Predicting Students' Academic Performance: A Review for the Attribute Used

Ismail Yusuf Panessai[1], Muhammad Modi Lakulu[2], Mohd Hishamuddin Abdul Rahman[3], Noor Anida Zaria Mohd Noor[4], Nur Iksan[5], Roznim Mohamad Rasli[6], Mohd Razimi Husin[7], Hishamuddin Ahmad[8], Shamsul Arrieya Ariffin[9], Asmara Alias[10], Siva Kumar Subramaniam[11], Sanusi Majid[12], M. Ansyar Bora[13], Aldrin Aran Bilong[14], Nur Maisyara Mazli[15]

[1,2,3,4,6,9,10,15]Department of Computing, Faculty of Art, Computing and Creative Industry. Universiti Pendidikan Sultan Idris, Tanjung Malim, Perak, Malaysia, [5]Department of Artificial Intelligence, Lamintang Education & Training Centre. Indonesia, [7,8]Department of Educational Studies, Faculty of Human Development. Universiti Pendidikan Sultan Idris, Malaysia, [11]Faculty of Electronics and Computer Engineering, Universiti Teknikal Malaysia Melaka, Malaysia, [12,13]Fakultas Teknik, Universitas Ibnu Sina, Batam. Indonesia, [14]Etiqa Life Insurance Berhad, Taman Bukit Indah, Johor Bahru, Johor, Malaysia
Email: ismailyusuf@fskik.upsi.edu.my

Abstract
This article is a review on attributes, models and tools used to solve the problem of predicting students' academic performance. Based on the reference throughout 2009, the attributes used include demographic, academic grades, school related and social attributes. The academic score attributes are the results of the percipient along their studies, while the school-related attributes are result for few subject taken in high school. Demographic and social attributes are family background and the daily interactions of respondents. These tools are supporting a lot of rule mining algorithms like clustering, classification and association, to use on datasets of different types.
Keywords: Lean Healthcare Management System, Healthcare Management, Healthcare Industry

Introduction
The focus objectives of the study are predicting the students' academic performance, CGPA, grades, score and performance (not related to education). There are 28 articles that are related to this study. However, the articles collected different type of data set to carry out the prediction process. In this subtopic, these articles will be discussed more specifically.

First of all, the articles with objective of predicting the students' academic performance. Articles by Czibulaa, Mihaia & Maria (2019); Almasri, Celebi & Rami (2019); Umar (2019); Graham et al., (2019); Yass, Abdoulaev & Mukesh (2019); Alarape& Adewole (2019); Yusuf & John (2019); Alsuwaiket, Mohammed & Blasi (2019); Anas & Ra'Fat (2019); John Lekan, Akinode. (2019); Tiwari, Sharma & Kashyap (2019); Thingbaijam & Lenin (2019); Editha (2019); Ranjeeth, Latchoumi, Paul (2019); Sana, Siddiqui, & Arain (2019). Harvey, Julie & Sathish (2019); Solanki & Amita (2019); Yaacob & Sobri (2019); Kumar & Salal (2019) and Panessai et. al (2019) has the same main objective. This objective can be achieved by using different types of data collected from data set or sample, based on different projects. There are few learning algorithm used as mentioned in Hachem & Alaou (2019); Anuraag, et al (2020); Yan & Abas (2020); Kangungu & Yatim (2020); Jianan Abas (2020); Maulana & Normalisa (2019); baba et al (2014); Desai & Singh(2020); Yusuf, et al, (2010); and Yusuf, et al, (2010).

Attribute Used

For article [1], the data set collected for the project is the grades got from the Computer Science undergraduate courses students of the chosen university in the research that got offered in the first, second and third semesters. Then, for article [2], data set used included the attributes that has information of preregistration students from a registration office. While for article [3], it took 61 students dataset of 2018 academic year that was obtained from the Department of Networking and System Security in KSITM.

