International Journal of Innovation, Creativity and Change. www.ijicc.net
Volume 5, Issue 6, 2019

Data Analytics for Higher Education – Promises & Limitations: A Case Study in Pelita Harapan University

Ihan Martoyo[a], **Lolla Manurung**[b], **Victoria Noble**[c], **Rijanto Purbojo**[d], **Adriani Gunawan**[e], [a,b,c,d,e]Universitas Pelita Harapan, Banten, Indonesia, Email: [a]ihan.martoyo@uph.edu, [b]lolla.manurung@uph.edu, [c]victoria.noble@uph.edu, [d]rijanto.purbojo@uph.edu, [e]adriani.gunawan@uph.edu

Universities sit on a huge databank, from admission data, student academic data, to the alumni data. However, unlike the financial or telecommunication industries, universities often do not see it as critical to harness these data for strategic decisions. In this paper, we explore data analytics by visualisation to enable a data dashboard for strategic decision in the case of Universitas Pelita Harapan (UPH). We perform initial interviews with selected deans and university leaders, then develop and present various academic data visualisations. Focus group discussions were then performed with deans and leaders to identify various ways they envision use of the data. The identified important indicators for the data dashboard include number of students, GPA, lecturer/student ratio, graduation time, class sizes, number of failing students, number of full time/part-time lecturers, and students' feedback on lecturers. Some examples of GPA and admission data are presented in this paper to illustrate the potential usefulness of data analytics. In discussion with deans and university leadership, we found potential uses of data analytics to include: increasing admission numbers/quality, providing early warnings for failing students, optimising teaching process and utilisation of resources, and optimising the faculty workload. The success of learning analytics seems to involve the careful design of observation data into the learning settings. The usefulness of data analytics must be balanced with the awareness of its limitations, as was shown, e.g., by the case of Google Flu Trends (GFT). Availability of skilled data scientists, wisdom in institutional strategic decisions, and awareness of ethical issues involved with data analytics is becoming more important with the implementation of data analytics.

Key words: *Data Analytics, Academic Analytics, Learning Analytics, Data Visualisation, Higher Education.*

Introduction

Every university-level strategic decision, such as opening a new program, starting an innovative new education scheme, or simply expanding existing programs, need to be guided by concrete data and facts. Data analytics can also be very useful to optimise resources or increase the quality of admission and education processes. This paper reports an exploration of how data analytics can be utilised for strategic decision making in the case of Universitas Pelita Harapan (UPH), Indonesia.

Data analytics can be performed in the learning management level to observe the performance of learners and to design a learning improvement intervention. With the availability of real-time insights into the learning experience, suitable learning activities and resources can be provided for students who are at risk (Siemens & Long, 2011). Such application of data analytics is often termed *learning analytics*. In US colleges and universities, graduation and retention rates are considered important indicators. For four-year degree programs, the average graduation rates at public and private universities in the US can be as low as 20% to as high as 65% (Picciano, 2012). Thus, learning analytics that can identify and give early warnings about students at-risk are deemed crucial.

On the other hand, data analytics can also be applied in the macro level of strategic decision making, where the dashboard shows historical or real-time data on how the businesses (university, units, schools, or departments) are performing. Such data analytics is often termed *academic analytics* or *business analytics* (Van Barneveld, Arnold & Campbell, 2012).

The challenges for making use of data analytics for education include the building of a reliable data sharing network for the whole university. Details such as relevant indicators and similar data format must be designed carefully to enable reasonable analysis and comparison (West, 2012). Other requirements for data analytics adoption in higher education include the support of university leaders, who are committed to evidence-based decision making, and the availability of skilled data analysts (Van Harmelen & Workman, 2012). Although the adoption of data analytics may imply considerable investment that includes the process of cleaning, integrating and visualising data, the scale of implementation can be chosen to start from a level that can be managed at the current context of the university. Localised implementation can be planned, e.g., data analytics for admission, or scheduling resources, or alumni donation management.

Besides the business/academic analytics vs the learning analytics model for higher education, data analytics approaches can also be classified into a *descriptive, predictive,* and *prescriptive* model (Daniel, 2015). The descriptive analytics describe and analyse historical university data to find and recognise useful patterns, such as student enrolment, graduation rates, etc. The predictive analytics estimate the likelihood of future events, risks and opportunities that could be caused by some hidden factors not yet revealed in the descriptive analytics, such as demographic factors. Predictive analytics may include categories such as data mining, predictive modelling, simulation, and forecasting (Rajni & Malaya, 2015). The prescriptive analytics combines descriptive and predictive analytics to provide informed choices for the university based on valid and consistent predictions.

Aldowah, Al-Samarraie & Fauzy (2019) reviewed 402 studies of educational data mining (EDM) and learning analytics (LA) in higher education and found that the most common data mining techniques used in the studies are: clustering, association rule, visual data mining, statistics and regression. Typical purposes of using EDM and LA include: predicting dropout and retention, monitoring and evaluation of students' learning, predicting students' performance and success, and evaluation of learning materials. Ekowo & Palmer (2016) claimed that predictive analytics could be used for targeted student advising, adaptive learning, and managing enrolment.

