INTERACTIVE DASHBOARD FOR TRACKING SYSTEM DASHBOARD USING POWER BI

Nur Nafiisah Najihah Noor Azuar[1], Syaheerah Lebai Lutfi[*1], Sharifah Darweena Syed Ahmad Amir Feisal[2]

[1]School of Computer Sciences, Universiti Sains Malaysia, Malaysia
[2]Alumni Development Centre, Universiti Sains Malaysia, Malaysia

nafisahazuar@student.usm.my, syaheerah@usm.my, darweena@usm.my

Highlights: Alumni are a good advertisement for the University and reflect the University's commitment to being a successful educational provider. However, it is difficult to keep in touch with alumni as little data has been updated. A dashboard was developed to track alumni progress. The existing dashboard was not very intuitive and did not provide descriptive analytics. To replace the existing dashboard, a new dashboard was developed to visualize statistics and results of key factors. The goal of this project is to improve the efficiency of the alumni dashboard system. Descriptive and predictive analysis was performed using Power BI. The current dashboard is more intuitive compared to the previous dashboard. As for the predictive part, Random Forest has proven to be the best classifier for tracking alumni.

Key words: dashboard, data visualization, business intelligence, power BI, alumni tracking

Introduction

The percentage of active alumni hardly changes when students graduate. This fact partly makes it difficult to track and update their data, resulting in a large number of inactive alumni. The percentage of active alumni was only about 17%, while the database contained more than half (about 74,000) missing values such as email addresses, phone numbers, and employment sectors. The available data was scattered and difficult to process. There was a lack of an overall view of graduate success. Each segment had fewer logical relationships. The existing dashboard was fairly basic and needed to be more thorough. What are the demographic statistics of alumni? How can we predict timely graduation? The dashboard also lacked descriptive analysis and needed more prescriptive analysis. The dashboard was mainly found on Bahasa Malaysia, which made it less attractive to international users. The goal of this project is to demographically display alumni statistics and use machine learning to predict students' timely graduation.

Review

Related Works

Valdez et. al. (2017) created a balanced scorecard project, and the project was developed in a higher education institution. It was based on the development of a customized dashboard that allows students to set their personal KPIs through Microsoft Office 365 Power BI. The results showed the acceptance of the students' admission and re-entry faculties. Based on the scorecard dashboard, they were able to design a strategy to overcome the challenges that increase the KPI. Deppi Linda (2015) conducted another study at Data Warehouse - student and graduate data were collected as part of a survey to support the completion of accreditation forms and to provide information on the number of prospective students and graduates. The faculty categorized graduates by index type as a percentage of cumulative performance. Business Life Cycle Methods and Online Analytical Processing are used in this research to create a data warehouse.

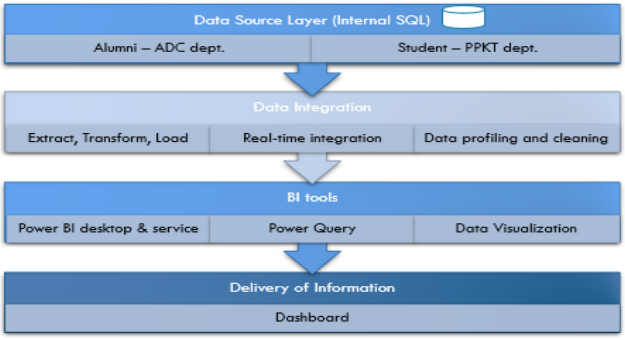
Framework Architechture of Business Intelligence

The first part of the framework architecture was the data source layer. The data is obtained from the two internal sources, whcih are the operational database and the database warehouse. The internal data in the dashboard was connected to student and alumni data from different departments of the university. The data was managed in a SQL database and then moved to the next phase, data integration. In the second phase, the heterogeneous data was extracted, transformed, and loaded (ETL) in real time. During the ETL process, the data was profiled and cleansed. Microsoft Power BI was selected as BI tool in this project, which includes Power Query features. Power Query provides functions to display table data in a variety of formats, from large databases to simple text files. Tables can be reformatted and merged until loaded as tables or into data with pivot tables and graphics software.

