Visual Analytics Design for Students Assessment Representation based on Supervised Learning Algorithms

Adlina Abdul Samad, Marina Md Arshad, Maheyzah Md Siraj, Nur Aishah Shamsudin
School of Computing, Faculty of Engineering,
Universiti Teknologi Malaysia
81310 UTM Johor Bahru, Malaysia
Email: adlina6@graduate.utm.my; marinama@utm.my; maheyzah@utm.my; nuraishahshamsudin97@gmail.com

Abstract—**Visual Analytics is very effective in many applications especially in education field and improved the decision making on enhancing the student assessment. Student assessment has become very important and is identified as a systematic process that measures and collects data such as marks and scores in a manner that enables the educator to analyze the achievement of the intended learning outcomes. The objective of this study is to investigate the suitable visual analytics design to represent the student assessment data with the suitable interaction techniques of the visual analytics approach. sheet. There are six types of analytical models, such as the Generalized Linear Model, Deep Learning, Decision Tree Model, Random Forest Model, Gradient Boosted Model, and Support Vector Machine were used to conduct this research. Our experimental results show that the Decision Tree Models were the fastest way to optimize the result. The Gradient Boosted Model was the best performance to optimize the result.**

Keywords—**Visual Analytics, Student Assessment, Assessment Report, Analytical Model**

I. INTRODUCTION

The higher rate of unemployed graduates was declared as one of the numerous issues in higher education institutions as the number of undergraduates has continued to grow with the expansion of colleges and universities in recent years. Universities, such as Universiti Teknologi Malaysia, a top 200 QS world ranking university [1], play a role for every student by providing professional training for high-level employment, as well as the requisite education for personal growth. Currently, there are 381 students with 76 international students who have enrolled in the Bachelor of Computer Science

(Computer Networks and Security) in the School of Computing at Universiti Teknologi Malaysia. The main purpose of this program was to produce computer scientists in the field of computer networking and security using correct principles and techniques in designing and building quality computer programs for the community and in industries.

There is a report that is widely used throughout education institutions aimed at providing the latest feedback to enhance the next semester academic, which is known as the Annual Program Assessment Report (APAR). This report contains the evaluation actions, data, and events that occurred during the academic year and any changes to the assessment plan program in the upcoming year. Hence, this raises academic awareness of the importance of continuing to enhance courses and programs to improve student learning [2]. Therefore, the understanding of the academic performance of students in universities or colleges is an essential topic in producing this report [3]. However, many challenges will affect the academic quality of undergraduates. One of the challenges is the use of big data in education and the technology that supports it [4]. The term "big data" refers to a vast volume of data created throughout a student's learning process [5]. The information generated is meant to aid top-level management in making decisions [6]. Moreover, educational institutions are grappling with the exponential increase in educational data, as well as how to improve the quality of administrative choices. [10]. Hence, the importance of analytics is a factor for better achievement of the goals in education, as well as offering new modern opportunities for improving education system effectiveness and achieving better learning outcomes [2].

According to [7], effective academic undergraduate academic quality evaluation has a positive role in promoting the construction of universities and improving the quality of talent cultivation. Besides, education programs now focus on career development for students after graduation. The student's academic performance is influenced by producing a high-quality education. It can occur after the completion of the academic program. It may also be the consequence of learning, which is a visible and observable presentation of knowledge in conjunction with skills and orientations. These factors can be determined to produce quality in higher education.

This research aims to investigate the suitable visual analytics design to represent the student assessment data with the suitable interaction techniques of the visual analytics approach, which is enabled for presenting the insight as well as identifying the interesting findings from the generated output. Besides, this research was also proposed to provide a suitable dashboard experience for the educator as the user of the Annual Program Assessment Report.

The following research question will be examined in this study, to focus on the most important aspects of the study and evaluate the findings.

• How to utilize the analytics models to help optimize results based on the cleaned data and presented in the suitable dashboard?

This paper is organized as follows: Section II represents the related research and methodology in Section III. Case studies are illustrated in Section IV. Finally, the findings and conclusions were made in Section V.