Research into LA to provide intervention for at-risk students seems to be popular. Such research usually utilises student's online behaviour data to monitor students' learning. However, an early study of the benefits of online interaction did not show any significant improvement in students' performance with passing grades (Davies, 2005). Thus, simply encouraging more online discussions does not automatically lead to higher grades. Jo, Yu, Lee & Kim (2015) also found that total studying time in LMS (Learning Management System), interactions with content, interactions with peers, and interaction with instructors, did not predict final grades. Macfadyen & Dawson (2012) also point out that unless learning analytics data are presented to those involved in strategic institutional planning who have the power to motivate cultural change and new policies, data analytics numbers alone might not be helpful. Furthermore, using students' data for monitoring might also pose some new ethical dilemmas (Lawson et al., 2016).

This paper focusses more on *academic analytics* or *business analytics* opportunities in Universitas Pelita Harapan (UPH). However, some *learning analytics* approaches will also be discussed in the discussion section below.

Methods

This paper reports the result of exploring data analytics in the context of Universitas Pelita Harapan (UPH), Indonesia. We performed initial interviews with the deans of several selected departments to identify meaningful indicators and their specific needs for the data analytics dashboard. The faculties include the School of Nursing (SoN), School of Music (SoM) and the School of Hospitality (FPar).
\

In this paper, we focus on *visual data analytics*. Selected indicators identified in the initial interviews were explored and visualised to be presented before the deans and university leaders. Confirmatory discussions were then performed with the deans and leaders to estimate the possible uses of the data dashboard. Samples of data visualisations are presented in this paper to illustrate the potential of applying data analytics for higher education.

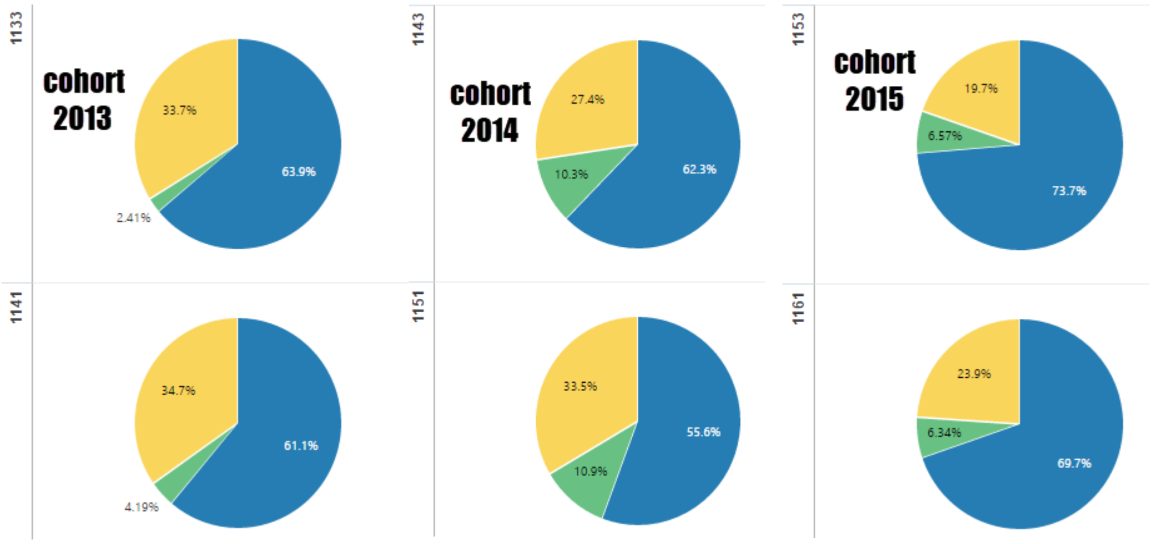
Results

The initial interviews with the deans and university leaders identified these important indicators: (1) Number of students, (2) GPA, (3) Faculty/student ratio, (4) Ratio of full time/part-time lecturers, (5) Lecturer's Workload, (6) Graduation time, (7) Admission numbers, (8) Number of faculty's research/publications/community service.