Data visualization was the visual representation of information and data. It is an open way to see trends, boundaries, and patterns in data with graphical elements such as tables, charts, and maps. Examples of visualizations include bar charts, stacked bar charts, pie charts, histograms, bubble charts, and others. A

Presentation for Power BI is a multi-page list of visualization forms. An overview of the architecture BI is shown in Figure 1.

Figure 1. Business Intelligence Architecture



Methodology

The first part was the analysis, which was achieved by studying the theories, approaches or models in general in the field of information systems, as well as the dashboard of business intelligence. Interviews were conducted with departments that deal with alumni monitoring issues and problems, such as the Alumni Development Center (ADC) and the University's Technology and Communication Center (PPKT). This project would be a detailed analysis of the information needs and would include analysis of the challenges and strategies of the ADC. Data collection, which includes an analysis of current reports and databases, was also included. The structure of the dashboard design was organized and microplanned. The workflow is presented in Figure 2 below. Access permission to the database was granted as per the objective, with read-only access.

Next came the planning stage where data was collected and loaded into Power BI from which the information was pulled and algorithms were added for principal component analysis. This phase involved more detailed interpretation of the preloaded, in-powered, and algorithm-based models, as well as the probability and expectation columns that displayed various charts and graphs within the dashboard. To fill in the missing values in the alumni database, the help of the data stewards, i.e., the technical team, was required. Our team had requested to access certain data from various departments such as personal and academic characteristics (name, marital status, gender, age, courses, school, department, CGPA and others).

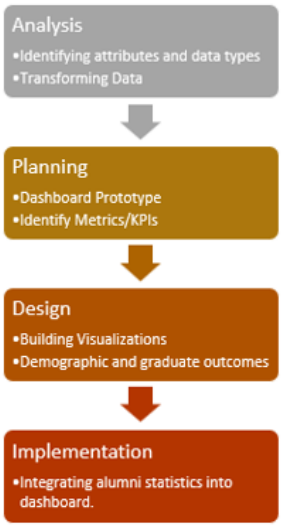


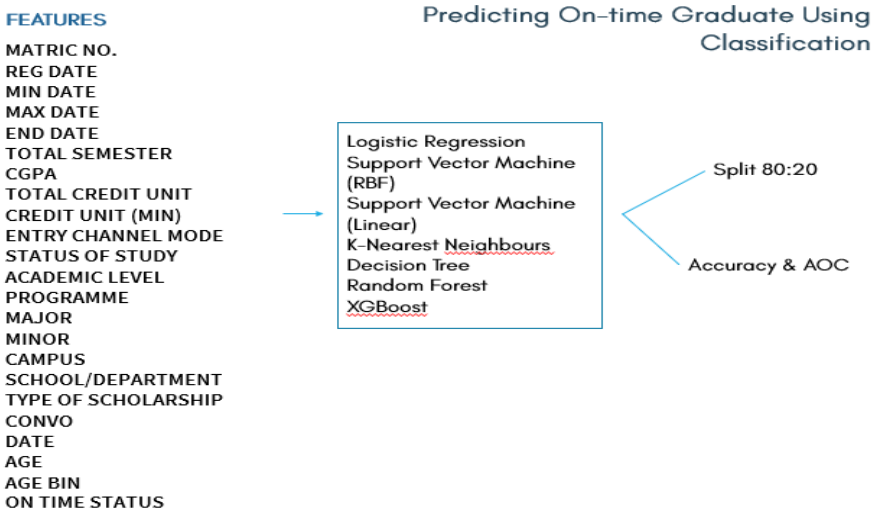
Figure 2. Procedure workflow

Based on these attributes, a descriptive analysis was performed using Microsoft Power BI. The program used for data processing and subsequent loading and deployment practices were developed during this process. During the design phase, data connectivity between the SQL database and BI tools was developed using large volumes of tabular data queries, which were then merged to meet the necessary requirements.