II. RELATED RESEARCH

A. Annual Programme Assessment Report (APAR)

As mentioned in the introduction, this report is widely used and conducted by all educational institutions. According to the University of Wisconsin-Madison in the United States, an annual academic programme assessment report documents which assessment activities from a program's assessment activities plan were carried out and describes how the programme plans to improve based on the results [8]. Besides this, based on University USA, a new program reporting procedure in which each program submits one report containing the information for all Program Learning Outcome (PLO) assessed in one academic year [9]. Furthermore, according to Oregon State University, the report should reflect assessment efforts and changes informed by data from the previous academic year [10].

APAR is conducted yearly at the end of every academic session at UTM. It contains all the achievements of program learning outcomes by analysing the marks in all courses, which helps the educators and administrators to identify the causes that occurred while conducting the program courses. Besides, it can also help educators to improve their teaching based on the comparison of the achievement from last semester.

The use of big data can create various ways for diverse fields to achieve faster insight to enhance the decision making in the report. Many approaches, such as reporting, dashboards, and data visualization, may be used to improve decision-making. With the rise of big data, these techniques have been more frequently employed. Each technique's objective is to provide the user with a better understanding of the information's contents.

B. Visual Analytics (VA) Model

Data analytics is quickly becoming a popular method that many businesses are implementing to extract useful information [11]. Based on [11], the important feature across the main stages of the framework is the sixth "V", representing visualization. This will help to identify what data is needed, which is often a combination of existing "small" data and newer, faster-moving "big" data. Visual analytics is an outgrowth of the field of scientific and information visualization [12].

Visual analytics technology has been increasingly used in the last decade because it enables users to generate insights through data visualization [13]. It connects the powers of visualization and analytical reasoning to help with decision making, which can support understanding the results of large-scale datasets in meaningful ways [11].
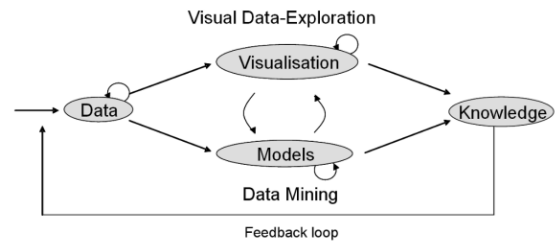

Fig. 1. Visual and autonomous data analysis approaches are combined with database technology to provide a scalable interactive decision support system. [14]

The objective of visual analytics, according to [14], is to make the process of processing data and information apparent for analytics disclosure. They claim that the information overload problem relates to the risk of becoming engrossed in material that is unrelated to the current job at hand, is processed and presented ineffectively. Rather than being left with the outcome, the depiction of these processes, as shown in Fig. 1, will provide a means of discussing them. Visual analytics can help companies evaluate, rectify, and enhance their processes, models, knowledge, and choices productively. It is extremely multidisciplinary and encompasses a wide range of relevant research topics, including visualization, data mining, data management, data fusion, statistics, and cognition science.
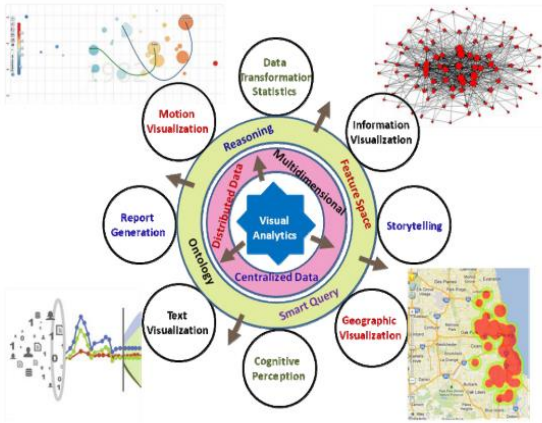
Fig. 2. A typical visual analytics model with managing data, techniques and visualization [15]

