 22, no. 9, pp. 1302-1309, 2013.
- [15] M. A. Hasan, M. D. Mamun, and M. Marufuzzaman, "Hardware approach of a novel algorithm of r-peak detection for the simultaneous measurement of fetal and maternal heart rates during pregnancy," *Revue Roumaine des Sciences Techniques Serie Electrotechnique et Energetique*, vol. 57, no. 4, pp. 432-443. 2012.
- [16] M. Marufuzzaman et al., "August. Classification and detection of intelligent house resident activities using multiagent," in Proc. of the 4th Int. Conf. on Computing and Informatics, pp. 195-200, 2013.
- [17] M. Marufuzzaman et al., "Hardware approach of two way conversion of floating point to fixed point for current dq PI controller of FOC PMSM drive," *Elektronika ir Elektrotechnika*, vol. 123, no. 7, pp. 79-82, 2012.

- [18] A. H. Al-Hamami, and A. A. Flayyih, "Enhancing Big Data Analysis by using Map-reduce Technique," *Bulletin of Electrical Engineering and Informatics (BEEI)*, vol. 7, no. 1, pp. 113-116, 2018.
- [19] E. García et al., "An architecture for making recommendations to courseware authors using association rule mining and collaborative filtering," *User Modeling and User-Adapted Interaction*. vol. 19, no. 1-2, pp. 99-132, 2009.
- [20] M. Zanker and M. Jessenitschnig, "Case-studies on exploiting explicit customer requirements in recommender systems," User Modeling and User-Adapted Interaction: The Journal of Personalization Research, A. Tuzhilin and B. Mobasher (Eds.): Special issue on Data Mining for Personalization, Springer, vol. 19, no. 1-2, pp. 133-166, 2009.
- [21] M. Ahmmed, U. Sharma and J. Deppeler, "Variables affecting teachers' attitudes towards inclusive education in Bangladesh," *Journal of Research in Special Educational Needs*, vol. 12, no. 3, pp. 132-40, 2012.
- [22] W. M. R. Tissera, R. I. Athauda and H. C. Fernando, "Discovery of Strongly Related Subjects in the Undergraduate Syllabi using Data Mining," 2006 International Conference on Information and Automation, Shandong, pp. 57-62, 2006.
- [23] C. Romero and S. Ventura, "Educational Data Mining: A Review of the State of the Art," in *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, vol. 40, no. 6, pp. 601-618, Nov. 2010.
- [24] H. Sun, "Research on student learning result system based on data mining," International Journal of Computer Science and Network Security, vol. 10, no. 4, pp. 203-205, 2010.
- [25] M. Olave, V. Rajkovic, and M. Bohanec, "An application for admission in public school systems," *Expert Systems in Public Administration*, vol. 1, pp. 145-160, 1989.

BIOGRAPHIES OF AUTHORS



Dr. Mohammad Marufuzzaman was born in Bangladesh in June 1983. He received his Ph.D. and M.Sc. degree in Electrical, Electronic and Systems Engineering from The National University of Malaysia (UKM), in 2017 and 2012 respectively. He was worked as an assistant professor at the Computer Science Department at American International University, Bangladesh (AIUB). Currently, he is working as a Post-Doctoral Researcher at the Institute of Energy Infrastructure, The National Energy University (UNITEN) in Malaysia. His research interest includes Algorithm Design, AI, Smart Homes, wireless networking, data communication, Embedded System Design and Hardware Realization. He is author and co-author of more than 50 research articles and reviewed numerous articles throughout his research career.



Dipta Justin Gomes was born on 29 March 1994 in Dhaka, Bangladesh. He has completed his B.Sc in Computer Science in 2018 and MSc in Computer Science in 2019 from American International University Bangladesh (AIUB). Currently, he is working as a lecturer at the Department of Computer Science at American International University-Bangladesh (AIUB). His research interest includes machine learning, data mining and image processing. His research interest includes Graph Theory Algorithms, Computer Vision, Machine Learning and Image Processing.



Aneem Al Ahsan Rupai was born on 16 January 1996 in Bangladesh. He has completed his B.Sc. and M.Sc. in Computer Science from American International University Bangladesh (AIUB). Currently, he is working as a lecturer at the Department of Computer Science at American International University-Bangladesh (AIUB). His research interest includes machine learning, data mining and image processing.



Datin Prof. Ir. Dr. Lariyah Mohd Sidek is a Professor of Civil Engineering Department at Universiti Tenaga Nasional (UNITEN). She has actively involved in research related to stormwater & urban drainage. She is one of the inventors of an innovative sustainable urban drainage system namely Bio-Ecological Drainage Systems (BIOECODS) and Principal Investigator for GPTs Expert System called DeGPT She is a Professional Engineer since 2003 (P. Eng. No.: 11742) and a corporate member of IEM (M25289) and a panel expert of detailed environmental impact assessment (DEIA) with DOE. She currently holds the position of Co-President in the Malaysian National Committee on Large Dams (MNCOLD). She is the author of more than 100 research papers and reviewed numerous research articles throughout her career.