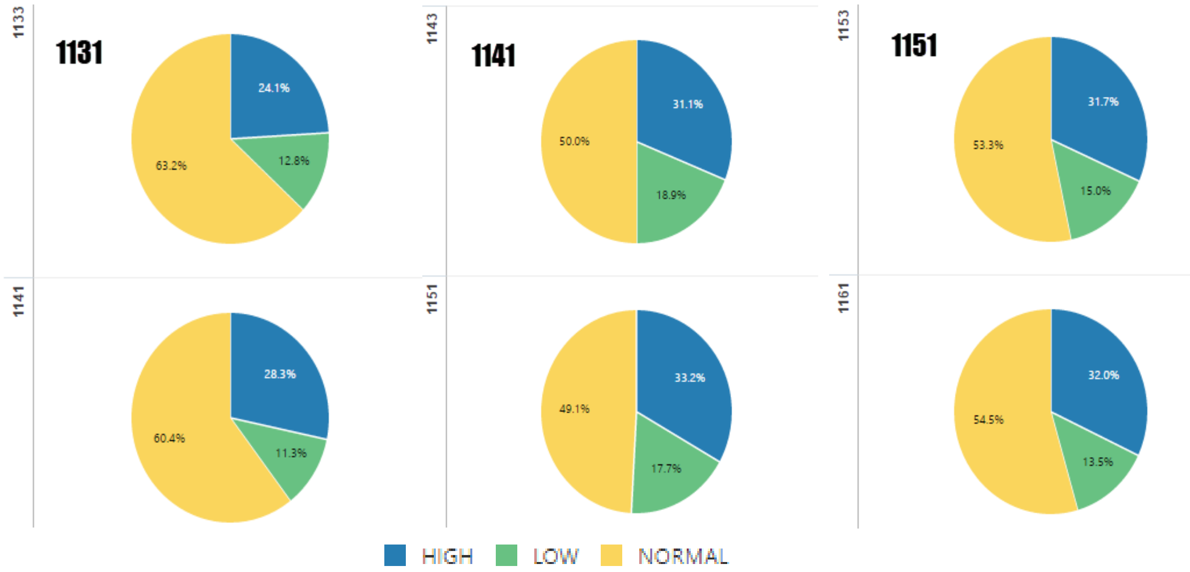
The UPH SoN is running a tight scholarship program for the training of nurses. Their primary concern involves preventing students from failing classes (or down-cohorting). FPar is also concerned about failing students, especially in the Liberal Arts core courses, which were deemed difficult for some students. Since the Liberal Arts courses and the SoN have already utilised a large portion of the Learning Management System (LMS, Moodle) in UPH, the application of some learning analytics to provide early warning for failing students might be possible.

The UPH SoM is interested in the analytics of alumni data since musicians can work in many branches or capacities: e.g., music educators, artists/performers, music creators, sound engineers, music therapists, advertisements, etc. The availability of data analytics based on the alumni database can provide guidelines for the SoM to tailor their curriculum for the job market needs, and perhaps also to organise a more effective alumni donation effort.

Figure 1 displays the visualisation of GPA percentage of Hospitality students (a) and Management students (b), which provide a quick overview of the performance of students from 3 different cohorts in two consecutive semesters (top: odd semester 2013, bottom: odd semester 2014). More than 50% of the Hospitality students show a "high" GPA (blue), compared to most (around 50%) of the Management students, who show a "normal" GPA (yellow). Management students also consistently show more percentage of students having a "low" GPA (green) compared to Hospitality students (although still well below 20%).



(a) Hospitality (Tourism) Department



(b) Management Department

Figure 1. GPA percentage of Hotel and Management Students from cohort 2013, 2014 & 2015 (Blue = high GPA, Yellow = normal GPA & Green = Low GPA)

Comparing the three cohorts (2013, 2014 & 2015), the cohort 2013 from both Hospitality and Management students show a consistently lower percentage of students with "low" GPA (green). Such descriptive patterns may give departments the impetus to analyse and adjust their learning processes. Are the students from 2013 with comparatively less low GPAs relatively better than other cohorts? Or is it because they have adapted longer to university life that they began achieving higher GPAs? On the other hand, are the higher GPAs (blue) of more senior students (cohort 2013) typically lower compared to cohort 2014 & 2015 because they have experienced more difficult courses?