Based on [15], Fig. 2 depicts a typical visual analytics approach for handling data, methodologies, and visualization. They claim that the VA with embedded computational intelligence can overcome this shortcoming and enable the development of a trustworthy human-centric system for efficient and effective data processing. They recognize that achieving robust VA is difficult, especially in terms of managing supporting ontologies, data analysis and data mining tools and methodologies. As a result, this Interactive Visual Analytics (IVA) model can handle the aspects of VA associated with human-human interaction (HHI), human vehicle interaction (HVA) and human-object interaction (HOI) by creating a human-centric environment that allows the user to explore and exploit feature spaces of new knowledge discovery for referencing and verifying hypotheses and improving performance. They stated that this technique is very useful for pattern generation because it is effective in classification and clustering of data for event matching, as well as potentially predicting events and activities based on the success of training or learning patterns, as well as hypothesis and concept development.
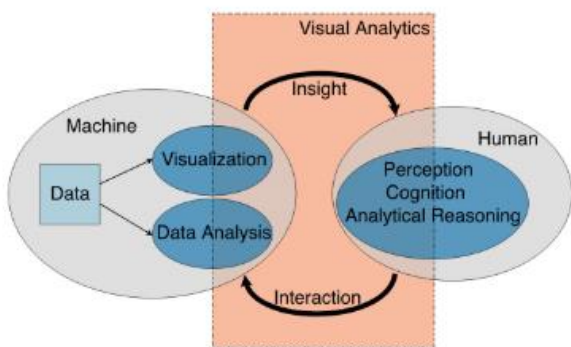

Fig. 3. A typical visual analytics model with managing data, techniques and visualization [12]

According to [12], the visual analytics process is inspired by Shneiderman's mantra that was created to focus on visual analytics, "Analyze/Overview first, interaction and

visualization repeatedly, insights into data." Based on the figure, the researcher explained that the typical steps in the visual-analytics process are summarized into six steps. Pre-processing the data to prepare it for future processing is the first stage. The algorithm analysis methods are applied to the data in the next phase. The final stage involves visualizing the processed data using the relevant visualization tools. The users can then employ human perception, cognition, and reasoning processes to create insightful knowledge through analysis and visualization through interactions based on the phase. Finally, depending on the interactions, they produce an updated graphic that reflects the user's understanding of the data. After the initial loop of the visual-analytic process, iterate from Step 4 to Step 6 until enough insight into the data has been obtained to make judgements or solve problems related to the data.

C. Visual Analytics Workflow

Visual analytics favours rapid access to large datasets, but the advantage it brings to education remains opaque [16]. The usage of visual analytics, on the other hand, is showing rising human engagement in the process of extracting information from data and building visual analytics workflows [17].
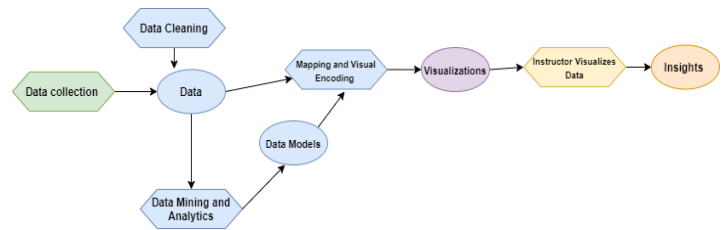

Fig. 4.   Visual Analytics flow chart phases [18]

To facilitate analytics on educational datasets, different steps of the visual analytics flow chart are augmented by operations, as described by [18]. Fig. 3 depicts a single iteration of an interactive visual exploration flow chart. The process began with data collection, which produced data, which was then followed by data cleaning processes, which were then applied to the data. Before producing visualization, data may be translated into visualization or further data mining and machine learning processes can be conducted.
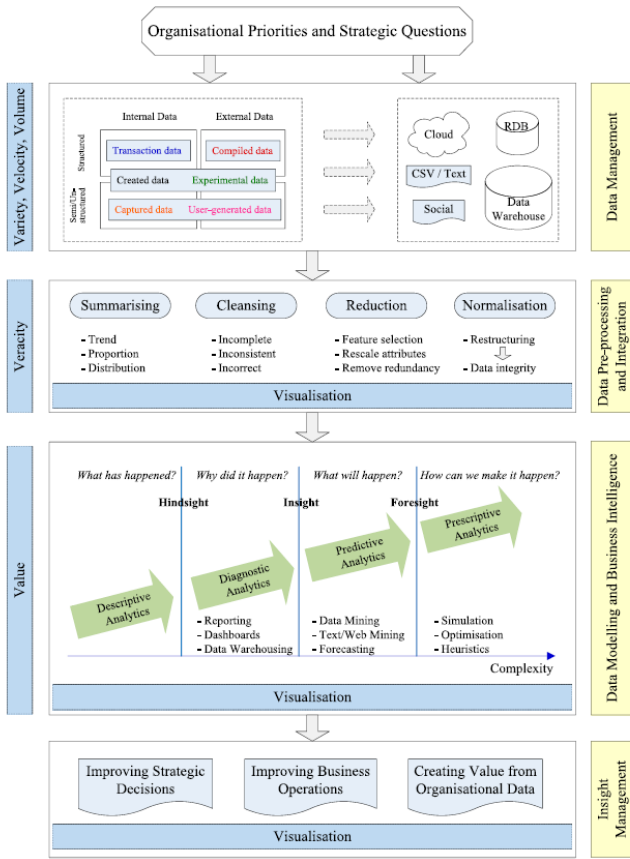
Fig. 5. Visual Analytics covers key aspects of the analytics lifecycle from data [11]

