CASE STUDY RESEARCH ASSIGNMENT

DATA MINING IN EDUCATION

Can Data Predict Your Grade?

Program Design Methods

Ir. TRI ASIH BUDIONO, M.I.T.

Bayu Hartho	2502013731
Carmen Cleosa	2502009601
Darryl Hilmy	2502009652
Ian Wirawan	2502009596
Kimberly Mazel	2502022250
Sandrian Wardana	2502016411

L1BC
Department of Computation and Media
Binus International University
November 2021

Table of Contents

Background

Data mining is essentially the analysis of large datasets, which involves the process of extracting and discovering anomalies, patterns, correlations and other valuable information from data points within the large dataset. Data mining is also known as “knowledge discovery in databases” and the concept itself has been around for over a century but the term “data mining” was only coined in the 1990’s. In today’s world, data mining is a field of computer science which combines the fields of statistics, artificial intelligence, neural networking and machine learning. Its purpose is to identify trends and extract information strictly in large datasets which would have otherwise remained hidden or overlooked.

In recent years, data mining has only gotten increasingly popular and prevalent alongside the growing prevalence of big data. This is due to datasets being too large and complex to analyse manually and therefore rely on machine learning algorithms in order to help analyse and extract meaningful information from them. This is useful as it significantly speeds up the process in which meaningful insights can be extracted from ever-growing datasets, especially as technologies improve the rate at which data and information is generated.

The process of data mining is generally done in standardized steps to ensure the quality of the extracted data is meaningful and accurate, choosing carefully and wisely the input datasets to be used for analysis, as bad data will only yield poor results:

1. **Business Understanding** - Understanding the objectives and parameters of the data mining project and setting what is appropriate in order to initiate and run a fruitful project.
2. **Data Understanding** - Determining the types of data which will be involved and carefully considering the source of required data.
3. **Data Preparation** - Preparing the data to be analysed, removing and fixing any issues such as duplicate or missing data. Also optionally reducing the depth or complexity of the dataset if deemed to be too computationally intensive.
4. **Modelling** - This is the most interesting step of the process where algorithms will be used to analyse and identify any meaningful relationships or deviations in the data.
5. **Evaluation** - Reviewing and interpreting the information gathered from the modelling phase and determining whether or not it will be useful and whether or not the goal of the project has been achieved. Multiple different algorithms may be used and determining which one is best suited or may have achieved the best results.
6. **Deployment** - Presenting the results and implementing solutions based on the data gathered.

Data mining can be applied and is useful in various industries such as Tech, Business, Education, Manufacturing, and many more. In tech, it can be used for analytics and to predict future outcomes and trends to better prepare for changes in the future. In business, it can help in sales forecasting and to help in adapting marketing strategies to better suit their target audiences or to expand their reach. As for education, it can be used in order to identify behavioural trends in students which

could help adapt teaching techniques and develop new methods in order to adjust and improve learning to be better suited to more students. There are many possible applications to data mining and it will only ever increase as the amount of scenarios where big data is applied increases as well.

Problem

Without the valuable data of a student's activity, the students won't have any way of knowing whether they are lacking in a course unless they take note of the grades they got and how well they are doing in the course themselves or are confronted by the lecturer themselves because of bad grades. Students may be unaware if they had a failing grade until it's too late, and it may cause anxiety when the student realizes this.

Data mining doesn't come without problems and flaws, with the amount of students attending university and the amount of courses and activities the university provides, the amount of data might be too much to handle and may strain the server's integrity which might cause glitches, high demand in power, and might damage the data. Even with the massive amount of data gathered every day, it would be near impossible to predict a student's scores with 100% accuracy, because there is still the probability of the student not attending a class or an exam, and it also depends on the student's behaviour and influence without monitoring student's activities 24/7. Which would lead to the lack of anonymity the system might have, because the university staff are monitoring every single one of each student's activities and grades.

Short literature review

Several research studies have identified many benefits and applications of data mining in education. Romero and Ventura (2010) have identified student's modelling, predicting students' performances, data visualization, social network analysis, feedback for support management, planning and scheduling, grouping students and detection of undesirable behaviors as applications of educational data mining. Many studies such as those done by Baradway and Pal (2012), Kabackijeva (2012), Mardikyan and Badur (2012), and Hsia et al. (2008) used decision trees as a support tool for classification of big data.