Data cleansing includes data authentication, data cleansing and data processing. In this step, the selected software for the ETL process was selected, formatted and displayed. Then, the data were normalized and only the graduates' data were isolated while the other data were removed. The graduate data was separated by type of study, subject, and matriculation number. After removing the data of graduates, the data of graduates have about 150000 rows. This data was then converted to English as the existing datasets were all in Malay. The cleaned data was then visualized in various charts. During the implementation, all the statistics and visualizations were combined into a three-page dashboard. This part also included configuring the dashboard design, presentation, measurement, metrics, dashboard vividness, deployment, and architecture.

To predict on-time graduation, the classification was created using machine learning features in Power BI and integrating a Python script. Only the data from the year 2016 was collected. The attributes were taken from the SQL database and transformed. The target predictor and attributes were assigned. The data types were modified and coded. The cleaned data was then used to split into training and testing data before going through the classification model. The performance of the models was evaluated based on the accuracy and area under the curve. The methods were simplified as shown in the figure below.

Figure 3. Classification using Machine Learning



Data Analysis

The initial undergraduate data included 145000 lines consisting of academic and personal data. Table 1 below shows some of the data extracted from the Alumni and Student Affairs database.

Table 1 : Attributes Chosen

Personal	Academic
Name, Date of Birth, Citizenship, Student Id, Marital Status, Gender, Email Address, Race and Religion	Program, School, Campus, Convocation Date, Major, Minor, Sponsorship and Entry Channel

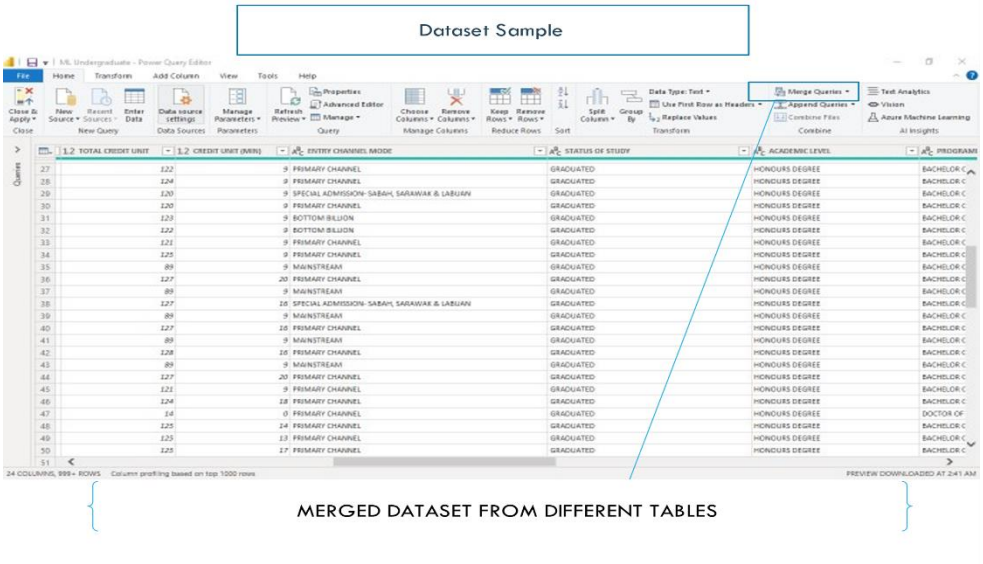


Figure 4. Sample Dataset to be cleaned

Figure 4 above shows part of the dataset in Power Query. In the part of the cleaned data, some of the tables were merged and further transformed by changing the data type, replacing blank values and selective attributive filtering. The original data contained both undergraduate and graduate data, so the rows were reduced by filtering type of study, program, and matriculation number. Thereafter, the data contained only first-year students. The results were selected based on consumer criteria after comparing the attributes. Appropriate chart types were selected to visualize the data. Once the data was collected and analyzed, dashboard model construction was performed.

A quick analysis was performed for a few demographic attributes. The following chart shows the percentage of alumni by gender. The number of female alumni is higher than male alumni.