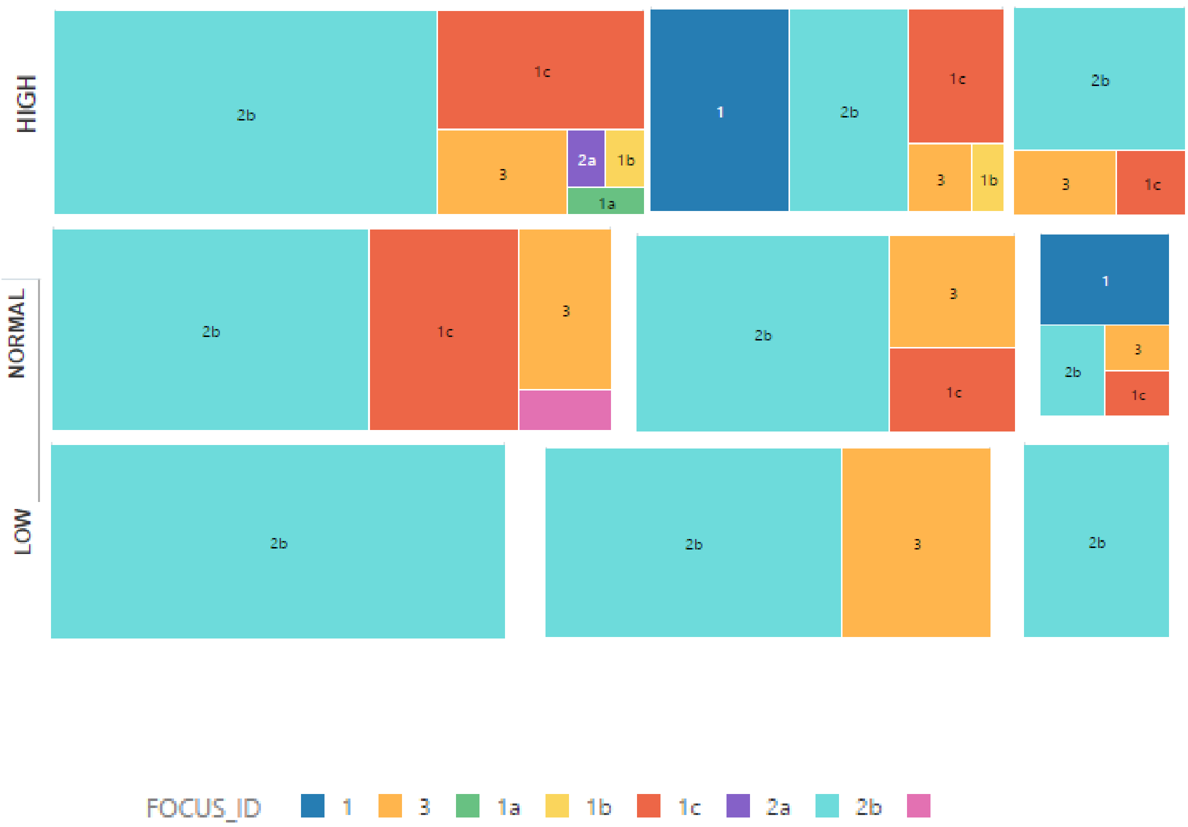


Figure 2. Students from groups of high school (admission data) who perform with "low", "normal" and "high" (bottom, middle, top) GPA in the Hospitality department.

Figure 2 displays the number of students performing with "low", "normal" and "high" (bottom, middle, top) GPA coming from certain groups of high school based on the admission data. Students from the high schools in group 1 (including group 1a, 1b & 1c) are performing consistently with "normal" or "high" GPA. Whereas students from group 2b & 3 are included among the students with the "low" GPA. Therefore, the data show that the grouping used by the admission office is reasonable for these group of students from the Hospitality department (group 1 schools are considered better schools than group 2b or group 3). If similar data for other departments show some deviations, for example, many students from group 1 schools deliver a "low" GPA, then perhaps the grouping for those schools can be analysed further or adjusted. This type of visualised data – although rather difficult to read – begins to promise certain predictive power, given they provide consistent patterns.

When the GPA of students is displayed according to the admission ranking of those students, we get Figure 3.