Table 1
Summary of selected students attributes

S/N	Attributes	Data Type
1	English Score	Float
2	Subject 2 Score	Float
3	Subject 3 Score	Float
4	Subject 4 Score	Float
5	English Grade	String
6	Subject 2 Grade	String
7	Subject 3 Grade	String
8	Subject 4 Grade	String
9	First Year CGPA	Float
10	Predicted Class of Graduation	Nominal

It contains some students' information and academic data such as Grade Point Average of first semester (as GPA_1) which is numeric variable and number of subjects failed from preceding semester. Article [4] took participants consist of undergraduate students that had been studying Engineering or Psychology, to complete answering some of the long-term memory tests for the process of collecting the data needed. For the successful of project of article [5], it used attributes that consist of some instances that also included the student grades, social, demographic and school related attributes. As for article [6], to predict the students' academic performance, it used the data set that includes attributes of the students such as demographic, academic grades, school related and social attributes. The article [8] used all the attributes from Table 1.

Other than that, article by Mohammed, et.al. (2019) used student records. For each record, a number of attributes that represent a student's academic accomplishments during three levels which are preparatory, 1st, and 2nd levels.

For article by Akinode (2019), some amount of data are used as variables which are English Result of the High school terminal Exam, Mathematics Result of the High school terminal Exam, Physics Result of the High school terminal Exam, Biology Result of the High school terminal Exam, Agric Agriculture Result of the High school terminal Exam, Economics Result of the High school terminal Exam, Chemistry Result of the High school terminal Exam, 1st Semester result of a four semesters program, 2nd Semester result of a four semesters program and 3rd Semester result of a four semester.

Besides, to create algorithm for predicting students' academic performance in Tiwari, et.al (2019), only student's academic data needed with other existed data algorithm. Other than that, article by Baijam & Lenin (2019), a total of 79 observation were selected consist of 11 different attributes. While, only need engineering students' final grades for the prediction data set (Editha, 2019). Another article, by Ranjeeth, et.al. (2019), collecting data by asking 42 different questions as shown in Table 2.

Table 2
Parameter list

Q1. Medium	Q22. Refer notes after school
Q2. Gender	Q23.Finishing test in time
Q3. Mother Education	Q24.Part of education feel stress
Q4. Father Education	Q25. Are you day scholar or hosteller?
Q5 Sibling Education	Q26. Distance from school to home?
Q6. Caretaker at home	Q27. Mode of transport
Q7. Caretaker at hostel	Q28. Financial problem
Q8. Care taker time spent	Q29.Coeducation disturbance?
Q9. Impact of advisor	Q30. Impact of friends
Q10. The reaction of parents on low marks	Q31. Number of times schools changed
Q11. The reaction of parents on high marks	Q32. Time spend on sports
Q12. Communication with family	Q33. Participation in extracurricular activities
Q13. Number of hours spent on study	Q34. Type of games interested
Q14.Time spend on tuition	Q35. Time spend with mobile
Q15.Number of tests conducted per week	Q36. Time spend with computer
Q16. Ask questions in class	Q37. Time spend on TV
Q17. Place prefer to study	Q38. Health issues
Q18. Types of study	Q39. Goal
Q19. Days before start preparation for exams	Q40. Physical exercises/ yoga
Q20. Time interval chooses for study	Q41. Role model
Q21. Making notes for study	Q42. Scored GPA

For the next article, in Siddiqui & Arain (2019), demographical features, academic background features, participation of parents on the whole learning process and parent behavioral feature of 500 students were collected. Continue with Julie & Sathish (2019) where they need data consist of 27 variables that can be refer at Table 3.

Table 2
Variables used in data analysis

Variable Abbreviation	Variable Description
% African American	Percentage of African American students
Average Salary	Average salary of the teachers
Avg_dist_exp	Average expenditure per student by district
Avg_exp	Average expenditure per student by school
% Non-Grad Completers	Percentage of student that did not graduate
% Males	Percentage of males
% Dropped Out	Percentage of student who dropped out of school
% GED	Percentage of students who completed a GED
% MA Community College	Percentage of students attending a MA community college
Econ_disadvantage	Number of students with an economic disadvantage
% Hispanic	Percentage of Hispanic students
% Attending College	Percentage of students attending college
% Private Four-Year	Percentage of students at a private four-year college
% Public Four-Year	Percentage of students at a public four-year college
% Graduated	Percentage of students who graduated from the school
% White	Percentage of White students
% AP_Score_3-5	Percentage of students with AP score of 3-5
Average SAT_Reading	Average SAT Reading score of students
Average SAT_Math	Average SAT Math score of students
% Females	Percentage of female students
Average Class Size	Average size of class in school
% Asian	Percentage of Asian students
TOTAL_Enrollment	Total number of students in attendance of the school
SAT_Tests_Taken	Number of students who took the SAT
First Language Not English	Students whose first language is not English
High Needs	Student requiring additional education accommodations
Economically Disadvantaged	Number of students with an economic disadvantage