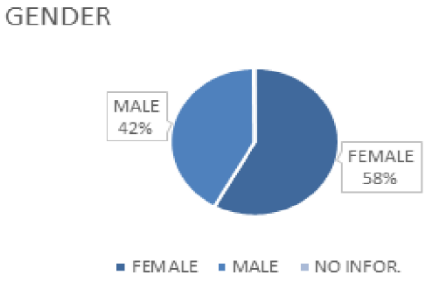


Figure 5. Percentage of Gender

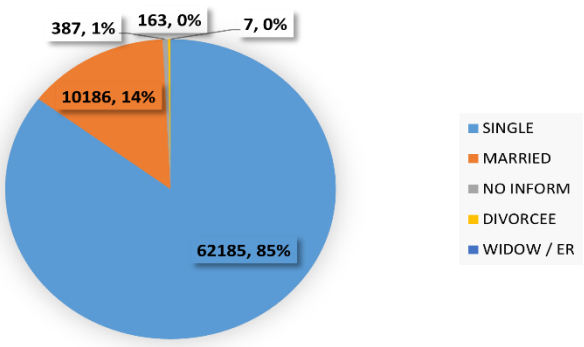


Figure 6. Percentage of Marital Status

Figure 6 describes the marital status of graduates, with 85.27% single, 14% married, 1% not specified, and only a small proportion divorced and widowed.

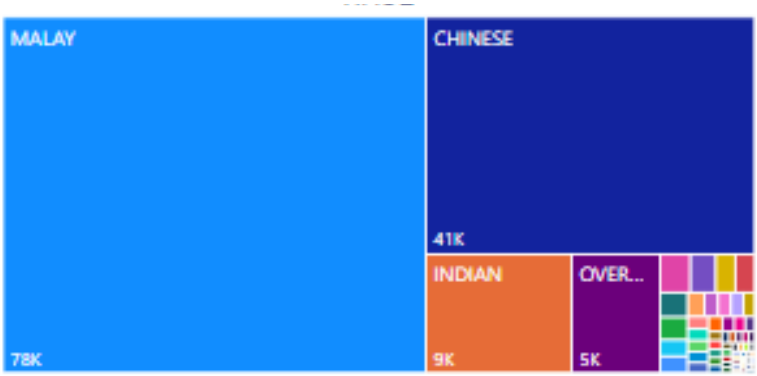


Figure 7. Treemap of Race

There were 114 races of graduates, most of whom were Malays (54%), followed by Chinese (28%) and Indians (6%).

Results and Discussion

Figure 8 shows the prototype front page accessible to the public and alumni to view alumni tracking updates. The page consists of three pages for visualization.