Support tools such as decision trees and rule learners allow big data to be classified and evaluated. This can then be used in producing predictive models and identifying groups of students. In order to guide students and instructors, the students' behavior needs to be understood. Without data mining, there is no use of collecting data as useful information has not yet been extracted and cannot be used to make decisions and conclusions. Multiple universities have taken advantage of data mining in order to enhance their educational system for their students. Purdue University and Austin Peay State University are reputable examples.

Purdue University's Course Signals make use of a student's grades, demographic characteristics, past academic history and effort to predict their performance. Instructors can then provide real-time feedback to a student without having to spend time analyzing each student's data on their own. Course Signals present the information through a stoplight system. A red light would indicate a high likelihood of failure; yellow indicates a possible problem while green indicates a high

likelihood of success within the course. Research indicates that there was a noticeable increase in satisfactory grades in courses that implement Course Signals, and a decrease in dissatisfactory grades and withdrawals. Thus, Purdue University has demonstrated that learning analytics provided by data mining has favorably impacted their students' success.

Austin Peay State University has successfully applied data mining in education through their course recommendation system, Degree Compass. Tristan Denley (2012) made use of big data from past students' grades and transcripts to pair current students with courses that are deemed best fitting to them. Degree Compass uses predictive analytics techniques to rank courses according to measured factors. Courses that apply directly to the student's program of study is ranked and modelled to predict the courses that the student will excel the most in. Additionally, Degree Compass provides strategic reports that are used by department advisors. These reports allow targeted interventions, for instance, identifying students that would benefit from mentoring. The project had shown a significant difference between the grade distributions in students who used the Degree Compass to select their courses and students who did not. Tristan Denley (2012) has proven that data mining successfully guided both staff and students in their education journey.

Findings and Discussions

Web-based tools provide an integrated environment to support both educators and learners alike with their needs via the Internet. Simply, a web based tool is a software that is used and utilized on a browser for specific purposes. Its name can also be referred to as a web based application or program. These software run on web servers which basically means that all you need is an internet connection and a browser, and you do not need to install anything into your computer. The purposes of a web based tool can vary from different things; it can come down to needing that certain functionality for a more convenient experience.

Examples of web based tools/applications include productivity tools such as Google Docs, Excel, and Slides. These productivity web based tools help more users to be able to work and collaborate on any device. On the business side, it can include the shopping cart function in e-stores, giving companies and businesses more ease of access to information between customers. Another example is Expert Electronic Coach (E2 Coach) which is a computer tailored communication system created by the University of Michigan. E2 Coach aims to provide personalized feedback and advice to all students in large introductory STEM courses. The coaching focused on improving study techniques and habits, providing encouragement at appropriate times, and providing matched advice from other students.

As previously stated, a benefit of web based tools is the convenience; these tools can be used on any device as long as it supports a browser and uses an

internet connection. This would allow users to work anywhere and anytime. For example, a student that needs to work on an essay. All they would need is their laptop and an internet connection to be able to fulfil their task.

In addition, web based tools use less resources and are easy to run. A traditional desktop application would require installation, therefore requiring space and would only be functional on one device. A web based tool does not require any space and runs on any device that has a browser; no installations or updates needed. Every user will have the same version as it would be based on the same server.

In the case of E2 Coach, students can figure out their strengths and weaknesses of the courses they are interested in. Surveys are done on their study habits such as how long it would take for them to study alone or with a partner. E2 Coach is beneficial to those who seek to improve their grades. Research results suggest that E2 Coach usage was not only positively associated with student performance, but even accounted for when the student had performed better than expected from their mathematics and physics backgrounds. This is similar to what Purdue University and Austin Peay State University had done with their respective methods.