According to [11], Fig. 5 depicts the essential aspects of the analytics lifecycle, which range from data management, data pre-processing, and integration through data modelling and business intelligence (BI). According to [11], having a defined data strategy informs organisational goals and can expose unsolved issues about the sort of data required, which is frequently a mix of current "small" data and newer, faster-moving "big" data. In multiple areas, such as visualization statistics-based data analysis, machine learning, data mining and perceptual and cognitive sciences, new tools and approaches have been studied to extract valuable information and generate reliable knowledge from unexplored data [12]. Visual analytics concepts can transform organisational knowledge gathered through semantic web technologies into an informative analytics experience. [22].

Based on [22] suggested a knowledge repository for analytics-related information that can be generated and managed independently of the organization's IT architecture and used to construct analytics apps with visual interfaces. The knowledge repository is intended to store and link domain knowledge about data, results about machine learning model. Besides, the IT infrastructure that are proposed which is mainly represents any software program to assists with the data processing operations. The Visual Analytics Applications are a set of end-user applications built and made available for analysts via graphical user interface (GUI) to help them use the

IT platform and knowledge repository to acquire and analyze high-quality data. The Knowledge Management platform is the user interface for inserting information about new IT infrastructure and domain knowledge to the organization's knowledge repository. This is to enable the end users can query and visualize knowledge stored in the repository.
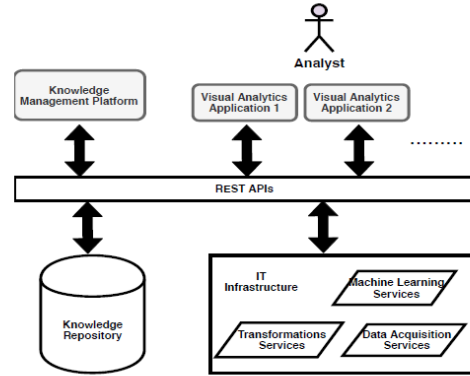

Fig. 6. Proposed Architecture that integrates a knowledge repository and current IT infrastructure as services via REST APIs [22]

III. METHODOLOGY

Four phases needed to be carried out in this research: data understanding, data preparation, data modelling and data presentation. The elaboration of the phases is included in the methodology based on Fig. 7.
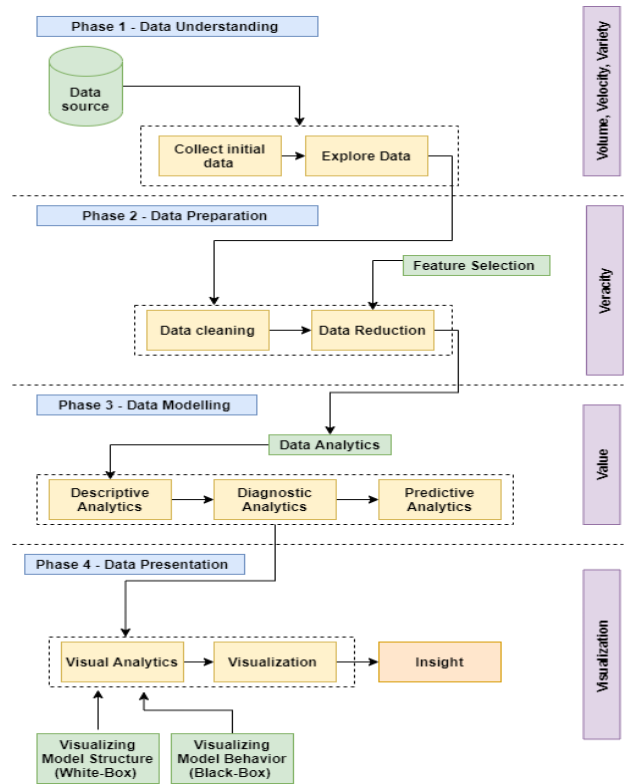