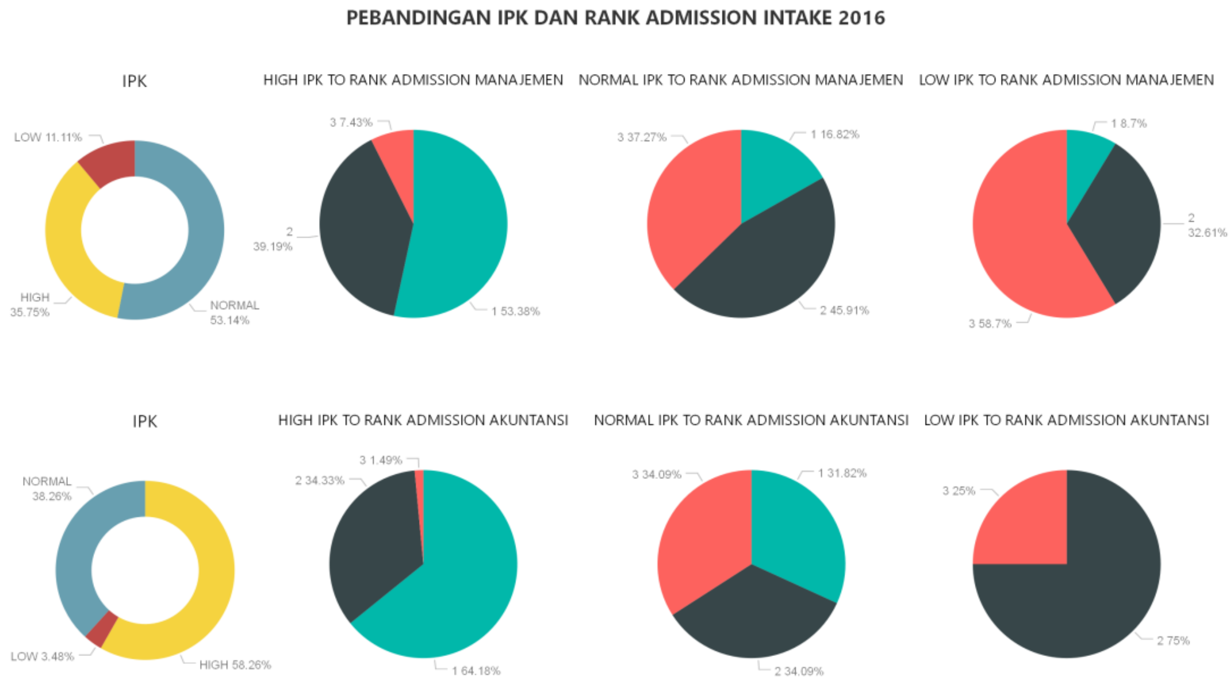


Figure 3. Students from admission ranking group 1, 2 & 3 (Green, Black & Red) who perform with "low", "normal" and "high" GPA in Management department (top) and Accounting department (bottom).

Figure 3 compares the data from the Management (top) and Accounting (bottom) students. The high GPA students are mostly coming from admission rank 1 & 2, and the low GPA students are mostly coming from admission rank 2 & 3. Moreover, the data shows that a higher percentage of students from admission rank 1 & 2 can achieve a higher GPA in the Accounting department than in the Management department.

Other data that have been explored and visualised include (1) full time vs part-time lecturer ratio, (2) lecturer's workload per semester, (3) size of classes and percentage of failing students in the class, (4) failing percentage in certain key courses, (5) and student questionnaires on lecturers. The full-time vs part-time lecturer pattern can give suggestions on recruitment policies. Lecturer's workload patterns can suggest in which semester lecturers can have more freedom in conducting research. Numbers on class sizes, failing students, and lecturer's evaluations by students can give ideas to create a better learning experience for students.

Discussion

In the affirmative discussions with the deans and university leaders, data analytics are seen to be useful to improve several areas: (1) increase admission numbers/quality, (2) optimise research and teaching (increase the number of research activities or the quality of student questionnaire results), (3) optimise resource utilisation (funding, classes, facilities, duration of the study), (4) optimise faculty workload, and (5) increase retention rates.

Although hopes are high for data analytics usefulness, some requirements are needed for the successful use of data analytics. Data always need sensible interpretation. An experienced reader may focus on the right features of the data and ignore the irrelevant artefacts to come to correct and useful conclusions. Handed to an untrained reader, data may be easily misinterpreted and become useless. The gain of data analytics is thus determined by the expertise of the data scientists interpreting the data.

Take the simplest form of student data: the GPA. The GPA might be averaged or sliced according to certain thresholds to indicate high, normal and low GPA, as shown in Figure 3. However, it cannot be generally concluded that higher GPA from one department is better than lower GPA of another. The department with the lower GPA might have been more rigorous than the department with a higher GPA. In Figure 3, for example, there can be two possible interpretations. Either the Accounting department provides a better learning environment (smaller class sizes, better lecturers, etc.) so that more percentage of students from rank admission 1 & 2 can achieve a higher GPA. Or the Management department is more rigorous in grading so that a lower percentage of students from rank admission 1 & 2 achieve a higher GPA.

Interpreting data requires good intuition into the business processes. Take, for example, the admission data. If we look at the admission data: e.g., 80% applicants follow through the process of applications, 20% are admitted (and fewer than 20% are finally enrolled), it seems that the efficiency for the admission process is very low, even lower than 20%. However, an experienced administrator would know that a good university usually has a very high number of applicants, and the admission rate can be even lower than 10%, for example, the Ivy League universities. So, every piece of visualised data can have multiple interpretations. It is thus imperative to discuss the interpretation of data with the stakeholders to arrive at good useful conclusions.

In our observations on many descriptive datasets, the more relations or additional variables explored between different datasets, the higher the potential to identify useful inference or predictions. For instance, comparing GPA data alone between different departments may not be extremely useful for interpretations. However, relating the GPA data with admission ranking or high school groups (from admission data) might facilitate better interpretation of the GPA data. Such data combination might also confirm or deny the efficacy of the admission (ranking) process. Another example is the data on classroom utilisation. Initiatives for the efficiency of classroom utilisation might be more sensible if the size of the rooms and the number of students are considered.

The discussions with deans and university leaders seem to suggest that the availability of data visualisation is only the first step towards the evolution in refining institutional processes and policies. As soon as deans and leaders see the data, they will come up with more ideas of potentially useful data and how to use them. On the other hand, to ensure the availability and integrity of certain data, some business processes and policies must be enforced or changed. So, it seems to be beneficial to start the iteration cycle early, rather than waiting for the perfect availability of data to reap the benefits of data analytics.

The UPH School of Nursing (SoN) and Hospitality (FPar) are very interested in using data analytics to help failing students. High interest in using learning analytics toward such application is also reflected in many research results. However, using learning analytics for at-risk students might not be as straightforward as it may seem. It was found that online interaction (Davies & Graff, 2005), as well as online interaction with peers, content and instructor (Jo, Yu, Lee & Kim, 2015), did not predict final grades. However, Davies & Graff (2005) also noted that students who failed courses tended to interact less frequently. Jo et al. (2015), on the other hand, also found that the total login frequency in LMS and the assignments and assessment composites showed a significant correlation with final grades.