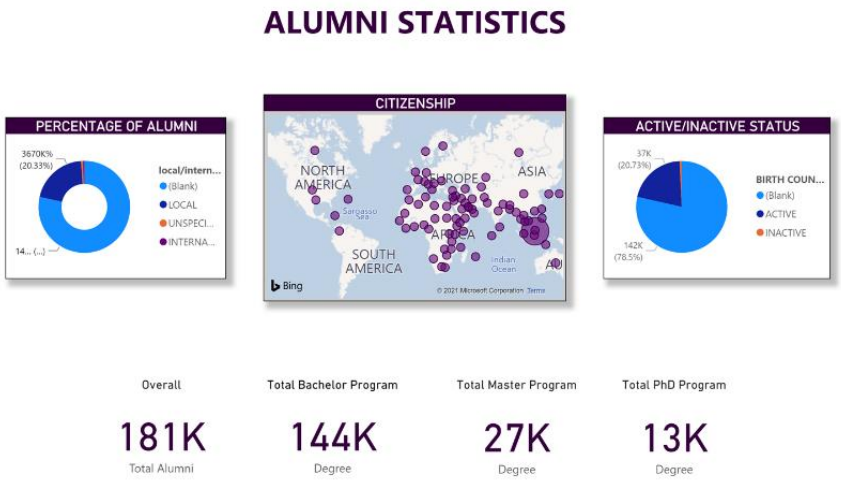


Figure 8. Front Page of Dashboard

On the left is a donut chart showing the percentage of graduates both locally and internationally. The next is a map chart for alumni nationalities and the right is a pie chart for active and inactive alumni. The bottom part shows the total number of alumni by educational level. The second side contains alumni demographics including gender, age group, marital status, religion, race, and average age. The right side and top left portion consist of slicers for age group, gender, school/department, program of study, major, minor, academic level, and on-time graduation (number of alumni graduating before the maximum graduation date). The third page presented the characteristics of graduation statistics related to academics. The page included the total number of graduates by school/department, entry route, and type of scholarship. This page also had slicers like the one on page 2 of the dashboard to allow the end user to view the data based on their filter of choice. This dashboard was improved from the previous version with additional features and slicers that allowed the end user to interactively scroll between pages. This made the data easier to be digested and presented to top management.