Though E2 Coach has its functionalities, it may not be desirable for all students. Some may think of E2 Coach as a breach of privacy as one of the factors needed for E2 Coach to function is personal information on the students' lifestyles. It monitors any changes in their lifestyle that may affect their learning experience which may be interpreted as invasive. E2 Coach makes assumptions based on data collected by the

survey and does not always promise a good outcome. At the end of the day, no matter how efficient E2 Coach is, changes will only apply to the students' grades and not their lifestyles. In order for positive changes to happen, students need to take it into their own hands and alter the way they work.

In addition, web applications usually require a monthly or annual payment as there is a cost to running the websites that host the web applications. Security may not be the best as data stored online would be more susceptible to hackers compared to when it is stored physically on a hard drive. Although there are methods to reduce the risk, for example, using SSL enforcement to protect the server and client from attacks, there is no guarantee that the data is completely safe.

Conclusions and Recommendations

Conclusively, data mining has been proven to be an effective method when utilized in an educational context. The use of technology grows every day and in turn, generates a growing amount of data. Data mining allows this raw data to be turned into useful information that can be analyzed and evaluated to predict future outcomes, such as a student's performance. Data mining in education greatly contributes to development methods such as extracting interpretable and useful information that can lead to a better understanding of both the students and the instructors. Educational data mining can be applied in many different areas such as identifying at-risk students, guiding students in course choices and providing models and patterns, effectively resulting in more substantial decision making and preventing further predicaments. It is clear that educational data mining promises a successful future.

Data should always be collected with the consent of the user. It is common to see cookie pop-ups on websites that inform the user that data is being collected for their algorithm and search recommendation engines. There should be transparency between the users and the data collectors. The users should be fully aware about what data is being collected and what it is collected for. There are certain data privacy laws that the collectors would have to abide to in order to gain the users' trust. As long as every principle is followed, concerns regarding privacy invasion can be diminished.

References

Admin. (2021, September 7). *What are the advantages and disadvantages of web applications?* iTrobes. Retrieved November 21, 2021, from https://www.itrobes.com/what-are-the-advantages-and-disadvantages-of-web-applications/.

Algarni, A. (2016). Data Mining in Education. *International Journal of Advanced Computer Science and Applications*, *7*(6). https://doi.org/10.14569/ijacsa.2016.070659

Arnold, K. E., & Pistilli, M. D. (2012). Course signals at Purdue. *Proceedings of the 2nd International Conference on Learning Analytics and Knowledge - LAK '12*. https://doi.org/10.1145/2330601.2330666

By: IBM Cloud Education. (n.d.). *What is data mining?* IBM. Retrieved November 21, 2021, from https://www.ibm.com/cloud/learn/data-mining.

Education analytics. SAS. (n.d.). Retrieved November 21, 2021, from https://www.sas.com/en_us/industry/education.html.

Encyclopædia Britannica, inc. (n.d.). *Data Mining*. Encyclopædia Britannica. Retrieved November 21, 2021, from https://www.britannica.com/technology/data-mining.

Huberth, M., Chen, P., Tritz, J., & McKay, T. A. (2015). Computer-tailored student support in introductory physics. *PLOS ONE*, *10*(9). https://doi.org/10.1371/journal.pone.0137001

Oblinger, D. (2012). *Game changers: Education and information technologies*. EDUCAUSE.

Roomi, M. (2021, April 17). *5 advantages and disadvantages of web application: Drawbacks & benefits of web application*. HitechWhizz. Retrieved November 21, 2021, from https://www.hitechwhizz.com/2021/04/5-advantages-and-disadvantages-drawbacks-benefits-of-web-application.html.

What is a web application? how it works, benefits and examples. Indeed Career Guide. (n.d.). Retrieved November 21, 2021, from https://www.indeed.com/career-advice/career-development/what-is-web-application.

What is data mining? definition and examples. Talend. (n.d.). Retrieved November 21, 2021, from https://www.talend.com/resources/what-is-data-mining/.

What is data mining? SAS. (n.d.). Retrieved November 21, 2021, from https://www.sas.com/en_us/insights/analytics/data-mining.html.

Zorić, A. B. (2020, November 30). *Benefits of educational data mining*. Research leap. Retrieved November 21, 2021, from http://dx.doi.org/10.18775/jibrm.1849-8558.2015.61.3002.