Yaacob et al (2019) predictive task had been carried out for students to collect the data. To predict student performance, we can also use students' academic information such as article in Kumar & Salal (2019) that collect data from a total number of transcripts of the year 2013 to 2016 from the students who had completed their degrees in academic.

For article Kumar & Salal (2019), it shows the review on several other articles and conclude that crucial factors to predict academic performance of students are personal attributes, family attributes, social attributes, student attributes, academic attributes, and school attributes.

Lastly, on article panessai, et. al. (2019) the researcher predicts the academic performance by using the data collected from the transcript of students from Universiti Pendidikan Sultan Idris that majoring in Software Engineering Program. This data is from the year of 2015.

The attributes for the data set is shown in Table 4.

Table 4
List of Features Used in the Study

Attributes	Type	Values
Student ID	Alphanumeric strings	
Pre-University academic qualification type	Varchar	
CGPA	Floating-point	[0.00-4.00]
Total of coursework mark	Floating-point	[0.00-100.00]
Final Grade	Varchar	[A, B, C, D, E, F]

The next objective that going to be discussed is predicting the students' academic CGPA (Aderibigbe & Odunayo, 2019). This article obtained the data needed by collecting the first three academic years GPA data of some students and their final CGPA from the year of 2002 to 2014, from some engineering departments to achieve the objective.

For the next objective is predicting the students' grades based on articles Sara & George (2019), Gamie, Abou El-Seoud, Salama & Hussein (2019), Dragana & Gabriela (2019), Álvaro & Bruno (2019). In the article Sara & George (2019), the grade prediction is carried out, by learning the weights of the prior courses towards predicting the grade of each target course. As for article Gamie et.al (2019), the grade prediction was made to identify the relationship between the multiple of inputs in the education procedure and the student's performance. Thus, the data of students used are students' attendance and grade, number of course login, and school leaving grade. Moreover, for article Dragana & Gabriela (2019), biographical essays are collected as the data used for prediction as its aim is to predict students' English grades that they had achieved for courses with Specific Purposes. Last article, Álvaro & Bruno (2019) for this objective used data that achieved from through an Analysis of logged data of online interactions.

As for the other main objective, which is predicting the students' academic score are related to articles Do, et.al. (n.d.), and Wang, et.al. (2019). the articles collect different data as article Do, et.al. (n.d.) focussed to predict score obtained from incomplete and optional courses in the third-fourth year that need data of students' grades. While article Wang, et.al. (2019), focussed on predicting score of the optional course which need the course-score records of the students to carry out the prediction.

The last objective for this subtopic is predicting the performance based on Meng, Jian, Guoxi & Kai. (2019). It is quite different from the first objective as it might not related to academic at all. To achieve this objective mentioned, data that had been used are those values that were obtained from product data management (PDM) system, which are the configuration parameters and performance parameters of each product.

Conclusions

One of the important things that need to be considered in solving the problem of Student Academic Achievement Prediction is Attributes.

The attributes used include demographics, academic grades, school-related and social attributes. The academic score attributes are the results of semester 1, semester 2 and semester 3, while the school-related attributes are high school English test results, high school mathematics results, high school physics exam results, high school biology exam results, agricultural test results, high school exams, test results Middle School Economics, High School Chemistry Exam Results. Demographic and social attributes are Gender, Mother's & Father's Education, Sibling Education, Financial problems, Advisory impact, parental reactions to low / high scores, Time spent on sports & games even on the type of play you are interested in.