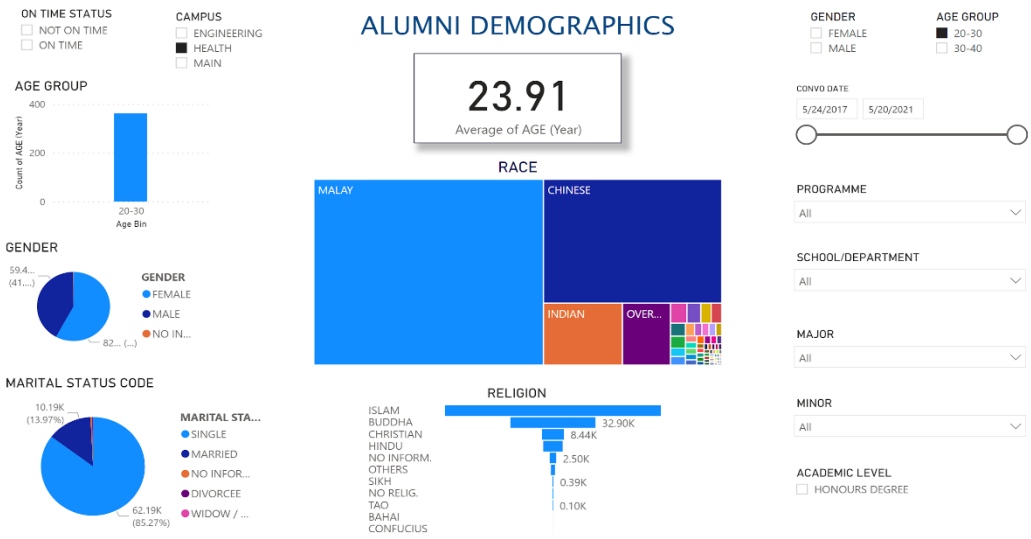


Figure 9. Second Page of Dashboard

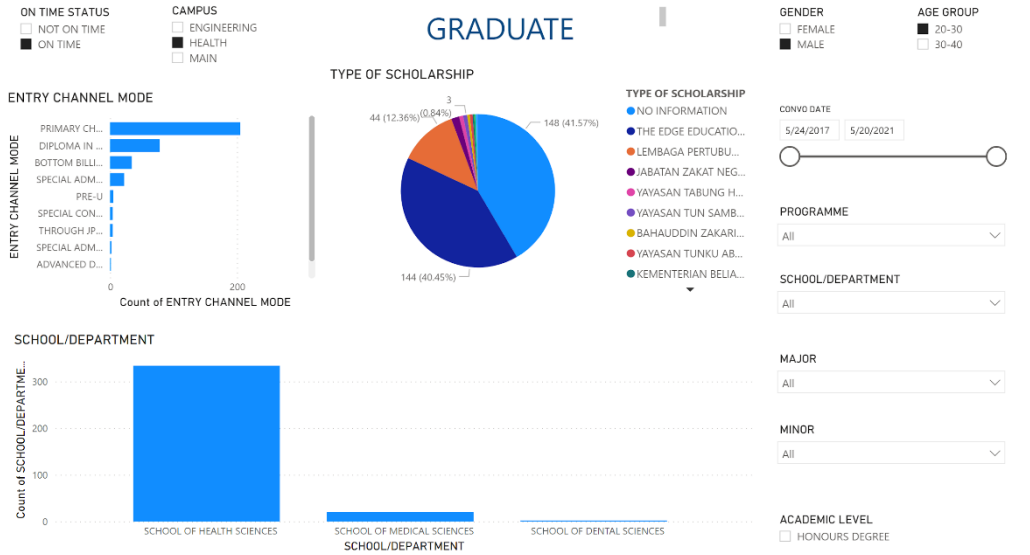


Figure 10. Third Page of Dashboard

Figure 11 shows side by side comparison between the old and new dashboards. The left side is the new dashboard, which contains more statistics and visualizations than the right side. The old dashboard had only four types of visualizations, indicating that the new dashboard is twice as good as the old version. The new dashboard facilitated the display of data pulled from the old dashboard to be presented better to the end user.

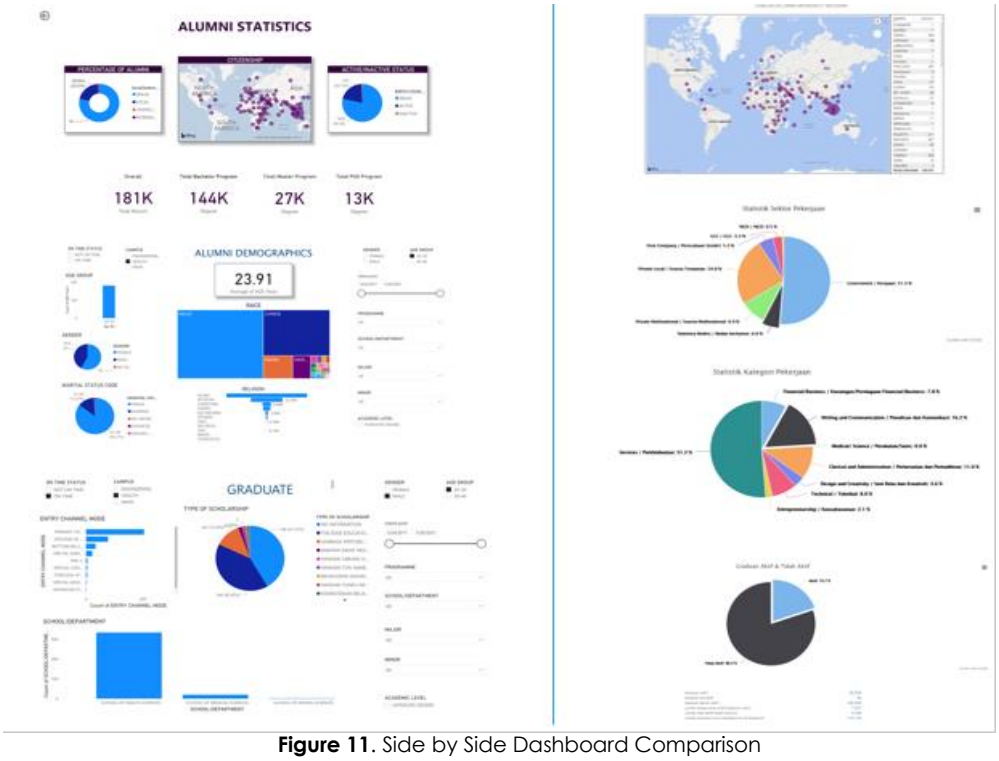


Figure 11. Side by Side Dashboard Comparison

In the predictive analysis, the classification was performed on the Jupyter notebook using the Panda script to run the test model. The model will be implemented and integrated into the dashboard in the future. The results based on accuracy and area under the curve (AUC) are shown in Figure 12. The best classifier is Random Forest with an accuracy of 0.832 and an AUC of 0.903. This classifier will be considered when using the latest data. However, the results were not completely accurate. This was due to the current challenges we faced which resulted in a delay in finalising the dashboard.