Fig. 7. Phases that were included to be carried out in this research

A. Data Understanding

This phase is a preparatory activity that starts with an initial data collection to get familiar with the data. Many challenges have been reported internally in previous APAR reports. One of the challenges that has been identified is the program coordinator having difficulty manually analyzing all the data that is obtained from the OBE system. All the data that is obtained from the OBE system must be separated according to a predetermined category manual. During this phase, three "V" were identified, which are "Volume", "Velocity" and "Variety". The first "Volume" was identified in collection data obtained from the records of Bachelor of Computer Science (Computer Network and Security) students from the system. The data was analyzed from cohort 2016 until cohort 2019. The second was "Velocity", which was referred to as the speed at which the data from the system was generated. The third is "Variety", which was based on the various courses that were taken by the students. In this study, we mainly focused on six subjects based on Table II from First Year to Second Year.

TABLE I. SUBJECT TAKEN BY THE NUMBER OF STUDENTS

No	Subject Code	Subject Name	No of Student Taken Subject
1	SCSR1013	Digital Logic	260
2	SCSR2033	Computer Organization & Architecture	243
3	SCSR1213	Network Communication	250
4	SCSSR3413	Computer Security	232

B. Data Preparation

After the collection and exploration of data was completed in the first phase, it was time to check the trustworthiness of the data, which will involve the fourth "V", veracity. The term "data cleaning" refers to the process of identifying, correcting, or eliminating erroneous or inaccurate values, and then replacing, altering, or deleting the wrong data as needed [11]. In this phase, we applied data cleaning using Rapid Miner by removing duplicated, erroneous, and missing values in the excel xlsx format. Fig. 8 shows the process of cleaning the data using Rapid Miner.
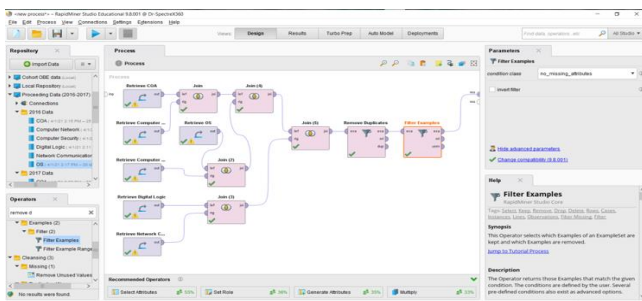

Fig. 8. Cleaning data process by using Rapid Miner

The act of detecting, correcting, or removing erroneous or inaccurate information, and then replacing, modifying, or deleting the incorrect data as appropriate, is referred to as "data cleaning." [11]. The technique involved in data reduction is feature selection. This process is performed to choose a subclass by removing non-prescient information. When the process is utilized, the efficiency of the prediction model is enhanced [19]. Based on [20], the purpose of using feature selection in this study is to reduce the training and application time of machine learning models. Besides, the machine learning algorithm's runtime is sped up by eliminating unnecessary features. We applied the filter-based feature selection by using a method of reducing non-informative variables from our data collection.

C. Data Modelling

Data modelling phases involve analytical techniques which are descriptive analytics, diagnostic analytics and predictive analytics. The fifth "V" refers to the value of the information that can be achieved by processing and analysing large datasets [11].

Descriptive analytics provide useful information. Diagnostic analytics, which takes descriptive data a step further, provides a deeper analysis of this research, which also includes processes such as data mining. Predictive analytics enables users to transform historical and present data into actionable knowledge, resulting in beneficial long-term outcomes.

D. Data Presentation

At this stage, which will provide the insight to present the data, two types of generic modalities can be used in model interpretation with visual analytics, such as Visualizing Model Structure (White-Box) and Visualizing Model Behavior (Black-Box) from [21]. Visualizing Model Structure is for transparent models such as decision trees, whereas Visualizing Model Behavior is for visualization as a way to look at the behavior of the model by looking exclusively at the relationship between the input and the output.