Acknowledgements

References

Abdulazeez, Y., & Ayuba, J. (2019). Classifiers ensemble and synthetic minority oversampling techniques for academic performance prediction. International Journal of Informatics and Communication Technology (IJ-ICT), Vol.8, No.3.

Abou, G., Abou, E. -S., Salama, S. M., & Hussein, W. (2019). Multi-Dimensional Analysis to Predict Students' Grades in Higher Education. International Journal of Emerging Technologies in Learning (IJET), 14(02), 4. https://doi.org/10.3991/ijet.v14i02.9905

Aderibigbe, I. A., & Odunayo, S. (2019). The impact of engineering students' performance in the first three years on their graduation result using educationaldata mining. Heliyon. Volume 5, Issue 2, Article e01250.

Alarape, M. A., & Adewole, K. S. (2019). Students' Academic Performance and Dropout Prediction. Malaysian Journal of Computing, 4(2), 278. https://doi.org/10.24191/mjoc.v4i2.6701

Alsuwaiket, M., Blasi, A., & Al-Msie'Deen, R. F. (2019). Formulating module assessment for improved academic performance predictability in higher education. Engineering, Technology and Applied Science Research, Vol. 9.

Álvaro, F., & Cabral, B. (2019). Preventing failures by predicting students' grades through an analysis of logged data of online interactions. 10.5220/0008356604910499.

Ammar, A., Erbug, C., & Rami, S. A. (2019). EMT: Ensemble Meta-Based Tree model for predicting student performance. Scientific Programming. 10.1155/2019/3610248.

Anuraag, M. A., Masood, & Mohanchandra, K. (2020). Brain Operated Wheelchair Using a Single Electrode EEG Device and BCI. International Journal of Artificial Intelligence, 7(1).

Baba,M. S., Panessai, I. Y., & Iksan, N. (2019). Solving Rich Vehicle Routing Problem Using Three Steps Heuristic. International Journal of Artificial Intelligence, 1(1).

Bozic, L., Dragana & Chmelíková, G. (2019). Applying predictive analytics in ESP courses based on students' writing. Zbornik Radova Filozofskog Fakulteta u Prištini, Volume 49.

Desai, SR, S., & Singh, M. (2020). Survey on Early Detection of Alzhiemer's Disease Using Capsule Neural Network. International Journal of Artificial Intelligence, 7(1).

Gabriela, C., Andrei, M., & Liana, M. (2019). S PRAR: A novel relational association rule mining classification model applied for academic performance prediction. Procedia Computer Science, Volume 159, 20-29.

Graham, P., Patricia, B. M., Claudio, E. M. G., Ana, F. T., & Angélica, M. U. A. (2019). Modulation of striatum based non-declarative and medial temporal lobe based declarative memory predicts academic achievement at university level. Trends in Neuroscience and Education, Volume 14, 1-10.

Harvey, J., & Kumar, S. (2019). A practical model for educators to predict student performance in k-12 education using machine learning. 3004-3011.10.1109/SSCI44817.2019.9003147.

Huang, L., Wang, C. -D., Chao, H. -Y., Lai, J. -H. & Y, P. S. (2019). A score prediction approach for optional course recommendation via cross-user-domain collaborative filtering. IEEE Access, 7, 19550–19563. https://doi.org/10.1109/access.2019.2897979

Imran, M., Shahzad, L., Danish, M., & Shah, M. S. (2019). Student academic performance prediction using supervised learning techniques. International Journal of Emerging Technologies in Learning, Vol 14, 14.

Jianan, L., & Abas, A. (2020). Development of Human-Computer Interactive Interface for Intelligent Automotive. International Journal of Artificial Intelligence, 7(2).

Jorda, E. (2019). Predictive model for the academic performance of the engineering students using CHAID and C5.0 algorithms. International Journal of Engineering Research and Technology, Volume 12, Number 6, pp. 917-928.

Kangungu, S. M., & Yatim, M. H. (2020). Teaching Programming Using the Robot-Based Learning Approach. International Journal of Artificial Intelligence, 7(2).