It has been observed that a student's performance improved after being approached for learning analytics-based intervention (Wong, 2017). However, the critical question might be, whether learning analytics allow lecturers to identify at-risk students that would be typically missed (Jayaprakashet al., 2014). That would be true for large classes with 100+ students. However, in small classes, the added value of learning analytics would seem to be smaller, especially if the cumulative GPA is shown to be the strongest predictor, overshadowing all other factors, including partial grades, forum posts and content read (Bainbridge et al., 2015). Lauria et al., (2013) also found that cumulative GPA is the element that is most predictive of student outcome. In that case, lecturers may simply pull up data of a student's last GPA to estimate which students might have difficulties with their class.

Gašević, Dawson & Siemens (2015) remind us that learning analytics is about learning. Thus, much available data that can be extracted from the LMS may not necessarily relate to the condition of learning itself. In their studies, Gašević et al. (2016) found different patterns of LMS trace data effect on academic success. Some variables were not significant predictors for academic success. In contrast, variables that were significant predictors in the general model were not found to be significant predictors in specific courses or vice versa. Furthermore, one variable showing a positive effect on academic success in one course might show a negative association in another course. Thus, attention to the instructional conditions (internal and external) is crucial for the productive use of learning analytics.

Jovanovic et al. (2017) observed students' learning in a flipped learning situation and did clustering on the learning sequence data to identify different patterns of learning strategy.

Students with *intensive* and *strategic* learning patterns, who showed higher engagement with learning materials, also showed higher exam performance compared to the larger group of students with *selective* learning strategy, who showed lower engagement. So, it is possible to design learning analytics that informs lecturers of students' different learning strategies: *deep approach* with critical evaluation and synthesis of information-driven by intrinsic motivation vs *surface approach* dominated by shallow cognitive strategies and extrinsic motivation. However, such learning analytics must be built into the instructional design, and cannot be done by simply observing login, clicking or posting behaviours.

UPH SoM (School of Music) is more eager to use data analytics to track the alumni's career. For the community of musicians, looking out for career opportunities and challenges seems to be more important than just concerns about grades or exam performance. It is possible to design clustering analytics or predictive modelling with the combination of academic data and alumni data that can be used for career advisement or curriculum evaluation.

The Limits of Big Data and Data Analytics

Data analytics is a powerful and useful tool, but not without limitations. Mayer-Schönberger & Cukier (2013) popularised big data predictive power with the example of Google Flu Trend (GFT) in their bestseller book. It is claimed that Google can predict the activity of influenza-like illness in the regions of the US within one day's lag by analysing large numbers of Google search queries (Ginsberg et al., 2009). The hype of big data can be summarised by what Mayer-Schönberger & Cukier wrote (2013):

> As humans, we have been conditioned to look for causes, even though searching for causality is often difficult and may lead us down the wrong paths. In a big-data world, by contrast, we won't have to be fixated on causality; instead, we can discover patterns and correlations in the data that offer us novel and invaluable insights. The correlations may not tell us precisely *why* something is happening, but they alert us *that* it is happening.

The optimism of big data can be stated like this: because of the sheer amount of data, we can detect patterns accurately in real-time (through correlations), without even knowing the causes of those patterns!

However, the optimism of the paper published by the Google scientists was quickly refuted by the failure of GFT to predict the swine flu pandemic that hit the world in 2009 (Salzberg, 2014; Olson et al., 2013). In February 2013, the GFT was predicting more than double the proportion of doctor visits for influenza-like illness than the Centers for Disease Control and Prevention (CDC), and the GFT has missed peaks for 100 out of 108 weeks in 2011 (Lazer et al., 2014).

The poster child of big data soon became the poster child of the foibles of big data (Lazer & Kennedy, 2015).

The problems with the GFT include: (1) Accidental strong correlation with other search terms unrelated to the flu (2) Dynamic changes in the Google search algorithm, and (3) Recommended search terms that increase the relative magnitude of specific searches (Lazer et al., 2014; Lazer & Kennedy, 2015). The main problem with the optimism on big data and data analytics is written by Lazer et al. (2014) as follows:

> However, the quantity of data does not mean that one can ignore foundational issues of measurement and construct validity and reliability and dependencies among data. The core challenge is that most big data that have received popular attention are not the output of instruments designed to produce valid and reliable data amenable for scientific analysis.

Thus, it is not true that the causality behind the observable correlation patterns can be ignored simply because of the immense quantity of data! Even more, observing live data means dealing with changing dynamics of causalities and noisy data. We can have overfitting phenomena, where our predictive models match too closely to specific noisy samples, that the models are highly predictive for those samples, but not for the general population (Provost & Fawcett, 2013).