IV. RESULT AND DISCUSSION

The study shows more useful information that had been interpreted based on the proposed methodology. We conducted the findings using the Rapid Miner as a tool in this experiment to identify the best way to utilize the analytics models that can help optimize results based on the cleaned data. The findings show that there are six types of analytical models which include Generalized Linear Model, Deep Learning, Decision Tree, Random Forest, Gradient Boosted Trees, and Support Vector Machine (SVM) that were utilized in this research based on Fig. 10. The comparison of the courses based on different years makes no changes between the differences in analytics models. We conducted the prediction of the PLO1 which is the first attribute of the PLO classifies as Knowledge Understanding (KW). The indicators of PLO achievement by the students are ability to acquire theory and principles of

Computer Science and Data Engineering and equip with social science and personal development knowledge. Computer Security based on 2017 and 2018. It is also one of the important criteria that consist in the APAR in order to ensure the students meet the criteria for the outcome of the courses. However, the results show the Decision Tree Model in Fig. 9 is the fastest total time to optimize the result. In conclusion, the Generalized Linear Model shows the best performance in the optimized result.
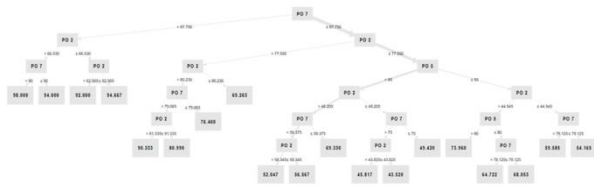

Fig. 9. Decision Tree Model from the prediction of PLO1 in computer science year 2017

TABLE II. Result From The Prediction of PLO1 in Computer Science Year 2017

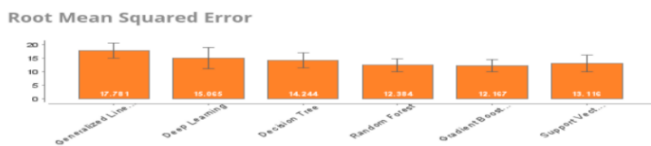| No	Model	RMSE	Standard Deviation	Total Time	Training Time	Scoring Time
1	Generaliz-ation Linear Model	15.9	3.7	80202.0	719.7	18.9
2	Deep Learning	16.9	2.8	102232.0	1621.2	37.7
3	Decision Tree	10.3	232	2686.0	15.2	18.9
4	Random Forest	10.2	227	12802.0	121.2	150.9
5	Gradient Boosted Trees	9.9	250	138210.0	2197.0	94.3
6	Support Vector Machine	12.9	232	12821.0	113.6	18.9

Fig. 10. Overview of comparison in analytics model result

V. Conclusion

In this study, we proposed the methodology of the visual analytics approach. Our experimental results show that the fastest total time that optimized the result is the Decision Tree Model. It can be a consideration of all possible outcomes of a decision and traces each path to the conclusion. For future work, we are trying to solve these issues with another type of analytical model with the extension of the usability of the visual analytics approach and conducting different courses in the same year. We are also considering continuing the experiment by adding a suitable dashboard to represent the assessment report with the prediction.

References

[1] QS WORLD UNIVERSITY RANKINGS 2021: UTM MADE IT TO TOP 200 | Department Deputy Vice Chancellor (Research & Innovation). (2021). Www.utm.my. https://www.utm.my/office-dvcri/2020/06/11/qs-world-university-rankings-2021-utm-made-it-to-top.

[2] Reich, A. Z., Collins, G. R., DeFranco, A. L. and Pieper, S. L. (2019). A Recommended Closed-loop Assessment of Learning Outcomes Process for Hospitality Programs. *International Hospitality Review,* 33(1), 53-66.

[3] Etemadpour, R., Zhu, Y., Zhao, Q., Hu, Y., Chen, B., Asif Sharier, M., Zheng, S. and Paiva, J. G. S. (2019). Academic Performance Analysis Supported by a Web-based Visual Analytics Tool. *Proceedings - IEEE 19th International Conference on Advanced Learning Technologies, ICALT 2019. IEEE*, 158-162.

[4] Bamiah, M. A., Brohi, S. N. and Rad, B. B. (2018). Big Data Technology in Education: Advantages, Implementations, and Challenges. *Journal of Engineering Science and Technology*, 13(Special Issue on ICCSIT 2018), 229-241.