Kumar, M., & Salal, Y. K. (2019). Systematic review of predicting student's performance in academics. International Journal of Engineering and Advanced Technology (IJEAT) ISSN: 2249 – 8958, Volume-8 Issue-3, February 2019.

Lekan, A. J. (2019). Student academic performance prediction using support vector machine. International Journal of Engineering Sciences & Research Technology, Vol.6.

Maulana, I., & Normalisa. (2019). Penerapan Algoritma Knuth-Morris-Pratt pada Fungsi Pencarian Dokumen untuk Sistem Informasi Administrasi Sekolah Berbasis Website. International Journal of Artificial Intelligence, 6(1).

Morsy, S., & Karypis, G. (2019). Sparse Neural Attentive Knowledge-based Models for Grade Prediction.

Mubarak, A. U. (2019). Student academic performance prediction using artificial neural networks: A case study. International Journal of Computer Applications. 10.5120/ijca2019919387

Panessai, I. Y., Lakulu, M. M., Abdul, M. H. R., Noor, N. A. Z. M., Salleh, N. S. M., & Bilong, A. A. (2019). PSAP: Improving Accuracy of Students' Final Grade Prediction using ID3 and C4.5. International Journal of Artificial Intelligence, 6(2), 125–133. https://doi.org/10.36079/lamintang.ijai-0602.42

Ranjeeth, S., Latchoumi, T. P., & Paul, P. V. (2019). Optimal stochastic gradient descent with multilayer perceptron based student's academic performance prediction model. Recent Advances in Computer Science and Communications, vol. 13.

Salal, Y., Abdoulaev, S., & Kumar, M. (2019). Educational Data Mining: Student performance prediction in academic. International Journal of Engineering and Advanced Technology, Volume 8, Issue 4C.

Sana, Siddiqui, I. F., & Arain, Q. A. (2019). Analyzing students' academic performance through educational data mining. 3C Tecnología. Glosas de innovación aplicadas a la pyme. Special Issue, May 2019, pp. 12–43. doi: http://dx.doi.org/10.17993/3ctecno.2019.

Solanki, S., & Dhankhar, A. (2019). Predicting student's performance by using classification methods. International Journal of Advanced Trends in Computer Science and Engineering, 1532–1536. https://doi.org/10.30534/ijatcse/2019/75842019

Thingbaijam & Lenin. (2019). Student's performance prediction modelling using classification technique in R. doi:10.35940/ijrte.B3259.078219

Thong, D. M., Thanh, C., Minh, T., & Nam. (n.d.). An apache spark-based platform for predicting the performance of undergraduate student. Retrieved from https://www.researchgate. net/publication/335219629_The_Performance_of_Undergraduate_Student

Tiwari, A. K., Ramakrishna, G., Sharma, L. K., & Kashyap, S. K. (2019). Academic performance prediction algorithm based on fuzzy data mining. IAES International Journal of Artificial Intelligence (IJ-AI), 8(1), 26. https://doi.org/10.11591/ijai.v8.i1.pp26-32

Yaacob, W. F. W., Nasir, S. A. M., Yaacob, W. F. W., & Sobri, N. M. (2019). Supervised data mining approach for predicting student performance. Indonesian Journal of Electrical Engineering and Computer Science, 16(3), 1584. 10.11591/ijeecs.v16.i3.pp1584-1592

Yusuf, I., Iksan, N., Herman, N. S. (2011). The application of genetic algorithms in designing fuzzy logic controllers for plastic extruders. Lecture Notes in Electrical Engineering, pp. 25–37

Yusuf, I., Iksan, N., Herman, N. S. (2010). Weight-feeder control for plastic extruder using fuzzy genetic algorithms. 2010 The 2nd International Conference on Computer and Automation Engineering, ICCAE 2010, 2010, 3, pp. 145–149.

Yan, X., & Abas, A. (2020). Preliminary on Human Driver Behavior: A Review. International Journal of Artificial Intelligence, 7(2), 29-34. https://doi.org/10.36079/lamintang.ijai-0702.146

Zhang, M., Zheng, J., Li, G. & Zhang, K. (2019). Performance prediction of a modular product variant with RS-SVM. Procedia CIRP, Vol. 79.