Gibson & Ifenthaler (2017) pointed out the preparation needed for education researchers to be able to deal with research in big data, which includes the awareness of the limitations of regression models for big data. For the regression models to be adequate, some assumptions about the population of data must be met: (1) Probability distribution of *independent* values must exist, (2) the variances of all the distributions must be equal, and (3) the means of the distributions *must all fall on the regression lines*. In a complex data environment with dynamic interdependencies, such assumptions for *linear*, *independent* and *well-behaved* regression models are rarely met. Research in big data can employ, for example, support vector machine (SVM) that functions as a non-probabilistic binary classifier, which can be trained (based on supervised machine learning) to map its inputs into high-dimensional feature spaces to perform a nonlinear classification.

Academic analytics is not exactly big data. Although the size is smaller, academic analytics retains many complications of big data analysis. Correlations and patterns imply complex causalities. The availability of data does not mean that correct or useful interpretations can be achieved automatically. Skilled data scientists and wisdom in strategic decisions by the university leadership are even more necessary with the availability of academic analytics.

The implementation of learning analytics (LA) may also entail complicated new ethical dilemmas. If LA involves location tracking and biometrics that come with new evolving technology, questions about data privacy and data usage in LA can become complicated. Especially if data belonging is not clear cut (does the data belong to individual, institution, or data collection system vendor?) (Reyes, 2015). Even when students are being made aware that their engagement in LMS is being monitored, do their consents only relate to data collection or also include data analysis and reuse (Lawson et al., 2016)? Could students opt out of this action? Care should also be taken in labelling students according to LA as at-risk students. Studies confirmed that teachers' biases towards students (positive or negative) could influence the performance of students (Gino, 2018). Higher education institutions must be aware of the ethical rights of students in implementing LA systems.

Conclusions

The interviews with several deans and university leaders of Universitas Pelita Harapan (UPH) indicate the potential usefulness of academic analytics pursued in Universitas Pelita Harapan. Direct comparison of GPA distributions throughout departments might already be valuable. The better departments might show curves more like the normal Bell curve, which shows that the department is good enough with a low number of struggling students but challenging enough that students must work hard to achieve high GPAs.

If the GPA data are combined with the admission demographic data, the admission ranking and high school grouping by reputations data used by the admission process, they might give additional interpretation angles for the comparison of GPA across different departments. Such a combination of data might also inform, confirm or deny the efficacy of the admission criteria/processes.

Other potential uses of academic analytics identified from data observation and interviews include (1) Early warning for failing students, (2) Mapping and analysis of alumni jobs, (3) optimising the number of full-time/part-time lecturers, and (4) optimising classroom utilisation.

Research results showed that the cumulative GPA is the strongest predictor of students' learning success. Therefore, to achieve reasonable added value than just observing students' cumulative GPA, successful implementation of learning analytics must involve a thoughtful design of data observation into the learning settings. Simply observing login, clicking or posting behaviours might not give a useful predictor for students' success. Awareness of ethical questions related to the implementation of learning analytics must also guide higher education policies.

The observation of many datasets and interviews with deans and university leadership in UPH also increased our awareness of the complicated task of data interpretations. Discussion with stakeholders and the availability of skilled data scientists are crucial for fruitful strategic decisions with academic analytics.

REFERENCES

Aldowah, H., Al-Samarraie, H., & Fauzy, W. M. (2019). Educational data mining and learning analytics for 21st century higher education: A review and synthesis. *Telematics and Informatics*, *37*, 13-49. https://doi.org/10.1016/j.tele.2019.01.007

Bainbridge, J., Melitski, J., Zahradnik, A., Lauría, E. J., Jayaprakash, S., & Baron, J. (2015).Using learning analytics to predict at-risk students in online graduate public affairs and administration education. *Journal of Public Affairs Education*, *21*(2), 247-262. https://doi.org/10.1080/15236803.2015.12001831

Daniel, B. (2015). Big Data and analytics in higher education: Opportunities and challenges. *British journal of educational technology*, *46*(5), 904-920. https://doi.org/10.1111/bjet.12230

Davies, J., & Graff, M. (2005). Performance in e-learning: online participation and student grades. *British Journal of Educational Technology*, *36*(4), 657-663. https://doi.org/10.1111/j.1467-8535.2005.00542.x

Ekowo, M., & Palmer, I. (2016). The Promise and Peril of Predictive Analytics in Higher Education: A Landscape Analysis. *New America*.

Gašević, D., Dawson, S., & Siemens, G. (2015). Let's not forget: Learning analytics are about learning. *TechTrends*, *59*(1), 64-71. https://doi.org/10.1007/s11528-014-0822-x

Gašević, D., Dawson, S., Rogers, T., & Gasevic, D. (2016). Learning analytics should not promote one size fits all: The effects of instructional conditions in predicting academic success. *The Internet and Higher Education*, *28*, 68-84. https://doi.org/10.1016/j.iheduc.2015.10.002

Gibson, D. C., & Ifenthaler, D. (2017). Preparing the next generation of education researchers for big data in higher education. In *Big data and learning analytics in higher education* (pp. 29-42).Springer, Cham. https://doi.org/10.1007/978-3-319-06520-5_4

Gino, F. (2018). *Rebel talent: Why it pays to break the rules at work and in life*. Pan Macmillan.