[5] Li, Y. and Zhai, X. (2018). Review and Prospect of Modern Education using Big Data. *Procedia Computer Science,* 129, 341-347.

[6] Santoso, L. W. and Yulia. (2017). Data Warehouse with Big Data Technology for Higher Education. *Procedia Computer Science.* Elsevier B.V., 124, 93-99.

[7] Shengnan, Q., Zhengxin, L. and Jingjing, L. (2020). Based on the OBE Concept Research and Construction of University Academic Information Feedback Mechanism. *Proceedings - 2020 International Conference on Big Data and Informatization Education, ICBDIE 2020*, 430-434.

[8] Criteria, A., 2021. Annual Assessment Reporting Criteria. [online] Student Learning Assessment. Available at: <https://assessment.provost.wisc.edu/annual-assessment-reporting-criteria/> [Accessed 1 April 2021].

[9] UMD. 2021. Program Assessment Report. [online] Available at: <https://www.d.umn.edu/assessment-student-learning/assessment-practices/program-assessment-report> [Accessed 1 April 2021].

[10] Office of Academic Programs and Assessment (APA). 2021. Annual Program Assessment Report FAQs. [online] Available at:<https://apa.oregonstate.edu/assessment/annual-program-assessment-report-faqs> [Accessed 1 April 2021].

[11] Lu, J. (2020). Data Analytics Research-informed Teaching in a Digital Technologies Curriculum Data Analytics Research-Informed Teaching in a Digital Technologies Curriculum. Informs Transactions on Education.

[12] Cui, W. (2019). Visual Analytics: A Comprehensive Overview. *IEEE Access*. IEEE, 7, 81555-81573.

[13] Zhang, Y. and Chan, K. K. (2020b). Infusing Visual Analytics Technology with Business Education: An Exploratory Investigation in Fostering Higher-order Thinking in China. *Innovations in Education and Teaching International.* Routledge, 00(00), 1-10.

[14] Keim, D., Andrienko, G., Fekete, J.-D., G¨org, C., Kohlhammer, J. and Guy Melan¸con. (2008). Visual Analytics: Definition, Process, and Challenges. *Information Visualization*, 34(2), 154-175.

[15] Habibi, M. S. and Shirkhodaie, A. (2013). Mining Patterns in Persistent Surveillance Systems with Smart Query and Visual Analytics. *Geospatial InfoFusion III*, 8747(May 2013), 87470P.

[16] Zhang, Y. and Chan, K. K. (2020a). Incorporating Visual Analytics with Knowledge Construction in Problem-based Learning: A Qualitative Study. *Interactive Learning Environments.* Taylor & Francis, 0(0), 1-13.

[17] Krak, I., Kruchynin, K., Barmak, O., Manziuk, E., & Kruchinin, S. P. (2020). Visual Analytics in Machine Training Systems for Effective Decision. *Advanced Nanomaterials for Detection of CBRN.* Springer, Dordrecht, 327-338.

[18] Gampa, R. and Baynes, A. (2019). Visual Analytic Workflow to Understand Students' Performance in Computer Science Courses. *Proceedings - Frontiers in Education Conference, FIE. IEEE*, 2018-Octob, 1-5.

[19] Ajibade, S. S. M., Ahmad, N. B. and Shamsuddin, S. M. (2019). An Heuristic Feature Selection Algorithm to Evaluate Academic Performance of Students. *ICSGRC 2019 - 2019 IEEE 10th Control and System Graduate Research Colloquium, Proceeding*. IEEE, (August), 110-114.

[20] Ghotra, B., McIntosh, S. and Hassan, A. E. (2017). A Large-scale Study of the Impact of Feature Selection Techniques on Defect Classification Models. *IEEE International Working Conference on Mining Software Repositories*, 146-157.

[21] Krause, J., Perer, A. and Bertini, E. (2016). Using Visual Analytics to Interpret Predictive Machine Learning Models. (Whi), 106-110.

[22] Bandara, M. and Rabhi, F. A. (2020). Knowledge-driven Framework for Designing Visual Analytics Applications. *Proceedings of the International Conference on Information Visualisation*, 2020-September(Iv), 515-520.