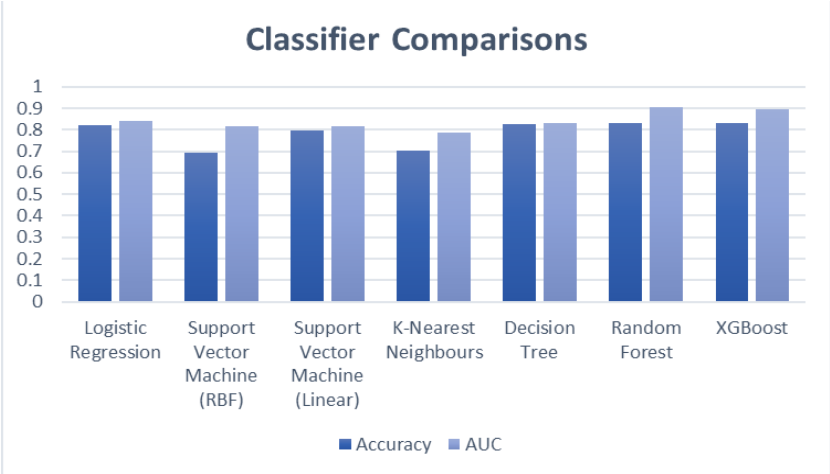


Figure 12. Comparisons of the Classifiers

Conclusion

Alumni are important because they are the ones who make the university to the world. In order to keep track of alumni, the database needed to be updated. However, the database had too many missing attributes, which were solved with the help of data organization. The existing dashboard was quite simple and the design was changed to visualize the data comprehensively using the Power BI tool. The current dashboard was more interactive and better than the previous one. Future plans include adding employment sector and graduation rate by year to the visualization and continuing the predictive analysis of this project.

Acknowledgement

Our deepest gratitude goes to the Alumni Development Centre from Universiti Sains Malaysia for supporting and sponsoring this work.

References

Valdez, A. Cortes, G. Castaneda, S. Vazquez, L. Medina, J. & Haces, G. (2017). Development and implementation of the balanced scorecard for a higher educational institution using business intelligence tools. *Development*, 8(10).

Linda, D. (2016). Pengisian Borang Akreditasi Standar 3 Menggunakan Model Data Warehouse (Studi Kasus Ibi Darmajaya). *Jurnal Informatika Darmajaya*, *15*(2), 89-101.

Sluijter, J. and Otten, M., 2017, March. Business intelligence (BI) for personalized student dashboards. In *Proceedings of the Seventh International Learning Analytics & Knowledge Conference*. 562-563.

Minelli, M. Chambers, M. & Dhiraj, A. (2013). Big data, big analytics: emerging business intelligence and analytic trends for today's businesses. Vol. 578. *John Wiley & Sons*.

Kajava, E. (2018). Improving company performance through implementation of Business Intelligence tools: Implementation of a Microsoft Power BI in a Case Study Company.

Khan, R.A. and Quadri, S.M.K., 2012. Business intelligence: an integrated approach. *Business Intelligence Journal*. 5(1). 64-70.

Ameer, M. Rahul, S.P. and Manne, S. 2020. Human Resource Analytics using Power Bi Visualization Tool. In *2020 4th International Conference on Intelligent Computing and Control Systems* (ICICCS). 1184-1189. IEEE.

Widjaja, S. and Mauritsius, T. 2019. The Development of Performance Dashboard Visualization with Power BI as Platform. *International Journal of Mechanical Engineering and Technology* (IJMET). 10(5). 235-249.

Data Visualization. Microsoft Power BI. (n.d.). Microsoft Power BI. Retrieved from: https://powerbi.microsoft.com/en-us/

Tableau (n.d.). Data visualization beginner's guide: a definition, examples, and learning resources. Retrieved from: https://www.tableau.com/learn/articles/data-visualization#:%7E:text=Data%20visualization%20is%20the%20graphical,outliers%2C%20and%20patterns%20in%20data.

Becker, L. T., & Gould, E. M. (2019). Microsoft power BI: extending excel to manipulate, analyze, and visualize diverse data. *Serials Review*. *45*(3). 184-188.

Becker, L. T., & Gould, E. M. (2019). Microsoft power BI: extending excel to manipulate, analyze, and visualize diverse data. *Serials Review*, *45*(3), 184-188.