Ginsberg, J., Mohebbi, M. H., Patel, R. S., Brammer, L., Smolinski, M. S., & Brilliant, L. (2009). Detecting influenza epidemics using search engine query data. *Nature*, *457*(7232), 1012. https://doi.org/10.1038/nature07634

Jayaprakash, S. M., Moody, E. W., Lauría, E. J., Regan, J. R., & Baron, J. D. (2014). Early alert of academically at-risk students: An open source analytics initiative. *Journal of Learning Analytics*, *1*(1), 6-47. https://doi.org/10.18608/jla.2014.11.3

Jo, I. H., Yu, T., Lee, H., & Kim, Y. (2015). Relations between student online learning behavior and academic achievement in higher education: A learning analytics approach. In *Emerging issues in smart learning* (pp. 275-287). Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-662-44188-6_38

Jovanović, J., Gašević, D., Dawson, S., Pardo, A., & Mirriahi, N. (2017). Learning analytics to unveil learning strategies in a flipped classroom. *The Internet and Higher Education*, *33*(4), 74-85. https://doi.org/10.1016/j.iheduc.2017.02.001

Lauría, E. J., Moody, E. W., Jayaprakash, S. M., Jonnalagadda, N., & Baron, J. D. (2013). Open academic analytics initiative: initial research findings. In *Proceedings of the Third International Conference on Learning Analytics and Knowledge* (pp. 150-154). https://doi.org/10.1145/2460296.2460325

Lawson, C., Beer, C., Rossi, D., Moore, T., & Fleming, J. (2016). Identification of 'at risk'students using learning analytics: the ethical dilemmas of intervention strategies in a higher education institution. *Educational Technology Research and Development*, *64*(5), 957-968. https://doi.org/10.1007/s11423-016-9459-0

Lazer, D. & Kennedy, R. (2015). What We Can Learn from the Epic Failure of Google Flu Trend. *Wired*. Retrieved from https://www.wired.com/2015/10/can-learn-epic-failure-google-flu-trends/

Lazer, D., Kennedy, R., King, G., & Vespignani, A. (2014). The parable of Google Flu: traps in big data analysis. *Science*, *343*(6176), 1203-1205. https://doi.org/10.1126/science.1248506

Macfadyen, L. P., & Dawson, S. (2012). Numbers are not enough. Why e-learning analytics failed to inform an institutional strategic plan. *Journal of Educational Technology & Society*, *15*(3), 149-163.

Mayer-Schönberger, V. & Cukier, K. (2013).*Big Data: A Revolution That Will Transform How We Live, Work and Think.* London: John Murray.

Olson, D. R., Konty, K. J., Paladini, M., Viboud, C., & Simonsen, L. (2013).Reassessing Google Flu Trends data for detection of seasonal and pandemic influenza: a comparative epidemiological study at three geographic scales. *PLoS computational biology*, *9*(10), e1003256. https://doi.org/10.1371/journal.pcbi.1003256

Picciano, A. G. (2012). The evolution of big data and learning analytics in American higher education. *Journal of Asynchronous Learning Networks*, *16*(3), 9-20. https://doi.org/10.24059/olj.v16i3.267

Provost, F., & Fawcett, T. (2013). *Data Science for Business: What you need to know about data mining and data-analytic thinking*. Sebastopol: O'Reilly Media, Inc.

Rajni, J., & Malaya, D. B. (2015). Predictive analytics in a higher education context. *IT Professional*, *17*(4), 24-33. https://doi.org/10.1109/MITP.2015.68

Reyes, J. A. (2015). The skinny on big data in education: Learning analytics simplified. *TechTrends*, *59*(2), 75-80. https://doi.org/10.1007/s11528-015-0842-1

Salzberg, S. (2014). Why Google Flu is a Failure. *Forbes*. Retrieved from https://www.forbes.com/sites/stevensalzberg/2014/03/23/why-google-flu-is-a-failure/#2d65019a5535

Siemens, G., & Long, P. (2011). Penetrating the fog: Analytics in learning and education. *EDUCAUSE Review*, *46*(5), 30.

Van Barneveld, A., Arnold, K. E., & Campbell, J. P. (2012). Analytics in higher education: Establishing a common language. *EDUCAUSE learning initiative*, *1*(1), l-ll.

Van Harmelen, M., & Workman, D. (2012).Analytics for learning and teaching. *CETIS Analytics Series*, *1*(3), 1-40.

West, D. M. (2012). Big data for education: Data mining, data analytics, and web dashboards. *Governance Studies at Brookings*, *4*, 1-0.

Wong, B. T. M. (2017). Learning analytics in higher education: an analysis of case studies. *Asian Association of Open Universities Journal*. https://doi.org/10.1108/AAOUJ-01-2017-0009